

UNIVERSIDAD NACIONAL DE CÓRDOBA
FACULTAD DE CIENCIAS EXACTAS, FÍSICAS Y
NATURALES



PRÁCTICA PROFESIONAL SUPERVISADA

Calibración y modelación hidrológica preliminar con énfasis
en el ordenamiento y control de datos

Aplicación a las cuencas de los ríos Salado (Santa Fe) y
Guauguay (Entre Ríos)

Nereo Jorge JUAREZ

DIRECTOR: Mg. Ing. Facundo José ALONSO

- 2013 -

*A mis padres Gladys y Nereo,
muchas gracias por todo su apoyo,
sin ellos, esto no hubiera sido posible.*

Calibración y modelación hidrológica preliminar con énfasis en el ordenamiento y control de datos

Aplicación a las cuencas de los ríos Salado (Santa Fe) y Gualeguay (Entre Ríos)

Resumen

El presente trabajo presenta los avances realizados durante el periodo de Práctica Supervisada (PS) en las instalaciones del Centro de Estudios y Tecnología del Agua (CETA) de la Universidad Nacional de Córdoba. Los trabajos estuvieron asociados, en primer lugar, a la recopilación de antecedentes hidrológicos históricos de las cuencas del río Salado (en la zona perteneciente a la Provincia de Santa Fe), y de río Gualeguay (Entre Ríos). En segundo lugar, se realizó una revisión exhaustiva de dicha información, depurando las series de variables hidrológicas para lograr la consistencia real y atribuir a los datos la confiabilidad necesaria para su futura modelación. A posteriori, se llevó a cabo la calibración automática de un modelo hidrológico de origen francés, denominado Génie Rural (GR), a través de una metodología de optimización multi-objetivo, denominada MOCOM-UA. Dicha metodología permitió calibrar el modelo en modo simulación y en modo pronóstico. Finalmente, se verificó la calidad del ajuste en periodos reservados para tal fin. Los resultados indican una buena performance del modelo hidrológico y del algoritmo de optimización, pero una deficiente calidad de algunos tramos de la información incluso depurada.

Calibration and preliminary hydrologic modeling with emphasis on planning and control of data

Application to the Salado river basin (Santa Fe) and Gualeguay (Entre Rios)

Abstract

This paper presents the progress made during the period of supervised practice (PS) at the premises of the Centre for Research and Water Technology (CETA) of the National University of Córdoba. The work was associated, first, to the collection of historical hydrological history of the Salado River basin (in the area belonging to the Province of Santa Fe), and river Gualeguay (Entre Rios). Second, a thorough review of this information was made by purifying series of hydrological variables to achieve real consistency and attribute data reliability needed for future modeling. In hindsight, was carried out automatic calibration of a hydrological model of French origin, called Génie Rural (GR), through a methodology of multi-objective optimization, called MOCOM-UA. This methodology allowed calibrating the model in simulation mode and prediction mode. Finally, the quality of the adjustment periods reserved for that purpose is fulfilled. The results indicate a good performance of the hydrological model and optimization algorithm, but poor quality of some sections of the even refined information.

Agradecimientos

En primer lugar, quiero agradecer a mis padres Gladys y Nereo, por la confianza que depositaron en mí, que si no fuera por su constante apoyo no hubiese logrado este objetivo de ser Ingeniero Civil.

A mis hermanas, Flavia y Lourdes, que desde su lado siempre se hicieron presentes, apoyando todas mis ideas en todos los aspectos de la vida. A mis abuelas, tíos y primas que siempre estuvieron a la expectativa de que me vaya bien. En fin agradezco a mi familia en general caracterizada por sus aportes de buenas vibras, llenándome de positivismo.

En segundo lugar agradezco a mis amigos de Alta Gracia que siempre estuvieron presentes y confiando plenamente que iba a ser algún día Ingeniero. A mis compañeros y amigos de la facultad que fueron de gran ayuda en diferentes aspectos, no solo los estrictamente académicos donde pasamos nervios, alegrías, y todo lo que requiere el hecho de estudiar full time durante 5 años, sino también los más importantes que son los humanos.

A mi director de Práctica Supervisada Ing. Facundo José Alonso, quien aparte de brindar al máximo sus conocimientos profesionales, siempre mantuvo firme la idea del compañerismo y trabajo en equipo, dejando sentado siempre su calidad de persona.

A mi país, por haberme brindado el apoyo económico a través del programa Nacional de Becas Bicentenario, en varias etapas de la carrera.

Al Ing. Carlos Marcelo García, director del Centro de Estudios y Tecnología del Agua (CETA), quien por medio de éste me brindó al posibilidad de realizar la PS.

Al Instituto Superior de Estudios Ambientales (ISEA) por brindarnos el lugar de trabajo y reuniones.

A todas aquellas personas que me facilitaron la información necesaria y contribuyeron al desarrollo del presente trabajo, Mg. Ing. Facundo Alonso, Dr. Ing. Juan Carlos Bertoni (UNC), Ing. Sergio Menajovski (UNC), Dr. Ing. Andrés Rodríguez, Ing. Érica Díaz, Ing. Nicolás Guillén, Ing. José Manuel Díaz Lozada, Mg. Ing. Juan Weber, Ing. Antoine Patalano, Mg. Ing. Marcelo

Baldisone, Ludovic Oudin, Juan Bianchi, entre otros que de alguna manera también contribuyeron.

Extiendo un agradecimiento especial al Servicio Meteorológico Nacional y Subsecretaría de Recursos Hídricos por brindar gran cantidad de información, no solo a través de páginas webs, sino por la constante ayuda a través del Ing. Luis Lenzi (Santa Fe) y Lic. Ana María Gómez.

INDICE

1. INTRODUCCIÓN	1
2. OBJETIVOS.....	3
2.1. Objetivo general	3
2.2. Objetivos particulares	3
3. REVISIÓN DE ANTECEDENTES.....	4
3.1. MODELOS HIDROLÓGICOS	4
3.2. SISTEMAS, CONCEPTOS Y CLASIFICACIÓN	6
3.3. EL SISTEMA HIDROLÓGICO	8
3.4. EL PROCESO DE TRANSFORMACIÓN LLUVIA-CAUDAL	9
3.5. MODELOS, CONCEPTOS Y CLASIFICACIÓN	11
3.6. EL MODELO LLUVIA-CAUDAL.....	16
3.7. OPTIMIZACIÓN – CONCEPTOS GENERALES	17
3.7.1. Métodos de optimización.....	18
3.7.2. Generación de números aleatorios.....	19
3.8. ALGORITMOS GENÉTICOS	22
3.8.1. Conceptos básicos de biología.....	23
3.8.2. La técnica de los algoritmos genéticos	28
3.9. TÉCNICAS DE CALIBRACIÓN	31
3.9.1. Clasificación de las técnicas de calibración	32
3.10. CALIBRACIÓN AUTOMÁTICA DE MODELOS P-Q.....	41
3.10.1. La función objetivo.....	41
3.10.2. Algoritmos de optimización para modelos P-Q.....	47
3.10.3. Criterio de parada.....	50
3.10.4. Datos de calibración.....	52
4. METODOLOGIA.....	55
4.1. DESCRIPCIÓN DEL MODELO HIDROLOGICO GR.....	55
4.1.1. Objetivos de su desarrollo.....	55
4.1.2. Principales características.....	55
4.1.3. Estructura del modelo	56

4.1.3.1.	Descripción matemática GR4J	57
4.2.	EL ALGORITMO MOCOM-UA.....	63
4.2.1.	Evolución competitiva compleja	64
4.2.1.1.	Atribución de pesos	64
4.2.1.2.	Selección de subcomplejos.....	65
4.2.1.3.	Evolución de los Subcomplejos.....	66
4.2.1.4.	Ordenamiento de los puntos del Complejo.....	67
4.2.2.	Algoritmo de Nelder y Mead	67
4.2.2.1.	Organización del Subcomplejo.....	67
4.2.2.2.	Cálculo del centro de gravedad.....	67
4.2.2.3.	Cálculo del punto de reflexión.....	68
4.2.2.4.	Evaluación del punto de reflexión.....	68
4.2.2.5.	Cálculo del punto de contracción.....	68
4.2.2.6.	Evaluación del punto de contracción.....	69
4.2.3.	Pasos del método.....	70
5. RECOPIACIÓN Y PROCESAMIENTO DE		
INFORMACIÓN.....		74
5.1.	TIEMPO DE RESPUESTA DE LAS CUENCAS	76
5.2.	CUENCA DEL RÍO SALADO	78
5.2.1.	Descripción de la cuenca	78
<i>Ubicación geográfica</i>	78	
<i>Características generales</i>	79	
<i>Dinámica Hídrica</i>	79	
<i>Zona de interés</i>	84	
5.2.2.	Precipitaciones.....	84
Información recolectada	84	
Procesamiento de información	86	
<i>Estaciones de medición de precipitaciones</i>	86	
<i>Ordenamiento</i>	88	
<i>Polígonos de Thiessen o Voronoi</i>	91	
<i>Determinación de precipitación media</i>	93	
5.2.3.	Caudales	94
5.2.4.	Evapotranspiración potencial.....	94
5.3.	CUENCA DEL RÍO GUALEGUAY.....	97

5.3.1. Descripción de la cuenca	97
5.3.2. Precipitaciones.....	98
5.3.3. Aforos, niveles y caudales – curva H-Q.....	100
5.3.4. Evapotranspiración potencial.....	106
5.3.5. Análisis de consistencia.....	107
6. ANÁLISIS ESTADÍSTICO	113
6.1. INTRODUCCIÓN	113
6.2. NIVEL DE SIGNIFICANCIA.....	113
6.3. PRUEBA DE ESTACIONALIDAD (Kendall)	115
6.3.1. Análisis de precipitaciones medias diarias	117
6.3.2. Análisis de caudales medios diarios.....	118
6.4. PRUEBA DE HOMOGENEIDAD (Wilcoxon)	119
6.4.1. Análisis de precipitaciones medias diarias	120
6.4.2. Análisis de caudales medios diarios.....	121
6.5. REVISIÓN FINAL DE LAS SERIES DE DATOS	122
7. ANÁLISIS DE RESULTADOS.....	123
7.1. ARCHIVOS PARA MODELACIÓN.....	123
7.2. ALGORITMO MOCOM-UA - OPTIMIZACIÓN	127
7.3. MODELO GR - SIMULACIÓN Y PRONÓSTICO.....	131
7.3.1. Modo Simulación	132
7.3.2. Modo Pronóstico.....	139
8. CONCLUSIONES	144
9. BIBLIOGRAFÍA	147
9.1. REFERENCIAS BIBLIOGRÁFICAS.....	147
9.2. PÁGINAS WEBS.....	151
10. ANEXO	152

LISTA DE FIGURAS

<i>Figura 3.1. El ciclo hidrológico (CONICYT, 2007)</i>	08
<i>Figura 3.2.: Modelo hidrológico de transformación lluvia-caudal (HEC-HMS, 2000)</i>	10
<i>Figura 3.3.: Etapas de la meiosis (Kuri y Galaviz, 2002)</i>	26
<i>Figura 3.4.: Esquema de trabajo de la enzima ADN polimerasa (Kuti y Galaviz, 2002)</i> ...27	
<i>Figura 3.5.: Esquema de la Región de Pareto para dos funciones objetivo</i>	35
<i>Figura 3.6.: Proyección plana de la superficie de respuesta de una FO</i>	37
<i>Figura 4.1. Estructura del modelo GR4J, (Perrin et al, 2003)</i>	57
<i>Figura 4.2. Modelo GR4J modo pronóstico</i>	62
<i>Figura 0.3. Zona de validez de los parámetros del modelo, (Alonso, 2008)</i>	70
<i>Figura 0.4. Puntos válidos y no válidos (Alonso, 2008)</i>	71
<i>Figura 5.1. Cuenca hidrologica, (Mijares, 1992)</i>	76
<i>Figura 5.2.:Cuenca del Río Salado, Argentina</i>	78
<i>Figura 5.3.: Caracterización de la dinámica hídrica del Río Salado</i>	80
<i>Figura 5.4.: Imagen escaneada de códigos de estaciones</i>	85
<i>Figura 5.5.: Localización de estaciones - Google Earth</i>	87
<i>Figura 5.6.: Localización de estaciones en plano Auto Cad</i>	88
<i>Figura 5.7.: estaciones cercanas vs tiempo</i>	89
<i>Figura 5.8a.: Polígono de Thiessen 1 (Software Arc Gis 10)</i>	92
<i>Figura 5.8b.: Polígono de Thiessen 2 (Software Arc Gis 10)</i>	93
<i>Figura 5.9.: Tabla para la determinación de la radiación extraterrestre</i>	96
<i>Figura 5.10: cuenca del Río Gualeguay en Entre Ríos, Argentina.</i>	97
<i>Figura 5.11.: Comparación de datos de aforo y mediciones de alturas y caudales</i> ...102	
<i>Figura 5.12.: Curva H-Q envolvente superior</i>	102
<i>Figura 5.13.: Comparación de caudales</i>	105
<i>Figura 5.14.: Gráficos de precipitaciones vs tiempo y caudales vs tiempo</i> <i>Figura 5.15a.: Comparación de caudales y precipitaciones, tiempo de retardo</i>	112
<i>Figura 5.15a.: Comparación de caudales y precipitaciones, tiempo de retardo</i>	110
<i>Figura 5.15b.: Comparación de caudales y precipitaciones, tiempo de retardo</i>	111
<i>Figura 6.1.: Funciones de frecuencia para la información de muestra y funciones de probabilidad para la población.</i>	115
<i>Figura 6.2.: Función de frecuencia para la información de muestra y funciones de probabilidad para la población.</i>	115
<i>Figura 6.3.: Salida HyFrinse prueba de estacionaridad de Kendall – precipitaciones</i>	117
<i>Figura 6.4.: Salida HyFrinse prueba de estacionaridad de Kendall – caudales</i>	118

Figura 6.5.: Salida HyFrinse prueba de homogeneidad de Wilcoxon – precipitaciones.....	120
Figura 6.6.: Salida HyFrinse prueba de homogeneidad de Wilcoxon – caudales.	121
Figura 7.1.: Archivos para la programación – MOCOM y GR.....	123
Figura 7.2.: Carpeta “Datos”	124
Figura 7.3.: Fragmento de archivo de texto de entrada de datos P-EP-Q.....	124
Figura 7.4.: Fragmento de archivo de texto de salida de datos.....	125
Figura 7.5.: Ventana del ejecutable – mensaje de terminación del programa.....	125
Figura 7.6a.: Archivo de entrada Ent_Main – parte 1.....	126
Figura 7.6b.: Archivo de entrada Ent_Main – parte 2.....	126
Figura 7.6c.: Archivo de entrada Ent_Main – parte 3.....	126
Figura 7.7.: Archivo de entrada Ent_Main – MOCOM.....	127
Figura 7.8.: Variación del coeficiente de Nash-Sutcliffe.....	129
Figura 7.9.: Variación del coeficiente de Persistencia.....	130
Figura 7.10.: Archivo de entrada Ent_Main – Modo Simulación.....	132
Figura 7.11a.: Salida modo simulación – primera mitad de los datos.....	134
Figura 7.11b.: Salida modo simulación – primeros dos años.....	135
Figura 7.11c.: Salida modo simulación – tercer y cuarto año.....	136
Figura 7.11d.: Salida modo simulación – último período.....	136
Figura 7.12.: Salida modo simulación – segunda mitad de los datos.....	138
Figura 7.13.: Archivo de entrada Ent_Main – Modo Previsión.....	139
Figura 7.14a.: Salida modo previsión – primera mitad de los datos.....	140
Figura 7.14b.: Salida modo previsión – pico central.....	142
Figura 7.15.: Salida modo pronóstico – segunda mitad de los datos.....	143
Figura 10.1.: Salida modo simulación – primera mitad de los datos- horizonte=1día.....	152
Figura 10.2.: Salida modo previsión – primera mitad de los datos – horizonte=1 día.....	153
Figura 10.3.: Salida modo simulación – primera mitad de los datos – horizonte=2 días.....	154
Figura 10.4.: Salida modo previsión – primera mitad de los datos – horizonte =2 días.....	155
Figura 10.5.: Salida modo simulación – primera mitad de los datos – horizonte=3 días.....	156
Figura 10.6.: Salida modo previsión – primera mitad de los datos – hzte=3 días.....	157

<i>Figura 10.7.: Salida modo simulación – primera mitad de los datos – horizonte=4 días.....</i>	<i>158</i>
<i>Figura 10.8.: Salida modo previsión – primera mitad de los datos – horizonte =4 días.....</i>	<i>159</i>
<i>Figura 10.9.: Salida modo simulación – primera mitad de los datos – horizonte=5 días.....</i>	<i>160</i>
<i>Figura 10.10.: Salida modo previsión – primera mitad de los datos – horizonte =5 días.....</i>	<i>161</i>
<i>Figura 10.11.: Salida modo simulación – primera mitad de los datos – horizonte =6 días.....</i>	<i>162</i>
<i>Figura 10.12.: Salida modo previsión – primera mitad de los datos – horizonte =6 días.....</i>	<i>163</i>
<i>Figura 10.13.: Salida modo simulación – segunda mitad de los datos – horizonte =1 día.....</i>	<i>164</i>
<i>Figura 10.14.: Salida modo previsión – segunda mitad de los datos – horizonte =1 día.....</i>	<i>165</i>
<i>Figura 10.15.: Salida modo simulación – segunda mitad de los datos – horizonte =2 días.....</i>	<i>166</i>
<i>Figura 10.16.: Salida modo previsión – segunda mitad de los datos – horizonte =2 días.....</i>	<i>167</i>
<i>Figura 10.17.: Salida modo simulación – segunda mitad de los datos – horizonte =3 días.....</i>	<i>168</i>
<i>Figura 10.18.: Salida modo previsión – segunda mitad de los datos – horizonte =3 días.....</i>	<i>169</i>
<i>Figura 10.19.: Salida modo simulación – segunda mitad de los datos – horizonte =4 días.....</i>	<i>170</i>
<i>Figura 10.20.: Salida modo previsión – segunda mitad de los datos – horizonte =4 días.....</i>	<i>171</i>
<i>Figura 10.21.: Salida modo simulación – segunda mitad de los datos – horizonte =5 días.....</i>	<i>172</i>
<i>Figura 10.22.: Salida modo previsión – segunda mitad de los datos – horizonte =5 días.....</i>	<i>173</i>
<i>Figura 10.23.: Salida modo simulación – segunda mitad de los datos – horizonte=6 días.....</i>	<i>174</i>
<i>Figura 10.24.: Salida modo previsión – segunda mitad de los datos – horizonte =6 días.....</i>	<i>175</i>

1. INTRODUCCIÓN

A pesar de que el agua constituye el recurso más elemental y necesario para la vida, todavía existen muchas incertidumbres en relación a los procesos que gobiernan su movimiento dentro de la hidrosfera. El sistema hidrológico global es sumamente complejo, debido a los múltiples ambientes en que el agua interacciona, pasando por los más diversos estados.

El hombre ha enfrentado esta problemática del movimiento del agua en el planeta a través de modelos simplificados de la realidad. De aquí que para análisis de cuencas hidrográficas resultan un factor fundamental para interpretar y simular la dinámica a la que responde.

La simulación hidrológica tiene aplicación en las tres principales funciones de la hidrología: la predicción, el proyecto y la planificación. En cuanto a la predicción, una de las principales aplicaciones se da en el campo del pronóstico de crecidas a tiempo real. En proyecto, los modelos se utilizan para estimar caudales de diseño de obras hidráulicas. Mientras que en planificación, se utilizan para realizar una estimación de la disponibilidad del recurso hídrico a mediano y largo plazo, análisis de la variabilidad hidrológica o estudios sobre los cambios del uso de suelo, entre otras.

Una de las etapas más críticas en la aplicación de modelos hidrológicos está asociada al proceso de calibración, es decir, la elección del juego de valores para los parámetros que mejor representen las condiciones locales de la cuenca hidrográfica.

Tradicionalmente, la calibración de un modelo matemático se ha venido realizando en forma manual. Pero en las últimas décadas se han desarrollado, a nivel mundial, nuevos métodos de calibración automática de modelos hidrológicos, heredados del campo de la optimización. En esencia, la calibración automática consiste en minimizar una expresión matemática, denominada función objetivo, que cuantifica el error de simulación.

Los últimos desarrollos en materia de optimización han tomado prestado conceptos de la biología para incrementar la variabilidad y la capacidad de búsqueda, originando los denominados métodos heurísticos.

Entre estos métodos, se destaca el método de los algoritmos genéticos (Holland, 1975) inspirados en los mecanismos de optimización de las especies, sobre la base de la selección natural y supervivencia del individuo más apto. Estos métodos han demostrado notable eficiencia y representan un considerable ahorro de consumo computacional.

En este informe se hace una presentación de los modelos matemáticos en forma general, abarcando una revisión bibliográfica del método de los algoritmos genéticos, métodos de optimización general y de calibración automática.

Posteriormente se presenta una descripción del modelo hidrológico de transformación lluvia-caudal Genie Rural (Perrin et al, 2007) y el método de calibración global multi-objetivo, basado en algoritmos genéticos, MOCOM-UA. Además previo a esto se realiza un análisis de procesamiento de datos de precipitaciones medias diarias, caudales medios diarios y los propios de temperatura media, siendo estos últimos necesarios para determinar la evapotranspiración potencial, todos pertenecientes a dos cuencas argentinas pertenecientes al sistema Paraná constituyendo la entrada para dichas metodologías.

Finalmente se muestran los resultados obtenidos efectuando una comparación entre los caudales observados, simulados y pronosticados.

2. OBJETIVOS

2.1. Objetivo general

Obtener series de datos de entrada confiables para el modelo hidrológico Génie Rural, a través de una metodología de procesamiento adecuada, para realizar la modelación sobre dos cuencas hidrográficas argentinas comparables.

2.2. Objetivos particulares

Se listan los siguientes:

✓ Recopilar información histórica de lluvias, temperaturas y niveles de las cuencas de los ríos Salado y Gualeguay, todas a nivel diario.

✓ Determinar valores promedios diarios de las variables analizadas.

✓ Determinar la curva H-Q de las cuencas analizadas en la sección de aforo.

✓ Estimar la serie de Caudales diarios para cada cuenca en la sección de aforo.

✓ Estimar la serie de Evapotranspiración potencial para cada cuenca.

✓ Analizar la confiabilidad de las series de datos obtenidas.

✓ Identificar los rangos confiables dentro de las series obtenidas.

✓ Realizar la calibración multiobjetivo del modelo Génie Rural.

✓ Evaluar la performance de la modelación a partir de las funciones objetivo.

3. REVISIÓN DE ANTECEDENTES

En este trabajo final se llevó a cabo un importante proceso de adecuación de datos para efectuar la calibración automática de un modelo hidrológico empírico de transformación lluvia-caudal en una cuenca argentina. En consecuencia, como parte de la revisión del estado del arte, se aborda inicialmente una introducción a la teoría de modelos matemáticos.

El proceso de calibración automática de un modelo hidrológico es un problema de la optimización de modelos. Existe una gran variedad de métodos de optimización que se han ido desarrollando para distintos problemas a optimizar. La Investigación de operaciones ha incursionado en el campo de la optimización para un gran número de casos particulares. Por esta razón, se presenta en esta práctica profesional supervisada una introducción de los conceptos básicos que gobiernan el problema de la optimización de modelos.

Entre los métodos desarrollados para resolver problemas de optimización uno de los que mayormente se adapta a los problemas de optimización de modelos hidrológicos son los Algoritmos Genéticos (AG). Dado que se trabajó con un algoritmo de calibración automática basado en los AG, resulta necesario presentar una revisión general de los conceptos asociados a éstos.

El desarrollo de este capítulo está basado en la tesis de maestría del Mg. Ing Facundo José Alonso, 2008.

3.1. MODELOS HIDROLÓGICOS

Un sistema es un conjunto de elementos que interactúan entre sí. Un modelo es una representación simplificada de un sistema real. Esta representación puede hacerse en forma física (modelo físico), o bien, a través de una expresión matemática (modelo matemático) de mayor o menor complejidad. De acuerdo a la Sociedad de Investigación de Operaciones de Gran Bretaña y de Estados Unidos (Zoraida Omaña, 2006), la esencia de la Investigación de Operaciones es el uso de modelos.

Independientemente del origen o estructura del modelo, se entiende que el mismo funciona correctamente si dado un conjunto de variables de entrada, genera la misma salida que el sistema al cual representa. El proceso por el cual un modelo imita al sistema se denomina simulación. Se dice que el modelo simula un proceso que ocurre en el sistema. Un modelo captura características selectas de un sistema, proceso o realidad, y luego las combina en una representación abstracta del original.

El conjunto de cantidades fijas que caracteriza al sistema recibe el nombre de parámetros del sistema (Tucci, 2006). Los parámetros de un sistema, normalmente no pueden medirse, razón por la cual deben estimarse. Por el contrario, las variables del sistema en general son cantidades que pueden ser medidas, al menos en forma aproximada, y además, como su nombre lo indica, varían durante el proceso. Existen tres tipos de variables: las variables de entrada, las de salida y las de estado. Estas últimas, definen el estado del sistema (Bertoni, 2003).

A los fines de ejemplificar los conceptos anteriores, bastaría considerar la cuenca hidrográfica como el sistema en el cual ocurre el proceso de transformación lluvia a caudal. Este proceso, a su vez, involucra principalmente los procesos de retención, infiltración, evaporación, escurrimiento superficial y flujo en canales. El modo de ocurrencia de cada uno de estos procesos estará gobernado por ciertos parámetros de la cuenca, como son el coeficiente de rugosidad, la permeabilidad del suelo, la pendiente de la cuenca, etc.

En consecuencia, el hidrograma generado a la salida de la cuenca (salida del sistema), depende de las variables de entrada a la cuenca (entrada al sistema), que es la precipitación, y de los parámetros que caracterizan al sistema. Por esta razón, todo modelo debería involucrar en su estructura, al menos los parámetros más significativos numéricamente que caracterizan el sistema que representa. La lluvia puede ser medida y también el caudal escurrido, aunque sea en forma aproximada. Sin embargo, los parámetros no pueden medirse en forma directa. Normalmente se procede a estimar estos parámetros en función de ciertas características observables. El gran desafío en la modelación matemática de un sistema se basa en la estimación de los parámetros que lo caracterizan.

Estos parámetros estimados y las variables de entrada medidas, se agregan al modelo y se simula el proceso que ocurre en el sistema, obteniendo, como resultado, las variables de salida. Las variables simuladas de salida deberían coincidir con las medidas, al menos dentro del rango de aplicación del modelo. En general, no se logra tal coincidencia, por lo cual es necesario modificar sensiblemente los valores de los parámetros hasta lograr un resultado aceptablemente cercano al buscado. Esta técnica de ajustar los parámetros del modelo recibe el nombre de calibración del modelo matemático (Bertoni, 2003). El éxito de un modelo hidrológico conceptual de transformación lluvia-caudal depende en gran medida de la bondad de su calibración (Duan et al., 1992).

El error cuantitativo de la simulación puede medirse pues con una función que dependa de la diferencia numérica entre el hidrograma simulado y el observado. A esta función se la denomina función objetivo (FO). De la mano de los objetivos, aparecen las restricciones, que consisten básicamente en las limitaciones asociadas a la aplicación misma del modelo, normalmente representado por un intervalo de validez de las diversas variables y parámetros que intervienen en el problema. Sin embargo, existen ciertas variables que no están sujetas a restricciones.

En síntesis, las FO son expresiones matemáticas que cuantifican el error de simulación de un modelo, como la diferencia entre los valores medidos y los simulados por el modelo. El proceso de calibración de un modelo consiste en estimar los valores de los parámetros que minimiza el valor de la función objetivo, dentro del límite definido por las restricciones. Debido al papel protagónico que desempeña la función objetivo en el proceso de calibración automática de modelos matemáticos, más adelante se presenta una sección dedicada a las funciones objetivo.

3.2. SISTEMAS, CONCEPTOS Y CLASIFICACIÓN

Se entiende por sistema a toda estructura, esquema o procedimiento real o abstracto que, en un tiempo de referencia dado interrelaciona una

entrada, causa o estímulo de energía, materia o información con una salida, efecto, o respuesta de energía, materia o información (Bertoni, 2003).

Matemáticamente se puede representar un sistema como:

$$y(t) = \mu(t)\psi x(t) \quad (3-1)$$

Donde:

$y(t)$: es la función de salida

$\mu(t)$: es la función que representa el proceso

$x(t)$: es la función de entrada

ψ : vincula la función de entrada con la del proceso

Los distintos tipos de problemas derivados del análisis de sistemas se asocian principalmente a la incógnita planteada. Según este criterio, los problemas pueden clasificarse en (Bertoni, 2003):

- I. problemas de predicción, en los cuales la incógnita es la salida $y(t)$.
- II. problemas de identificación, cuya incógnita es la función que representa el proceso $m(t)$.
- III. problemas de detección, que tienen por incógnita la entrada $x(t)$.
- IV. problemas de síntesis, donde la incógnita es el sistema mismo.

El problema de síntesis es poco común y el más difícil de imaginar. En este tipo de problemas, se presenta el desafío de diseñar el sistema mismo, con el objetivo de que para una cierta entrada se obtenga una determinada salida, conociendo la función de los procesos que ocurren en el sistema. Por ejemplo, el proyectista de un sistema de drenaje urbano, que debe modelar una cuenca, con sus materiales, pendientes, etc, se enfrenta a un auténtico problema de síntesis.

Los sistemas pueden clasificarse como Lineales, cuando la salida de un proceso es lineal a la entrada, o No lineales, cuando no lo sea. Los lineales cumplen con las propiedades de proporcionalidad y superposición.

Pueden ser Concentrados o Distribuidos, de acuerdo a que las variables que intervienen en el proceso que ocurre en el sistema, no varíen en el espacio, o sí lo hagan, respectivamente.

También pueden clasificarse entre Continuos y Discretos. Siendo los primeros aquellos sistemas cuyos procesos incluyen variables de variación continua, y los segundos, cuyos procesos incluyen variables de variación discreta.

Finalmente, se habla de los sistemas Causales y No causales. Siendo los causales aquellos sistemas en los cuales el efecto ocurre en un tiempo posterior o en el mismo tiempo que la causa que le dio origen. No causales son aquellos en los cuales lo hacen en un tiempo anterior.

3.3. EL SISTEMA HIDROLÓGICO

De los innumerables sistemas que ofrece la naturaleza, en particular nos interesa el denominado: ciclo hidrológico (*Figura 3.1.*), considerado el concepto fundamental de la hidrología. De acuerdo con la clasificación mencionada anteriormente, se puede decir que el Ciclo Hidrológico, constituye un sistema de tipo no lineal, distribuido, continuo y casual. Como se verá posteriormente, las limitaciones inherentes al hombre obligan a representar el alto grado de complejidad que propone la naturaleza mediante un modelo lineal, concentrado y discreto.

Como todo ciclo, no tiene principio ni fin, y su descripción puede comenzar en cualquier punto, se elige arbitrariamente un punto de comienzo para su explicación.

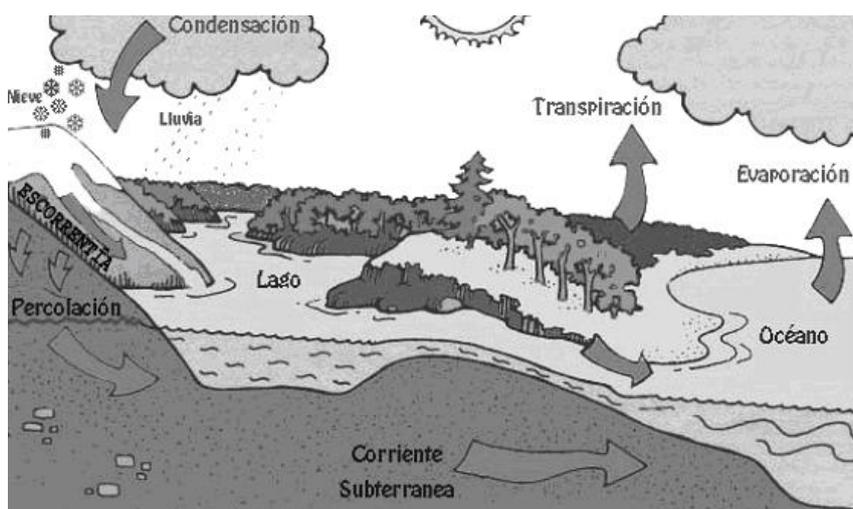


Figura 3.1. El ciclo hidrológico (CONICYT, 2007)

El agua que se encuentra sobre la superficie terrestre o muy cerca de ella se evapora bajo el efecto de la radiación solar y el viento. El vapor de agua, que así se forma, se eleva y se transporta por la atmósfera en forma de nubes hasta que se condensa y cae hacia la tierra en forma de precipitación. Durante su trayecto hacia la superficie de la tierra, el agua precipitada puede volver a evaporarse o ser interceptada por las plantas o las construcciones, luego fluye por la superficie hasta las corrientes o se infiltra. El agua interceptada, más una parte de la infiltrada y de la que corre por la superficie se evapora nuevamente. De la precipitación que llega a las corrientes, una parte se infiltra y otra llega hasta los océanos y otros grandes cuerpos de agua, como embalses y lagos. Del agua infiltrada, una fracción es absorbida por las plantas y posteriormente es transpirada, casi en su totalidad, hacia la atmósfera, y otra fluye bajo la superficie de la tierra hacia las corrientes, el mar u otros cuerpos de agua, o bien hacia zonas profundas del suelo (percolación) para ser almacenada como agua subterránea y después aflorar en manantiales, ríos o el mar. Una parte de esta agua es nuevamente evaporada, y así comienza nuevamente el ciclo hidrológico.

3.4. EL PROCESO DE TRANSFORMACIÓN LLUVIA-CAUDAL

El ciclo hidrológico descrito anteriormente, se desarrolla a nivel global, alimentado por una innumerable cantidad de subprocesos que ocurren en forma simultánea en todo el planeta. En la ingeniería hidrológica, rara vez interesa un estudio global del ciclo hidrológico. En cambio tiene mucha aplicación el estudio hidrológico de cuencas, dentro de las cuales ocurre el proceso de transformación lluvia-caudal. En otras palabras, la ingeniería hidrológica, prefiere estudiar el proceso por el cual el agua de lluvia que cae sobre la cuenca, se mueve sobre y dentro de ella, generando hidrogramas de flujo subterráneo y superficial, cuyos volúmenes y picos totales interesa conocer.

El escurrimiento se define como el agua proveniente de la precipitación que circula sobre o bajo la superficie terrestre y que llega a una corriente para finalmente ser drenada hasta la salida de la cuenca. El agua proveniente de la precipitación que llega hasta la superficie terrestre, una vez

que una parte ha sido interceptada y evaporada, sigue diversos caminos hasta llegar a la salida de la cuenca. Conviene dividir estos caminos en tres clases: escurrimiento superficial, escurrimiento subsuperficial y escurrimiento subterráneo. La Figura 3.2. ilustra el proceso.

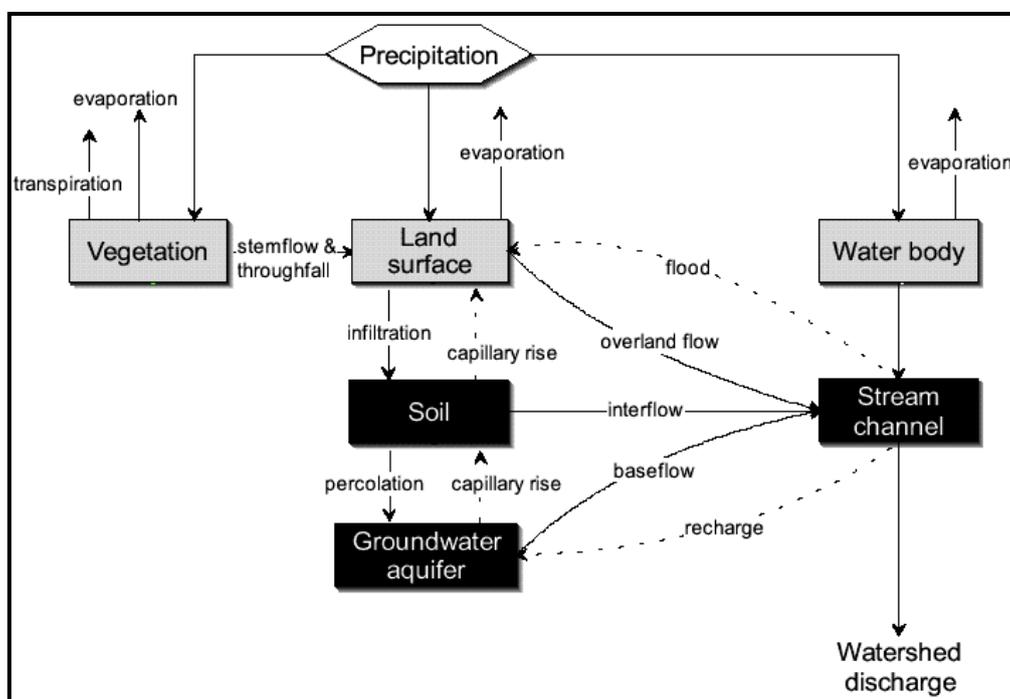


Figura 3.2.: Modelo hidrológico de transformación lluvia-caudal (HEC-HMS, 2000)

Una vez que la precipitación alcanza la superficie del suelo, se infiltra hasta que las capas superiores del mismo se saturan. Posteriormente, se comienzan a llenar las depresiones del terreno y, al mismo tiempo, el agua comienza a escurrir sobre su superficie. Este escurrimiento, llamado flujo en la superficie del terreno, se produce mientras el agua no llegue a cauces bien definidos (se entiende por cauces definidos a aquellos que no desaparecen entre dos tormentas sucesivas). En su trayectoria hacia la corriente más próxima, el agua que fluye sobre el terreno se sigue infiltrando, e incluso se evapora en pequeñas cantidades. Una vez que llega a un cauce bien definido se convierte en escurrimiento en corrientes.

El flujo sobre el terreno, junto con el escurrimiento en corrientes, forma el escurrimiento superficial. Una parte del agua de precipitación que se infiltra escurre cerca de la superficie del suelo y más o menos paralelamente a él. A esta parte del escurrimiento se le llama escurrimiento subsuperficial; la

otra parte, que se infiltra hasta niveles inferiores al freático, se denomina escurrimiento subterráneo.

De los tres tipos de escurrimiento, el superficial es el que llega más rápido hasta la salida de la cuenca. Por ello está relacionado directamente con una tormenta particular y entonces se dice que proviene de la precipitación en exceso o efectiva y que constituye el escurrimiento directo. El escurrimiento subterráneo es el que de manera más lenta llega hasta la salida de la cuenca (puede tardar años en llegar), y, en general, difícilmente se le puede relacionar con una tormenta particular, a menos que la cuenca sea demasiado pequeña y su suelo muy permeable. Debido a que se produce bajo el nivel freático, es el único que alimenta a las corrientes cuando no hay lluvias y por eso se dice que forma el escurrimiento base.

El escurrimiento subsuperficial puede ser casi tan rápido como el superficial o casi tan lento como el subterráneo, dependiendo de la permeabilidad de los estratos superiores del suelo; por ello es difícil distinguirlo de los otros dos. Cuando es relativamente rápido se le trata junto con el escurrimiento superficial, y cuando es relativamente lento se le considera parte del subterráneo.

La clasificación anterior, aunque ilustrativa, no deja de ser arbitraria. El agua puede comenzar su viaje hacia la corriente como flujo superficial e infiltrarse en el camino, terminando como escurrimiento subsuperficial o subterráneo. A la inversa, el escurrimiento subsuperficial puede emerger a la superficie si en su camino se encuentra con un estrato muy permeable que aflora en una ladera (vertiente). Lo importante en realidad es la rapidez con que una cuenca responde a una tormenta, pues esto es lo que determina la magnitud de las correspondientes avenidas.

3.5. MODELOS, CONCEPTOS Y CLASIFICACIÓN

Un modelo es una representación simplificada de un sistema real. Existen básicamente tres tipos de modelos, según su naturaleza. El modelo físico, que es una maqueta a escala que se construye a los fines de representar físicamente el fenómeno o proceso. El modelo analógico, es la

representación de un sistema físico (prototipo), por un sistema físico análogo, de modo que su comportamiento simule de forma aproximada (o exacta) el del prototipo. Un ejemplo de modelos analógico típico de la hidrología, es la simulación del flujo de agua en un cauce (o a través de un medio poroso), mediante el flujo de una corriente eléctrica en un circuito compuesto por resistencias y condensadores. Finalmente, el modelo matemático, consiste en un sistema de ecuaciones o formulaciones matemáticas que representan el sistema en estudio. Estos últimos, se han convertido en una poderosa herramienta de trabajo, fundamentalmente desde el vertiginoso desarrollo computacional que se comenzó a verificar en las últimas décadas.

Algunas de las ventajas principales asociadas al uso de modelos matemáticos son:

- ✓ Permite estudiar y experimentar en las complejas relaciones e interacciones que existen y ocurren en el interior de un sistema real.

- ✓ Permite experimentar sobre un sistema que puede no existir en la realidad (de otra manera sería imposible) y en este caso asiste a la decisión de construirlo o no.

- ✓ Se ahorra tiempo y dinero, y se evitan riesgos.

- ✓ Se logra independencia de la duración real del evento. Por ejemplo se puede simular procesos muy largos como muy cortos. Por ejemplo, se simula en pocos minutos una crecida ocurrida sobre una cuenca, que en la realidad podría tardar días en manifestarse.

- ✓ Se puede estudiar cómo reacciona el sistema al introducir variantes, sin correr riesgos (estudio de sensibilidad).

- ✓ Se puede aplicar para verificar soluciones analíticas (por ejemplo, mediante la metodología denominada de Montecarlo). Ofrece mayor facilidad para estudiar sistemas dinámicos y no determinísticos.

Con respecto a este tema, cabe aclarar que el análisis cuantitativo que resulta de la aplicación de un modelo matemático, no reemplaza, sino que complementa al cualitativo, producto del sentido común y la experiencia del modelador. Los modelos cuantitativos auxilian, pero no sustituyen a los encargados de tomar decisiones.

Un modelo matemático representa a un sistema de modo simplificado. En dicha simplificación que se hace de la realidad, se conservan

algunas características propias del sistema, y evidentemente, otras, se pierden. Estas características pueden ser, por ejemplo, la linealidad, su propiedad de ser concentrado o distribuido, etc. En base a esto, podría pensarse para los modelos, una clasificación semejante a la que se realizó con los sistemas.

Sin embargo, no necesariamente un sistema de ciertas características es representado con un modelo de las mismas características. En efecto, por definición un modelo matemático es más simple que el sistema real al cual representa. La función que cumple el modelo, es convertirse en una herramienta que represente al sistema real pero del modo más simple posible. Esta simplificación, que facilita la modelación, se realiza a expensas de sacrificar algunas propiedades del sistema. Así, por ejemplo, un sistema no lineal puede representarse mediante un modelo lineal. Tal es el típico caso del hidrograma unitario. El sistema genera una transformación lluvia-caudal no lineal, sin embargo el modelo del hidrograma unitario si lo es. Otro de los muchos ejemplos que podrían citarse es el modelo de las ecuaciones de Saint Venant, que representa el flujo a superficie libre en una formulación matemática unidimensional, aunque el flujo a superficie libre, naturalmente, es tridimensional. Por último, resulta ilustrativo citar el caso de la lluvia. En el sistema real, la lluvia cae en forma continua, sin embargo, en un modelo, la lluvia es incorporada como una variable de entrada que se computa en forma discreta, ya que la misma se mide a intervalos de tiempo mediante los pluviógrafos.

De acuerdo a la función que cumple el modelo, y a la filosofía que subyace el desarrollo del mismo, el grupo EURYDICE 92 (Chocat, 1997), los clasifica del siguiente modo:

- a. Según el propósito del modelo:
 - ✓ Modelos Cognitivos
 - ✓ Modelos Previsionales
 - ✓ Modelos de toma de decisión
 - ✓ Modelos normativos
- b. Según la concepción planteada en su formulación:
 - ✓ Modelos axiomático o heurístico
 - ✓ Modelos determinísticos o estadísticos (o estocástico)
 - ✓ Modelos reduccionista u holístico

- ✓ Modelos empírico o conceptuales
- ✓ Modelos global o distribuido

Los *Modelos Cognitivos*, tienen por objeto dar una explicación del funcionamiento o estado del sistema. Dentro de los modelos cognitivos, pueden diferenciarse dos grupos, que son los *modelos explicativos* y los *descriptivos*, de acuerdo a que permitan o no, respectivamente, comprender la influencia de las variables de entrada sobre las de salida. Por ejemplo, en la representación de las pérdidas por infiltración, el modelo de Richard, es mucho más explicativo que el modelo de Green y Ampt o que el de Horton. Pero, a su vez, estos últimos son más explicativos que el modelo del “índice Phi”. Es importante destacar que el mayor número de los modelos empleados en hidrología son de tipo Cognitivo.

Los *Modelos Provisionales*, tienen por objeto prever y simular el estado de funcionamiento de un sistema. Los *Modelos de Toma de Decisión*, se emplean para determinar las variables de entrada (o de concepción) a proveer al modelo para obtener las variables de salida conocidas y previamente impuestas. Los *Modelos Normativos*, tienen por objeto establecer las normas que relacionan la entrada y la salida. Formulan las relaciones que deben existir para que, con cierta entrada, se obtenga una salida también conocida. Se diferencian de los sistemas cognitivos en que estos modelos no analizan el proceso.

Los *modelos axiomáticos* son aquellos que se desarrollan partiendo rigurosamente de los principios y reglas que dominan el sistema, basados en hipótesis sobre el mismo. Los *modelos heurísticos* plantean principios que reflejan parcialmente el comportamiento del sistema que luego articulan en forma aproximada para obtener el modelo.

Un *modelo determinístico* es aquel que produce siempre una misma salida para una cierta entrada. Es decir que no presentan aleatoriedad sus variables (Chow, 1994). Un modelo estadístico, es aquel en el cual las salidas son aleatorias. Cuando en un modelo la probabilidad de ocurrencia de sus variables es tenida en cuenta, o bien, el concepto de probabilidad es introducido en el modelo mismo, se dice que el modelo es de naturaleza estadística (Clarke, 1973). Cuando el proceso evoluciona en el espacio y en el

tiempo de un modo que es parcialmente predecible y parcialmente aleatorio se denomina estocástico (Chow, 1994), entonces el modelo que lo representa manifiesta una estructura parcialmente determinística y parcialmente aleatoria y se denomina también estocástico.

La mayoría de los procesos hidrológicos son de tipo estocásticos, y dan nacimiento a una rama importante de la hidrología denominada Hidrología Estocástica. Cuando una variable de salida de un modelo hidrológico tiene un grado de aleatoriedad, un problema de rutina para la hidrología estocástica es identificar en qué grado influye su componente aleatorio y en qué grado su determinístico (Chow, 1994). Si la influencia del componente probabilístico no es significativa, es apropiado un modelo de tipo determinístico; pero si tiene importancia significativa, será necesario un modelo de tipo estocástico, pues de lo contrario, la salida real será muy distinta a la salida única que arrojaría el modelo determinístico.

Es importante destacar que, un modelo de tipo determinístico, puede tener variables de entrada o salida de tipo aleatorias, y que un modelo estocástico, puede tener variables determinísticas. Un ejemplo de esto, es un modelo determinístico que arroja para cada valor de entrada P (lluvia), su correspondiente valor de salida Q (caudal). Aún cuando P y Q son variables aleatorias, el modelo de transformación lluvia-caudal es determinístico.

El enfoque Reduccionista (también llamado microscópico, mecanicista o de caja blanca), incluye aquellos modelos que se desarrollan a partir de descomponer previamente el sistema real en muchos microprocesos, y luego se formula el modelo teniéndolos a todos en cuenta. En esencia se trata de un proceso de diferenciación e integración. Por ejemplo, las ecuaciones de Saint Venant. Mientras que el enfoque Holístico (o macroscópico, o de caja gris, o conceptual), consideran el sistema en su conjunto, y se desarrolla el modelo basado en el comportamiento global del mismo. Por ejemplo, el modelo de Muskingum.

Los *modelos Empíricos* (o de caja negra) se basan en la experimentación. Pueden no guardar relación alguna con los fenómenos físicos y conceptuales del proceso analizado. El único criterio de formulación consiste en reducir las diferencias entre las variables observadas y las calculadas. Los modelos conceptuales (o de enfoque teórico), en cambio, se apoyan sobre el

concepto de abstracción y el de razonamiento. En este caso el modelo es construido a priori, sobre un análisis teórico, y la experimentación posterior contribuye a validar el modelo.

Los modelos concentrados (o globales, o puntuales, o de orden cero), no consideran la variación espacial de los parámetros y las variables. Por ejemplo el método de la Curva Número de infiltración del SCS. Los modelos distribuidos (o de malla, o matriciales, o de discretización espacial) consideran la heterogeneidad del cuadro físico modelado. Pero siempre es necesario realizar alguna discretización sobre el dominio para poder analizar la variación dentro del campo de estudio. Esta discretización acota su concepción de distribuidos.

3.6. EL MODELO LLUVIA-CAUDAL

Tal como ha sido indicado anteriormente, el proceso de transformación lluvia caudal, es sumamente complejo y requiere de una abstracción a través de un modelo que simplifique suficientemente el sistema como para poder simular conceptualmente sus partes.

Dentro del proceso de transformación lluvia-caudal pueden identificarse dos procesos claramente independientes desde el punto de vista conceptual, como son el proceso de producción y el de transferencia. El primero, está asociado al mecanismo por el cual sólo una parte de la precipitación llega en forma rápida a la salida de la cuenca, ya que el resto queda retenido, se evapora o se infiltra y llega a la salida de la cuenca, pero ya como flujo subterráneo. El segundo, está relacionado con el recorrido que el agua realiza desde los distintos puntos de la cuenca hasta la salida de la misma.

En síntesis, el componente de producción en un modelo de transformación lluvia-caudal se encarga de estimar el porcentaje del volumen total caído sobre la cuenca como precipitación que se convierte en caudal a la salida de la misma. Por su parte, el componente de transferencia se ocupa de establecer la forma en la cual llega repartido en el tiempo ese volumen a la

salida de la cuenca, obteniendo así, el hidrograma de escurrimiento directo en la sección de salida de la cuenca.

En cuencas intermedias y grandes, para realizar el análisis del proceso de transformación lluvia caudal, normalmente, se subdividen en subcuencas. Entonces, se determina el hidrograma de salida de cada subcuenca y luego, se transita la creciente hasta las salida de la cuenca. Este proceso se denomina tránsito de creciente. Esto es, el proceso por el cual el caudal de agua a la salida de la subcuenca viaja hacia aguas abajo del río, alterándose el hidrograma de una sección del mismo, a la sección siguiente. Para este tipo de casos se implementan los modelos de tránsito de crecientes. Los mismos tienen por objetivo calcular el hidrograma en una sección del río, conociendo el hidrograma en otra sección, aguas arriba del mismo.

Hasta la actualidad, se ha desarrollado un gran número de modelos aplicados a estos procesos, barriendo un campo que abarca desde modelos simples de producción o transferencias a modelos más complejos que comprenden ambos procesos, más el de tránsito de crecidas. En esta práctica se utilizó el modelo Génie Rural (GR) de Charles Perrin et al, 2007.

3.7. OPTIMIZACIÓN – CONCEPTOS GENERALES

La optimización de modelos matemáticos, tradicionalmente ha sido tomada como una de las etapas de un proceso más amplio, que se ha dado en llamar la Investigación de Operaciones, la cual se refiere a la conducción y coordinación de operaciones (o actividades) dentro de una organización. Actualmente, el campo de la optimización ha crecido tanto que puede tratarse como un campo de investigación en sí mismo, que cumple la función de desarrollar herramientas que sean útiles en un proceso más complejo, que le otorga utilidad real.

El proceso involucrado en la Investigación de Operaciones comienza por la observación cuidadosa y la formulación del problema incluyendo la recolección de los datos pertinentes. El siguiente paso es la construcción de un modelo científico (por lo general matemático) que intenta abstraer la esencia del problema real. En este punto se propone la hipótesis de que el modelo es

una representación lo suficientemente precisa de las características esenciales de la situación como para que las conclusiones (soluciones) obtenidas sean válidas también para el problema real. Después, se llevan a cabo los experimentos adecuados para probar esta hipótesis, modificarla si es necesario y eventualmente verificarla. Con frecuencia este paso se conoce como validación del modelo. A continuación se analizan alternativas de decisión, y se procura seleccionar la solución óptima al problema planteado a partir de un criterio de jerarquía previamente establecido. Finalmente, se realiza la toma de decisión e implementación de las medidas adoptadas, en función de los resultados de la modelación.

Resumidamente casi todos los autores (Castillo et al., 2002; Zoraida Omaña, 2006) parecen coincidir con estas etapas, cuyo orden no es necesariamente el mencionado arriba, sino que las etapas se van desarrollando en paralelo y retroalimentándose unas con otras. Conviene remarcar que una de las características principales de la investigación de operaciones es la de intentar encontrar una mejor solución (llamada solución óptima) para el problema bajo consideración. Se dice una mejor solución, y no la mejor solución, porque pueden existir muchas soluciones que empaten como la mejor.

3.7.1. Métodos de optimización

Las metodologías de optimización nacen de la mano de la Investigación de Operaciones, por lo cual tienen un origen común como ciencia aplicada. La mayoría de los textos introductorias a la Investigación de Operaciones describen un desarrollo histórico parecido de esta ciencia. Información más detallada se puede encontrar en literatura especializada (Alonso, 2008)

A los fines de este informe interesa señalar los avances que se han logrado desde la década de los sesenta, en cuanto a los conocidos como métodos heurísticos. Estos se aplican a aquellos problemas de optimización que no pueden ser abordados por métodos exactos, ya sea, por su alto grado combinatorio o por la dificultad de generar un modelo basado en programación matemática que represente exactamente una situación real. En los primeros tiempos se generaron métodos orientados específicamente a la resolución de

cada problema, pero a partir de la década del 80, se han generado una familia de métodos conocidos como meta-heurísticos que ahora tienen la capacidad de ser aplicables a problemas de diversa naturaleza. Es decir, una misma plantilla algorítmica puede ser utilizada para resolver problemas que provienen de diversos sectores.

Los métodos meta-heurísticos más conocidos son: Búsqueda Tabú, Simulated Annealing, Algoritmos Genéticos, Redes Neuronales, Métodos basados en Trayectoria de Hormiga y Métodos basados en Inteligencia Artificial. En este informe, sólo se profundiza el método de los Algoritmos Genéticos (AG). Una explicación detallada de cualquiera de estos métodos, pueden encontrarse en la literatura especializada (Palahí y Pukkala, 2004; Kuri y Galaviz, 2002, Viveros et al., 1998; USACH, 2007, entre otros).

El formato general del problema de optimización abordable mediante métodos heurísticos es el siguiente: La función $f(x)$, es una función matemática que se desea minimizar y las soluciones posibles deben pertenecer a un espacio W , los métodos de búsqueda heurística recorren el espacio W tratando de identificar la solución que genera el mejor valor para $f(x)$, el menor en la caso de minimización o el mayor en el caso de maximización.

3.7.2. Generación de números aleatorios

Los métodos heurísticos constituyen la propuesta de los nuevos paradigmas computacionales de optimización, entre los cuales se encuentran los métodos de los algoritmos genéticos. La mayoría de estos métodos tienen entre sus rutinas la generación de números aleatorios con distintas distribuciones probabilísticas (Goldberg, 1989, Duan et al., 1992, Yapo et al., 1998, Vrugt et al., 2003, Viveros, et al. 1998). Se presentarán en este informe solamente los métodos clásicos de generación de números aleatorios con distribución uniforme. Para generar números con distintas distribuciones probabilísticas, se recurre al método de Montecarlo, desarrollado principalmente por Stan Ulam y John Von Neumann, por el año 1944, cuya presentación formal puede encontrarse, entre otros, en Eckhardt (1987) y Herring y Palmore (1989).

La generación de variables aleatorias tiene un rol preponderante en el proceso de simulación. Para generar variables aleatorias es necesario contar con un generador de números aleatorios. Para que un número sea considerado aleatorio, debe tener las siguientes características (UROU, 2006): 1) Distribución Uniforme. Lo que significa que cualquier número que pertenezca al rango de interés debe tener la misma chance de resultar sorteado. 2) No exista Correlación Serial. Lo que significa que la aparición de un número no afecta de ninguna manera la chance de aparecer de otro número o de sí mismo. Existen pruebas para comprobar que un generador cumpla estas dos condiciones.

La forma más perfecta de generación de números aleatorios es a través de algún modelo físico, como por ejemplo la ruleta, que es el modelo por excelencia. También pueden conseguirse números aleatorios a través de tablas que han sido construidas con modelos físicos (en 1955 la compañía RAND publicó una tabla con un millón de números aleatorios). Matemáticamente, se puede mencionar dos formas diferentes de generar números aleatorios:

I. Métodos de generación aleatoria. En base a algún circuito o característica de la máquina se generan números que son totalmente aleatorios. El método básicamente consiste en interrumpir un proceso uniforme en forma aleatoria (que es esencialmente lo que ocurre cuando la bola cae en un casillero de la ruleta).

II. Números pseudoaleatorios. Se generan a través de una fórmula e imitan los valores de una variable aleatoria. Se llaman pseudoaleatorios porque se obtienen realizando un conjunto de operaciones sobre el número generado antes (recurrencia) por lo que no son realmente aleatorios según lo expuesto previamente. A pesar de esto muchos de los métodos de generación de números pseudoaleatorios se comportan correctamente y pasan todos los tests como si fueran una variable aleatoria. Presenta la gran ventaja de ser un método muy veloz y barato y la mayor desventaja es que son de período finito.

En 1950, Von Neumann propuso el conocido Método de Centros de Cuadrados. El mismo, se basa en tomar un número, elevarlo al cuadrado y tomar los dígitos del centro como nuevo número, luego repetir el procedimiento. Por ejemplo, el número de 4 dígitos: 2061, elevado al cuadrado resulta en

4247721, del cual los cuatro números centrales son: 2477, que elevado al cuadrado resulta en 6135529, y así sucesivamente. La desventaja es que la secuencia generada por lo general es corta. El ejemplo anterior luego de 34 números degenera a 0; y si en lugar de empezar con 2061 se inicia con 2500, se repite 2500 en la primer iteración (y por lo tanto, en todas las restantes). Aún así si se toman números muy largo se puede llegar a secuencias de 100.000 números diferentes.

En 1951, Lehmer propuso el Método Congruencial Lineal, que se ha convertido en el generador pseudo-aleatorio estándar (Herring y Palmore, 1989), ya que es el método utilizado por excelencia, y se basa en la siguiente recurrencia (Eckhardt, 1987):

$$Z_i = (A * Z_{i-1} + C) \bmod M \quad (3-2)$$

Donde: Z_i es el i -ésimo número de la secuencia. A , C y M son parámetros que pueden ser ajustados por conveniencia y para asegurar la naturaleza pseudo-aleatoria de la secuencia. Por ejemplo, el módulo M , es normalmente tomado como la longitud de la palabra de la computadora, y el multiplicador A , se elige para producir tanto un periodo largo de la secuencia pseudo-aleatoria, como para asegurar buenas propiedades estadísticas. La expresión $y = z \pmod{m}$, significa que existe un entero k tal que $mk = y - z$.

La ecuación 3-2, puede ser llevada sencillamente a la forma:

$$Z_i = A * Z_{i-1} + C - [(A * Z_{i-1} + C) / M] * M \quad (3-3)$$

Donde la notación: $[x]$ significa el mayor entero menor o igual a x , y se puede demostrar por inducción completa (UROU, 2006) que:

$$Z_n = [Z_0 * A^n + (A^n - 1) / (A - 1) * C] \bmod M \quad (3-4)$$

Si se quieren obtener número Uniformes $(0,1)$ se normaliza el resultado dividiendo por el módulo: $U_i = Z_i / M$.

A continuación se presentan dos ejemplos:

a) Para los valores: $A = 3$, $C = 0$, $M = 5$ y $Z_0 = 4$; se obtiene la serie:
2,1,3,4,2,1,...

b) Para los valores: $A = 3$, $C = 0$, $M = 9$ y $Z_0 = 4$; se obtiene la serie:
3,0,0,...

Se desprende de estos ejemplos que, en este método, si se repite un número ya se repite toda la secuencia. Presenta la ventaja de consumir

poca memoria y ser tener gran velocidad de cálculo. Además, es muy fácil volver a generar la misma secuencia guardando un solo número, ya que alcanza con partir de la misma semilla Z_0 .

Es muy importante la elección de las constantes A , C , M y la semilla Z_0 ya sea para la velocidad de generación como sobre todo para el largo de la secuencia. En la práctica para aumentar la velocidad de cálculo se suele tomar $M=ab$ donde a es el número base de operación de la máquina, por lo que en máquinas binarias $a=2$, esto permite calcular muy fácilmente el resto y efectuar la normalización a $U(0,1)$ operando con desplazamientos y no con divisiones y multiplicaciones.

Con respecto al largo del período de la secuencia siempre será menor o igual a M y es deseable obtener el largo máximo (o sea M) aunque no siempre es posible. Cuando se obtiene largo máximo se dice que el método es de Período Completo.

Se distinguen dos casos según el valor de C :

- ✓ $C = 0$, se llama Método Congruencial Multiplicativo Puro.
- ✓ $C \neq 0$, se llama Método Congruencial Multiplicativo Mixto.

3.8. ALGORITMOS GENÉTICOS

Los algoritmos genéticos (AG) fueron desarrollados por John Holland, sus colegas, y sus estudiantes en la Universidad de Michigan. Los principales logros de sus investigaciones fueron: (1) lograr una explicación abstracta y rigurosa de los procesos de adaptación de los sistemas naturales, y (2) diseñar sistemas artificiales que retengan los mecanismos importantes de los sistemas naturales. El foco principal de investigación de los AG ha sido la robustez, y el balance entre la eficiencia y la eficacia necesaria para sobrevivir en diferentes ambientes.

Según establece Goldberg (1989), los algoritmos genéticos son algoritmos de búsqueda basados en los mecanismos de selección natural y genética natural. Estos combinan, continúa Goldberg (1989), la supervivencia del más apto entre estructuras de cadenas (que representan distintos individuos) con alguna otra estructura de información cambiante de manera

aleatoria, para formar un algoritmo de búsqueda con ciertos instintos de búsqueda innovadores respecto de la búsqueda humana.

En cada generación, se crea un nuevo juego de individuos artificiales (cadenas) que emplean algunos elementos y piezas que eligen de entre los más aptos de los antiguos individuos, probando ocasionalmente alguna pieza nueva que pudiera resultar buena. Aunque es una técnica aleatoria, aclara Goldberg (1989), los AG no son un simple camino aleatorio, sino que exploran eficientemente toda la información histórica para especular en el siguiente punto de búsqueda con una performance mejorada.

3.8.1. Conceptos básicos de biología

A continuación se abordan algunos conceptos básicos de la genética que son estrictamente necesarios para situar al lector en el contexto preciso en el cual se gestan los algoritmos genéticos. El texto presentado posee limitaciones, ya que de ninguna manera se pretende en esta breve sección describir completamente el complejísimo y riquísimo misterio de la vida y consecución de las especies vivientes.

El cuerpo humano está formado por tejidos, los cuales se componen de millones de células. Cada célula del cuerpo humano contiene toda la información genética de esa persona. Hay dos tipos de células básicamente: las somáticas, que son células diploides, denominadas así porque poseen la dotación completa del número de cromosomas de la especie; y los gametos o células sexuales, que son células haploides, porque poseen la mitad de la dotación. En el caso del ser humano, las células diploides poseen 23 pares de cromosomas, mientras que las haploides sólo 23 cromosomas. Un cromosoma no es otra cosa que una larga molécula de ADN en su máximo estado de empaquetamiento (o compresión). El ADN es una molécula muy compleja de la cual -a los fines de esta tesis- sólo interesa destacar que allí está codificada la información genética de la persona. Cada cromosoma posee innumerables características de la persona; entre los 23 cromosomas incluyen la totalidad de las características de la persona (Alberts et al., 2002).

En las células diploides, cada par de cromosomas, está formado por uno heredado de la madre y otro heredado del padre. En cada par de

cromosomas, la información de ambos es respecto de la misma característica, sin embargo, esta información puede ser diferente entre sí. Por ejemplo, si en un par de cromosomas está almacenado el color de ojos de la persona (entre otras características), entonces uno de los cromosomas puede indicar color azul (por ej. el de la madre), y el otro negro (el del padre). De la confrontación entre la información que tiene cada cromosoma respecto de cada característica, surgen las características dominantes y las recesivas. Siguiendo con el ejemplo, si se confronta el color de ojos azules que aporta la madre, con el negro del padre, triunfará el negro por orden de jerarquía. Es importante aclarar que el hecho de que el padre haya aportado negro no significa que tenga ojos negros, ni la madre azules.

La molécula de ADN (Ácido Desoxirribonucleico) está formada por cuatro distintos compuestos más simples llamados bases o nucleótidos: adenina (A), guanina (G), citosina (C) y timina (T). Cada subcadena de tres nucleótidos codifica un aminoácido diferente. A las subcadenas de tres nucleótidos se les llama codones. Al unirse varios codones, forman una proteína. Al conjunto de nucleótidos que codifican una proteína completa se les llama genes. El valor que posee un gen determinado se denomina alelo (Alberts et al., 2002). Por ejemplo, un gen puede tener el valor “azul” para la característica “color de ojos”.

Cuando en un par de cromosomas de una célula diploide, ambos tienen el mismo alelo se dice que son homocigos, si no, son heterocigos, y en este último caso sólo uno de los alelos se manifestará en el fenotipo. El alelo que se manifiesta se denomina dominante, y su homólogo que no se manifiesta, se llama recesivo. El conjunto de todos los cromosomas, es decir, toda la información genética de un individuo se llama genoma y el conjunto de genes contenidos en el genoma, genotipo. Es éste el que determina, en buena medida, el fenotipo del individuo.

En síntesis, el fenotipo de cada individuo está determinado por las proteínas que produce, y esto a su vez está definido en la información genética de cada una de sus células.

La reproducción comienza con el encuentro entre el óvulo y el espermatozoide. Ambos son células de tipo haploides, por lo tanto poseen solamente 23 cromosomas cada una de ellas. Al unirse en la fecundación, se

fusionan el núcleo del espermatozoide con el núcleo del óvulo. De este modo, se obtiene una nueva célula denominada huevo o cigota, de tipo diploide, ya que posee los 23 cromosomas de la madre y los 23 del padre.

Esa célula es la primera expresión de vida y se denomina embrión. En ese momento se ha producido la fecundación. La célula comenzará a efectuar el proceso de mitosis, que es el fenómeno por el cual a partir de una célula madre, se originan dos células hijas con la misma información genética. Resumidamente, la mitosis comienza con la duplicación de la información genética de cada cromosoma, y luego se parte en dos células que se llevan, cada una de ellas, uno de cada cromosoma duplicado, logrando así dos células exactamente iguales a la célula que le dio origen. Luego, se producirá nuevamente la mitosis en cada una de las dos células recientemente formadas, con lo cual se tendrán cuatro células iguales. En cada una de las cuatro, también ocurrirá la mitosis para volverse ocho, y luego dieciséis, y así sucesivamente. De una célula original, se habrán logrado millones de copias exactamente iguales a ella.

Este proceso de sucesivas mitosis continuará hasta que se alcance la etapa de mórula. Hasta la mórula, el embrión tiene la totalidad de las células exactamente iguales entre sí. A partir de aquí comienza la etapa de diferenciación, en la cual se comienzan a manifestar diferentes genes en cada célula.

Cuando comienza esta etapa, algunas de estas células dan origen a las gametas. Es decir, siendo diploides aún, darán origen a la formación de células haploides, a través de un proceso denominado Meiosis. La Meiosis es el proceso de división celular mediante el cual se originan, a partir de una célula madre, cuatro células hijas con la mitad del número cromosómico que la célula madre que les dio origen.

Ese tipo de células (diploides), se denomina espermatogonia (cuando se producen en el hombre) y ovogonia (mujer). Las espermatogonias (diploides) generarán los espermatozoides (haploides), y las otras, los óvulos (o más precisamente los ovocitos secundarios, ya que el término óvulo debe aplicarse cuando el ovocito ha sido fecundado). Las etapas de la Meiosis (Kuri y Galaviz, 2002) se grafican en la Figura 3.3., y son:

1. Se observa una célula diploide que dará origen a dos haploides a través del proceso de Meiosis. Esta célula tiene los 23 cromosomas del padre, y los 23 de la madre. En este esquema sólo se representa lo que ocurre en cada par de cromosomas. Así por ejemplo se puede suponer que la línea color negro representa uno de los 23 cromosomas que aporta el padre, y la verde uno de los 23 cromosomas que aporta la madre. En cada par de cromosomas se producirá lo que se describe aquí para uno sólo. Esta célula recibe el nombre de espermatogonia (u ovogonia) y dará origen a dos gametas.

2. Se produjo la duplicación del material genético en cada cromosoma.

3. Comienza el proceso de Crossing Over, en el cual hay un entrecruzamiento entre uno de los cromosomas duplicados de la madre en contacto con uno de los cromosomas duplicados del padre de cada par. Se denomina quiasma a los puntos de contacto entre los cromosomas.

4. Ya están entrecruzados. De este modo quedará cada par de cromosomas formado por dos pares de cromátidas hermanas. Una dada por un cromosoma puro del padre y otro del padre con componentes de la madre, y la otra dada por un cromosoma puro de la madre, y otro de la madre con componentes del padre.

5. Comienza a producirse una polarización dentro de la célula que atrae a los cromosomas del padre a un polo, y a los de la madre, al otro.

6. La célula se divide en dos células nuevas, cada una de las cuales posee 23 cromosomas duplicados, dando finalización a la llamada Meiosis 1.

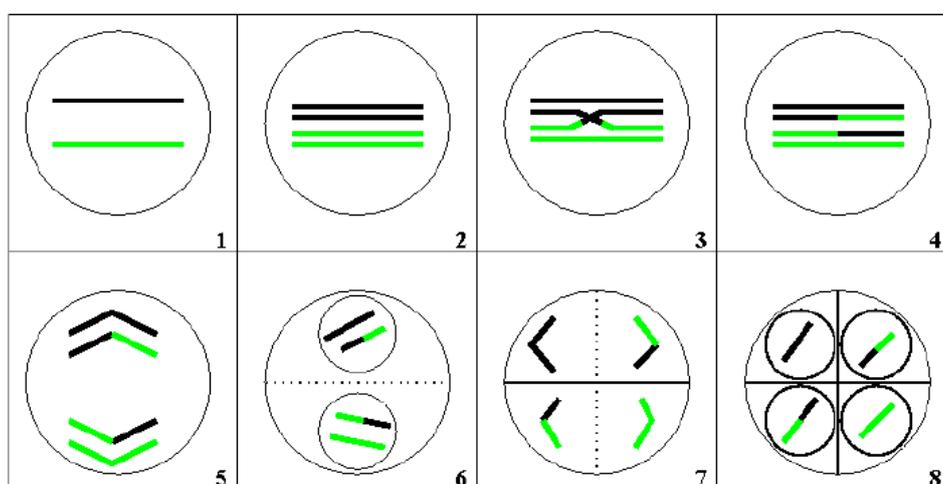


Figura 3.3.: Etapas de la meiosis (Kuri y Galaviz, 2002)

7. Se separan las cromátidas hermanas, dando origen cada célula recientemente formada, a dos nuevas células, con la mitad del número de cromosomas. Por lo que resultan cuatro células.

8. Se han formado las cuatro células hijas.

Es preciso remarcar que entre el paso uno y el dos, se ha producido una réplica de los cromosomas existentes. Para ello existe una enzima encargada de copiarlos, denominada ADN polimerasa. La molécula de ADN tiene forma de una doble hélice, como una escalera de caracol. La enzima abre por en medio los “escalones” de esta hélice y ensambla en cada mitad los nucleótidos que debe ensamblar (Figura 3.4.).

Ocasionalmente esta enzima comete un error, que puede ser causado por radiaciones energéticas externas o sustancias extrañas. La alteración de la molécula de ADN original constituye una mutación que puede manifestarse en el fenotipo y hacer al individuo diferente del resto de sus congéneres. Es muy poco probable que cambiar al azar un trozo de información que la naturaleza ha refinado cuidadosamente a lo largo de millones de años resulte en algo bueno. Por lo general las mutaciones son desfavorables, incluso letales, para el organismo mutante. Pero ocasionalmente pueden no serlo y conferirle a dicho organismo alguna ventaja que le permita sobrevivir más fácilmente en su medio. Esta característica será transmitida a sus descendientes y un pequeño paso evolutivo se habrá dado (Kuri y Galaviz, 2002).

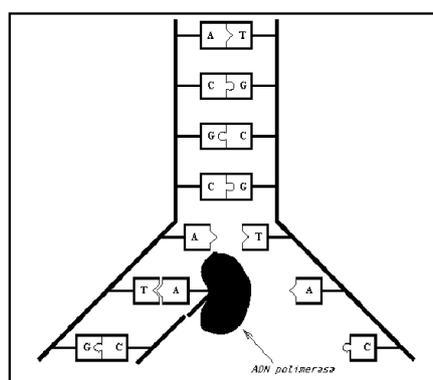


Figura 3.4.: Esquema de trabajo de la enzima ADN polimerasa (Kuti y Galaviz, 2002)

3.8.2. La técnica de los algoritmos genéticos

Los algoritmos genéticos son algoritmos computacionales inspirados en los procesos de evolución natural. Por eso, cuando se realiza la descripción de los mecanismos de optimización basados en algoritmos genéticos, se introducen muchos términos que han sido tomados de la biología genética.

A continuación, se presentan los mecanismos más destacados de los algoritmos genéticos, que tienen base en la biología, conforme fueron descritos por Kuri et al. (2002), y por Gómez Rocha et al. (2006):

Codificación del dominio: como se ha expresado anteriormente, las características de cada individuo dependen de las proteínas. Las proteínas son codificadas por los genes, que constituyen una combinación de nucleótidos. Existe una innumerable cantidad de combinaciones posibles entre los millones de nucleótidos que forman cada molécula de ADN, cada una de las cuales sería capaz de codificar una proteína distinta. En otras palabras, las características de cada individuo responden a un código genético. A lo largo de miles y millones de generaciones, ese código genético ha ido optimizando las especies para su perfecta adaptación al medio. Se puede decir que todos los posibles códigos genéticos, son posibles soluciones a la formación de un nuevo individuo. Este conjunto de todas las posibles soluciones constituye el dominio dentro del cual el código puede tomar valores.

Del mismo modo, en un algoritmo genético lo primero que se requiere es determinar en qué espacio se encuentran las posibles soluciones al problema que se pretende resolver. Es decir, es necesario codificar de alguna manera el dominio del problema para obtener estructuras manejables que puedan ser manipuladas por el AG. Cada una de estas estructuras constituye el equivalente al genotipo de un individuo en términos biológicos. El individuo del dominio del problema al que se mapea este genotipo es el análogo al fenotipo. Es frecuente que el código de los elementos del dominio del problema utilice un alfabeto binario (ceros y unos).

Una vez que se ha definido la manera de codificar los individuos del dominio del problema y se conoce la forma de pasar de un individuo a su código y viceversa, es necesario fijar un punto de partida. Los algoritmos

genéticos manipulan conjuntos de códigos en generaciones sucesivas. Nuevamente haciendo una analogía manipulan poblaciones de códigos. En éstas un código puede aparecer más de una vez. El algoritmo se encargará de favorecer la aparición en la población de códigos que correspondan a elementos del dominio que estén próximos a resolver el problema. En resumen, el algoritmo recibirá como entrada una población de códigos y a partir de ésta generará nuevas poblaciones, donde algunos códigos desaparecerán mientras que otros, que representan mejores soluciones posibles, aparecen con más frecuencia hasta que se encuentra una satisfactoria o hasta que se cumple alguna otra condición de terminación.

Adaptación: en una población, hay algunos individuos más hábiles para sobrevivir en el medio que otros. Estos individuos tienen mejor adaptación al medio (fitness) respecto del resto de los individuos. De la misma manera, en los AG existen códigos que resultan en individuos mejor adaptados para resolver el problema que se pretende optimizar. En general, este calificativo de adaptación se mide a través de un real no negativo. El objetivo de este número es que permita distinguir propuestas de solución buenas de aquéllas que no lo son. Si el problema a resolver consiste en minimizar una función, entonces tanto mejor calificado estará el individuo cuanto menor sea dicho valor.

Evidentemente, al hablar de que a cada individuo de la población se le asigna una y sólo una calificación, se está hablando de una función que se denomina función de adaptación, cuya evaluación puede no ser sencilla y es, de hecho, lo que en la mayoría de los casos consume más tiempo en la ejecución de un algoritmo genético. Hay que tener en cuenta que se evalúa una vez en cada individuo de cada generación. Si un AG es ejecutado con una población de tamaño 100 durante 100 generaciones, la función es evaluada 10.000 veces. Además, puede darse el caso de que la función de evaluación no tenga una regla de correspondencia explícita, esto es, una expresión algebraica, y puede ocurrir incluso que la función cambie de generación en generación.

Selección: una vez calificados todos los individuos de una generación, el algoritmo debe, al igual que lo hacen la naturaleza y el hombre,

seleccionar a los individuos más calificados, mejor adaptados al medio, para que tengan mayor oportunidad de reproducción. De esta forma se incrementa la probabilidad de tener individuos “buenos” (con alta calificación) en el futuro.

En principio podría parecer que es conveniente tener una estrategia de selección estricta para que mejore rápidamente la población y converja el algoritmo, es decir, que la población se acumule alrededor de un genotipo óptimo. Esto no es cierto. Lo que ocurrirá es que la población se acumulará rápidamente alrededor de algún individuo que sea bueno, comparativamente con el resto de los individuos considerados a lo largo de la ejecución del algoritmo, pero este individuo puede no ser el mejor posible. A esto se le suele llamar convergencia prematura. Se habrá obtenido un óptimo local. Por esta razón, los algoritmos genéticos están diseñados para fomentar la supervivencia de los individuos aptos, pero no elimina la posibilidad de supervivencia y procreación por parte de los malos. Esta característica permite que se incursione una nueva región del dominio que podría converger hacia un óptimo mejor que los anteriores. En la naturaleza ocurre igual. Los individuos menos favorecidos tienen menor probabilidad de sobrevivir y procrear, pero nada impide que ocurra cada tanto e impongan sus genes, en principio, desfavorables. En los algoritmos genéticos debe buscarse en todo momento la variabilidad de individuos. Mientras que se tiende al óptimo debe permitirse que sobrevivan varios malos o poco aptos.

Cruzamiento: durante la meiosis ocurre el proceso de producción de gametos. El código genético de los padres de un individuo se mezcla para producir gametos cuyo contenido genético es híbrido, es decir, una mezcla. De esta manera es posible que un individuo herede a sus descendientes las características mezcladas de sus propios padres, por ejemplo: el color de ojos del padre y el de cabello de la madre. La cruce de los códigos genéticos de individuos exitosos favorece la aparición de nuevos individuos que hereden de sus ancestros características deseables.

En el contexto de los algoritmos genéticos reproducirse significa que, dados dos individuos seleccionados en función de su grado de adaptación, éstos pasen a formar parte de la siguiente generación o, al menos, mezclen sus códigos genéticos para generar hijos que posean un código híbrido. Es

decir, los códigos genéticos de los individuos se cruzan. Existen muchos mecanismos de cruzamiento.

Mutación: algunas veces, muy pocas de hecho, la ADN-polimerasa (la enzima encargada de replicar el código genético), se equivoca y produce una mutación, una alteración accidental en el código genético de los seres vivos. Ocasionalmente algunos elementos del código de ciertos individuos de un algoritmo genético se alteran a propósito. Éstos se seleccionan aleatoriamente en lo que constituye el símil de una mutación. El objetivo es generar nuevos individuos, que exploren regiones del dominio del problema que probablemente no se han visitado aún. Esta exploración no presupone conocimiento alguno, no es sesgada.

Aleatoriamente se buscan nuevas soluciones posibles que quizá superen las encontradas hasta el momento. Esta es una de las características que hacen aplicables los algoritmos genéticos a gran variedad de problemas: no presuponer conocimiento previo acerca del problema a resolver ni de su dominio, no sólo en la mutación sino en el proceso total. De hecho, el problema a resolver sólo determina la función de evaluación y la manera de codificar las soluciones posibles (la semántica de los códigos genéticos de los individuos). El resto de los subprocesos que constituyen el algoritmo son independientes y universalmente aplicables.

3.9. TÉCNICAS DE CALIBRACIÓN

Una de las etapas fundamentales para la aplicación de un modelo matemático de tipo lluvia-caudal a una cuenca dada es la calibración de sus parámetros tomando como información de ajuste los datos de caudales observados. Esta calibración es realizada modificando los valores de los parámetros hasta que el hidrograma calculado reproduzca al hidrograma observado con buena precisión. Sin embargo, un ajuste cuantitativo del modelo, dado por una minimización aceptable de la FO, no basta para considerar buena la simulación, sino que es necesario agregar la contribución del método cualitativo. Este último consiste básicamente en una evaluación

visual de los resultados, que se presentan usualmente en gráficos. En esta evaluación visual del ajuste, se presta especial atención a la parte del hidrograma que presenta mayor interés para el propósito del modelo, como puede ser el pico o la curva de recesión del hidrograma, o por ejemplo, el volumen total escurrido.

Sumado al ajuste cualitativo-cuantitativo, resulta necesaria una interpretación de los valores obtenidos para los parámetros por parte de un hidrólogo experimentado. Esto se debe a que no existe un conjunto único de parámetros capaz de representar todos los procesos hidrológicos involucrados en la cuenca, debido a las incertidumbres inherentes de los datos, a las simplificaciones introducidas en el desarrollo de los modelos y a la limitada representatividad de los parámetros involucrados (Collischonn et al., 2001). Existen varias soluciones posibles que, de forma adecuada darán la respuesta deseada (Beven y Binley, 1992), dentro del concepto de equifinalidad. Es decir, existen varios juegos de parámetros que satisfacen las ecuaciones planteadas, debido a que el número de incógnitas supera al de las ecuaciones. Sin embargo, solo algunos de esos juegos de parámetros tendrán un sentido físico lógico en el problema. A la inversa, será posible establecer varios juegos de parámetros con sentido físico lógico en el problema, pero que sólo algunos satisfacen las ecuaciones planteadas. A la vista de los resultados, la experiencia del modelador debe dar la última palabra respecto del juego de parámetros que mejor representa las condiciones modeladas.

3.9.1. Clasificación de las técnicas de calibración

Las técnicas de calibración empleadas en modelos hidrológicos se pueden clasificar de acuerdo a los siguientes criterios (Bertoni, 2003):

- I. desde el punto de vista operativo: técnicas de calibración automática y manual.
- II. desde el punto de vista de la cantidad de objetivos que se propone optimizar: técnicas de calibración que utilizan funciones mono y multi-objetivo.
- III. desde el punto de vista del enfoque sobre el dominio analizado: técnicas de calibración local y global.

Calibración automática y calibración manual

Básicamente existen dos formas de calibrar un modelo matemático hidrológico desde el punto de vista operativo: (a) ajuste manual mediante prueba y error y (b) ajuste o calibración automática. La primera es la manera más clásica, que ha sido empleada por los hidrólogos desde los comienzos de la aplicación de las técnicas de calibración. Esta consiste básicamente en la aplicación del conocido método de prueba y error. Cuando el modelo utilizado tiene un gran número de parámetros y el usuario tiene poca experiencia en el proceso de calibración, este proceso puede volverse demasiado lento y tedioso. Además de ello, esta técnica no brinda al usuario un criterio objetivo de parada (Sorooshian y Gupta, 1995), sino que, por lo general, usuarios diferentes generalmente encuentran un conjunto diferente de parámetros con el mismo modelo, en la misma cuenca y en el mismo periodo de tiempo (Beven y Binley, 1992).

La calibración automática consiste en el empleo de las computadoras digitales para llevar a cabo el proceso de ajuste de los parámetros. Surgió como una alternativa para acotar el grado de error de la calibración manual ejecutada por los usuarios sin experiencia, para acelerar el proceso de calibración y para restarle subjetividad (Sorooshian y Gupta, 1995). De modo que se convierte en una herramienta que intenta determinar el conjunto óptimo de los valores de los parámetros de un modelo hidrológico con la mínima interferencia del usuario. La experiencia del usuario es sustituida, en gran parte, por una función objetivo sometida a una técnica iterativa de ajuste. El objetivo de la calibración automática es, por lo tanto, minimizar (o maximizar) el valor de una función objetivo que mide la calidad del ajuste de los hidrogramas. Sin embargo, enfatizan Sorooshian y Gupta (1995), el estado del arte de los métodos de calibración automática no han madurado al punto de sustituir enteramente a la calibración manual.

Calibración con funciones mono y multiobjetivo

Como se explicó previamente, generalmente la FO representa la diferencia entre los valores medidos y los calculados por el modelo. Pero también se aclaró que esa diferencia debía venir dada por la característica del

hidrograma que se pretenda ajustar con mayor precisión, en virtud del propósito para el cual será empleado el modelo una vez que se haya calibrado. Así por ejemplo, si el modelo hidrológico ha sido pensado para predecir caudales máximos de diseño, la FO debería marcar la diferencia entre los picos de los hidrogramas medidos y simulados. Si el propósito del modelo hidrológico estuviese asociado al balance hídrico, convendría asegurar un correcto ajuste de volúmenes escurridos, por lo que la FO debería representar la diferencia entre el volumen escurrido medido y el simulado.

Ahora bien, cuando se desean ajustar varias características del hidrograma de salida, como caudal pico, volumen escurrido, curva de recesión, etc., es necesario proponer la minimización simultánea de varias funciones objetivos, cada una de las cuales representa la diferencia entre la variable medida y la simulada de cada una de las características del hidrograma de salida. Este tipo de ajuste se denomina multiobjetivo. En una calibración automática multiobjetivo el propósito es optimizar, al mismo tiempo, varias funciones que evalúan varios aspectos de calidad de ajuste entre hidrogramas calculados y observados.

En numerosos procesos hidrológicos resulta insuficiente el ajuste de mono-objetivo. Por ejemplo, en la simulación continua se requiere de una característica multi-objetivo para el uso del modelo hidrológico, que permita representar correctamente los periodos lluviosos y de estiaje y no simplemente de uno de ellos en detrimento de los otros. La optimización multi-objetivo es un nuevo estado del uso de estas técnicas en el ajuste de los modelos hidrológicos.

Dadas las imperfecciones de los modelos y de los datos de entrada, es prácticamente imposible encontrar un conjunto de parámetros que produzca, al mismo tiempo, valores óptimos de varias funciones objetivo diferentes. Casi siempre, la optimización adecuada de una función objetivo se logra a expensas de la inadecuada optimización de otra. Esta es la situación que normalmente se enfrenta en la calibración de modelos lluvia-caudal, cuando un conjunto de parámetros resulta en un buen ajuste de los picos de caudal, pero como consecuencia presenta resultados mediocres durante los periodos de estiaje (o viceversa).

La principal característica de un problema de calibración multiobjetivo, es que la solución, en general, no será única (Collischonn et al., 2003). Es decir, un conjunto A de valores de los parámetros, podrá corresponder al óptimo de la función F1, y un conjunto B (distinto de A), corresponderá al óptimo de la función F2. Existirán muchos conjuntos de parámetros entre A y B que no optimizan F1 ni F2, pero pueden ser considerados aceptables. Por ejemplo, si C y D son dos conjuntos de valores de parámetros entre A y B, serían considerados aceptables si, partiendo de la solución C hacia la solución D, un cambio en mejoría para una función objetivo será acompañada de un deterioro en la otra, y no existirá ninguna solución E que sea al mismo tiempo mejor que C y D en ambas funciones objetivo F1 y F2. La Figura 3.5. ilustra este concepto para el supuesto caso de un modelo de un parámetro. Como puede verse, no existe un valor de ese parámetro que minimice simultáneamente ambas FO, sino que cuando el parámetro toma el valor de 6 minimiza FO1, pero cuando toma el valor de 10, minimiza FO2. Entre 6 y 10 existe un rango de valores para el parámetro tal que, a medida que se acerca a 6, mejora FO1, pero empeora FO2, y viceversa. Si se toman dos valores en el rango entre 6 y 10, como por ejemplo, 7 y 9 puede verse claramente que no existe un valor entre ellos que sea mejor que ambos en ambas FO. Entonces 7 y 9 son considerados aceptables. Del mismo modo, todos los valores entre 6 y 10 también lo son.

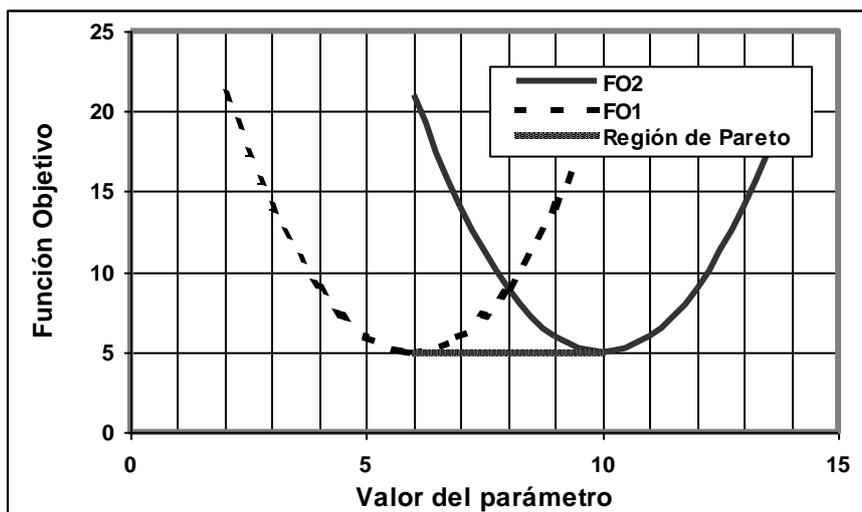


Figura 3.5.: Esquema de la Región de Pareto para dos funciones objetivo

Si no hubiere ninguna información adicional que permita dar mayor importancia a una de las dos funciones objetivo (Collischonn y Tucci, 2003), todas las soluciones intermedias deben ser consideradas buenas. Debido a estas características, la solución de un problema de optimización multi-objetivo no es un punto, sino que se trata de una región de soluciones posibles, todas ellas aceptables, denominada región de Pareto, o región de soluciones no inferiores o no dominadas (Yapo et al., 1998).

Este último enfoque responde al concepto actual en la calibración de modelos matemáticos que se orienta a buscar un conjunto de soluciones posibles y aceptables, en vez de una solución óptima. Esto se debe a que existe una gran incerteza asociada a la estimación de los parámetros del modelo. Las técnicas recientes en calibración de modelos matemáticos prefieren asignar un conjunto de soluciones posibles con una cierta probabilidad de éxito en la calibración. De este moderno concepto proviene la importancia del análisis de incertidumbre en las técnicas de calibración.

Métodos de Calibración local y global

Dado que las técnicas de calibración automática son casos particulares de los métodos de Optimización, estas técnicas también pueden clasificarse entre local y global, de acuerdo al tipo de búsqueda que haga. El gráfico de la Figura 3.6. muestra el caso hipotético de un modelo con un solo parámetro. Como puede verse, la FO tiene varios mínimos locales. Se denomina mínimo global, al mínimo valor de los mínimos locales. En el ejemplo, los puntos A, B y C son mínimos locales, ya que en la vecindad de dichos puntos los demás son superiores a ellos, pero sólo el punto B es mínimo global, ya que es el menor de los mínimos locales, dentro del dominio de validez del único parámetro del modelo. Este concepto es fácilmente extrapolado a modelos de varios parámetros, como son la mayoría de los modelos matemáticos utilizados en hidrología.

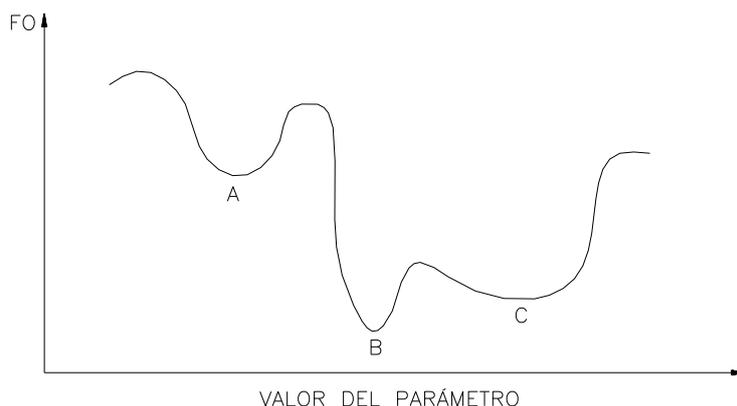


Figura 3.6.: Proyección plana de la superficie de respuesta de una FO

Las primeras técnicas desarrolladas que se encuentran en la bibliografía estaban basadas en algoritmos de búsqueda local. Debido a la complejidad de los modelos hidrológicos, las funciones objetivos, son expresiones implícitas. En otras palabras, no es posible despejar los valores de los parámetros asociados al mínimo valor de la FO, sino que resulta necesario tender a éste mínimo mediante aproximaciones sucesivas basadas en iteraciones direccionadas de prueba y error.

Estos métodos iterativos de búsqueda local poseen cuatro características básicas (Tucci, 2005), que son:

- ✓ definición del punto de partida,
- ✓ dirección de la búsqueda,
- ✓ espaciamiento de cada prueba,
- ✓ y criterio de parada.

La bibliografía es pródiga en metodologías de optimización local (Beveridge y Schechter, 1979), sin embargo sólo algunos de estos métodos, fueron preferidos para calibración de modelos hidrológicos.

Una de las primeras propuestas metodológicas de calibración local, que posteriormente fueron empleadas en calibración de modelos hidrológicos (Beard, 1966), surge de la mano de dos economistas (Friedman y Savage, 1947) con el método Univariacional. Para ejemplificar este método resulta ilustrativo suponer un modelo con dos parámetros y luego extrapolar la idea a n parámetros.

Supóngase que se tienen dos parámetros X_1 y X_2 . La superficie de respuesta de la FO se desarrolla en tres dimensiones, dos de las cuales son los valores de ambos parámetros del modelo, siendo la tercera, el valor de la FO para cada par de valores $FO=f(X_1, X_2)$. El método se inicia en un punto de partida elegido por el usuario, y luego se elige una dirección de búsqueda, por ejemplo, la dirección de X_1 . Se realiza un paso DX_1 , en algún sentido del eje X_1 . Si en la nueva ubicación se obtiene un valor de FO menor que la anterior, realiza un nuevo paso en este sentido, de lo contrario, prueba con un DX_1 en sentido contrario. Si en esta nueva ubicación la FO es inferior continúa en este sentido. Una vez que se ha verificado que para la dirección X_1 , en cualquier sentido se obtiene un valor superior al actual, entonces se comienza la búsqueda a lo largo de X_2 , para lo cual se elige un sentido de búsqueda y se da un paso DX_2 . Nuevamente se procura reducir el valor de FO, de lo contrario se invierte el sentido.

En cualquier caso, se continúa el sentido de búsqueda siempre que la FO siga reduciéndose con cada nuevo paso, y se invierte en caso de que aumente. Cuando en ambos sentidos FO verifica un aumento, se toma la siguiente dirección. Así se continúa hasta que se agoten todas las direcciones, posterior a lo cual se dice que se ha cumplido un ciclo, y se comienza con la primera dirección nuevamente. En este ejemplo hemos supuesto dos direcciones, pero si se tratara de un modelo de n parámetros, tendríamos un espacio hiperdimensional con n ejes ortogonales entre sí. Si bien no puede graficarse este espacio, la idea del método es exactamente igual. Así se continúa, hasta que la FO no perciba reducciones en ningún sentido de las n direcciones, o bien, las reducciones sean aceptablemente despreciables para el caso en estudio.

El método Univariacional tiene una estructura muy convincente pero poco eficiente, ya que no tiene ninguna dirección preferencial de búsqueda. Simplemente prueba en cada una de las direcciones correspondientes al sistema de coordenadas de referencia de la superficie de respuesta. Rosenbrock (1960) propuso el método de Rotación de Coordenadas, que comienza del mismo modo que el método Univariacional, hasta el final del primer ciclo. Posteriormente, traza una recta imaginaria entre el primer punto y el último del primer ciclo, y la establece como dirección preferencial de

búsqueda para el parámetro X1, ajustando las restantes direcciones a esta rotación. Las pruebas son realizadas del mismo modo en cada ciclo, provocando una rotación de los ejes de coordenadas al cabo de cada uno. Otra diferencia con respecto al método Univariacional es la incorporación de un par de coeficientes de aceleración (A valor mayor que 1) y retardo (B valores entre 0 y 1). Estos coeficientes de aceleración y retraso se multiplican por el valor del espaciamiento del paso de avance, de acuerdo a la secuencia de éxitos (reducción de la FO) o fracasos (aumento de la FO) que experimente el avance en cierta dirección.

Ibbitt (1970) modificó el procedimiento de búsqueda en cada dirección del método de rotación de coordenadas. El criterio adoptado incluye una aproximación cuadrática para estimar la tendencia de minimización de la superficie de respuesta. Una explicación más detallada puede encontrarse en Tucci (2005).

El método Simplex (Nelder y Mead, 1965) dio un paso considerablemente superior en la eficiencia de búsqueda, ya que este incursiona libremente en cualquier dirección del espacio hiperdimensional de validez de los parámetros. En otras palabras, no se limita a las direcciones de los ejes de coordenadas de los parámetros, sino que elige la dirección en la cual parece más probable que se minimice la FO. Además de eso, no parte de un punto, sino de un conjunto de puntos, lo que permite un direccionamiento más preciso y eficiente en la búsqueda. Este algoritmo se explica en detalle en el capítulo referido a la Metodología.

Todos estos métodos de optimización local, dependen fuertemente del punto de partida o conjunto de puntos elegidos inicialmente (Johnston y Pilgrim, 1976), y por lo tanto caen comúnmente en punto óptimos locales. Diferentes puntos de partida, podrían conducir a diferentes óptimos locales, lo que genera diferentes juegos de valores de los parámetros del modelo, aún cuando se utilice la misma serie de datos en la misma cuenca, para todos los diferentes puntos de partida. Esto se debe a que las superficies de respuesta de las funciones objetivos no siempre son suaves y continuas, sino que presentan discontinuidades, y varios óptimos locales (Duan et al., 1992). Un complemento a estos métodos de búsqueda local podría estar dado por una

iniciación múltiple del método, en diferentes puntos de partida, y la selección del mínimo de los mínimos locales encontrados.

Sorooshian y Gupta (1995) probaron la eficiencia del algoritmo Simplex, mediante una implementación de múltiples partidas, denominado MSX. En esta prueba quedó demostrado que ni siquiera este algoritmo de búsqueda local con múltiples partidas, fue capaz de encontrar eficazmente el óptimo global de una superficie de respuesta. Inicializando 50 veces para cada optimización, el algoritmo MSX no fue capaz de encontrar el óptimo global de los parámetros del modelo Sacramento (Burnash, 1995) en ninguna de las diez pruebas realizadas.

Dadas las limitaciones de las metodologías de búsqueda local, surgen las técnicas de optimización global. En sus versiones primitivas, estos métodos consistían esencialmente en la aplicación sucesiva de la técnica de búsqueda local. Surgen los Métodos de Búsqueda Aleatoria (Random Search, RS). Entre otros, se comienza a desarrollar la técnica de Búsqueda Aleatoria Uniforme (URS) que consiste en asignar una distribución probabilística uniforme para cada variable a calibrar. Como ejemplo, se puede suponer dos variables dadas X_1 y X_2 a calibrar en el modelo. Las mismas se van arrojando aleatoriamente según una distribución probabilística uniforme para cada variable y luego con cada par X_1 , X_2 se evalúa la función objetivo. A continuación se comienzan a aplicar las técnicas de optimización local en cada punto X_1 , X_2 . Una vez obtenido un óptimo para cada par X_1 , X_2 se elige el óptimo entre éstos, que corresponde al óptimo global. Brazil (1988) propone la Técnica de búsqueda aleatoria adaptativa, que asigna una distribución de probabilidades a cada variable. Por lo demás, es igual al método anteriormente descrito. Kuczera (1997), aporta el Método de búsqueda local con múltiples partidas (o método combinado – Múltiples start local search, MSLS), que utiliza métodos aleatorios para fijar los puntos de comienzo de búsqueda.

Dando principio a una nueva generación de métodos de optimización, Holland (1975), formaliza los principios de la técnica de los Algoritmos Genéticos (Genetic Algorithm, GA) que han sido tratados en sección aparte.

3.10. CALIBRACIÓN AUTOMÁTICA DE MODELOS P-Q.

Un procedimiento típico de estimación automática de los parámetros de un modelo hidrológico comprende los siguientes cuatro elementos principales (Sorooshian y Gupta, 1995):

- I. Función objetivo, también llamada criterio de calibración.
- II. Algoritmo de optimización.
- III. Criterio de parada.
- IV. Datos de calibración.

3.10.1. La función objetivo

Todos los autores concuerdan con la destacada importancia de las funciones objetivos en el proceso de calibración automática de los modelos hidrológicos. Sorooshian y Gupta (1995) apuntan una sintética definición estableciéndola como “una ecuación usada para computar una medida numérica de la diferencia entre la salida simulada por el modelo (típicamente el hidrograma de escurrimiento) y el hidrograma observado (medido)”. Agregan estos autores que el propósito de la calibración del modelo es, por consiguiente: “encontrar esos valores de los parámetros del modelo que optimizan (minimizan o maximizan, como sea apropiado) el valor numérico de la función objetivo”.

Debido a que la FO es la expresión matemática que evalúa la calidad de la calibración, debe ser elegida con cuidado, de modo que represente fielmente el error de la calibración con una estructura matemática lo más sencilla posible.

El proceso de calibración del modelo matemático puede estar sometido a restricciones o no. En el primer caso, el valor de los parámetros debe moverse en un intervalo previamente establecido por las restricciones, por lo que la FO puede tomar valores en una región definida por los rangos de validez de los parámetros del modelo. En el segundo caso, los parámetros del modelo pueden tomar cualquier valor en el proceso de calibración. Tanto la FO, como sus restricciones son funciones matemáticas que podrán tener forma lineal o no lineal (cuadrática, logarítmica, etc.). Las restricciones vienen

expresadas, normalmente, como desigualdades. Las FO deben tener una estructura matemática sencilla, para poder identificar con claridad el sentido físico del número que nos arroja.

Existen diversos tipos de FO, de acuerdo a su finalidad. Sin embargo la FO mayormente empleada en la teoría de optimización (Sorooshian y Gupta, 1995), y por lo tanto en calibración automática, ha sido diferentes formas de la Función de Mínimos Cuadrados Ponderada, presentada en la ecuación 3-5.

$$FO = \sum_{i=1}^n w_i (QO_i - QC_i)^2 \quad (3-5)$$

Que puede simplificarse a la denominada Función Cuadrática, como la denomina Tucci (2005), simplemente igualando los pesos de todos los intervalos, como se muestra en la ecuación 3-6.

$$FO = \sum_{i=1}^n (QO_i - QC_i)^2 \quad (3-6)$$

Donde:

QO_i: valor de caudal observado en el intervalo i.

QC_i: valor de caudal calculado en el intervalo i.

w_i: peso relativo de la diferencia cuadrática en el intervalo i.

n: número de intervalos de los hidrogramas comparados.

Según Tucci (2005), esta función asigna mayor peso a los caudales mayores. En el ajuste de un periodo continuo de caudales la tendencia del ajuste será el de ajustar mejor las crecidas. Esto se debe a que, para una diferencia de caudales iguales, el porcentaje de error es mayor cuando los caudales son más pequeños. Por ejemplo, si se estima 18 m³/s, para un caudal observado de estiaje de 20 m³/s, y 198 m³/s, para un caudal de crecida de 200 m³/s, en ambos casos se estará cometiendo un error de 2 m³/s, sin embargo, en estiaje representa el 10% del caudal observado, mientras que en crecida, sólo el 1%. Estas diferencias todavía se marcan más, cuando consideramos que se elevan al cuadrado. Entonces, si nuestro método de ajuste asegura un error del 5%, por ejemplo, para el ajuste de los hidrogramas, en los picos es insignificante, pero en estiaje se puede estar cometiendo un error importante.

La FO de mínimos cuadrados tiene gran aplicación por su sencilla formulación y fácil interpretación. Sin embargo, recuerdan Sorooshian y Gupta (1995) que la elección de alguna forma de la Función de Mínimos Cuadrados está típicamente basada en un criterio subjetivo, de acuerdo al cual se asumen dos hipótesis referidas a la distribución probabilística de los errores, que se desprenden de la teoría de Máxima Verosimilitud:

- ✓ La probabilidad conjunta de los errores sobre la serie de datos disponibles, es Gaussiana con media cero.
- ✓ Los errores son independientes unos de otros.

Adicionalmente, Clarke (1973) enfatizó que, dado que los datos hidrológicos pueden frecuentemente violar esas hipótesis, los efectos de asumirlas pueden resultar bastante serios. En consecuencia, Sorooshian y Dracup (1980) desarrollaron funciones objetivo basadas en la Máxima Verosimilitud, para contabilizar apropiadamente tanto la auto correlación (carencia de independencia), como la heteroscedasticidad (varianza cambiante) de las series de datos de caudales. En muchos casos, el criterio de Máxima Verosimilitud (MV) dio pruebas de superioridad con respecto a la Función de Mínimos Cuadrados (FMC). Estudios posteriores de Sorooshian y Gupta (1983) revelaron que entre las razones que justifican el éxito del criterio de MV, fueron las propiedades de convergencia superiores de las superficies de las funciones respuestas generadas por esta.

La forma más exitosa del criterio de Máxima Verosimilitud (Sorooshian y Gupta, 1995) ha sido una función llamada HMLE (Heteroscedastic Maximum Likelihood Estimator) que contabiliza la no estacionariedad de la varianza en los errores de caudal medidos. El estimador HMLE es el estimador de máxima verosimilitud, mínima varianza, asintóticamente no sesgado, cuando los errores en los datos de salida son Gaussianos, con media cero, no correlacionada, y tiene una varianza relativa a la magnitud de los caudales generados. La no estacionariedad de la varianza de los errores se cree que son comunes en los datos de caudal. Además, el estimador HMLE está relacionado con la FMC, ya que aquella se reduce a esta en el caso especial en que la varianza se hace constante en los errores de los datos de caudal.

El estimador HMLE, tiene la forma de la ecuación 3-7.

$$\min HMLE = \frac{\frac{1}{n} \sum_{i=1}^n w_i \varepsilon_i^2}{\left[\prod_{i=1}^n w_i \right]^{1/n}} \quad (3-7)$$

Donde:

$\varepsilon_i = QO_i - QC_i$, es el residual del modelo en el tiempo i , siendo QO y QC los caudales observados y calculados por el modelo, respectivamente.

w_i = es el peso asignado al tiempo i , computado como sigue:

$$w_i = f_i^{2(\lambda-1)}$$

Donde:

f_i = es el valor verdadero de caudal en el intervalo de tiempo i . Se aproxima mediante QO_i .

n es el número de datos de la serie.

λ es el parámetro desconocido de transformación que estabiliza la varianza.

Tucci (2005), propone como las principales funciones empleadas en la hidrología para calibración automática de modelos hidrológicos, además de la Función de Mínimos Cuadrados: la Función Módulo, la Función Inversa y la Función Relativa, presentadas respectivamente en las ecuaciones 3-8, 3-9, y 3-10. Indica Tucci (2005), al respecto de estas FO, que la Función Módulo también asigna mayor peso a los caudales mayores, mientras que la Función Inversa busca ajustar mejor los caudales menores. Mientras que la Función Relativa, empareja las prioridades dividiendo por el valor del caudal observado en cada intervalo.

$$FO = \sum_{i=1}^n |QO_i - QC_i| \quad (3-8)$$

$$FO = \sum_{i=1}^n \left(\frac{1}{QO_i} - \frac{1}{QC_i} \right)^2 \quad (3-9)$$

$$FO = \sum_{i=1}^n \left(\frac{QO_i - QC_i}{QO_i} \right)^2 \quad (3-10)$$

Tassi et al. (2006), presentan las tres FO anteriores, más la Función Cuadrática (ecuación 3-6), pero dividiéndolas por el número de datos n , y obteniéndole la raíz cuadrada. Con lo cual, le asigna un sentido conceptual al valor numérico hallado, convirtiendo la Función Cuadrática en la expresión de la Raíz Media Cuadrática, o sea, una forma de error promedio. Además, Tassi et al. (2006), agregan una nueva FO, que se muestra en la ecuación 3-11. Esta forma de FO también es elegida por Yapo et al. (1998), quienes adoptan además el criterio de HMLE para presentar y probar el desempeño del algoritmo multiobjetivo MOCOM-UA.

$$FO = \sum_{i=1}^n \left(\frac{QO_i - QC_i}{QO_i} \right) \quad (3-11)$$

Como puede, La FO adoptada para calibrar el modelo hidrológico es altamente dependiente del uso al cual será destinado el mismo. Si el objetivo es ajustar el modelo para simulación de crecidas, debe procurarse minimizar los errores en la simulación de los caudales mayores, para lo cual sería adecuado aplicar la función cuadrática o módulo. Si, en cambio, el modelo será usado para simulación de una secuencia de caudales, para la operación o para dimensionamiento de reservorios, el ajuste busca que las frecuencias de los caudales observados y calculados sean semejantes, en cuyo caso sería recomendable la función relativa (Tucci, 2005).

Madsen (2000) resume los objetivos mayormente considerados en una apropiada calibración automática, en los siguientes cuatro:

- I. Un buen acuerdo entre los volúmenes totales escurridos y simulados por el modelo.
- II. Una concordancia global aceptable entre las formas de ambos hidrogramas.
- III. Una buena concordancia entre los picos de hidrogramas, en términos de tiempo de pico, pendiente de la curva y volumen.
- IV. Una buena aproximación de los caudales bajos.

Agrega Madsen (2000) que resulta necesario traducir el objetivo global de la calibración en términos más operacionales, de donde se desprenden las formas más comunes de las funciones objetivos. Propone en

su trabajo, las FO presentadas en las ecuaciones de la 3-12 a la 3-15, que asignan cierto protagonismo al peso de los errores en cada intervalo de la serie de datos, y las denomina respectivamente: error global de volumen, raíz media cuadrática global de los errores, raíz media cuadrática promedio de los errores al pico de los eventos, raíz media cuadrática promedio de los errores en los caudales bajos de los eventos.

$$FO = \left| \frac{\sum_{i=1}^n w_i (QO_i - QC_i)}{\sum_{i=1}^n w_i} \right| \quad (3-12)$$

$$FO = \left| \frac{\sum_{i=1}^n w_i^2 (QO_i - QC_i)^2}{\sum_{i=1}^n w_i^2} \right|^{1/2} \quad (3-13)$$

$$FO = \frac{1}{M_p} \sum_{j=1}^{M_p} \left| \frac{\sum_{i=1}^{n_j} w_i^2 (QO_i - QC_i)^2}{\sum_{i=1}^n w_i^2} \right|^{1/2} \quad (3-14)$$

$$FO = \frac{1}{M_l} \sum_{j=1}^{M_l} \left| \frac{\sum_{i=1}^{n_j} w_i^2 (QO_i - QC_i)^2}{\sum_{i=1}^n w_i^2} \right|^{1/2} \quad (3-15)$$

Donde:

M_p es el número de eventos con caudales pico.

M_l es el número de eventos con caudales bajos.

n_j es el número de intervalos de tiempo con caudales de pico, para la ecuación 3-14, y de caudales bajos, para la ecuación 3-15.

Los eventos con caudales pico se definen como los periodos donde el caudal observado está por encima de un dado nivel prefijado. Del mismo modo, los eventos con caudales bajos se definen como los periodos donde el caudal observado está por debajo de un dado nivel prefijado.

Los pesos, siempre según Madsen (2000), representan la importancia a darle a una porción particular del hidrograma, reflejando el error en los datos y en el modelo. Es decir, se les asigna poco peso a los periodos en los cuales se presentan largas series de datos, o errores del modelo.

Para verificar la bondad de la calibración realizada, se adopta normalmente el famoso coeficiente de Nash y Stutcliffe (1970), presentado en la ecuación 3-16, que representa una reinterpretación del coeficiente de determinación, empleado en regresión lineal.

$$R^2 = 1 - \frac{\sum_{i=1}^n w_i^2 (QO_i - QC_i)^2}{\sum_{i=1}^n w_i^2 (QO_i - \overline{QO})^2} \quad (3-16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (QO_i - QC_i)^2}{\sum_{i=1}^n (QO_i - \overline{QO})^2} \quad (3-17)$$

Donde:

\overline{QO} es el promedio de los caudales del hidrograma observado.

AQUÍ FACU TIENE QUE AGREGAR LA EXPLICACIÓN DEL COEFICIENTE DE PERSISTENCIA

3.10.2. Algoritmos de optimización para modelos P-Q

La superficie descrita por la FO en el espacio de validez de los parámetros, recibe el nombre de superficie de respuesta. Un algoritmo de optimización es un procedimiento lógico empleado para buscar dentro de la superficie de respuesta, restringida a los rangos permitidos para los parámetros, el juego de valores de los parámetros del modelo que optimizan (minimizan o maximizan, según corresponda) el valor numérico de la FO (Sorooshian y Gupta, 1995). El procedimiento, agregan Sorooshian y Gupta (1995), es típicamente implementado en una computadora digital para permitir una búsqueda rápida.

Una descripción detallada de la historia en cuanto al estudio de los algoritmos de optimización puede encontrarse en la tesis de maestría del Mg. Ing. Facundo José Alonso. Acá solo se abarcará lo que es de estrictamente de interés para esta práctica.

Una de los primeros antecedentes que se encuentran en la bibliografía, respecto del empleo de las técnicas de optimización antes citadas para llevar a cabo la calibración automática de modelos hidrológicos, se remonta a la década del 60. Beard (1966) empleó la calibración automática para ajustar el modelo hidrológico HEC1, predecesor del conocido modelo hidrológico HEC-HMS. Ibbitt (1970) empleó la calibración automática para ajustar un modelo hidrológico para diferentes combinaciones de series con y libre de errores, y estudió los efectos de los errores sobre el ajuste. Posteriormente, (Ibbitt, 1972) utilizó un modelo conceptual para generar datos de caudales sintéticos libre de errores de información de precipitación y evaporación potencial. Luego introdujo errores en las tres series de datos. Mediante ajuste automático y objetivo, ajustó el modelo para diferente combinación de series libre de errores y con errores, y estudió los efectos de los errores sobre el ajuste. Encontró que los valores del criterio de ajuste dependen fuertemente de los errores en los datos de caudal, ya que las variaciones en los valores finales de los parámetros para el ajuste a los datos con errores mostraron no ser tan grandes como a los datos libres de errores.

Duan et al. (1992) presentan claramente la naturaleza de los problemas que asocia a la calibración la existencia de múltiples óptimos locales, en el intento de calibración del modelo hidrológico PQ denominado SIXPAR. Estos resultados sugieren que la optimización de los modelos PQ es más difícil de lo que se pensaba previamente y que los procedimientos corrientes de búsqueda local tienen una muy baja probabilidad de éxito a la hora de encontrar el juego óptimo de parámetros. Por consiguiente, evalúan la performance de tres algoritmos de búsqueda local aplicados sobre el modelo SIXPAR. Finalmente, presentan un nuevo procedimiento de optimización global, denominado SCE-UA (Shuffled Complex Evolution method), el cual fue capaz de encontrar consistentemente el óptimo global del modelo SIXPAR, y aparece como capaz de resolver efectiva y eficientemente el problema de optimización de un modelo PQ.

Yapo et al. (1998) introducen la necesidad de calibrar con múltiples objetivos. Indican que las experiencias prácticas realizadas con calibración de modelos sugieren que ninguna función mono-objetivo es adecuada para identificar las formas en que el modelo falla para reunir las características

importantes de los datos observados. Dado que algunos de los últimos modelos hidrológicos simulan varias características del flujo de salida (como por ej., agua, energía, constituyentes químicos), sugieren que existe una necesidad de procedimientos de calibración multiobjetivo eficientes y efectivos a los fines de explotar toda la información útil respecto del sistema físico contenida en series de datos medidos. Entonces, presentan el algoritmo MOCOM-UA, que es una metodología eficiente y efectiva para resolver el problema de la optimización global multiobjetivo. El método es una extensión del exitoso algoritmo mono-objetivo de calibración global SCE-UA. A través de un simple modelo hidrológico estudian las características y capacidades de calibración del MOCOM-UA.

Vrugt et al. (2003) propusieron una versión modificada del algoritmo original de optimización global SCE-UA desarrollado por Duan et al. (1992). Se trata de un manipulador de muestras (sampler) del tipo Markov Monte Carlo (MCMC), denominado Shuffled Complex Evolution Metropolis (SCEM-UA), ya que incorpora conceptos del método de Metrópolis. Este algoritmo está diseñado para inferir la distribución posterior de los parámetros del modelo hidrológico. Proponen combinar las destrezas del exitoso algoritmo SCE-UA con la ventaja de los métodos de cadenas de Markov Monte Carlo, que han ido aumentando su popularidad debido a que estiman la distribución de probabilidades posterior de los parámetros en los modelos hidrológicos. En este trabajo Vrugt et al. (2003), presentan tres casos de estudio, en los cuales demuestran que las capacidades de adaptación del SCEM-UA reduce significativamente el número de simulaciones del modelo necesarias para inferir la distribución posterior de los parámetros cuando se comparan con el tradicional manipulador de muestras Metropoli-Hastings.

Un nuevo equipo de trabajo (Vrugt et al., 2003), encabezado por Vrugt, lanza el mismo año, la versión multiobjetivo del SCEM-UA, denominado MOSCEM-UA (Multiobjective Shuffled Complex Evolution Metropolis). Los autores insisten en que la experiencia práctica con la calibración de modelos hidrológicos sugiere que ninguna función mono-objetivo, independientemente de cuan cuidadosamente haya sido elegida, es capaz de representar por sí misma, adecuadamente todas las características de los datos considerados importantes. Entonces proponen el MOSCEM, como un algoritmo capaz de

resolver el problema de optimización multiobjetivo para modelos hidrológicos, empleando el concepto de dominancia de Pareto (más que el de evaluación directa de función mono-objetivo) para evolucionar la población inicial de puntos hacia un conjunto de soluciones estancadas con distribución estable (set de Pareto). Por último, los autores comparan la eficacia del MOSCEM-UA con la del algoritmo MOCOM-UA para tres caso de estudio de modelaciones hidrológicas que se incrementan en complejidad.

3.10.3. Criterio de parada

Las estrategias de optimización que se presentan en este trabajo, son todos procedimientos iterativos que buscan el juego óptimo de valores para los parámetros mediante pasos de aproximación. En consecuencia, como lo establecen Sorooshian y Gupta (1995), resulta necesario establecer un criterio que determine cuando debe detenerse la búsqueda. En principio, la solución existe en aquel punto de la superficie de respuesta en el cual la pendiente es nula y el valor de la función, un mínimo. En la práctica, es virtualmente imposible, agregan Sorooshian y Gupta (1995), conocer en qué momento exacto ese punto ha sido alcanzado, en consecuencia, estos autores proponen los criterios de terminación mayormente usados en la práctica, del modo que se transcriben a continuación.

Convergencia de la FO. Una manera sencilla de terminar la búsqueda es detenerla cuando el algoritmo es incapaz de apreciar una mejora del valor de la FO en dos iteraciones seguidas. Aunque este criterio puede indicar que se ha alcanzado el óptimo, también puede representar simplemente que se ha llegado a una zona plana de la superficie de respuesta. Si no fuese importante el valor preciso del óptimo, la convergencia de la FO puede ser muy útil como criterio de parada. Una implementación típica es detenerse cuando se cumple con la condición:

$$\frac{(FO_{i-1} - FO_i)}{FO_i} \leq \varepsilon_{FO} \quad (3-18)$$

Donde:

FO_i: valor de la FO en el paso iterativo i.

eFO : error de la FO prefijado. Por ejemplo eFO = 10-3.

Convergencia de los parámetros. Otra forma de terminar la búsqueda es cuando el algoritmo es incapaz de generar un cambio apreciable en el valor de los parámetros, y simultáneamente mejorar el valor de la FO. Aunque este criterio puede indicar que se ha alcanzado el óptimo, también puede representar simplemente que se ha alcanzado una región con alta interacción paramétrica en la superficie de respuesta, como por ejemplo, una región al tipo de un estrecho valle. Sorooshian y Gupta (1995) comentan que se han obtenido mejores resultados con este criterio que con el anterior. Una implementación típica de este criterio se muestra en la siguiente ecuación:

$$\frac{(\theta(j)_{i-1} - \theta(j)_i)}{(\theta(j)_{\max} - \theta(j)_{\min})} \leq \varepsilon_{\theta} \quad (3-19)$$

Donde:

$\theta(j)_i$: valor del parámetro j, en el paso iterativo i.

e θ : error paramétrico prefijado. Por ejemplo e θ = 10-3.

Número máximo de iteraciones. Si el tiempo de computadora es limitado, y para asegurar que el algoritmo no quedará atrapado en un ciclo infinito, se emplea como práctica normal, programar la terminación del algoritmo si se ha alcanzado un número máximo preestablecido de iteraciones, a menos que se haya cumplido previamente con alguno de los criterios anteriores. Para métodos de búsqueda aleatoria, este es el procedimiento normal de terminación. No es posible dar una regla fija respecto del número de iteraciones convenientes, ya que es dependiente, tanto del algoritmo como del problema particular tratado.

Limitaciones

Ninguno de estos criterios son confiables en términos de dar una indicación precisa respecto del arribo al óptimo global, excepto en el caso más trivial en el cual la FO presenta una superficie de respuesta convexa y de comportamiento suave. Sin embargo, en la experiencia de Sorooshian y Gupta (1995), el criterio de terminación por convergencia de los parámetros es más

adecuado para los estudios de calibración, ya que termina la calibración cuando el parámetro no está cambiando significativamente. El criterio de máximo número de iteraciones se emplea como respaldo para evitar consumo computacional innecesario. Si el algoritmo no concluye dentro de un número “razonable” de iteraciones, el código computacional debería ser examinado de “bugs”. Finalizan los autores citados, indicando que la mejor manera de establecer valores apropiados como criterios de terminación consiste en correr el algoritmo empleando series sintéticas de datos sobre el modelo.

3.10.4. Datos de calibración

Para realizar la calibración automática de un modelo hidrológico de transformación lluvia-caudal, ajustado para el funcionamiento en una cuenca específica, resulta necesario contar con una serie de datos buenos y confiables. Sin embargo, según ironizan Sorooshian y Gupta (1995), poco se conoce objetivamente respecto de aquello que constituye una “buena” serie de datos de calibración. Por esta razón, resulta necesario analizar cuántos y qué tipo de datos darán los mejores o aceptables resultados de calibración automática. Los autores citados, establecen los siguientes criterios que se transcriben a continuación.

Cantidad de datos. La práctica común consiste en utilizar la mayor cantidad de los datos disponibles para la calibración, luego de apartar una cierta cantidad de datos para la verificación de los resultados. Sin embargo, los estudios de Sorooshian et al. (1983) indicaron que, cuando el procedimiento de calibración había sido correctamente diseñado, sólo un año de datos diarios eran suficientes como para obtener estimaciones conceptualmente realistas para los parámetros del modelo de pronóstico de crecidas de la US-National Weather Services (SMA-NWSRFS) bien presentado en Burnash (1995). Más aún, el uso de series de datos de mayor longitud sólo sirven para mejorar marginalmente la estimación de los parámetros. En general, desde un punto de vista estadístico, la serie de datos debería ser al menos unas 20 veces el número de parámetros a ser estimados (por ejemplo, si hay 7 parámetros, entonces al menos se debería contar con 140 datos diarios de caudal para computar la función). Esto es, aclaran los autores, simplemente una

aproximación de dedo gordo. Gupta y Sorooshian (1985) mostraron que el error estándar ($\sigma(j)$) en la estimación del parámetro (j) decrece con el tamaño n de la muestra aproximadamente de acuerdo a la fórmula de la ecuación:

$$\sigma(j) \propto \frac{1}{\sqrt{n}} \quad (3-20)$$

Como la mejora marginal en $1/n$ se convierte en pequeña después de 500 o 100 puntos de dato, esto sugiere que dos a tres años de datos de calibración deberían ser suficientes, siempre y cuando los datos sean de los “buenos”. Lo cual nos conduce a una discusión acerca de la calidad de los datos.

Calidad de datos: desde el punto de vista de la calibración del modelo, la calidad de los datos dependen de la información (relacionada con los parámetros) contenida en los datos y el ruido (errores) en los datos. Claramente, es deseable que el contenido de información paramétrica en los datos sea tan larga como sea posible, y el ruido tan pequeño como sea posible.

Respecto de la información, cualitativamente, una serie de datos informativos son aquellos tales que contienen o representan tanta variabilidad en el comportamiento del escurrimiento, que los distintos modos de operación del proceso hidrológico son apropiadamente representados. Esto maximiza las chances de los datos de contener información suficiente como para identificar apropiadamente el valor de los parámetros del modelo (es decir, variaciones de los parámetros del modelo, resultarán en variaciones detectables en las salidas del modelo, fuera del hidrograma observado). Por ejemplo, si los datos elegidos corresponden a un año relativamente seco, ciertos procesos del modelo, no serán activados; en consecuencia, la respuesta del modelo será insensible a algunos de los parámetros del modelo que determinan el fraccionamiento de la humedad entre las distintas capas subterráneas y el agua superficial. Sin embargo, si la serie de datos corresponde a un año tan húmedo que mantiene la capa superior del suelo saturada la mayor parte del tiempo, la respuesta del modelo puede ser insensible a variaciones en los parámetros que controlan las distintas capas subterráneas del modelo.

La mejor elección parece ser una serie de datos que contenga una gran “variabilidad hidrológica”. Por ejemplo, Gupta y Sorooshian (1985b) mostraron que la identificabilidad del parámetro correspondiente a la capa

superior de humedad de un modelo de escurrimiento típico, depende no de cuán larga sea el flujo excedente en esta capa, sino de cuantas veces el flujo excede este reservorio. Por consiguiente, la mayoría de las veces en que el régimen hidrológico intercambia entre modos secos, medios y húmedos, es probable que los datos sean más informativo aún.

Con respecto al error en los datos, la presencia de errores en la medida y escritura de los datos causa un deterioro en la calidad, resultando en un desmedro de confianza en la estimación de los parámetros. El error en los datos tiene distintas fuentes, como registro incorrecto de la máquina, mal funcionamiento de la máquina, métodos de transferencia de datos a la computadora, aproximaciones en la curva de transformación, cambios en el fondo del río, aproximaciones en las fórmulas usadas para convertir registros de campo, negligencia, fraude, etc. Aunque se ha sugerido que los efectos combinados de todos esos errores deberían tender hacia una distribución normal, esto no ha sido demostrado suficientemente. En la práctica, los errores en los datos probablemente contengan sesgos, particularmente en el caso de datos de precipitación, que tienden a subestimar el valor real de lluvia. Los datos de caudal son en general, los más confiables, pero contienen típicamente un 10 % de error. En la selección de los datos para calibración del modelo, sería deseable que la serie sea cuidadosamente examinada para identificar primeramente errores groseros, por ejemplo periodos donde el caudal se eleva, pero no se observan datos de precipitación. Mientras que los errores en los datos de caudal pueden ser manejados mediante el uso de la máxima verosimilitud y otras teorías estadísticas, todavía deben desarrollarse métodos satisfactorios para el manejo apropiado de los errores de precipitación.

4. METODOLOGIA

En el presente capítulo se describe el modelo hidrológico Genie rural (GR) (Perrin et al., 2007), y el método de optimización MOCOM-UA (Yapo et al., 1998), empleado para realizar la calibración automática multi-objetivo, del modelo.

4.1. DESCRIPCIÓN DEL MODELO HIDROLOGICO GR

4.1.1. Objetivos de su desarrollo

La simulación del caudal en la salida de una cuenca hidrográfica es necesaria, y en ocasiones, indispensable para las numerosas aplicaciones de ingeniería y de gestión de los recursos hídricos, tales como el dimensionamiento y la gestión de obras, el pronóstico de crecidas o de estiaje, la detección del impacto en el uso del suelo o en los cambios climáticos. Para responder a esta cuestión, el instituto IRSTEA (ex CEMAGREF) comenzó a desarrollar a comienzos de los años 1980 (Michel, 1983) una generación de modelos hidrológicos denominados genéricamente “Génie Rural” (GR).

4.1.2. Principales características

Como todo modelo hidrológico, los modelos GR operan una triple globalización:

✓ *A nivel de proceso:* se propone una relación comportamental simple a la escala de cuenca hidrográfica, puesta a punto empírica, y sin vínculo conceptual con los procesos físicos subyacentes. Es decir, la puesta a punto se focaliza en el juego de parámetros que mejor ajusta al modelo para las condiciones planteadas.

✓ *A nivel espacial:* los modelos GR, son de tipo global, es decir que trabajan con información promediada de toda la cuenca hidrográfica. A pesar de ello, algunas de sus versiones han sido calibradas con éxito en cuencas hidrográficas desde unos pocos Km², hasta varios miles de Km² (Perrin, 2000),

como por ejemplo, la cuenca hidrográfica del Río Sena a la entrada a París, de unos 43.800 km².

✓ *A nivel temporal:* los modelos GR han sido desarrollados para distintos pasos de tiempo específicos: (i) Versión anual denominado GR1A (Mouelhi et al., 2006a), (ii) Versión mensual denominado GR2M (Mouelhi et al., 2006a), (iii) Versión con paso de tiempo diario, denominada GR4J (Perrin et al., 2003), (iv) Versión horaria, GR3H (Berthet, 2010). Adicionalmente, existe una versión reformulada para la aplicación en pronóstico hidrológico, denominada GRP, orientada al paso de tiempo diario (Tangara, 2005), y horario (Berthet, 2010).

4.1.3. Estructura del modelo

La filosofía general de construcción de modelos hidrológicos que sigue el equipo de trabajo que desarrolló la familia de modelos “GR”, consiste en reducir tanto como sea posible la complejidad de la estructura del modelo, procurando conservar la performance y eficiencia del mismo.

Por esta razón, las versiones indicadas previamente no difieren mucho entre sí, a pesar de estar orientadas a distintos pasos de tiempo. La estructura general de este tipo de modelos “GR” se conserva, y sólo existen pequeñas diferencias según se trate de un modelo orientado al paso de tiempo anual, diario u horario.

A continuación se presenta el desarrollo sólo de la versión GR4J, es decir, con paso de tiempo diario, empleada en este trabajo.

El modelo GR4J (Génie Rural à quatre paramètres journaliste), tiene paso de tiempo diario y dispone de 4 parámetros ajustables. Es un modelo de precipitación-caudal global. Su desarrollo se inició en el instituto “Cemagref” (hoy “Irstea”) a principios de 1980, con el objetivo de desarrollar un modelo de simulación robusta y fiable para utilizar en problemas de lluvia-caudal aplicados a la gestión de los recursos hídricos en ingeniería (diseño, pronóstico de inundaciones y flujos bajos, manejo de reservorios, detección de impacto, etc) .

Este modelo ha sido adecuado en varias versiones en los trabajos de Edijatno y Michel (1989), Edijatno (1991), Nascimento (1995), Edijatno et al. (1999), Perrin (2000), Perrin (2002) y Perrin et al. (2003). Estos han mejorado

progresivamente el rendimiento del modelo. La versión que aquí se presenta es la de Perrin et al. (2003).

Aunque GR4J es un modelo empírico, su estructura es similar a la de los modelos conceptuales que utilizan tanques, con un procedimiento de control de las condiciones de humedad del reservorio que tiene en cuenta las condiciones anteriores y para garantizar el funcionamiento en simulación continua. Su estructura combina un reservorio de producción y uno de transferencia, unos hidrogramas unitarios, así como una función de apertura hacia el exterior no atmosférico (es decir, subterráneo), permitiendo simular el comportamiento global de la gran cuenca hidrológica, que incluye el flujo subterráneo.

Descripción matemática GR4J

Un diagrama de la estructura del modelo se muestra en la Figura 4.1.

Donde:

P_k es la lluvia diaria de la jornada k

E es la evapotranspiración potencial media del mismo día calendario.

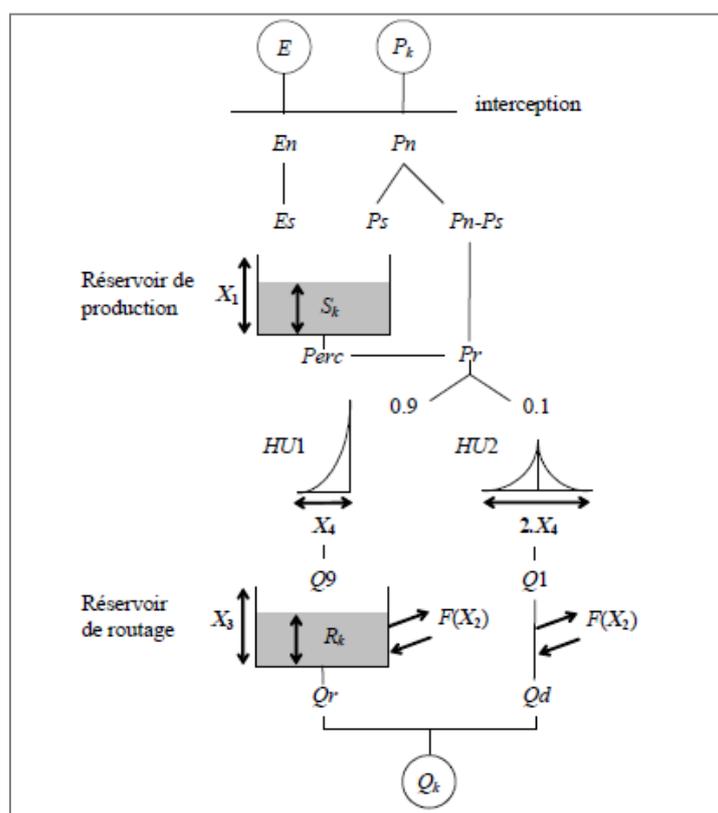


Figura 4.1. Estructura del modelo GR4J, (Perrin et al, 2003)

Neutralización

El primer paso es la neutralización de P_k y E para determinar una lluvia neta P_n y la evapotranspiración neta E_n . Estas se calculan por las siguientes ecuaciones, según corresponda:

$$\begin{aligned} \text{Si } P_k \geq E &\longrightarrow P_n = P_k - E \quad \text{y} \quad E_n = 0 \\ \text{Si } P_k < E &\longrightarrow P_n = 0 \quad \text{y} \quad E_n = E - P_k \end{aligned}$$

Función de salida

En el caso en que P_n es diferente de cero, una parte de P_n , denominada P_s , es asignada al tanque de producción. Esta porción es calculada por:

$$P_s = \frac{X_1 \left(1 - \left(\frac{S_k}{X_1} \right)^2 \right) * \tanh \left(\frac{P_n}{X_1} \right)}{1 + \frac{S_k}{X_1} * \tanh \left(\frac{P_n}{X_1} \right)}$$

Donde:

X_1 (mm) es la capacidad máxima del reservorio de producción S_k al comienzo de la jornada k .

Cuando E_n es distinto de cero, una cantidad de E_n , denominada E_s , abandona el tanque de producción. Esta cantidad se calcula según la siguiente ecuación:

$$E_s = \frac{S_k \left(2 - \frac{S_k}{X_1} \right) * \tanh \left(\frac{E_n}{X_1} \right)}{1 + \left(1 - \frac{S_k}{X_1} \right) * \tanh \left(\frac{E_n}{X_1} \right)}$$

El contenido del reservorio resultante, está dado por:

$$S' = S_k + P_s - E_s$$

Percolación

La Percolación $Perc$ que se produce desde el tanque de producción se calcula con la siguiente ecuación:

$$Perc = S' * \left\{ 1 - \left[1 + \left(\frac{4}{9} * \frac{S'}{X_1} \right)^4 \right]^{\frac{1}{4}} \right\}$$

Como resultado, el contenido del tanque se convierte en:

$$Sk + 1 = S' - Perc$$

La cantidad de agua Pr que finalmente llega a la parte de transferencia del modelo está dada por:

$$Pr = Perc + (Pn - Ps)$$

Hidrogramas unitarios

Pr se divide en dos componentes de flujo, el 90% describe un hidrograma unitario, HU1, y se almacena en un depósito y el 10 % describe otro hidrograma unitario, UH2, el cual es simétrico. UH1 Y UH2 dependen del mismo parámetro X4, que es el tiempo en la base de HU1 expresado en días.

Las ordenadas hidrograma se calculan a partir de la curva S denominadas SH1 y SH2, respectivamente, que corresponden a las funciones acumuladas de los hidrogramas.

SH1 se define por una función de tiempo:

- ✓ Para $t \leq 0$ \longrightarrow $SH1(t) = 0$
- ✓ Para $0 < t < X_4$ \longrightarrow $SH1(t) = \left(\frac{t}{X_4} \right)^{\frac{5}{2}}$
- ✓ Para $t \geq X_4$ \longrightarrow $SH1(t) = 1$

SH2 se define de manera similar:

- ✓ Para $t \leq 0$, \longrightarrow $SH2(t) = 0$
- ✓ Para $0 < t < X_4$, \longrightarrow $SH2(t) = \frac{1}{2} * \left(\frac{t}{X_4} \right)^{\frac{5}{2}}$
- ✓ Para $0 < t < 2 X_4$, \longrightarrow $SH2(t) = 1 - \frac{1}{2} * \left(2 - \frac{t}{X_4} \right)^{\frac{5}{2}}$
- ✓ Para $t \geq X_4$ \longrightarrow $SH2(t) = 1$

Los valores en las ordenadas de UH1 Y UH2 se calculan por medio de:

$$UH1 (j) = SH1 (j) - SH1 (j - 1) \quad (4.16)$$

$$UH2 (j) = SH2 (j) - SH (j - 1) \quad \text{con } j = \text{número entero.}$$

En cada paso de tiempo k, los caudales Q9 y Q1 salidos de los dos hidrogramas corresponden a la convolución de las lluvias anteriores por el esquema de repartición dado por el hidrograma discretizado y son calculados por:

$$Q9 (k) = 0.9 * \sum_{j=1}^l UH1 (j) * Pr (k - j+1)$$

$$Q1 (k) = 0.1 * \sum_{j=1}^m UH2 (j) * Pr (k - j+1)$$

Donde:

$$l = \text{int}(X4) + 1 \text{ y } m = \text{int}(2X4) + 1, \text{ int}(\cdot) \text{ denota la parte entera.}$$

Función de intercambio

El intercambio de agua subterránea se calcula a través de:

$$F = X_2 * \left(\frac{R_K}{X_3} \right)^{\frac{7}{2}}$$

Donde:

Rk: nivel en el reservorio al inicio del paso de tiempo,

X3: capacidad del tanque a un día

X2: coeficiente de intercambio de agua que puede ser positivo en el caso que aporte o negativo en el caso de pérdidas a hacia las aguas subterráneas o bien puede ser nulo. La interpretación física de esta función de cambio no es sencilla.

Reservorio de transferencia

El nivel en el depósito de transferencia se modifica mediante la adición de Q9 que proviene del hidrograma HU1 y F:

$$R' = \max(0; RK + Q9(k) + F)$$

El caudal Q_r que se obtiene a la salida del depósito de transferencia es:

$$Q_r = R' * \left\{ 1 - \left[1 + \left(\frac{R'}{X_3} \right)^4 \right]^{\frac{1}{4}} \right\}$$

Entonces el nivel del reservorio se convierte en:

$$RK + 1 = R' - Q_r$$

Total

La salida Q_1 del hidrograma HU2 es sometida al intercambio para dar la componente de flujo Q_d :

$$Q_d = \max(0; Q_1(k) + F)$$

El caudal total Q resulta entonces:

$$Q_k = Q_r + Q_d$$

Parámetros

El modelo GR4J se compone de cuatro parámetros de calibración:

- ✓ X1: Es la capacidad del reservorio de producción en mm.
- ✓ X2: Es el coeficiente de intercambio subterráneo, en mm.
- ✓ X3: Es la capacidad del reservorio de rutaje en mm
- ✓ X4: Es el tiempo de base del hidrograma unitario HU1.

En una amplia muestra de las cuencas hidrográficas, los valores de parámetros que se obtuvieron se entran en la tabla 4.1.

Paramètre	Médiane	Intervalle de confiance à 80%
X_1 (mm)	350	100 à 1200
X_2 (mm)	0	-5 à 3
X_3 (mm)	90	20 à 300
X_4 (jours)	1.7	1.1 à 2.9

Modo pronóstico

El algoritmo funcionando en modo pronóstico incluye un proceso de corrección de los caudales simulados, según los caudales medidos en tiempo real. Un esquema de esto se presenta en la figura 4.2:

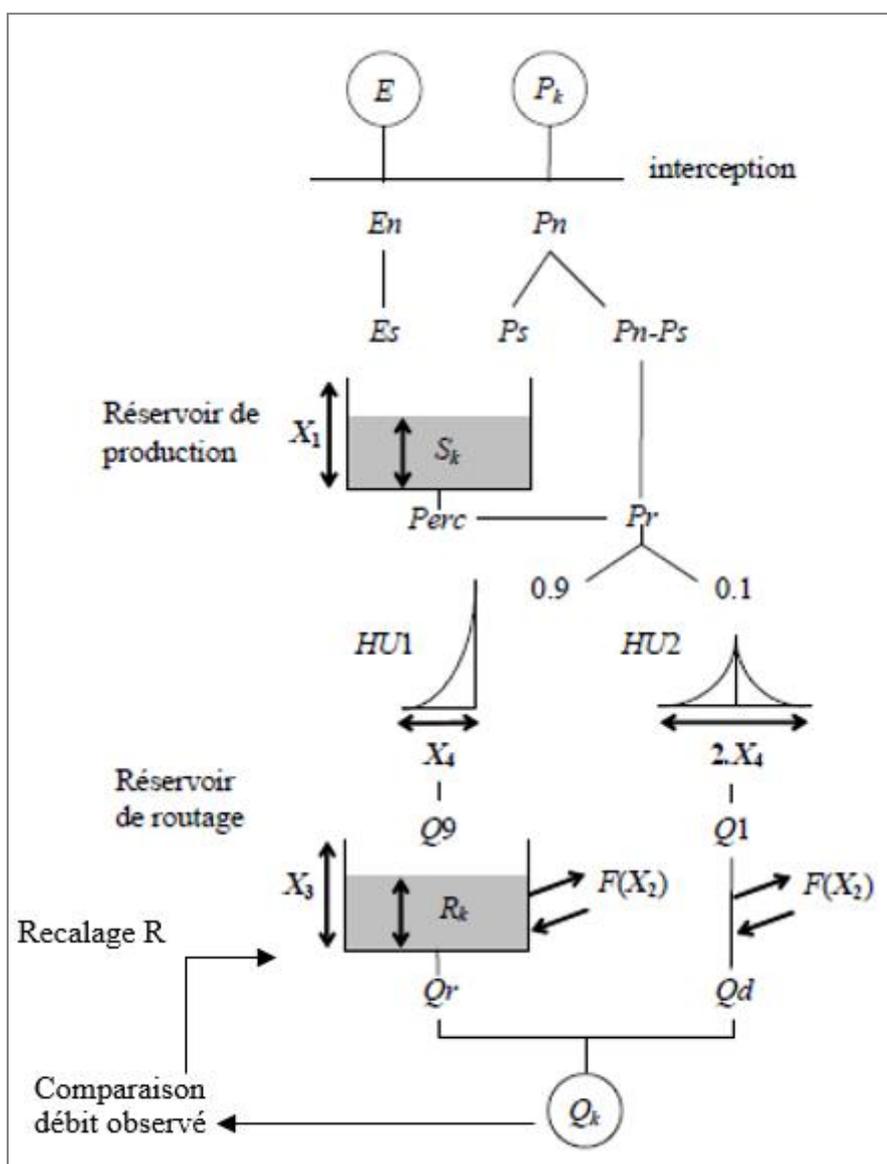


Figura 4.2. Modelo GR4J modo pronóstico.

4.2. EL ALGORITMO MOCOM-UA

Como se ha presentado previamente, el algoritmo MOCOM-UA es un algoritmo de optimización multi-objetivo, ya que está diseñado para optimizar un modelo matemático empleando más de una función objetivo.

Se ha dicho que una función objetivo es una expresión matemática que cuantifica la diferencia entre la salida del modelo hidrológico (hidrograma simulado, si se trata de un modelo de transformación lluvia-caudal), y los datos medidos de referencia (hidrograma medido). Cuanto menor sea dicha función objetivo, mejor habrá sido el ajuste.

Ahora bien, esa “diferencia” que cuantifica la FO, puede ser diferencia de caudales máximos de los hidrogramas, diferencia de volúmenes de escurrimiento directo, diferencias de los caudales de recesión, diferencia de los tiempos en que se produjeron los picos, etc. De acuerdo a la FO elegida para realizar la calibración automática, el juego de valores de parámetros que la minimice será la que mejor representa cierta parte del proceso que simula el modelo.

Por ejemplo, si se elige como FO, aquella que minimiza la diferencia entre caudal máximo simulado y el medido, entonces el juego de valores de parámetros que la minimice, será el que mejor ajusta los picos de los hidrogramas, es decir, probablemente sea el que mejor represente el proceso de transferencia. Pero si se elige la FO que cuantifica la diferencia entre volúmenes de escurrimiento directo de hidrogramas simulado y medido, el juego de parámetros que la minimice será el que mejor representa el proceso de infiltración. Sin embargo, este último juego de valores de los parámetros no tiene porqué coincidir con el anterior. En otras palabras, el conjunto de valores de parámetros que mejor representa la transferencia en la cuenca, no necesariamente sea el que mejor represente la infiltración. En la mayoría de los casos esos valores no sólo no coinciden sino que se diferencian notablemente, al menos en algunos de sus parámetros.

Como consecuencia del análisis anterior, el hidrólogo que va a realizar la calibración automática del modelo de simulación, deberá elegir un

criterio de calibración, el cual dependerá de los objetivos de la simulación. Si por ejemplo, el objetivo de la simulación está asociado al proyecto del vertedero de un embalse, indudablemente que deberá elegirse una FO que conduzca a un juego de valores para los parámetros que permita una simulación adecuada de los picos (y de los volúmenes) de las crecidas. Por el contrario, si se está realizando un estudio de balance hídrico, la FO debe calibrar el modelo para adecuarse a simular con precisión los volúmenes escurridos.

A la vista de la experimentación, ha quedado ampliamente demostrado que un mismo juego de parámetros puede ser una excelente solución para un objetivo y pésima para otro, es decir que puede minimizar notablemente una FO y resultar en valores elevados para otra. En efecto, el juego de parámetros que minimiza una FO, normalmente tiene un valor elevado en la otra, y viceversa. Entre el mejor punto para una FO y el mejor para la otra, existe un conjunto de puntos intermedios que conforman una región de puntos tales que ninguno de ellos es mejor que el otro en ambas FO al mismo tiempo. Esta región se denomina Región de Pareto. Frente a este panorama de divergencias, surgen los métodos de optimización multi-objetivo, que se emplean para obtener la Región de Pareto mediante la minimización simultánea de dos o más FO.

Antes de abarcar la descripción paso a paso del método se introduce el proceso de evolución competitiva y el algoritmo de Nelder y Mead, los cuales son empleados, como se verá más adelante, para llevar a cabo la evolución del complejo.

4.2.1. Evolución competitiva compleja

La Evolución Competitiva Compleja es un proceso de evolución que se lleva a cabo sobre un conjunto de individuos dados. En el contexto del Algoritmo SCE-UA, el conjunto de individuos sometido a esta evolución es el complejo. A continuación se describe su metodología.

4.2.1.1. Atribución de pesos

En primer lugar, se atribuye a cada individuo del complejo un peso tal que los mejores puntos del complejo (menor valor de FO) tienen mayor

peso. Ese peso podría estar dado por la expresión que proponen Duan et al. (1992):

$$\rho_i = \frac{2(m+1-i)}{m(m+1)}, \quad i = 1, 2, 3, m$$

Donde:

- ri: valor del peso asignado al individuo i.
- m: número de puntos en cada complejo.
- i: posición del punto dentro del complejo.

De acuerdo a esta expresión, la suma de los pesos de los m puntos del complejo debe ser igual a 1.

4.2.1.2. Selección de subcomplejos.

Luego se elige aleatoriamente un conjunto de puntos dentro del complejo. Ese conjunto de puntos recibe el nombre de subcomplejo, y naturalmente, debe tener un número de puntos menor o igual al tamaño del complejo. Por ejemplo, si el complejo tiene 18 individuos, el tamaño del subcomplejo debe ser menor o igual a 18. Los puntos del subcomplejo son elegidos de modo tal que tengan mayor probabilidad de ser elegidos aquellos puntos del complejo con mayor peso, aún cuando cualquiera pueda serlo. O sea que el método de selección de los puntos del complejo que pasarán a formar parte del subcomplejo, debe dar mayor probabilidad de ser elegidos a los mejores, aunque los peores también podrán formar parte del subcomplejo.

Siguiendo con nuestra analogía, la selección del subcomplejo es análoga a la selección de los padres (o reproductores) que serán los que den origen a la siguiente generación. De todos los individuos del complejo, interesa elegir como reproductores a los mejores, por eso se le otorga mayor peso a los puntos con menor FO, para que tengan más probabilidad de formar parte del subcomplejo, o sea de reproducir. Sin embargo, los puntos malos también pueden ser elegidos, aunque tienen menor probabilidad de serlo. La única diferencia que existe con la reproducción natural es que, en la natural, los

padres son sólo dos, en cambio, en el subcomplejo (los reproductores) pueden ser más de dos.

Es necesario remarcar los siguientes puntos:

- ✓ El número de individuos de la población puede ser cualquiera, aunque conviene que sea grande.
- ✓ El número de individuos del complejo puede ser cualquiera, siempre que sea un submúltiplo del tamaño de la población.
- ✓ El número de individuos del subcomplejo puede ser cualquiera, siempre que sea menor o igual que el tamaño del Complejo.

Por ejemplo, de una población de 100 individuos pueden elegirse 5 complejos de 20 individuos cada uno. A su vez, de cada complejo pueden tomarse subcomplejos de 13 individuos.

4.2.1.3. Evolución de los Subcomplejos.

Una vez que se elige el subcomplejo, el mismo evoluciona mediante el Algoritmo de Nelder y Mead un número “a” de veces. Luego de esta evolución, el subcomplejo habrá llegado a un nuevo conjunto de puntos. El Algoritmo de Nelder y Mead es un algoritmo de búsqueda local, que se explica más adelante. El subcomplejo con sus nuevos puntos es devuelto al complejo, en donde se reúnen con los que no habían sido elegidos para formar parte del subcomplejo. Es decir que el complejo vuelve a tener la misma cantidad de puntos, pero ahora tiene algunos que han evolucionado y son distintos a los que partieron hacia el subcomplejo antes de que este fuese sometido al algoritmo de Nelder y Mead, y por lo tanto, tienen distinto valor de FO. En consecuencia los puntos están desordenados. Si por ejemplo, de un complejo de 20 individuos, se toman 13 del modo explicado para formar parte del subcomplejo, estos 13 puntos evolucionarán mediante el Algoritmo de Nelder y Mead, pero los otros 7 puntos esperarán pacientes y en su estado inicial el regreso de los 13 que partieron. Una vez que los 13 evolucionaron, se juntan con los otros 7 y vuelven a sumar 20 individuos en el complejo. Pero 13 de esos 20 individuos del complejo ya no son los de antes, por lo tanto tienen distinto valor de FO y en consecuencia es necesario volver a ordenarlos para poderlos pesar nuevamente y continuar con el proceso.

4.2.1.4. Ordenamiento de los puntos del Complejo.

Los puntos del nuevo complejo son reordenados de mejor a peor, o sea, de menor a mayor valor de la FO. Una vez reordenados, se les vuelve a asignar un peso, se vuelve a elegir un subcomplejo, y se vuelve a someter al subcomplejo al algoritmo de Nelder y Mead, y así sucesivamente un número “b” de veces. En otras palabras, se repite del paso “a” al paso “d” un número “b” de veces.

Durante la Evolución Competitiva Compleja, se han elegido y vuelto a mezclar en el complejo “b” veces subcomplejos, cada uno de los cuales fue sometido “a” veces al algoritmo de Nelder y Mead. Al cabo de la misma, el complejo ha evolucionado hacia nuevos individuos distintos a los que salieron.

4.2.2. Algoritmo de Nelder y Mead

El Algoritmo de Nelder y Mead es un método de búsqueda local que se lleva a cabo dirigiendo la evolución de un conjunto de individuos hacia el óptimo local. Una vez que el subcomplejo ha sido seleccionado del modo que se explicó previamente, será sometido a este algoritmo siguiendo los pasos que se explican a continuación.

4.2.2.1. Organización del Subcomplejo.

Los puntos del subcomplejo son organizados de mejor a peor, o sea, de menor a mayor valor de la FO. Si suponemos que nuestro subcomplejo tiene “q” individuos, entonces el primer punto será el de mejor valor de FO, y el punto q-ésimo será el peor.

4.2.2.2. Cálculo del centro de gravedad.

Se calcula el centro de gravedad de los (q-1) mejores puntos del subcomplejo, mediante la expresión:

$$g = \frac{1}{q-1} \sum_{j=1}^{q-1} u_j$$

Donde:

u_j es el vector posición del individuo j-ésimo

g es el vector de coordenadas del centro de gravedad de los $q-1$ mejores puntos.

Notar que si fuesen dos parámetros del modelo, u_j sería un vector con dos coordenadas, o sea, cada punto del subcomplejo sería un punto en el plano cuya evolución podría visualizarse. Si fuesen tres parámetros, sería un subcomplejo que evoluciona en el espacio. Más de tres coordenadas no puede visualizarse, ni imaginarse, pero sí pueden realizarse proyecciones en dos dimensiones del hiperespacio, y analizar las variaciones de las funciones objetivos en estas proyecciones.

4.2.2.3. Cálculo del punto de reflexión.

Si trazamos un segmento que parta del peor punto del complejo, que pase por el centro de gravedad de los $(q-1)$ mejores puntos, y luego se prolongue más allá del centro de gravedad una distancia igual a la distancia que lo separa del peor punto, obtendremos un punto “ r ” equidistante al centro de gravedad con el peor punto. Ese punto así obtenido, se denomina punto de reflexión, y su vector de coordenadas de ubicación puede calcularse como:

$$r = 2g - u_q$$

4.2.2.4. Evaluación del punto de reflexión.

En el caso de que el punto r estuviera contenido dentro del hipercubo de validez de los parámetros, se calcula el valor de su FO y se compara con el valor de la FO del peor punto. Si es menor que aquel, el punto r reemplaza al peor, caso contrario, sigue en el paso siguiente. Ahora bien, si el punto r cayera fuera del hipercubo de validez de los parámetros, es definido el menor hipercubo que contiene todos los puntos del complejo, y se genera un nuevo punto (z) aleatoriamente en el interior de este hipercubo, el cual reemplaza al peor punto.

4.2.2.5. Cálculo del punto de contracción.

Como se dijo en el paso anterior, si r fuese peor que el peor punto del subcomplejo, se calculan las coordenadas de un nuevo punto “ c ”

denominado punto de contracción, y sus coordenadas son calculadas de acuerdo con la ecuación:

$$c = \frac{(g + u_q)}{2}$$

Este punto está ubicado sobre el segmento que une el centro de gravedad de los (q-1) mejores puntos y el peor punto, equidistante a ambos.

4.2.2.6. Evaluación del punto de contracción.

Se evalúa la función objetivo del punto c. Si es menor que el peor punto del subcomplejo, reemplazará a este, en caso contrario, es definido el menor hipercubo que contiene todos los puntos del complejo, y se genera aleatoriamente en su interior un nuevo punto (z).

En síntesis, el algoritmo consiste en sustituir el peor punto del subcomplejo por otro punto. Para lo cual, calcula un punto de reflexión en dirección al baricentro de los mejores puntos. Si se escapa del hipercubo, genera aleatoriamente un punto en su interior y reemplaza al peor. Si no, se analiza si el punto de reflexión es peor que el peor del subcomplejo, siendo así, genera el de contracción. Si este también es peor, define el menor hipercubo que contiene a los puntos del subcomplejo y genera aleatoriamente un punto en su interior. Pero, si el de reflexión (prioritariamente) o el de contracción, son mejores que el peor del subcomplejo, reemplaza al peor, por uno de estos.

Debe notarse que en el método descrito de búsqueda del punto óptimo, se da la oportunidad de subsistir y reproducirse a todos los individuos de la población. Por ejemplo, en la construcción del Complejo, no se descarta ningún punto, sino que en cada complejo hay puntos buenos y malos. En la selección del subcomplejo, se da la oportunidad de salir elegido a todos los puntos del complejo, aún a los peores. De acuerdo al criterio de selección del subcomplejo, podría elegirse como reproductor al peor punto inclusive. Estos criterios, a primera vista contradictorios (ya que permiten la supervivencia de los puntos malos, cuando se buscan los buenos), tienen una propiedad muy interesante que se resume en una palabra: variabilidad. A través de este método se incursiona en todo el hiperespacio de validez de los parámetros,

dando la posibilidad que el modelo se optimice localmente aún en las proximidades de los puntos malos.

4.2.3. Pasos del método

Generación de la población inicial

Se establece el rango de validez de los parámetros del modelo matemático, que definen un hipercubo de validez de los parámetros. Luego, se genera aleatoriamente una población de puntos con distribución probabilística uniforme en su interior.

Por ejemplo, si se estuviese calibrando un modelo de dos parámetros (P1 y P2), cada uno de ellos tendría su rango de validez, de tal modo que se cumpla para cualquier punto (P1, P2) de la población inicial, que:

$$P1_{\min} \leq P1 \leq P1_{\max}$$

$$P2_{\min} \leq P2 \leq P2_{\max}$$

De tal modo que la zona de validez de los valores de los parámetros está dada por un espacio de dos dimensiones (plano), como lo muestra la Figura 4.3.

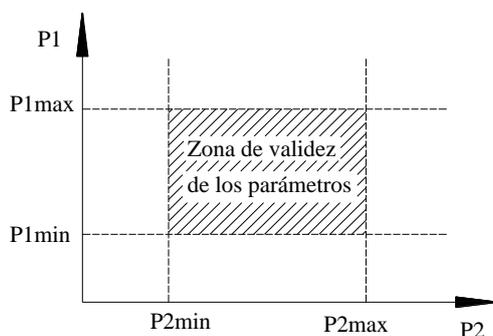


Figura 4.3. Zona de validez de los parámetros del modelo, (Alonso, 2008).

En la Figura 4.4, pueden observarse cuatro puntos A, B, C y D generados aleatoriamente con distribución uniforme en este espacio bidimensional asociado al supuesto modelo hidrológico de dos parámetros P1 y P2. Cada punto generado tiene asociado un valor de cada parámetro, así el punto B por ejemplo tiene asociado un valor de P1, que se ha denominado P1B, y uno de P2, denominado P2B. Como puede verse, el punto B satisface las

desigualdades antes expuestas, que constituyen las restricciones del problema en cuestión, del mismo modo que lo hace el punto C. Mientras que los puntos A y D, están fuera de la zona de validez, por lo tanto no son puntos válidos y no pueden ser considerados en la población.

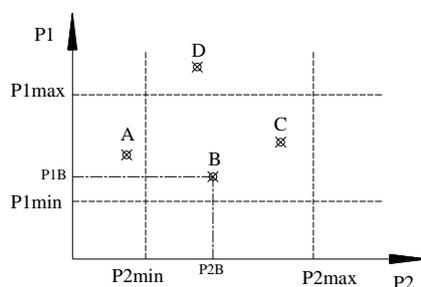


Figura 4.4. Puntos válidos y no válidos (Alonso, 2008).

Si en vez de tratarse de un modelo de dos parámetros, fuese de tres, la zona de validez de los parámetros estaría definida entonces por un espacio de tres dimensiones, tal que los puntos en su interior cumplirían con las dos condiciones anteriores, más la siguiente:

$$P3_{\min} \leq P3 \leq P3_{\max}$$

Extrapolando esta idea a cuatro, cinco y más parámetros, se puede imaginar la existencia de un hipercubo de n dimensiones, que naturalmente no puede visualizarse, dentro del cual pueden caer los puntos aleatoriamente generados. Por ejemplo, el modelo hidrológico GR tiene cuatro parámetros, cada uno de los cuales tiene un rango dentro del cual puede tomar valores posibles o con sentido físico. Una generación aleatoria de puntos para los parámetros del modelo GR, debería caer dentro del hipercubo limitado por los rangos de variación admisibles de estos cuatro parámetros.

De este modo, cada punto generado en el interior del hipercubo de validez de los parámetros, tiene asociado un juego de valores de los parámetros del modelo a calibrar, que constituyen las coordenadas de este punto. Por ejemplo, si el modelo tiene cuatro parámetros, cada punto tiene cuatro valores asociados (que son sus coordenadas en este espacio hiperdimensional), cada uno de los cuales corresponde a cada uno de los parámetros. Si hemos asignado a cada parámetro un valor dentro de su rango de validez, entonces los cuatro valores definirán un punto que caerá indudablemente en el interior del hipercubo imaginario.

La generación de puntos en el hiperespacio sigue, en este método, una ley uniforme de distribución probabilística, pero también podría pensarse en una generación aleatoria con distribución normal, o triangular por ejemplo. Para ello se emplea, por ejemplo, el Método de Monte Carlo.

Cada punto, de este modo generado, es un individuo de una población primitiva y original que se dispone a evolucionar. Es decir, esta población irá evolucionando mediante un reemplazo de sus individuos por otros generados de algún modo. Luego de la evolución, la población resultará ser otra población nueva con otros, pero igual número, de individuos finales.

Clasificación y ordenamiento de los individuos de la población. Para realizar la clasificación se evalúa cada una de las FO para cada uno de los puntos de la población. Luego de la evaluación resultarán puntos que tienen menor valor que otros en todas las funciones objetivos, en cuyo caso no cabe duda que estos puntos son mejores que los otros. Pero el problema se plantea cuando se observa que hay puntos que tienen menor valor que otro en algunas FO, pero mayor en otras FO. No puede establecerse un orden de prioridad si hay puntos mejores que otros con el criterio de una de las FO, pero peores con el criterio de la otra. Por ejemplo, supóngase que se tiene dos FO, denominadas FO1 y FO2. Si un punto A tiene menor valor de FO1 que otro B, pero mayor valor de FO2, entonces no puede decirse que A es mejor que B, ni tampoco que B es mejor que A. Entonces surgen los conceptos de puntos Dominados y puntos No Dominados.

Son puntos Dominados aquellos tales que existe otro punto que es mejor que ellos en todas las FO. Y por consiguiente, los restantes puntos son No Dominados. Entonces, se eligen todos los puntos no dominados de la población y se dice que forman el Nivel 1. Una vez que se retiran de la población esos puntos no dominados, habrá ahora, otros puntos que son dominados, y otros no dominados. Entonces se retiran los nuevos puntos no dominados y se dice que pertenecen al Nivel 2. Una vez retirados estos puntos no dominados, se toman los restantes y se identifican los nuevos puntos no dominados, y así sucesivamente, hasta que resulte un conjunto de puntos tal que todos ellos son no dominados.

De acuerdo al esquema de jerarquización antes explicado, en cada nivel habrá un conjunto de puntos. Se dice que los puntos del nivel 1, son todos

mejores que los puntos del nivel 2, y estos mejores que los del nivel 3, y así sucesivamente. Pero no hay manera de identificar cuál de los puntos del nivel 1 es el mejor. Puede suponerse que todos los puntos de un mismo nivel tienen la misma jerarquía. El ranking de este modo logrado se denomina Ranking de Pareto.

Construcción de complejos. Una vez que los puntos de la población se han jerarquizado en niveles, se les atribuye peso a cada uno de los puntos. Todos los puntos de un mismo nivel tendrán el mismo peso, pero mayor peso que los puntos de niveles inferiores. Por ejemplo, los puntos del nivel 4 tendrán mayor peso que los del nivel 5.

Luego se eligen tantos complejos como puntos tiene el último nivel. Por ejemplo, supongamos que luego de la jerarquización, se obtengan 8 niveles, y que en el nivel octavo (el último) hayan quedado 25 puntos de la población. Entonces, deben elegirse 25 complejos de la población, tales que, cada complejo debe tener uno de los puntos del último nivel y el resto de los puntos del complejo se eligen de la población aleatoriamente, pero dándole mayor probabilidad de ser elegido a los que tienen mayor peso, o sea, a los mejores. Si, por ejemplo, cada complejo debe tener 12 puntos, entonces, uno de esos será uno de los 25 correspondientes al último nivel (en nuestro ejemplo, al octavo nivel), y cada uno de los restantes 11 puntos que falta agregar, se eligen de la población en forma aleatoria, pero ponderada, dándole mayor probabilidad de ser elegido a los de mejor nivel.

Evolución independiente de cada complejo. Una vez que la muestra ha sido separada en p complejos, se da a cada complejo la oportunidad de evolucionar en forma independiente mediante la Evolución Competitiva Compleja. Con la diferencia de que no se eligen subcomplejos, sino que evolucionan los complejos mismos de acuerdo a Nelder y Mead. Se repite para cada complejo varias veces el algoritmo de Nelder y Mead. Pero como hay dos funciones objetivos el criterio de aceptación de un punto en el algoritmo de Nelder y Mead es que sea no dominado. Por ejemplo, cuando se genera el punto de reflexión, se verifica que no haya ningún punto que tenga menor valor en ambas FO; si es así, entonces se acepta el punto, puesto que es no dominado. Cuando los complejos han evolucionado independientemente, se

devuelven a la población que ya está más optimizada, y se vuelve al paso 2, es decir, se vuelven a ordenar y jerarquizar, etc.

Los pasos del 1 al 4 descriptos se repiten hasta que no sea posible encontrar puntos dominados. Es decir que todos los puntos de la población se encuentren en el nivel 1 del ranking de Pareto. Como consecuencia, el resultado final de este algoritmo no es un punto óptimo, como lo era el SCE, sino una población óptima de puntos que forma la región de Pareto.

5. RECOPIACIÓN Y PROCESAMIENTO DE INFORMACIÓN

En este apartado se recopila y ordena todas las informaciones necesarias para conformar los datos de entrada a los modelos hidrológicos descriptos con anterioridad, pertenecientes a dos cuencas hidrográficas argentinas:

Cuenca del Río Salado, más precisamente la porción que se encuentra en la zona de Santa Fe, ya que como más adelante se va a tratar, es la extensión que realmente interesa para el análisis.

Cuenca del Río Gualeguay, ubicada en la provincia de Entre Ríos.

Durante el periodo de esta práctica supervisada, sólo fue posible llevar la calidad de los datos al estado final requerido para modelación, en el caso del río Gualeguay. En el caso del Salado, no fue posible concretar esta tarea debido a causas que se exponen a continuación. Sin embargo, se dejó planteado el procedimiento a seguir para la depuración de esta información, por lo que podrá avanzarse sin dificultad en futuras instancias, y llegar finalmente a la calidad requerida.

El procesamiento de los datos para ambas cuencas, consistió en la estimación de las variables de entrada del modelo hidrológico Génie Rural, que como se manifestó en el capítulo anterior, es necesario contar con series continuas de Precipitación (P), Evapotranspiración (E) y Caudal observado (Q), para permitir la calibración del mismo, y su posterior verificación.

Ahora bien, existen algunas razones que obligan a un tratamiento de la información para la obtención de la serie definitiva, como son las siguientes:

✓ Las variables no siempre son medidas en forma directa. Normalmente, sólo la precipitación se mide en la misma unidad en que se requiere en el modelo. El caudal en cambio, se deduce en general de los niveles, y la Evapotranspiración potencial, se deduce al menos de la temperatura.

✓ Las variables medidas no siempre son registradas en el paso de tiempo requerido. A veces, se mide en pasos de tiempo mayor o menor, lo cual obliga a interpolar, o agrupar respectivamente. Dependiendo del tipo de variable será distinto el tratamiento.

✓ Normalmente hay periodos no registrados, que hay que rellenar con algún criterio, o simplemente no se dispone, y produce una interrupción en la serie.

✓ Existe un alto grado de incertidumbre asociado a los datos. Esto se manifiesta normalmente cuando se contrasta precipitación con caudal. Se supone que estas dos variables tienen implícitas una relación causa-efecto que debe observarse en la mayor parte de los eventos al menos. Estas falencias en los datos son originados por múltiples causas que no interesa abordar a los fines de este estudio, pero sí purgar los errores para llevar adelante la calibración con datos confiables.

En consecuencia, se procedió a analizar la información y trabajar en cada uno de los aspectos manifestados, a los fines de propender a series, o tramos de estas, confiables.

5.1. TIEMPO DE RESPUESTA DE LAS CUENCAS

El tiempo de respuesta de una cuenca depende de factores tales como la pendiente y los cauces que las conforman, la bifurcación del sistema de drenaje, la permeabilidad del suelo, la longitud de su cauce principal, etc.

En cuanto a la pendiente se puede señalar que cuanto mayor sea esta más rápida será la respuesta de la cuenca.

La bifurcación del sistema de drenaje se puede indicar con el orden de corrientes. El orden de corrientes se determina como se muestra en la figura 5.1. Una corriente de orden 1 es un tributario sin ramificaciones, una de orden 2 tiene solo tributarios de primer orden, etc. Dos corrientes de orden 1 forman una de orden 2, dos corrientes de orden 3 forman una de orden 4, etc., pero por ejemplo una corriente de orden 2 y una de orden 3 forman otra de orden 3. El orden de una cuenca es el mismo que el de la corriente principal en su salida; así, por ejemplo el orden de la cuenca de la Figura 5.1. es 4.

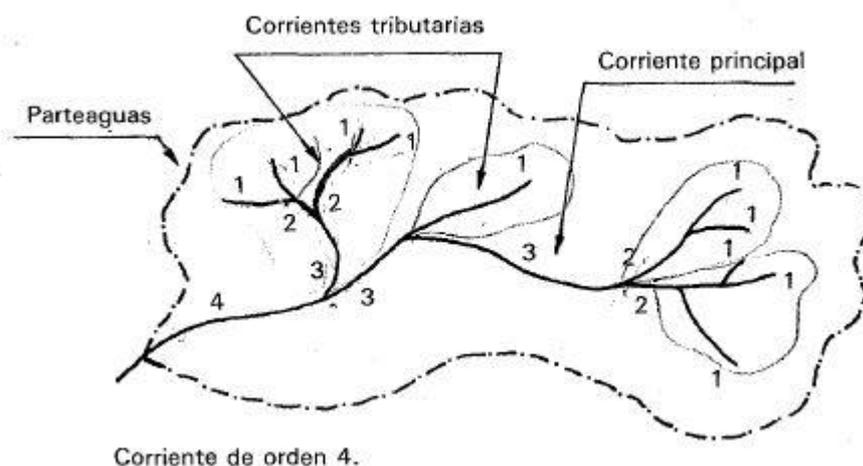


Figura 5.1. Cuenca hidrológica, (Mijares, 1992)

Otro indicador del grado de bifurcación son: la densidad de corrientes D_s definida como el número de corrientes perennes e intermitentes por unidad de área y la densidad de drenaje D_d , definida como la longitud de corrientes por unidad de área.

Un orden de corriente alto o una densidad elevada refleja una cuenca altamente disectada, que responde rápidamente a una tormenta.

La elección de una escala de tiempo adecuada entre datos sucesivos de la serie continua es de suma importancia. Si el paso de tiempo elegido es más largo que el necesario, se corre el riesgo de perder información,

o en caso contrario se estaría detallando en vano para un tiempo más corto, implicando elevados costos computacionales sin utilidad.

Para cuencas de grandes extensiones, la variación de los caudales de sus ríos se aprecia para períodos de varios días, es decir que diariamente la magnitud de la variación de los caudales es importante en relación a los picos del hidrograma, ahora bien resulta evidente que si se eligiese un dato horario, no se perdería información, pero no tendría sentido un análisis tan detallado porque la variación horaria del caudal en relación a la magnitud de los mismos picos sería insignificante.

En conclusión en función de las características propias de las cuencas en estudio se optó por un paso de tiempo diario justamente por cubrir ambas grandes superficies de extensión, con el consecuente elevado orden de magnitud de los caudales de sus ríos principales.

5.2. CUENCA DEL RÍO SALADO

5.2.1. Descripción de la cuenca

Ubicación geográfica

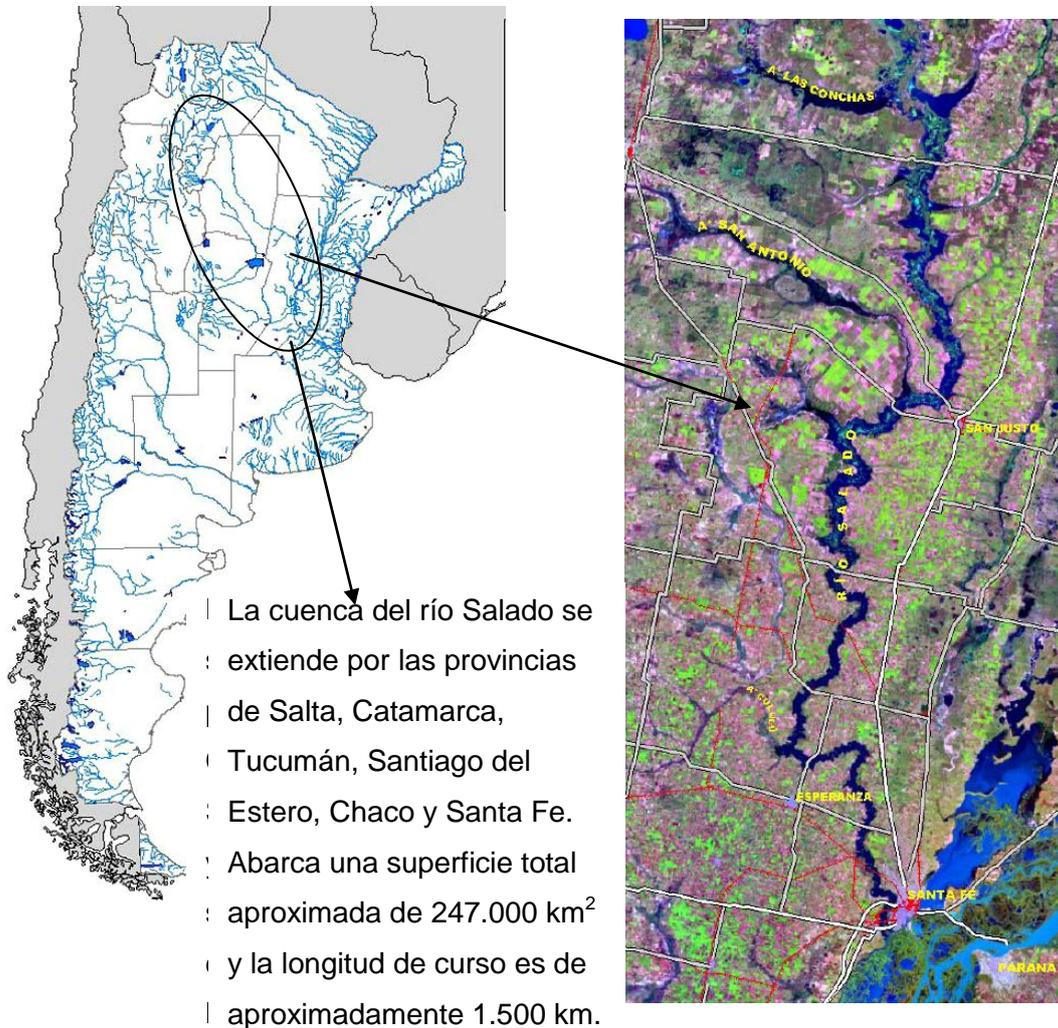


Figura 5.2.:Cuenca del Río Salado, Argentina

Características generales

La cuenca puede caracterizarse en 3 grandes componentes, de funcionamiento hidrológico diferenciado: superior, medio e inferior:

✓ La *cuenca superior* está comprendida entre las nacientes, en el borde oriental de la Puna, y la zona de El Tunal (localidad de J.V. González). En esta región tanto el curso principal (denominado Pasaje o Juramento), como sus afluentes poseen características de río de montaña y de pedemonte.

✓ La *cuenca media* se extiende entre la localidad de J.V. González y el ingreso a la provincia de Santa Fe. Al ingresar en la provincia de Santiago del Estero el curso toma el nombre de Salado y adopta las características generales de un río de llanura, sin recibir afluentes.

✓ Se considera como *cuenca inferior* del río Salado al área que se desarrolla en la provincia de Santa Fe a partir del punto en el cual el río ingresa a la misma, a la altura de la ciudad de Tostado. Dicha área posee una superficie de 55.950 km². El escurrimiento en la cuenca inferior del río Salado se compone de los aportes provenientes de la cuenca media, que son de baja significación, de los aportes del río Calchaquí (significativos) y de los excesos de lluvia colectados fundamentalmente en las sub cuencas de margen derecha.

Dinámica Hídrica

Para interpretar la dinámica hídrica del río Salado es preciso considerar su funcionamiento como unidad integrada, considerando su vinculación con el espacio geográfico que lo rodea.

Es posible distinguir tres áreas, que caracterizarían la dinámica hídrica del río, además de una zona de transición (ver figura 5.3.), y son:

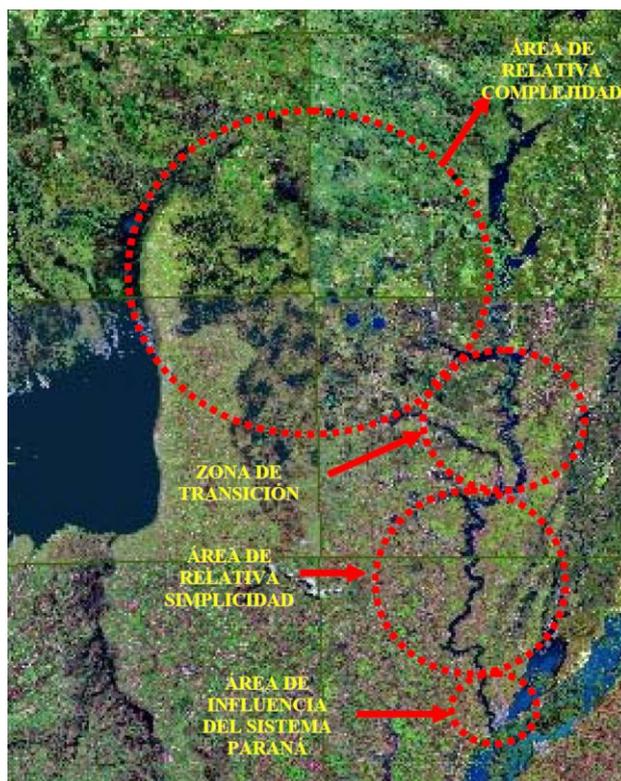


Figura 5.3.: Caracterización de la dinámica hídrica del Río Salado

✓ *Áreas con procesos de relativa complejidad:* Esta contiene, como tributarios, dos componentes fundamentales: los ríos Calchaquí y Salado respectivamente.

El primero es el de mayor relevancia ya que resulta el desagüe de la región de los Bajos Submeridionales. La complejidad radica en las características de este espacio geográfico.

La región de los Bajos Submeridionales consiste en un área muy plana, cuya dinámica está determinada por los escurrimientos denominados areales o mantiformes. Los mismos conforman sistemas hídricos denominados “no típicos”, SHNT.

Los SHNT están caracterizados por interrelaciones muy complejas. En los mismos las condiciones de manejo y uso del suelo poseen consecuencias que no siempre es posible advertir a priori.

Esta región de Bajos Submeridionales constituye un sistema prácticamente asimétrico. El mismo reúne a los aportes provenientes de la provincia del Chaco, conducidos mediante la cañada de las Víboras y otros escurrimientos menores, a los provenientes de la provincia del

Santiago del Estero, conducidos en este caso por la Cañada de los Saladillos, y a los provenientes de la dorsal Occidental. Se suman, a su vez, los aportes generados dentro de la misma región de los Bajos, en la propia provincia. Todos ellos son tributados a lo largo del sistema de encadenamiento del arroyo Golondrinas, el cual se recuesta sobre el borde Oeste de la dorsal Oriental.

Esta planicie ha sido barrida, históricamente, bajo diferentes condiciones climáticas y estructurales, por un sistema fluvial cuyos rasgos se manifiestan por relictos enmascarados por los diferentes procesos resultantes y por un sinnúmero de lagunas acordes con la morfometría de esos sistemas no funcionales.

Los aportes son muy lentos y se generan a partir de cierto umbral de acumulación. Este subsistema almacena importantes volúmenes. Cuando su capacidad máxima se satisface, se producen escurrimientos laminares o relativamente encausados, aportando volúmenes muy importantes al punto de convergencia (caracterizados por hidrogramas muy amortiguados, pero de gran permanencia). Esto hecho es muy importante a tener en cuenta, ya que dicha área constituye la de mayor aporte al río Salado en territorio santafesino. De acuerdo a sus condiciones de humedad antecedente y a la magnitud de las lluvias, la misma puede ser la responsable de la generación de importantes volúmenes de escurrimiento que aportan a la región de su desembocadura en el sistema Paraná.

El segundo componente inicialmente citado, el río Salado, se caracteriza por desarrollarse antes de su ingreso a la región de transición antes mencionada, en un contexto muy disímil.

Desde el punto de vista morfogenético y morfométrico el río Salado es muy diferente a lo anteriormente descripto. Su ingreso a la provincia de Santa Fe reconoce un condicionamiento geológico, presentando un trazo muy particular al sur de Tostado. Su diseño está inserto en la llanura aluvial que el mismo construyó. Presenta un cauce definido, muy meandroso. Se caracteriza por presentar albardones naturales de aproximadamente 150 m de ancho y 1 m de altura. En este sector el río carece de un valle aluvial. Ello implica que posee gran capacidad para divagar en amplios rangos espaciales.

Este proceso de divagación se manifiesta en el abandono de meandros, ya sea por extrangulamiento o por desbordamiento, y en el fenómeno de avulsión. Este último significa que en cortos períodos de tiempo (en la escala natural de los procesos), el río es capaz de cambiar su trayectoria, abandonando su cauce a partir de cierto punto. Las causas de estos comportamientos son variadas, incluyéndose, entre otras, a los problemas sedimentológicos y a los cambios de pendiente general provocados por impedimentos de origen tectónico o antrópicos.

Lo más destacado de este tramo del río Salado es que el mismo no presenta una trascendencia, en términos comparativos y cuantitativos, como tributario a la zona de transición antes mencionada.

✓ *Zona de Transición:* Determinada por una llanura fluvial de alta complejidad dinámica, que se va simplificando en función del incremento de los aportes. Este tramo recibe, fundamentalmente, los aportes del A⁰ Las Conchas. El derrame del sistema Las Conchas en este valle es el que da origen al cauce del río Salado propiamente dicho hacia aguas debajo de su confluencia.

Dentro de esta zona de transición, y en forma paralela a la conformación del río Salado propiamente dicho, el A⁰ Calchaquí pierde sus condiciones morfológicas. A partir de allí existe un único cauce cuyo conjunto de cauce y valle recibe la denominación de Salado.

Hacia aguas debajo de la zona de transición se presenta un esquema de funcionamiento relativamente más simple, ya que el río en sus diferentes estadios (en aguas bajas, medias o altas) se linealiza, en función de este orden, adquiriendo jerarquía y organización definible.

✓ *Áreas con procesos de relativa simplicidad:* En esta área, el río Salado se desarrolla en un cauce y su respectivo valle con contornos bien definidos. Su estructura espacial continúa con una relación asimétrica en lo que se refiere a los tributarios de ambos márgenes, prevaleciendo la derecha, donde existen importantes arroyos (Las vizcacheras, San Antonio, Arizmendi y Cululú-Las Prusianas).

Además de estas contribuciones, a lo largo de todo su recorrido existen innumerables aportes de pequeña extensión. En sentido perpendicular,

colectando una importante faja areal, se observan problemas de ablación generalizada.

El valle, al estar condicionado tectónicamente, es bastante encajonado. Se observa un importante gradiente topográfico hacia la planicie o valle de inundación, hecho que implica una importante capacidad de evacuación y una muy baja probabilidad de desborde de esta última.

Los arroyos afluentes antes citados tienen comportamientos intrínsecos diferenciales, pero conservan un patrón definido en la relación con el curso del Salado. Son arroyos que presentan pendientes superiores a éste último, por lo que determinan una evacuación y contribución que, en términos relativos, es rápida.

Otra particularidad del sistema analizado radica en que el proceso de tributación se produce tipo “cauce con cauce”, dentro de los valles, en el punto de confluencia. Esto significa que, en la relación entre cauces, existe un sistema organizado de régimen permanente, en equilibrio dinámico.

Todos los arroyos presentan, en general, fajas de erosión generalizada e importantes problemas salinos.

✓ *Área de influencia del sistema Paraná:* Este tramo presenta dos razones fundamentales, para ser discretizado del conjunto. Es el tramo de incidencia directa a la ciudad de Santa Fe y, por otra parte, está directamente influenciado por la actividad propia del valle del Paraná. Esto hace que en su tramo final, el río Salado presente comportamientos no vinculados exclusivamente a su propia dinámica.

El límite norte de esta área se ubica, en términos prácticos, en proximidades de la Ruta Provincial 70. La determinación de su límite sur es, quizás, lo más significativo. Este tramo está planteado desde el punto de vista de la influencia del sistema del Paraná, siendo por lo tanto necesario realizar una distinción entre “efecto de remanso” y “espacio ambiental”, de acuerdo a la génesis correspondiente.

En este sentido el fin de la diferenciación evidente entre estos dos ambientes (Salado y Paraná) se da aproximadamente a la altura de las vías del FFCC Mitre y el límite Norte del ejido urbano de Santo Tomé. Allí se observa

que dentro del propio valle del río Salado, existen (al sur) depósitos aluviales y una morfometría que corresponde, en realidad, al sistema del Paraná.

Zona de interés

La zona que se le va a dar importancia en cuanto a la recolección de información va a ser la que de relativa simplicidad, es decir la que se encuentra en la provincia de Santa Fe, ya que las demás zonas presentan características que a priori no justifican su análisis en el presente trabajo.

5.2.2. Precipitaciones

Información recolectada

La información disponible consta de series de precipitaciones medias diarias pertenecientes a 19 estaciones posicionadas en la región de la cuenca del Río Salado en la Provincia de Santa Fe, ellas son:

✓	Garabato	✓	Santa Fe
✓	Logroño	✓	Moisés Ville
✓	Vera	✓	Eusebia
✓	Calchaquí	✓	Casablanca
✓	Huanqueros	✓	Rafaela
✓	La Cabral	✓	Grutly
✓	San Cristóbal	✓	Laguna Paiva
✓	Ñanducita	✓	Pilar
✓	Marcelino Escalada	✓	Angélica
✓	Frank		

Se recopilaron 19 archivos en formato Excel nombrados a través de una serie numérica que indica un código de estación, cada uno contiene una matriz rellena con las mediciones efectuadas, donde cada columna representa los días del mes y las filas los meses con su respectivo año.

A priori no se identificaba a que estación representaba el nombre de tipo numérico, de aquí que se dió uso a una tabla en formato imagen (escaneada, figura 5.4.) que contenía los códigos necesarios y el nombre apropiado con el fin de identificar cada una de las estaciones por medio de su nombre (lugar de medición), departamento, datos de coordenadas geográficas, cantidad de datos recabados, entre otras.

De esta manera se pueden localizar en una imagen geográfica y poder posicionarlas respecto a la configuración de la cuenca

DISKETTE								
N 21								
CODIGO	NOMBRE	DPTO.	PERIODO	TOTAL	DATOS	% DATOS	DATOS	% DATOS
EST.				DATOS	FALTAN.	FALTAN.	OBSERV.	OBSERV.
550201	INDEPENDENCIA	9 DE JULIO	'56-88	396	99	25	297	75
550202	TOSTADO MET	9 DE JULIO	'59-66	96	69	71.875	27	28.125
550204	TOSTADO FC	9 DE JULIO	'56-90	420	104	24.7619	316	75.2381
550206	LOS CHARABON	9 DE JULIO	'56-87	384	173	45.05208	211	54.94792
550208	EST. EL ORDEN	9 DE JULIO	'56-62	84	15	17.85714	69	82.14286
550210	ANTONIO PINI	9 DE JULIO	'82-88	84	13	15.47619	71	84.52381
550301	LAS AVISPAS	9 DE JULIO	'56-90	420	142	33.80952	278	66.19048
550302	NUÉVA ITALIA	9 DE JULIO	'56-90	420	144	34.28571	276	65.71429
551101	POZO BORRADO	9 DE JULIO	'56-90	420	113	26.90476	307	73.09524
551102	VILLA MINETI	9 DE JULIO	'56-90	420	166	39.52381	254	60.47619
551103	FOR. ATAHUALPA	9 DE JULIO	'68-88	252	103	40.87302	149	59.12698
551201	STA. MARGARITA	9 DE JULIO	'56-86	372	107	28.76344	265	71.23656
551202	EL NOCHERO	9 DE JULIO	'56-90	420	261	62.14286	159	37.85714
551203	PADRE ITURRAL	9 DE JULIO	'84-88	60	9	15	51	85
551401	KM 486	9 DE JULIO	'56-90	420	169	40.2381	251	59.7619
				4668	1687		2981	
DISKETTE								
N 22								
CODIGO	NOMBRE	DPTO.	PERIODO	TOTAL	DATOS	% DATOS	DATOS	% DATOS
EST.				DATOS	FALTAN.	FALTAN.	OBSERV.	OBSERV.
555301	CALCHAQUÍ	VERA	'56-90	420	69	16.42857	351	83.57143
555302	MARGARITA	VERA	'56-88	396	110	27.77778	286	72.22222
555303	LA GALLARETA	VERA	'56-95	480	152	31.66667	328	68.33333
555304	ESPIN	VERA	'57-78	264	47	17.80303	217	82.19697
555401	VERA MET	VERA	'56-61	72	12	16.66667	60	83.33333
555402	VERA FC	VERA	'56-90	420	43	10.2381	377	89.7619
555403	OCLIVIE	VERA	'56-69	168	31	18.45238	137	81.54762
555404	GAYCURÚ	VERA	'56-77	264	112	42.42424	152	57.57576
555405	FOR. OLMOS	VERA	'76-79	48	21	43.75	27	56.25
555406	TOBA	VERA	'82-90	108	48	44.44444	60	55.55556
556301	GARABATO	VERA	'56-88	396	59	14.89899	337	85.10101
556302	COLMENA	VERA	'56-78	276	84	30.43478	192	69.56522
556303	INTIYACO	VERA	'56-88	396	52	13.13131	344	86.86869
556304	FLORIDA	VERA	'56-69	168	41	24.40476	127	75.59524
556305	GOLONDRINA	VERA	'56-77	264	105	39.77273	159	60.22727
556306	KM 320	VERA	'82-88	84	62	73.80952	22	26.19048
556401	CAÑADA OMBÚA	VERA	'56-78	276	150	54.34783	126	45.65217
				3684	1019		2665	

Figura 5.4.: Imagen escaneada de códigos de estaciones
Constituye una de las 12 imágenes disponibles.

Lo recabado pertenece a la base de datos del Servicio Meteorológico Nacional, a través del Ing. Luis Lenzi.

Procesamiento de información

Estaciones de medición de precipitaciones

A partir del reconocimiento de las 19 estaciones (remitirse al fragmento anterior) se procedió a la sus localizaciones en Google Earth, en buscas de clarificar la distribución de las mismas a lo largo de extensión de la cuenca (véase figura 5.5).

Haciendo uso de las coordenadas correspondientes, se las ubicó en un plano (figura 5.6.) a través del Software Auto Cad 2010, para proceder a la determinación de los polígonos de Thiessen, necesarios para la obtención de las precipitaciones medias que conformarían el archivo de entrada para la programación.

Una vez localizadas las estaciones, se procedió al desglose de la cada una de las series de matrices en una única tabla, con la finalidad de abordar a un panorama claro del conjunto de mediciones diarias y establecer conclusiones en cuanto a su depuración y/u ordenamiento.

Las series contaban con innumerables datos, en incluso con días donde se efectuó más de una medición, para ello se debió procesar la información de manera tal de dejar en cada día un único dato (el mayor) siguiendo criterios de relleno que se exponen en el desarrollo del apartado, a continuación.



Figura 5.5.: Localización de estaciones - Google Earth



Figura 5.6.: Localización de estaciones en plano Auto Cad

Ordenamiento

✓ Primera etapa:

La serie de datos abarcaba períodos de una cierta extensión pero con rangos donde no se disponía de ninguna medición, presentándose la incertidumbre lógica de pensar si en esos períodos realmente se efectuaron trabajos de medición o no.

Como lo que realmente se necesita es una serie continua se creó una macro en Excel combinada con funciones propias del programa, donde su finalidad era en principio trasponer los datos para disponer de una serie dispuesta en una columna y luego en orden rellenar los períodos sin datos de manera tal que quede la serie completa y continua con un único dato diario.

Esta metodología se repitió para cada estación en particular, llegando a una tabla final de datos (con mediciones o no) “estaciones vs tiempo (fecha)”, la figura 5.7. representa un fragmento.

✓ Segunda etapa:

Con la tabla a la que se llegó en la etapa anterior se procedió a ordenar según la cercanía entre las estaciones, para obtener una grafica de las estaciones cercanas entre sí vs tiempo. Se utilizó una metodología de asignarle colores a las estaciones que conformaban un grupo de cercanía (fila 1), y a su vez otro color a pares de estaciones muy cercanas entre sí (fila 2), para de esta manera encontrar criterios aceptables de relleno de datos según el faltante en cada estación.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
1		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
2		GARABATO	VERA	CALCHAQUI	LOGROÑO	HUANQUEROS	LA CABRAL	SAN CRISTOBAL	MANUCITA	MARCELINO ESCALADA	MOISES VILLE	EUSEBIA	CASABLANCA	RAFAELA	GRUTLY	LAGUNA PAIVA	SANTA FE	FRANCK	PILAR	ANGELICA	
3		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	T
4	FECHA	ESTACIONES																			
5		556301	555402	555301	550105	454407	454406	454310	454403	459309	454307	454103	454102	453208	453407	458203	458114	458207	453401	453112	
13	08/01/1956	9,00	7,00	12,00	0,00	0,00	1,00	17,00	3,00	13,00	0,00	0,00	3,00	5,00	0,00	6,00	6,00	17,50	16,00	0,00	
14	09/01/1956	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
15	10/01/1956	0,00	0,00	0,00	25,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
16	11/01/1956	33,00	24,00	21,00	0,00	0,00	3,00	10,00	0,00	9,00	0,00	0,00	0,00	0,00	0,00	6,00	17,00	0,00	9,00	0,00	

Figura 5.7.: estaciones cercanas vs tiempo

Se ve en la figura que las estaciones 5, 6, 7 y 8 conforman un grupo de cuatro estaciones cercanas y a su vez la 5-6 y 7-8 conforman pares de aún mayor cercanía.

Es importante aclarar que tras ordenar los datos de esta manera queda definidas las zonas del norte, central y sur en correspondencia con el orden numérico de menor a mayor de cada estación respectivamente.

✓ Tercera etapa:

En esta etapa se procedió al relleno de las series en base a criterios aceptables para lograr la menor cantidad de tipologías de combinación de datos diarios, que en definitiva van a conformar las tipologías de polígonos de Thiessen o Voronoi necesarios para determinar las precipitaciones medias.

Se enuncian a continuación los criterios que se utilizaron para rellenar la serie:

1. El primer criterio es uno de los más simples, consiste en rellenar la serie con ceros si en la mayoría de las estaciones predomina dicha nulidad, ya que resultaría ilógico que si en la mayor parte de la región no ha llovido justamente en esa estación puntual se disponga de un dato no nulo.

03/04/1956	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-1,00	0,00	-1,00	0,00	0,00
04/04/1956	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-1,00	0,00	-1,00	0,00	0,00

Antes de rellenar

	GARABATO	VERA	CALCHAQUI	LOGROÑO	HUANQUIPOS	LA CABRAL	SAN CRISTOBAL	MANDUCITA	MARCELINO ESCALADA	MOISES VILLE	EUSEBIA	CASABLANCA	RAFAELA	GRUTLY	LAGUNA PAIVA	SANTAFE	FRANCK	PILAR	ANGELICA
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
FECHA	ESTACIONES																		
	556301	555402	555301	550105	454407	454406	454310	454403	459309	454307	454103	454102	453208	453407	458203	458114	458207	453401	453112
03/04/1956	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
04/04/1956	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Luego de aplicar el criterio

2. Se considera aquí cambiar datos de valores inferiores a 3mm por ceros (no ha llovido), siempre y cuando predomine en la región tal circunstancia de escasas de lluvia, es lógico pensar que un valor muy cercano al nulo corresponda a un error en la medición o bien no se lo utilice porque en definitiva tras efectuar la media no representaría un orden de precipitación que haga que los caudales se vean modificados.

3. En parte es bastante similar al criterio anterior, consiste en primera instancia determinar la sumatoria de los valores de toda serie diaria, y desechar aquellos días donde esta sumatoria no supere el valor de 10mm, reemplazando la serie por valores nulos, siendo bastante lógico ya que un valor inferior a ese para semejante magnitud areal no correspondería a uno representativo.

4. Este criterio consiste en rellenar con un valor promedio de los datos de estaciones aledañas, siempre y cuando sea un solo dato y pertenezca a un mismo grupo estaciones según la cercanía, utilizando para esto lo propio de la segunda etapa. Por ejemplo si en el grupo de las estaciones 5-6-7-8 (ver

figura 1), las últimas tres tienen valores nulos, sería lógico suponer que la 5 sería también nula, por más que en estaciones lejanas haya llovido, de la misma manera suponer un valor promedio entraría en el mismo criterio.

5. Este corresponde al grupo de estaciones ubicadas en la zona sur (color gris en la figura 1), ya que estas siete estaciones se encuentran muy cercanas entre sí conformando una zona aislada claramente definida, es por esto que se considera un buen criterio rellenar más de un dato según la preponderancia de valores de las estaciones del grupo.

6. En días donde no se disponía de datos en la mayoría de las estaciones, se relleno toda la serie con valores nulos, a pesar de no ser muy criterioso desde el punto de vista conceptual se consideró como la única opción disponible, ya que en definitiva el polígono correspondiente no representaría correctamente la precipitación media de la región.

7. Este criterio surge porque al seguir los anteriores, las tipologías seguían siendo numerosas, lo cual escapaba al objetivo de esta etapa, es por ello que a través de un criterio poco conservador se decidió por llenar datos en base a valores promedio en las estaciones 1-2-3-4 que conforman la zona norte pero no presentan una cercanía tal para aplicar el criterio de la misma forma que los de la zona sur (paso 5).

Polígonos de Thiessen o Voronoi

Una vez rellena la serie se determinan las distintas tipologías de polígonos de Thiessen considerando solo las estaciones con mediciones.

Se hicieron 22 polígonos de Thiessen a través del software ArcGis y Auto Cad, para realizarlos se necesitó la polilínea de la extensión de cuenca y los puntos de las estaciones en formato Auto Cad para que sean transformados en archivos tipo Gis, acto seguido se procede a la determinación automática de los polígonos para luego rectificarlos utilizando nuevamente Auto Cad, dicha rectificación apunta a completar áreas de algunos polígonos que quedan sin determinar para no arrastrar errores de cálculo (véase en las figuras 5.8. como en la parte izquierda hay polígonos incompletos), finalmente a través de una

tabla de Excel se determinan los coeficientes de influencia areal de cada estación.

A continuación se presentan algunas tipologías de polígonos para que se pueda ver en forma clara como cambian las influencias areales para cada tipología:



Figura 5.8a.: Polígono de Thiessen 1 (Software Arc Gis 10)

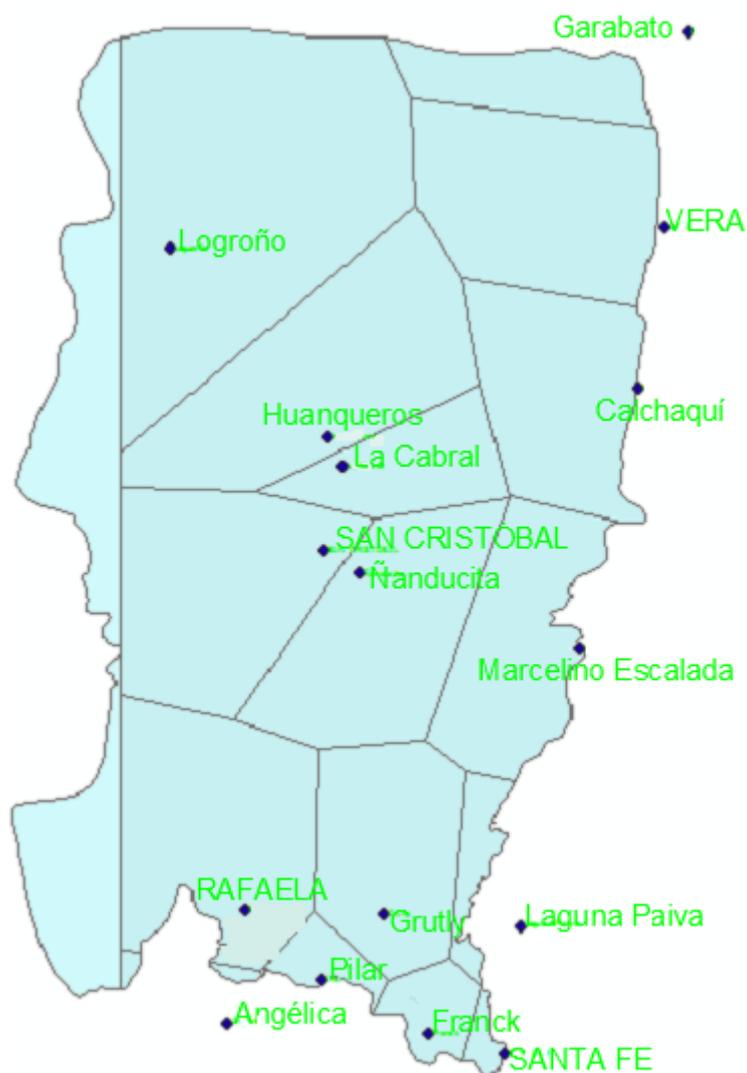


Figura 5.8b.: Polígono de Thiessen 2 (Software Arc Gis 10)

Determinación de precipitación media

De cada tipología de polígonos se obtuvieron las áreas de influencia de cada estación y se determinó cada coeficiente de participación o influencia areal, para la obtención de la precipitación media.

Así se conformó una tabla para cada tipología con cada coeficiente y precipitación media, una vez terminado este paso se procedió nuevamente a reacomodar cada día con su precipitación correspondiente en una tabla final, que es básicamente el objetivo de la etapa.

Para efectuar las tareas anteriores se usaron una serie de software, tales como:

- *Microsoft Excel*: se utilizaron macros propias del programa, con combinación de funciones para ordenamiento de series.
- *Arc Gis 10*: para la determinación automática de los polígonos de Thiessen, utilizando funciones: polígono, polígonos thiessen y clip.
- *Auto Cad*: se lo utilizo para el posicionamiento de las estaciones y para complementar el uso de Arc Gis en la determinación de los polígonos de Thiessen.

5.2.3. Caudales

Los caudales del Río Salado presentan la particularidad de estar influenciados por la dinámica del Río Paraná (efecto de remanso), por esta razón se hizo imposible disponer del listado de caudales diarios que en la actualidad no se encuentran corregidos, es decir que para que la serie se pueda ser usada se necesitaría efectuar la tarea de corrección que no formaría parte del objetivo del presente. A pesar de esto la lista está disponible para que en un futuro, una vez corregida pueda efectuarse la modelación correspondiente.

Actualmente el Ing. Sergio Menajovsky dispone de estos datos y se encuentra trabajando en ellos para su tesis de maestría.

5.2.4. Evapotranspiración potencial

Información recolectada

Existen innumerables metodologías para determinar la evapotranspiración potencial de una cuenca, en el presente se va a seguir una donde se requieren solo las series de temperatura y radiación extraterrestre como únicas variables, en el inciso siguiente se va a explicar en qué consiste dicho método.

Lo que respecta a este apartado es dejar sentado que se encontraron datos de temperaturas de bulbo seco pertenecientes a las estaciones de Saladillo en San Justo, San Javier en Colonia Mascias y en

Pueblo Libertador. Además se recopiló una tabla que presenta los datos de Radiación extraterrestre.

Las series de temperatura y radiación extraterrestre fueron descargadas de la página web oficial de la subsecretaría de recursos hídricos y de la FAO (Food and Agriculture Organization).

Procesamiento de información

La evapotranspiración potencial diaria se la determinó según el modelo planteado por Ludovic Oudin (Oudin-2005-Evapotranspiration-Part-2), basado en Jensen-Haise y McGuinness, el cual necesita solamente temperaturas medias diarias climáticas y radiación extraterrestre.

Se utilizó la siguiente fórmula:

$$PE = \frac{R_e T_a + 5}{\lambda \rho \cdot 100} \quad \text{if } T_a + 5 > 0$$

Caso contrario, $PE = 0$

Donde:

PE: Tasa de evapotranspiración potencial [mm/día],

Re: Radiación extraterrestre [MJ/m² día],

λ : Flujo de calor latente,

ρ : Densidad del agua [kg/m³],

Ta: Temperatura media diaria del aire [°C],

K1 [°C] y K2 [°C].

Como se plantea en las líneas superiores, para determinar la PE es necesario contar con una serie de datos a parte de la temperatura, además de algunas constantes, a continuación se muestran:

Radiación extraterrestre (Re): para utilizar los valores correctos es necesario tener presente las latitudes de la cuenca y su hemisferio (Google Earth), con cuatro puntos bien dispersos es suficiente (se utilizaron las latitudes sur: 28, 30, 32 y 34), para que a partir de ellos se efectúe un promedio, para el cálculo es necesario usar aquellos valores que correspondan al mes en cuestión (véase figura 5.9.).

Northern Hemisphere												Lat.	Southern Hemisphere											
Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec	deg.	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
0.0	2.6	10.4	23.0	35.2	42.5	39.4	28.0	14.9	4.9	0.1	0.0	70	41.4	28.6	15.8	4.9	0.2	0.0	0.0	2.2	10.7	23.5	37.3	45.3
0.1	3.7	11.7	23.9	35.3	42.0	38.9	28.6	16.1	6.0	0.7	0.0	68	41.0	29.3	16.9	6.0	0.8	0.0	0.0	3.2	11.9	24.4	37.4	44.7
0.6	4.8	12.9	24.8	35.6	41.4	38.8	29.3	17.3	7.2	1.5	0.1	66	40.9	30.0	18.1	7.2	1.5	0.1	0.5	4.2	13.1	25.4	37.6	44.1
1.4	5.9	14.1	25.8	35.9	41.2	38.8	30.0	18.4	8.5	2.4	0.6	64	41.0	30.8	19.3	8.4	2.4	0.6	1.2	5.3	14.4	26.3	38.0	43.9
2.3	7.1	15.4	26.6	36.3	41.2	39.0	30.6	19.5	9.7	3.4	1.3	62	41.2	31.5	20.4	9.6	3.4	1.2	2.0	6.4	15.5	27.2	38.3	43.9
3.3	8.3	16.6	27.5	36.6	41.2	39.2	31.3	20.6	10.9	4.4	2.2	60	41.5	32.3	21.5	10.8	4.4	2.0	2.9	7.6	16.7	28.1	38.7	43.9
4.3	9.6	17.7	28.4	37.0	41.3	39.4	32.0	21.7	12.1	5.5	3.1	58	41.7	33.0	22.6	12.0	5.5	2.9	3.9	8.7	17.9	28.9	39.1	44.0
5.4	10.8	18.9	29.2	37.4	41.4	39.6	32.6	22.7	13.3	6.7	4.2	56	42.0	33.7	23.6	13.2	6.6	3.9	4.9	9.9	19.0	29.8	39.5	44.1
6.5	12.0	20.0	30.0	37.8	41.5	39.8	33.2	23.7	14.5	7.8	5.2	54	42.2	34.3	24.6	14.4	7.7	4.9	6.0	11.1	20.1	30.6	39.9	44.3
7.7	13.2	21.1	30.8	38.2	41.6	40.1	33.8	24.7	15.7	9.0	6.4	52	42.5	35.0	25.6	15.6	8.8	6.0	7.1	12.2	21.2	31.4	40.2	44.4
8.9	14.4	22.2	31.5	38.5	41.7	40.2	34.4	25.7	16.9	10.2	7.5	50	42.7	35.6	26.6	16.7	10.0	7.1	8.2	13.4	22.2	32.1	40.6	44.5
10.1	15.7	23.3	32.2	38.8	41.8	40.4	34.9	26.6	18.1	11.4	8.7	48	42.9	36.2	27.5	17.9	11.1	8.2	9.3	14.6	23.3	32.8	40.9	44.5
11.3	16.9	24.3	32.9	39.1	41.9	40.6	35.4	27.5	19.2	12.6	9.9	46	43.0	36.7	28.4	19.0	12.3	9.3	10.4	15.7	24.3	33.5	41.1	44.6
12.5	18.0	25.3	33.5	39.3	41.9	40.7	35.9	28.4	20.3	13.9	11.1	44	43.2	37.2	29.3	20.1	13.5	10.5	11.6	16.8	25.2	34.1	41.4	44.6
13.8	19.2	26.3	34.1	39.5	41.9	40.8	36.3	29.2	21.4	15.1	12.4	42	43.3	37.7	30.1	21.2	14.6	11.6	12.8	18.0	26.2	34.7	41.6	44.6
15.0	20.4	27.2	34.7	39.7	41.9	40.8	36.7	30.0	22.5	16.3	13.6	40	43.4	38.1	30.9	22.3	15.8	12.8	13.9	19.1	27.1	35.3	41.8	44.6
16.2	21.5	28.1	35.2	39.9	41.8	40.8	37.0	30.7	23.6	17.5	14.8	38	43.4	38.5	31.7	23.3	16.9	13.9	15.1	20.2	28.0	35.8	41.9	44.5
17.5	22.6	29.0	35.7	40.0	41.7	40.8	37.4	31.5	24.6	18.7	16.1	36	43.4	38.9	32.4	24.3	18.1	15.1	16.2	21.2	28.8	36.3	42.0	44.4
18.7	23.7	29.9	36.1	40.0	41.6	40.8	37.6	32.1	25.6	19.9	17.3	34	43.4	39.2	33.0	25.3	19.2	16.2	17.4	22.3	29.6	36.7	42.0	44.3
19.9	24.8	30.7	36.5	40.0	41.4	40.7	37.9	32.8	26.6	21.1	18.5	32	43.3	39.4	33.7	26.3	20.3	17.4	18.5	23.3	30.4	37.1	42.0	44.1
21.1	25.8	31.4	36.8	40.0	41.2	40.6	38.0	33.4	27.6	22.2	19.8	30	43.1	39.6	34.3	27.2	21.4	18.5	19.6	24.3	31.1	37.5	42.0	43.9
22.3	26.8	32.2	37.1	40.0	40.9	40.4	38.2	33.9	28.5	23.3	21.0	28	43.0	39.8	34.8	28.1	22.5	19.7	20.7	25.3	31.8	37.8	41.9	43.6
23.4	27.8	32.8	37.4	39.9	40.6	40.2	38.3	34.5	29.3	24.5	22.2	26	42.8	39.9	35.3	29.0	23.5	20.8	21.8	26.3	32.5	38.0	41.8	43.3
24.6	28.8	33.5	37.6	39.7	40.3	39.9	38.3	34.9	30.2	25.5	23.3	24	42.5	40.0	35.8	29.8	24.6	21.9	22.9	27.2	33.1	38.3	41.7	43.0
25.7	29.7	34.1	37.8	39.5	40.0	39.6	38.4	35.4	31.0	26.6	24.5	22	42.2	40.1	36.2	30.6	25.6	23.0	24.0	28.1	33.7	38.4	41.4	42.6
26.8	30.6	34.7	37.9	39.3	39.5	39.3	38.3	35.8	31.8	27.7	25.6	20	41.9	40.0	36.6	31.3	26.6	24.1	25.0	28.9	34.2	38.6	41.2	42.1
27.9	31.5	35.2	38.0	39.0	39.1	38.9	38.2	36.1	32.5	28.7	26.8	18	41.5	40.0	37.0	32.1	27.5	25.1	26.0	29.8	34.7	38.7	40.9	41.7
28.9	32.3	35.7	38.1	38.7	38.6	38.5	38.1	36.4	33.2	29.6	27.9	16	41.1	39.9	37.2	32.8	28.5	26.2	27.0	30.6	35.2	38.7	40.6	41.2
29.9	33.1	36.1	38.1	38.4	38.1	38.1	38.0	36.7	33.9	30.6	28.9	14	40.6	39.7	37.5	33.4	29.4	27.2	27.9	31.3	35.6	38.7	40.2	40.6
30.9	33.8	36.5	38.0	38.0	37.6	37.6	37.8	36.9	34.5	31.5	30.0	12	40.1	39.6	37.7	34.0	30.2	28.1	28.9	32.1	36.0	38.6	39.8	40.0
31.9	34.5	36.9	37.9	37.6	37.0	37.1	37.5	37.1	35.1	32.4	31.0	10	39.5	39.3	37.8	34.6	31.1	29.1	29.8	32.8	36.3	38.5	39.3	39.4
32.8	35.2	37.2	37.8	37.1	36.3	36.5	37.2	37.2	35.6	33.3	32.0	8	38.9	39.0	37.9	35.1	31.9	30.0	30.7	33.4	36.6	38.4	38.8	38.7
33.7	35.8	37.4	37.6	36.6	35.7	35.9	36.9	37.3	36.1	34.1	32.9	6	38.3	38.7	38.0	35.6	32.7	30.9	31.5	34.0	36.8	38.2	38.2	38.0
34.6	36.4	37.6	37.4	36.0	35.0	35.3	36.5	37.3	36.6	34.9	33.9	4	37.6	38.3	38.0	36.0	33.4	31.8	32.3	34.6	37.0	38.0	37.6	37.2
35.4	37.0	37.8	37.1	35.4	34.2	34.6	36.1	37.3	37.0	35.6	34.8	2	36.9	37.9	38.0	36.4	34.1	32.6	33.1	35.2	37.1	37.7	37.0	36.4
36.2	37.5	37.9	36.8	34.8	33.4	33.9	35.7	37.2	37.4	36.3	35.6	0	36.2	37.5	37.9	36.8	34.8	33.4	33.9	35.7	37.2	37.4	36.3	35.6

Figura 5.9.: Tabla para la determinación de la radiación extraterrestre

Flujo de calor latente (λ): constante de 2.45 MJ/ kg.

Densidad del agua (ρ): es de 1000 kg/m³

Temperatura media diaria del aire (T_a): a partir de la información recabada, se consideró correcto utilizar las temperaturas de bulbo seco como una media del día en cuestión, para que a partir de ellas se efectúe un promedio e incluso se rellenen las series para fechas donde no se disponían datos.

Las series de temperaturas presentaban la particularidad de tener más de un dato por cada día de medición, por esto se aplicó una macro de Excel para definir un único dato por día, siendo este el mayor de los correspondientes a ese día. De esta manera, la serie sigue cubriendo el periodo 1979 – 2013, pero con un único dato diario, ya que es lo que se necesita a fines prácticos para proseguir con el desarrollo del trabajo.

La acción de relleno se torna interesante plantearla, ésta se basa en la conclusión a la que llegó Ludovic Oudin (Oudin-2005-Evapotranspiration-Part-2), donde puntualiza que la serie de PE se comporta de manera semejante que una rellena con los promedios de los días de épocas anteriores, es decir para rellenar un 01/01/2013 sería necesario efectuar el promedio de todos los 01/01 de los años anteriores de los que se tengan datos.

área de aproximadamente 15.000 km² (representando un 18% del total del área de la provincia) y una pendiente media del 1%.

La precipitación anual media es de 1311 mm, mientras que la evapotranspiración potencial anual es de aproximadamente 900 mm. El coeficiente de escurrimiento global para el periodo 1993-2010 es 0,27 (Giordano, 2011). El caudal medio anual es 165 m³/s, mientras que la descarga media mensual máxima es de 320 m³/s (mayo) (Goniadzki et al., 2011). El régimen hidrológico muestra dos crecidas anuales. El tiempo de concentración de la cuenca en la sección de interés varía aproximadamente entre 5 y 10 días según el estado de humedad.

La principal fuente de agua de los cursos de la cuenca proviene de las precipitaciones. El relieve de la zona favorece un buen escurrimiento debido a uno de los principales componente edafológicos que son las arcillas; estas aguas de lluvia son recolectadas por uno de los tantos cursos que tiene la cuenca y circula en los cauces de características meandrosas.

El curso principal puede dividirse en cuatro tramos bien definidos, con un recorrido de 486 km. El tramo del curso superior, cuyo origen se encuentra en las nacientes del río hasta un paraje llamado El Quebracho. El agua fluye de noreste a sudoeste y entre los tributarios más caudalosos se encuentran los arroyos Federal, Ortiz, Del Medio, Sauce de Luna, Moreira y Chañar. En el curso medio, que va desde el paraje El Quebracho hasta la ciudad de Rosario de Tala, la orientación cambia de norte a sur y se pueden mencionar los arroyos Mojones, Tigre, Raíces, Obispo, Tala, Bergara, El Tigre y Villaguay Chico. El curso inferior va de la ciudad de Rosario de Tala hasta la ciudad de Gualeguay y recibe gran número de afluentes de poco caudal manteniendo la orientación norte-sur. Por último, en el tramo final que va desde Gualeguay hasta la desembocadura en el río Paraná, el curso se torna meandroso debido al relieve propio de la zona.

5.3.2. Precipitaciones

Información recolectada

La serie de las precipitaciones medias se obtienen tras promediar series de estaciones de Villaguay, Paraná, Gualeguaychú, Monte Caseros y Concordia.

Se utilizan los mismos datos de precipitación que los de la presentación de la modelación hidrológica propuesta por Juan Bianchi, Dora Goniadzki.

Para llevarla a cabo dispusieron de precipitaciones medias areales bajo estimaciones satelitales (TRMM 3B42 V6), que correspondían a un período desde el 01/01/1998 al 30/03/2013, serie continua con un único dato diario. También contaron con datos pluviométricos de campo obtenidos de las estaciones meteorológicas de Villaguay, Paraná, Gualeguaychú, Monte Caseros y Concordia, de las cuales sólo Villaguay se encuentra dentro de la cuenca. Para ponderar cada estación se computaron polígonos de Thiessen (propios del desarrollo de otro trabajo, en el presente no se requirió hacerlos) y así se obtuvo la serie definitiva de precipitaciones medias areales diarias, la misma se extiende desde el 01/01/1998 al 10/11/2010. En la presentación se pone de manifiesto un análisis de ambas series, llegando a conclusiones que más adelante se expondrán.

Procesamiento de la información

Como se había anticipado se obtuvieron para la modelación hidrológica (*Juan Bianchi y Dora Goniadzki*), por un lado la intensidad de precipitación media areal (PMA) diaria [mm/día] utilizando estimaciones satelitales a través del producto TRMM 3B42 V6 (Huffman et al., 2009), el mismo se agregó a paso diario y se realizaron los cálculos para la cuenca, para el periodo 1998-2010. Por otro lado, calcularon la PMA a partir de datos pluviométricos de campo ponderando cada estación a través de polígonos de Thiessen. Se compararon ambas fuentes y se observó que las mismas son consistentes entre sí, pero que el algoritmo de precipitación de TRMM estima en promedio un 25% por encima del dato de campo.

Se presentaba la incertidumbre de decidir por la utilización de una de las dos series, si bien lo correcto sería utilizar la serie de precipitaciones de campo, se apreció a través de un análisis simple que las propias de radar se

asemejan mucho más con el cambio de caudales. Esto se verá más adelante con un análisis más pormenorizado, pero dejando asentado que las precipitaciones de campo y caudales no se correlacionan.

5.3.3. Aforos, niveles y caudales – curva H-Q

Información recolectada

Se recabaron datos de aforos pertenecientes a la Subsecretaría de Recursos Hídricos de la Nación, su estación de medición se localiza en Rosario del Tala, siendo esta localidad el punto de cierre de la cuenca.

✓ Aforos: se contaron con 126 datos de aforos de altura y caudal para un período que se extiende desde el 23/11/1988 al 09/05/2013.

✓ Niveles: se dispuso de una serie de mediciones efectuadas desde el 01/11/1992 al 30/05/2013, se destaca que en algunos días de la misma se disponían más de una medición.

✓ Caudales medios diarios: con características similares a la anterior, cubriendo un período de medición que se extiende desde el 01/11/1992 al 31/08/2012.

Procesamiento de la información

Como se aclaró en líneas anteriores, las series tanto de niveles y caudales estaban conformadas en principio por varios datos para cada día de medición, por esta razón a través de una macro en Excel se ordenó la serie de tal manera que para cada día le correspondiese una sola medición, precisamente el mayor valor.

El procesamiento de la información también apunta a la determinación de la curva H-Q a partir de los datos de aforos, niveles y caudales recabados.

Como las tres series de datos corresponden a la misma estación, se procedió a buscar una relación que las vincule y de esta manera corroborar la

validez de los datos de caudales, que en definitiva son los que sirven para el desarrollo del presente trabajo.

Para esto se correlacionaron los datos de niveles y caudales en una gráfica, acto seguido se realizó la misma tarea con los de aforos, incluso en el mismo gráfico con el objetivo de comparar ambas series (véase figura 5.10), donde se observa que se correlacionan para niveles inferiores a los 6 m, pero para niveles superiores no solo que no se correlacionan, sino que incluso no responden a un patrón definido de curva H-Q, ni siquiera viéndolas en forma independiente.

De aquí se desprende que es necesario definir un criterio para la determinación de una curva H-Q, para que a través de ésta se pueda obtener una serie de caudales más confiable de la que se dispone, justamente porque esta última no responde a una función acorde con los datos de niveles. El dato fundamental que llevó a la determinación de la curva, sin lugar a dudas fue el referido a experiencias de trabajos realizados en el río por expertos, donde para niveles cercanos a 8 m se tienen caudales próximos a los 1000 m³/s. Cabe preguntarse, ¿por qué con un único dato de estas características se puede definir una relación H/Q confiable?, la respuesta a esta pregunta se la puede ver con claridad en la figura 5.11, donde para niveles inferiores a los 6 m ambas series se correlacionan y además tienden a responder a un patrón definido, por lo que no habría problemas, pero para niveles superiores si los hay, ahora bien si se utiliza el dato experimental, este se corresponde la envolvente superior de la nube de puntos en azul o en forma de rombo (mediciones de niveles y caudales), es decir tomar cualquier otro par de valores que no formen parte de dicha envolvente estaría contradiciendo a la información recabada, es por esto que se considera criterioso tomar la envolvente superior de la nube de puntos como curva H-Q ya que es la única forma que responda a la realidad (dato experimental).

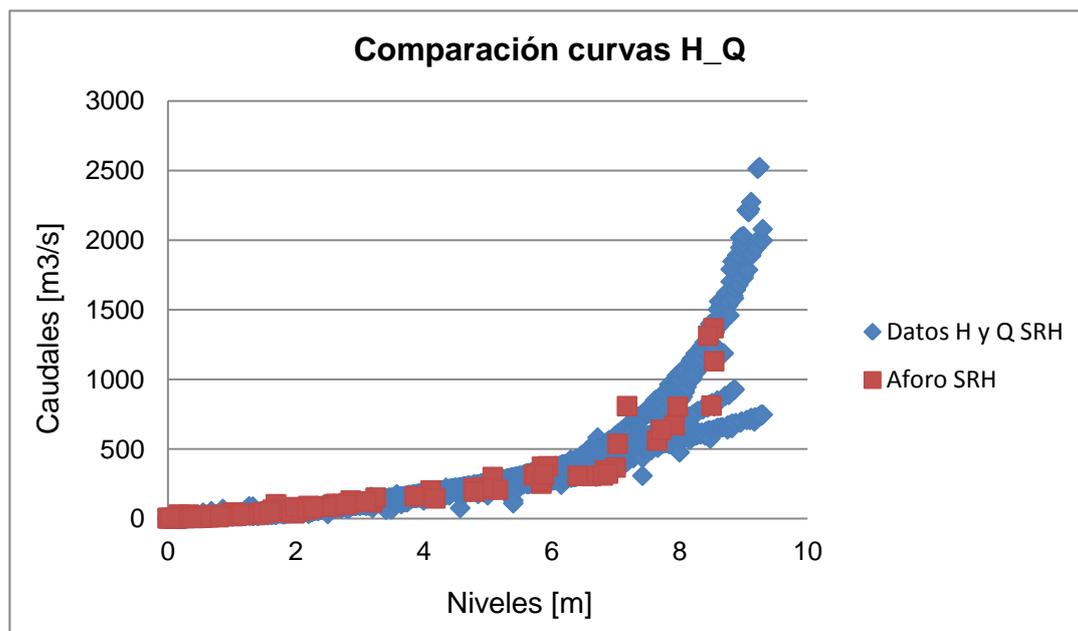


Figura 5.11.: Comparación de datos de aforo y mediciones de alturas y caudales

De esta manera mediante una sustracción de datos de altura que no conformaban la envolvente superior se llegó a una curva H-Q de tendencia muy bien ajustada de orden 6 (figura 5.12.), que se justifica su uso ya que en definitiva sirve de modo comparativo y de corroboración de datos, entonces a partir de ésta curva definitiva corroborada se determinaron los caudales.

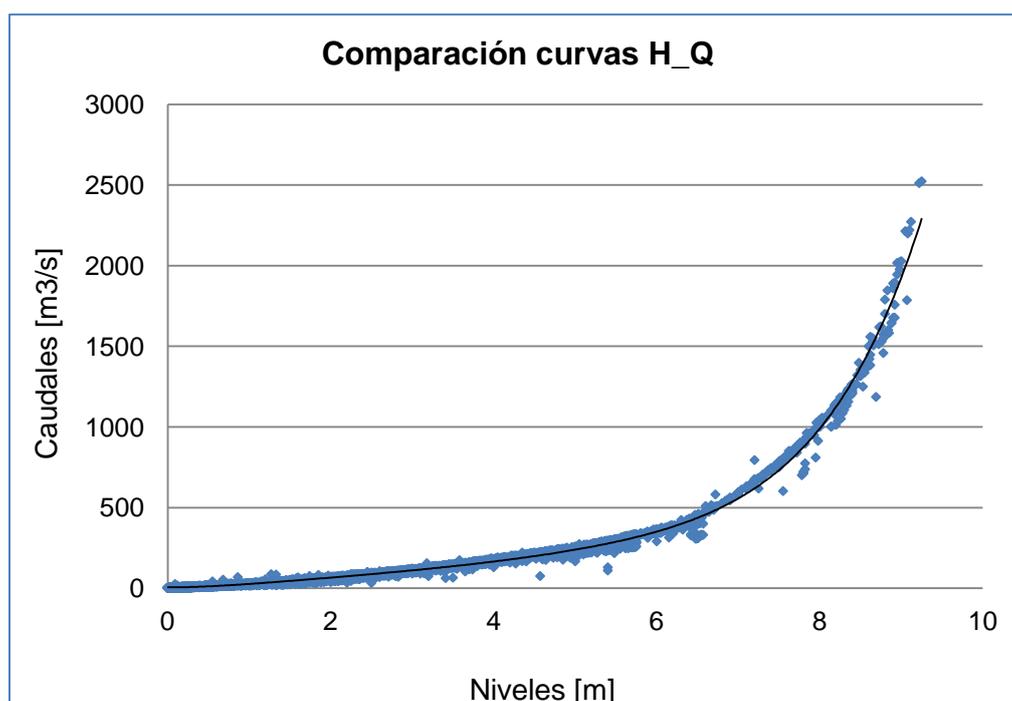


Figura 5.12.: Curva H-Q envolvente superior

Luego de culminado el análisis anterior se obtuvieron nuevos datos referidos a la determinación de la curva H-Q, que si bien es información que se obtuvo con cierta tardanza, no por eso pasa a ser menos importantes, provienen de subsecretaría de recursos hídricos a través del Ing. Luis Lenzi, la misma consiste en diversas curvas H-Q correspondiente a diferentes períodos e incluso para distintos momentos del día.

Listado de Fórmulas H-Q

Red Hidrometeorológica Nacional

3004 - ROSARIO DEL TALA - GUALEGUAY

- ✓ Válida desde: 01/11/1992 00:00
Desde: 0.00 Hasta: 9.50:
$$Q = 6.680638 * HM^2 + 17.8632614 * HM + 4.83797613$$
- ✓ Válida desde: 01/09/2001 00:00
Desde: 0.00 Hasta: 4.27:
$$Q = 18.657855911 * (HM + 0.1316888)^{1.62949165}$$

Desde: 4.27 Hasta: 9.00
$$Q = 51.9862385 * 1.38484184^{HM}$$
- ✓ Válida desde: 01/09/2002 00:00
Desde: 0.00 Hasta: 1.73
$$Q = 15.831481 * HM^2 - 25.948779 * HM + 27.5422299$$

Desde: 1.73 Hasta: 5.68
$$Q = 4.499639 * HM^2 + 32.0926799 * HM - 38.9938954$$

Desde: 5.68 Hasta: 9.50
$$Q = 12.76992932 * 1.731537915^{HM}$$
- ✓ Válida desde: 26/10/2003 00:00
Desde: 0.00 Hasta: 0.69
$$Q = 3.42302129224186 * 8.89176714006981^{HM}$$

Desde: 0.69 Hasta: 2.03
$$Q = 0.4483341 * HM^2 + 20.5128912357333 * HM + 1.1129$$

Desde: 2.03 Hasta: 5.68

$$Q = 4.499639 * HM^2 + 32.0926799 * HM - 38.9938954$$

Desde: 5.68 Hasta: 9.50

$$Q = 12.76992932 * 1.731537915^{HM}$$

✓ Válida desde: 22/10/2004 00:00

Desde: 0.00 Hasta: 1.34

$$Q = 3.9290116 * 5.8678728^{HM}$$

Desde: 1.34 Hasta: 4.39

$$Q = 30.56705 * (HM - 0.04935)^{1.25809}$$

Desde: 4.39 Hasta: 6.02

$$Q = 8.127450249E - 12 * (HM + 21.26)^{9.493669}$$

Desde: 6.02 Hasta: 9.50

$$Q = 12.76992932 * 1.731537915^{HM}$$

✓ Válida desde: 04/12/2005 00:00

Desde: 0.00 Hasta: 0.57

$$Q = 2.33194 * 25.63431^{HM}$$

Desde: 0.57 Hasta: 7.03

$$Q = 25.94123 * (HM + 0.093458)^{1.3493}$$

Desde: 7.03 Hasta: 9.30

$$Q = 0.758406 * 2.40991^{HM}$$

✓ Válida desde: 12/09/2007 00:00

Desde: 0.00 Hasta: 0.22

$$Q = 17.28566 * (HM + .14)^{1.09392}$$

Desde: 0.22 Hasta: 7.02

$$Q = 22.58094 * (HM + 0.15564)^{1.4112}$$

Desde: 7.02 Hasta: 9.20

$$Q = 0.758406 * 2.40991^{HM}$$

✓ Válida desde: 01/09/2010 00:00

Desde: -0.14 Hasta: 9.30

$$Q = 17.646 * (HM + 0.18226)^{1.5115}$$

- ✓ Válida desde: 01/09/2011 00:00
Desde: -0.14 Hasta: 7.02
 $Q = 17.646 * (HM + 0.18226)^{1.5115}$
Desde: 7.02 Hasta: 9.50
 $Q = 0.75841 * 2.4099^{HM}$

Se puede ver que para cada período se tiene más de una curva que corresponde a diferentes horarios del día, esto presenta una problemática para el análisis al que se apunta con este trabajo final, ya que el paso de tiempo de las cuencas en estudio es diario, motivo por el cual se hace necesario utilizar solo una de ellas para cada día, es decir en definitiva surgen análisis nuevos que escapan del central justamente por falta de tiempo, sería muy interesante analizarlos en trabajos posteriores a fin de abordar a nuevas conclusiones.

A pesar de esto se realizó un análisis tomando solo una de las curvas (la perteneciente a la última medición diaria), se obtuvieron nuevos caudales y se los comparó con los deducidos anteriormente a través de un gráfico de correlación, llegando al siguiente:

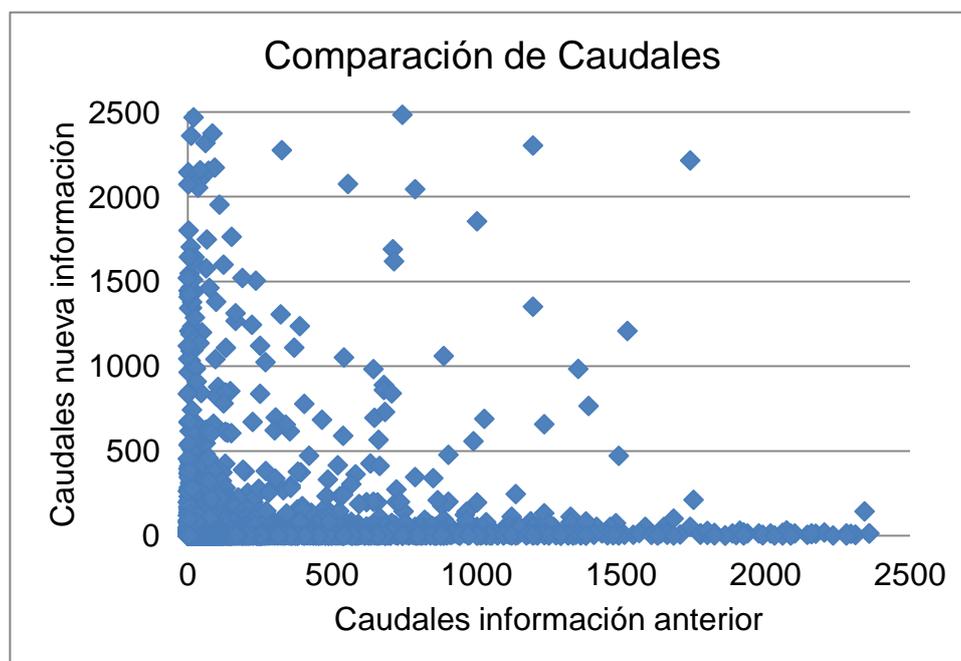


Figura 5.13.: Comparación de caudales

No existe ninguna correlación (figura 5.13.) entre los datos, es decir a priori pareciera que ambas fuentes arrastran algún tipo de error, pero para llegar a una conclusión certera sería necesario analizar con más detalle estas fuentes.

Sin embargo se procede al desarrollo con las series que en principio se analizaron, dejando a análisis futuro la validez de alguna de las series.

5.3.4. Evapotranspiración potencial

Información recolectada

Tal como se expuso en la cuenca del Salado, para efectuar el cálculo de la evapotranspiración potencial se requieren como únicas variables las series de temperatura y radiación extraterrestre.

Se dispusieron datos de temperaturas de las estaciones de Gualeguaychú (10134), Sauce Viejo (10451) y Concordia (10477), todas se encuentran en la zona aledaña a la cuenca en estudio y correspondían a diferentes lapsos de tiempo de medición.

La información fue provista por el Servicio Meteorológico Nacional a través de la Lic. Ana Teresa Gómez.

Procesamiento de la información

De la misma manera que se planteó este tema para la cuenca del Salado, la evapotranspiración potencial diaria se la determinó según el modelo planteado por Ludovic Oudin (Oudin-2005-Evapotranspiration-Part-2), basado en Jensen-Haise y McGuinness, el cual necesita solamente temperaturas medias diarias climáticas y radiación extraterrestre.

Se utilizó también:

$$PE = \frac{R_e T_a + 5}{\lambda \rho \cdot 100} \quad \text{if } T_a + 5 > 0$$

Caso contrario, $PE = 0$

Donde:

PE: Tasa de evapotranspiración potencial [mm/día],

R_e : Radiación extraterrestre [$\text{MJ}/\text{m}^2 \text{ día}$],

λ : Flujo de calor latente,

ρ : Densidad del agua [kg/m^3],

T_a : Temperatura media diaria del aire [$^{\circ}\text{C}$],

K_1 [$^{\circ}\text{C}$] y K_2 [$^{\circ}\text{C}$].

Es necesario contar con una serie de datos a parte de la temperatura, a continuación se muestran dichos datos:

Radiación extraterrestre (R_e): para determinarla es necesario saber las latitudes de la cuenca (Google Earth), con tres puntos bien dispersos fue suficiente (se utilizaron las latitudes sur: 30, 32 y 34), para que a partir de ellos se efectúe un promedio teniendo en cuenta el hemisferio donde se encuentra la cuenca, ahora bien para el cálculo propiamente dicho es necesario usar aquellos valores que correspondan al mes en cuestión (véase figura 5.9)

Flujo de calor latente (λ): se tomó un valor constante de $2.45 \text{ MJ}/\text{kg}$.

Densidad del agua (ρ): es de $1000 \text{ kg}/\text{m}^3$

Temperatura media diaria del aire (T_a): a partir de la información recabada, se efectuó un promedio de las tres series, todas presentaban un único dato diario (no se necesitó realizar un ordenamiento con alguna macro tal como se hizo con el Salado) y además cubrían el lapso desde el 01/01/1998 al 10/11/2010 que en definitiva es el de interés ya que es aquel donde se tienen datos de los tres parámetros a utilizar como entrada a la modelación (no se necesitó abordar a la conclusión de Ludovic Oudin - Oudin-2005-Evapotranspiration-Part-2).

K_1 y K_2 : Parámetros para ajustar el modelo de modo tal que el total anual coincida con el valor anual calculado, según Oudin, se considera correcto que $K_1=100^{\circ}\text{C}$ y $K_2=5^{\circ}\text{C}$.

5.3.5. Análisis de consistencia

Una vez ordenados los datos es necesario analizarlos con el objetivo de encontrar cierta consistencia, especialmente entre las series de precipitaciones y caudales, es decir para una precipitación debería haber un

consecuente crecimiento en los caudales. Este análisis se realizó a través de un programa en MatLab (Ing. Antoine Patalano).

Además se debe encontrar el tiempo de retardo de la serie, entendiéndose como tal al período que transcurre entre el pico de la precipitación y el pico de los caudales.

Reconocimiento visual

El programa de MatLab arroja dos gráficas correlacionadas que ordenan las precipitaciones (radar) por un lado y los caudales por el otro, ambos en función del tiempo.

De esta manera se tiene una herramienta para analizar ambos gráficos y determinar si son consistentes.

A continuación se presentan gráficas (figura 5.14.) de ciertos períodos para mostrar dicha consistencia, en la parte superior se presentan las precipitaciones vs tiempo, mientras que el inferior los caudales vs tiempo.

Se distinguen cuatro picos bien definidos en ambos gráficos que se corresponden bastante bien, es decir ante una precipitación (gráfica superior), genera un pico de caudales (gráfica inferior).

Es necesario aclarar que solo se realizó el análisis para ese lapso únicamente ya que se apreciaba la misma correlación en toda la serie, con lapsos muy cortos con falta de sincronización, donde no se justifica eliminarlos ya que la serie en principio consta de pocos datos. En conclusión se usa toda la serie de datos para conformar la entrada a la programación.

Es importante aclarar que el análisis se realizó solo a modo de prueba para las precipitaciones de campo y no se observaba ninguna correlación con los caudales, es por esto que se descartó por completo este análisis, incluso dejando de lado una salida que no tendría sentido exponerla.

Determinación del tiempo de retardo

Esta determinación se la realizó de dos formas, la primera haciendo un reconocimiento visual con alguno de los picos para que se pueda ver con claridad una medición gráfica y la otra a través de un análisis de correlación de ambas series con Excel.

La metodología con Excel se basa en hacer un análisis de correlación de las series para diferentes tiempos de retardo de tal manera de encontrar aquel que sea el óptimo, es decir aquel donde se presente el mayor coeficiente de correlación:

Retardo	
Días	Coeficiente de correlación
0	0.038
5	0.244
6	0.266
7	0.272
8	0.271
9	0.262

Se ve en la tabla anterior que el tiempo de retardo es de 7 días.

Se realizó este mismo análisis para la serie de precipitaciones de campo y se encontró que el tiempo de retardo era de 6 días, pero con la particularidad de que el coeficiente de correlación es del de orden 0.011, es decir muy lejano al valor por excelencia de 1, dejando a la vista que las serie de caudales y la de precipitaciones medidas en campo no se correlacionan, ratificando la idea de utilizar solo la serie de precipitaciones de radar.

En cuanto a la metodología gráfica se ve claramente en la figuras 5.15a. que el tiempo de retardo ronda los 7 días, esto es porque como cada división representa 10 días, los picos máximos se distancian en una cantidad semejante a 7 unidades aproximadamente.

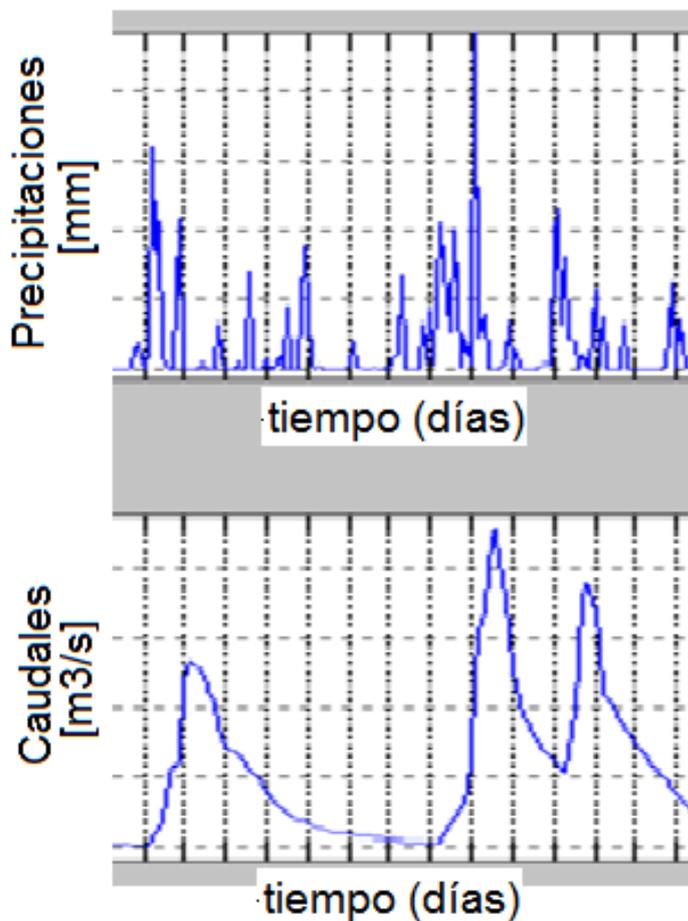


Figura 5.15a.: Comparación de caudales y precipitaciones, tiempo de retardo (intervalo eje de abscisas = 10 días)

Se presenta otro gráfico (figura 5.15b.) similar al anterior donde se acota el tiempo de retardo:

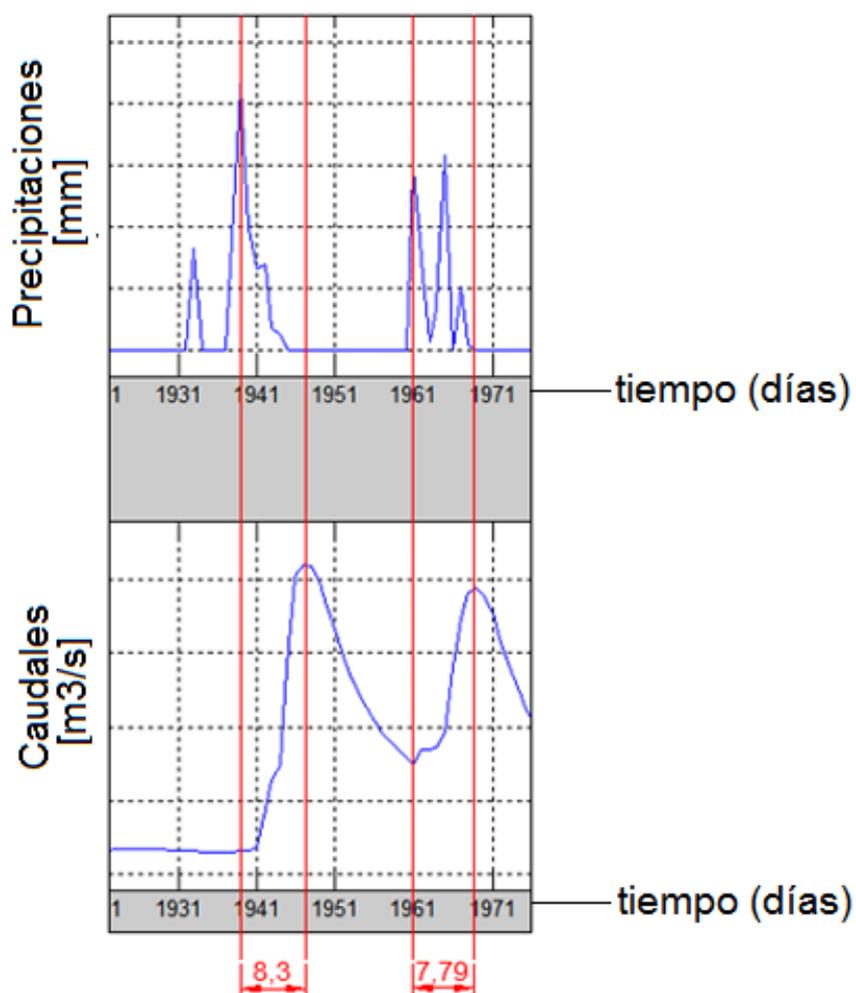


Figura 5.15b.: Comparación de caudales y precipitaciones, tiempo de retardo

Días: 1931 = 15/04/2003 al 1971 = 25/05/2003

El tiempo de retardo entre los picos en el gráfico está entre 7 y 8 días aproximadamente, por esta razón se sigue sosteniendo que el tiempo de retardo que presentan las series es de 7 días.

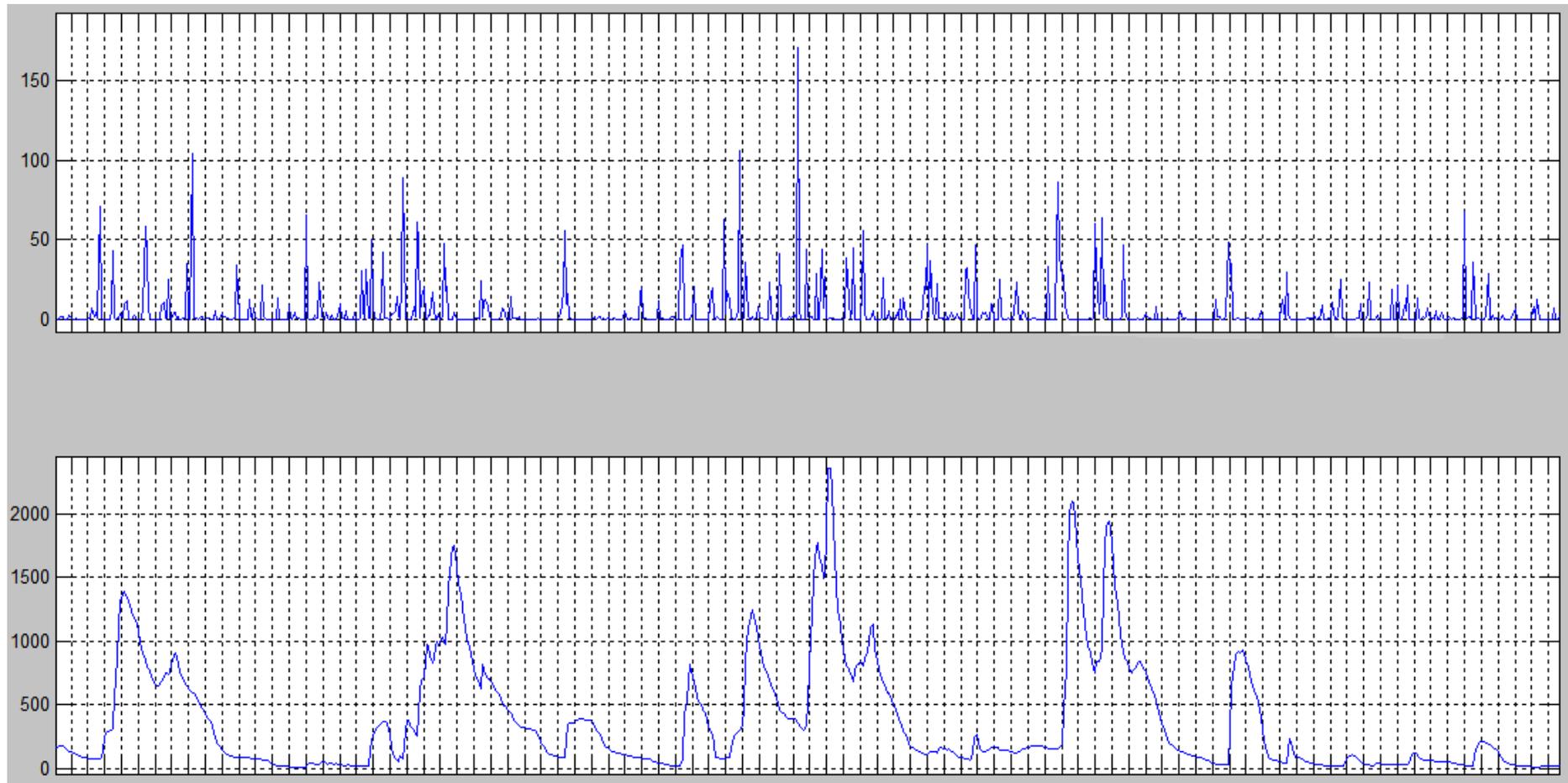


Figura 5.14.: Gráficos de precipitaciones vs tiempo y caudales vs tiempo

6. ANÁLISIS ESTADÍSTICO

6.1. INTRODUCCIÓN

Las series de datos de caudales y precipitaciones fueron corroboradas para que cumplan ciertas condiciones o pruebas estadísticas. Las mismas deben ser superadas con un determinado nivel de significancia. Las pruebas estadísticas a las que se sometieron las series y que fueron superadas fueron:

- ✓ Prueba de estacionalidad (Kendall).
- ✓ Prueba de homogeneidad a escala anual (Wilcoxon).

La aplicación de estas pruebas se realizó mediante la herramienta computacional Hyfran.

No se realizó el análisis estadístico sobre las series de evapotranspiración no es una variable que afecte de manera significativa a los resultados arrojados por el modelo.

Además es menester aclarar que el análisis solo se realizó para los datos de la cuenca del Río Gualeguay ya que es la que dispone de todos los necesarios para la programación, no obstante se seguiría la misma metodología para los propios del Río Salado.

A continuación se va a hacer el análisis de los datos, previamente introduciendo teóricamente la prueba en cuestión.

6.2. NIVEL DE SIGNIFICANCIA

En las distintas pruebas estadísticas se plantea una hipótesis nula (H_0) la cual se mantiene como verdadera a menos que en la serie de la muestra se encuentre evidencia de lo contrario.

A esta hipótesis nula se la contrasta con una hipótesis alternativa (H_1). En los casos en los que no se especifica H_1 de manera explícita, se puede considerar que la misma ha quedado definida implícitamente como “ H_0 es falsa”.

Al analizar una serie de datos, los mismos pueden ser relacionados con una determinada distribución de probabilidad la cual esté relacionada con

la hipótesis de estudio. Luego, pueden tomarse un conjunto o un rango de valores para el cual si un valor estadístico obtenido pertenece a él se rechazara la hipótesis nula. Dicho rango se conoce como región de rechazo y el mismo se puede modificarse para plantear condiciones más estrictas o más relajadas.

Cuando se contrastan las hipótesis, se determina el rechazo o no de la hipótesis nula H_0 . Esta decisión puede ser errónea o no coincidir con lo que sucede en la realidad. Se pueden dar cuatro casos los cuales se visualizan en la siguiente tabla:

	<i>H₀ es cierta</i>	<i>H₁ es cierta</i>
<i>Se acepta H₀</i>	<i>No hay error</i>	<i>Error tipo II</i>
<i>Se acepta H₁</i>	<i>Error tipo I</i>	<i>No hay error</i>

La probabilidad de cometer un error de tipo I es conocido como nivel de significancia de la prueba y se denota con la letra griega " α ".

Por otro lado la probabilidad de cometer un error de tipo II se denota con la letra griega " β ". Se denomina potencia del contraste al valor $1-\beta$, esto es, a la probabilidad de escoger H_1 cuando esta es cierta.

Considerado un nivel de significación α (error que estamos dispuestos a asumir en el caso de optar por H_1 y equivocarnos) y una muestra de tamaño n . Se pueden plantear dos alternativas representadas en las figuras 6.1. y figura 6.2.

Los valores, t_α o $t_{2\alpha}$ se encuentran tabulados para cada valor de α . Si la media de la muestra cae en la región de aceptación se acepta la hipótesis H_0 ; si cae en la región crítica, se acepta la hipótesis H_1 .

En el presente trabajo, se utilizó un nivel de significancia del 5% (0.05) en la mayoría de las series. En los casos en que no se cumplía alguna de las pruebas estadísticas se procedió a analizarlas con un nivel de significancia del 1%. Luego si las series eran rechazadas bajo este nivel de significancia se las descartaba para el análisis y dicha estación no era tomada en cuenta.

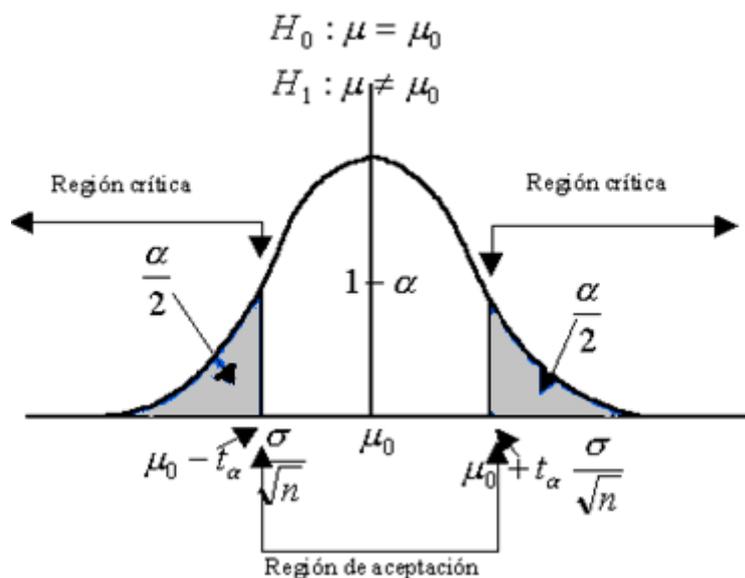


Figura 6.1.: Funciones de frecuencia para la información de muestra y funciones de probabilidad para la población.

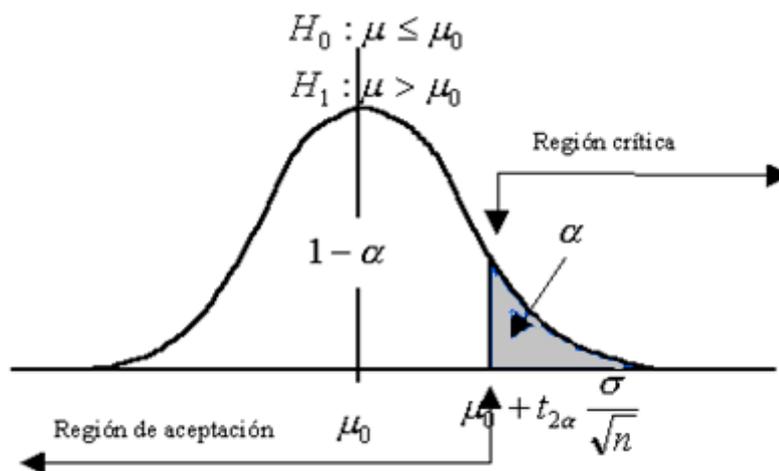


Figura 6.2.: Función de frecuencia para la información de muestra y funciones de probabilidad para la población.

6.3. PRUEBA DE ESTACIONALIDAD (Kendall)

Mediante la prueba estadística de estacionalidad de Kendall se busca que en la serie de datos no se presente una tendencia importante.

En la misma, para cada dato de la serie, la cual se encuentra ordenada cronológicamente, se calcula cuantos datos posteriores son superiores. De esta forma:

$$S = \sum_{i=1}^{n-1} s(i)$$

Donde:

$s(i)$ es el número de los valores de $X_j > X_i$, siendo $i < j \leq n$.

Por otro lado se determinan cuantos valores posteriores son menores. Los mismos se determinan de la siguiente manera:

$$T = \sum_{i=1}^{n-1} t(i)$$

Donde:

$t(i)$ es el número de valores $X_j < X_i$, siendo $i < j \leq n$.

Con los valores de S y T calculados, se define el índice $I = S - T$, que debe ser próximo a cero si la hipótesis nula es verdadera.

Por lo tanto se deben verificar las siguientes condiciones:

✓ $I < I_{cri}(\alpha)$, o sea que el I calculado sea menor que un valor de I crítico tabulado en función del tamaño de la muestra y según diferentes niveles de significación.

✓ Para tamaños de muestra mayores a 10, el valor del estadístico se aproxima a una distribución normal. Por lo que se debe utilizar una forma estandarizada del estadístico en donde se transforma el índice I en el índice K, definido como:

$$K = \frac{I - 1}{\left(\frac{N * (N - 1) * (2N + 5)}{18}\right)^{0.5}} \quad I > 0$$

$$K = 0 \quad I = 0$$

$$K = \frac{I + 1}{\left(\frac{N * (N - 1) * (2N + 5)}{18}\right)^{0.5}} \quad I < 0$$

Los valores de K_{cri} son obtenidos de una distribución normal, los cuales son presentados en la tabla siguiente:

α	0,01	0,05	0,1
K_{cri}	2,33	1,64	1,28

6.3.1. Análisis de precipitaciones medias diarias

A partir de las precipitaciones diarias, se obtuvo la serie en precipitaciones mensuales acumuladas mediante una tabla dinámica de Excel para someter los datos a la prueba.

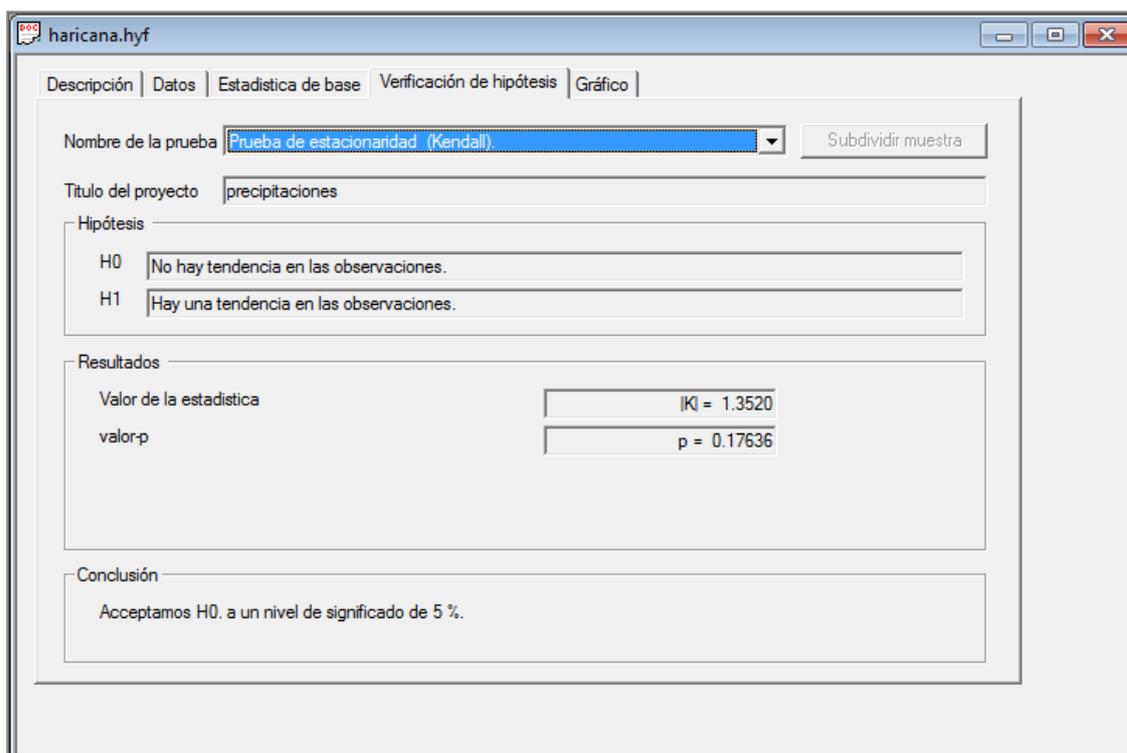
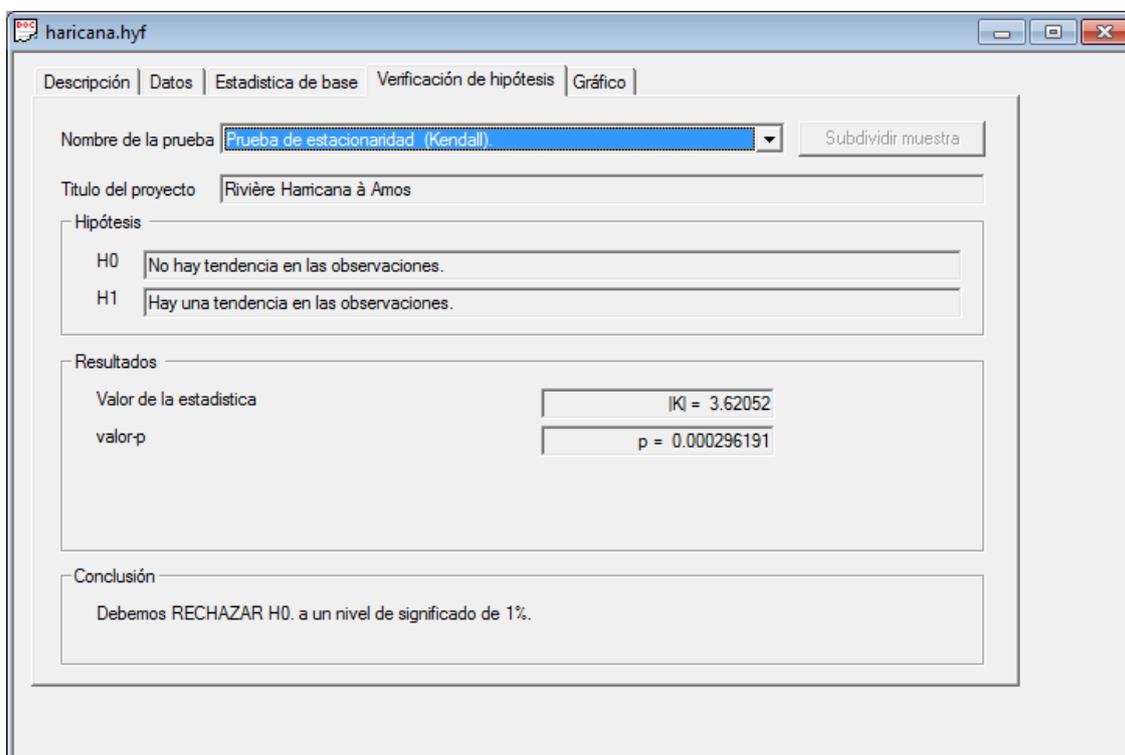


Figura 6.3.: Salida HyFrinse prueba de estacionaridad de Kendall – precipitaciones.

Como hipótesis H_0 se planteó que no existe tendencia entre las observaciones, resultando que la hipótesis deba aceptarse, de esta manera decimos que no hay ninguna tendencia en la serie.

6.3.2. Análisis de caudales medios diarios

Se obtuvo al igual que el caso anterior a partir de la serie de caudales diarios, los propios medios mensuales por medio de una tabla dinámica de Excel para someter los datos a la prueba.



haricana.hyf

Descripción | Datos | Estadística de base | Verificación de hipótesis | Gráfico

Nombre de la prueba: Prueba de estacionaridad (Kendall) [Subdividir muestra]

Título del proyecto: Rivière Hamicana à Amos

Hipótesis

H0: No hay tendencia en las observaciones.

H1: Hay una tendencia en las observaciones.

Resultados

Valor de la estadística	K = 3.62052
valor-p	p = 0.000296191

Conclusión

Debemos RECHAZAR H0. a un nivel de significado de 1%.

Figura 6.4.: Salida HyFrinse prueba de estacionaridad de Kendall – caudales.

De la misma manera que en las precipitaciones se planteó una hipótesis H0 de que no hay tendencia en la serie, como resultado se debió rechazar H0 a un nivel de significancia del 1%, la manera correcta de justificarlo se puede deber a que la serie no es lo suficientemente larga, pudiendo estar influenciada por los ciclos multidecimos comunes. Sería de suma importancia investigar si en la cuenca existió algún antecedente que explique dicha influencia, pudiendo ser el causante por ejemplo una variabilidad climática.

6.4. PRUEBA DE HOMOGENEIDAD (Wilcoxon)

La prueba de homogeneidad a escala anual de Wilcoxon permite comprobar si el promedio de un primer subconjunto de la muestra difiere significativamente del promedio de un segundo subconjunto de la muestra.

Para efectuar la prueba se debe sustituir las observaciones de las dos muestras de tamaños n_1 y n_2 , por sus rangos combinados. Los rangos son asignados de tal manera que el rango 1 se asigna a la más pequeña de las $n = n_1 + n_2$ observaciones combinadas, el rango 2 se le asigna a la siguiente más alta y así sucesivamente, de modo que el rango n queda asignado a la observación más grande. Si varias observaciones tienen el mismo valor, se asigna a cada una de estas el promedio de los rangos que, en otra circunstancia, se les habría asignado.

La estadística de la prueba de Wilcoxon T_1 , es simplemente la suma de los rangos asignados a las n_1 observaciones de la muestra más pequeña.

Para cualquier valor entero n , la suma de los n primeros enteros consecutivos pueden calcularse fácilmente como $n(n+1)/2$. La estadística de prueba T_1 , más la suma de los rangos asignados a los n_2 elementos de la segunda muestra, T_2 , por consiguiente, debe ser igual a este valor, es decir;

$$T_1 + T_2 = \frac{n(n+1)}{2}$$

De modo que esta ecuación puede ser empleada como verificación del procedimiento de asignación de rangos.

La prueba de la hipótesis nula puede ser de dos extremos o de un extremo, dependiendo de si se está probando si las dos medianas de población son exclusivamente diferentes o si una es mayor que la otra.

La estadística de prueba, T_1 , está distribuida de manera aproximadamente normal. Se puede emplear la siguiente fórmula de aproximación de muestra grande para probar la hipótesis nula:

$$|Z| = \frac{T_1 - \mu_{T_1}}{\sigma_{T_1}}$$

El valor medio de la estadística de prueba T_1 , puede calcularse con:

$$\mu_{T_1} = \frac{n_1(n+1)}{2}$$

y la desviación estándar de la estadística de prueba T_1 , se puede calcular con:

$$\sigma_{T_1} = \sqrt{\frac{n_1 \cdot n_2 (n+1)}{12}}$$

Basándose en α , el nivel de significación seleccionado, la hipótesis nula puede ser rechazada si el valor Z calculado cae en la región de rechazo apropiada.

6.4.1. Análisis de precipitaciones medias diarias

A partir de las precipitaciones diarias, se obtuvo la serie en precipitaciones mensuales acumuladas mediante una tabla dinámica de Excel para someter los datos a la prueba, al igual que la prueba de Kendall.

The screenshot shows the 'Verificación de hipótesis' (Hypothesis Verification) tab in the HyFrinse software. The test selected is 'Prueba de homogeneidad a escala anual (Wilcoxon)'. The project title is 'precipitaciones'. The null hypothesis (H0) is 'Los promedios de las muestras son iguales.' and the alternative hypothesis (H1) is 'Los promedios de las muestras son diferentes.' The results section displays: 'Valor de la estadística' (Value of the statistic) as $|W| = 1.4031$, 'valor-p' (p-value) as $p = 0.16058$, 'Fecha inicial del primer' (Initial date of the first) as 1, and 'Fin del primer grupo' (End of the first group) as 78. The conclusion states: 'Aceptamos H0. a un nivel de significado de 5 %. No se puede concluir que los promedios de las dos sub-muestras son diferentes.'

Figura 6.5.: Salida HyFrinse prueba de homogeneidad de Wilcoxon – precipitaciones.

Como hipótesis H0 se plantea que los promedios de las series eran iguales. Como resultado se obtuvo que si bien se debía aceptar dicha hipótesis, de esta manera podemos afirmar que las dos sub-series son homogéneas y provienen de la misma población.

6.4.2. Análisis de caudales medios diarios

haricana.hyf

Descripción | Datos | Estadística de base | Verificación de hipótesis | Gráfico

Nombre de la prueba: Prueba de homogeneidad a escala anual (Wilcoxon) [Subdividir muestra]

Título del proyecto: Rivière Haricana à Amos

Hipótesis

H0: Los promedios de las muestras son iguales.

H1: Los promedios de las muestras son diferentes.

Resultados

Valor de la estadística	W = 4.30529
valor-p	p = 1.68878E-005
Fecha inicial del primer	1
Fin del primer grupo	78

Conclusión

Debemos RECHAZAR H0. a un nivel de significado de 1%.
Se puede concluir que las dos sub-muestras NO SON homogéneas.

Figura 6.6.: Salida HyFrinse prueba de homogeneidad de Wilcoxon – caudales.

De la misma manera que en las precipitaciones se planteó una hipótesis H0 de que el promedio de las muestras son iguales, como resultado se debió rechazar H0 a un nivel de significancia del 1%, la manera correcta de justificarlo se puede deber a que la serie no es lo suficientemente larga, pudiendo estar influenciada por los ciclos multidecimos comunes. De la misma manera que se dijo en la prueba de Kendall sería de suma importancia investigar si en la cuenca existió algún antecedente que explique dicha influencia, pudiendo ser el causante por ejemplo una variabilidad climática.

6.5. REVISIÓN FINAL DE LAS SERIES DE DATOS

Debido a que la revisión estadística de los datos tiene como objetivo la verificación de las pruebas, no se plantea ninguna modificación en las series, además como se había concluido en la revisión visual a través del programa en MatLab de que no se iba a recortar la serie, podemos decir en conclusión final que la serie se usa con todos los datos procesados donde los tres parámetros (precipitaciones, caudales y temperaturas) coinciden respecto al tiempo, extendiéndose desde el 01 de enero de 1998 hasta el 10 de noviembre de 2010, con una cantidad total de 4697 datos.

7. ANÁLISIS DE RESULTADOS

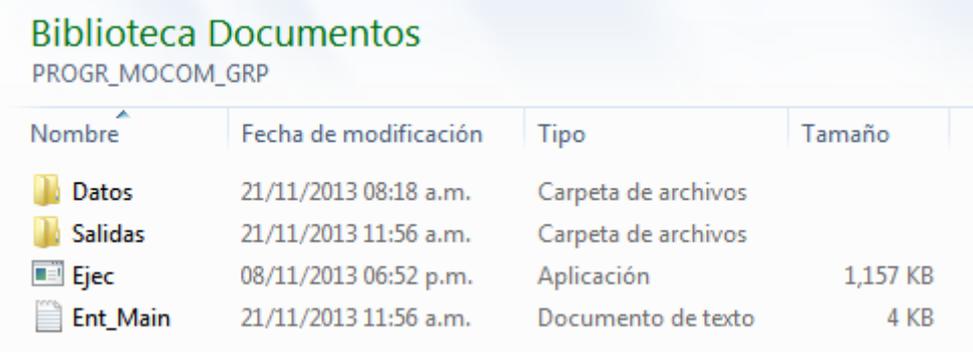
Como se había anticipado la calibración se llevo a cabo solamente para los datos de la cuenca del Río Gualeguay, ya que no se cuentan con datos de caudales para el Salado.

Además es importante remarcar que la serie se usa completa (sin recortes), tema que está debidamente justificado en capítulos anteriores.

7.1. ARCHIVOS PARA MODELACIÓN

La programación del modelo a cargo de Mg. Ing. Facundo José Alonso se efectuó a través del lenguaje de programación de alto nivel fortran (Formula Translating System).

Se cuenta con los siguientes archivos necesarios para llevar a cabo la modelación, una carpeta “Datos”, otra “Salidas”, el ejecutable “Ejec” y el archivo de entrada “Ent_Main”, (véase figura 7.1.), a continuación se los explicará a modo aclaratorio para una posible modelación futura:



Nombre	Fecha de modificación	Tipo	Tamaño
Datos	21/11/2013 08:18 a.m.	Carpeta de archivos	
Salidas	21/11/2013 11:56 a.m.	Carpeta de archivos	
Ejec	08/11/2013 06:52 p.m.	Aplicación	1,157 KB
Ent_Main	21/11/2013 11:56 a.m.	Documento de texto	4 KB

Figura 7.1.: Archivos para la programación – MOCOM y GR

La carpeta “Datos” debe contener un archivo de texto que constituye los datos de entrada que se procesaron con anterioridad, este archivo necesariamente debe nombrarse como se ve a continuación (figura 7.2.), de lo contrario el archivo no puede ser leído:

Nombre	Fecha de modificación	Tipo	Tamaño
Gualeg00_ArD	08/11/2013 11:09 p.m.	Documento de texto	94 KB
Gualeg01_ArD	20/11/2013 05:55 p.m.	Documento de texto	46 KB

Figura 7.2.: Carpeta “Datos”

Estos archivos de texto que conforman la carpeta Datos contienen la siguiente información:

Archivo	Edición	Formato	Ver	Ayuda
	15000.			
2.400	3.967	1120.890		
0.765	4.461	1032.760		
0.000	5.066	941.293		
20.393	5.389	885.851		
66.260	5.636	854.556		
11.302	4.843	885.851		
0.000	4.378	907.523		
0.000	4.572	1013.620		
0.000	4.878	1626.370		
0.549	5.260	2058.780		
0.000	5.542	2086.900		
0.000	5.519	1990.180		
16.897	5.536	1739.020		
2.134	4.925	1481.840		
0.000	5.178	1291.130		
0.253	5.172	1106.760		

Figura 7.3.: Fragmento de archivo de texto de entrada de datos P-EP-Q

La primera fila contiene el área de la cuenca en Km², en este caso corresponde a la propia de Gualeguay con una extensión de 15000 Km². A partir de la segunda fila se ordenan de izquierda a derecha los datos de Precipitación, Evapotranspiración Potencial y Caudal, de toda la serie.

La carpeta “Salidas” (remítase nuevamente a la figura 7.1.), debe estar creada de manera obligatoria tal como lo indica la figura para que las salidas de la programación se guarden con éxito.

A continuación se ve una de las múltiples salidas que pueden presentarse:

Gualeg01_ArD_Sal: Bloc de notas		
Archivo	Edición	Formato Ver Ay
	Qprev 1Dt,	Qprev
	21.6508	21.2830
	16.3319	17.9970
	16.2794	16.3970
	16.2271	14.8310
	16.1756	13.6030
	16.1243	13.6030
	16.0731	13.3000
	8.8119	10.6460
	8.7970	11.2240
	8.7821	10.6460
	8.7673	10.3600
	8.7525	10.3600
	8.7379	10.0760
	8.7239	9.7940
	8.7105	10.6460

Figura 7.4.: Fragmento de archivo de texto de salida de datos

El archivo contiene encolumnado de izquierda a derecha los caudales previstos y los reales, aclarando que es una salida del modo previsión del algoritmo MOCOM.

El ejecutable “Ejec”, como su nombre lo indica ejecuta el programa tanto GR como el algoritmo MOCOM en sus dos modos simulación y previsión, es una ventana que cuando termina la ejecución da a conocer el mensaje que puede verse en la siguiente (figura 7.5.):

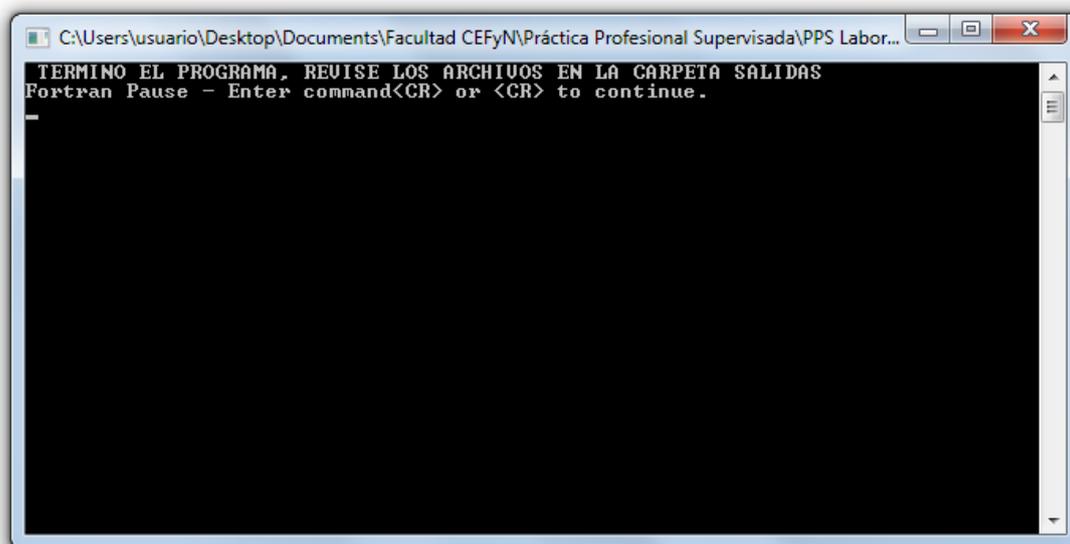


Figura 7.5.: Ventana del ejecutable – mensaje de terminación del programa

El archivo de entrada “Ent_Main”, es aquel que contiene todo lo necesario para correr el GR y el algoritmo MOCOM, incluso se especifica el

archivo de entrada de datos que se ve en la figura 7.2. y 7.3., a continuación se ven figuras que conforman el archivo de texto (véase figuras 7.6a., 7.6b. y 7.6c.):

```

Ent_Main: Bloc de notas
Archivo Edición Formato Ver Ayuda
      2      1      2      10
      6      300
1069.23 2.86 295.40 2.832 12. 16.2 5. 1.5
      1      2      3      4      5      6      36      48
      100      10 5000000      1      10
      200.      -5.      20.      1.1      10.      6.      4.      1.3
      1200.      3.      300.      2.9      13.      7.5      6.2      1.6
Gualleg00_ArD      2349
A3422010_FrH      82791
Y2102010_FrH      83994
V7135010_FrH      84000
!
!A continuación se da detalle de la lectura que se hace del archivo de entrada en esta subrutina:
!
!Primer fila (variables):
!Opc: 1 (Modelo IPH2), 2 (Modelo GR4), 3 (Mocom-IPH2), 4 (Mocom-GR4).
!Ncu: Número de cuencas que se quieren correr (en cualquiera de los cuatro modos anteriores)
!Msp: 1 (Modo Simulación), 2 (Modo Previsión). (Sólo usa este valor si "Opc" toma el valor 1 o 2)
!Salnum: ver notas en el programa.
!
!Segunda fila: se leen los parámetros:
! Ntprev: Es el núm de veces que se almacenarán salidas en previsión.
! Ldes: Número de Dt del primer tramo que se descarta para el cálculo de la FO.
!

```

Figura 7.6a.: Archivo de entrada Ent_Main – parte 1

```

!
!Tercer fila: Todos los valores de los parámetros del modelo elegido (IPH2 o GR):
!
! Si es IPH2, estos parámetros son:
! Io: Máxima capacidad de infiltración del suelo del algoritmo de Horton (mm/Dt)
! Ib: Mínima capacidad de infiltración del suelo del algoritmo de Horton (mm/Dt)
! h = eλ(-kt), donde k es un parámetro de decaimiento de la ley de infiltración
! ks: Parámetro de decaimiento del flujo superficial (unidades de Dt)
! ksub: Parámetro de decaim del flujo subterráneo (unidades de Dt)
! Rm: Capacidad máxima de almacenamiento del suelo (mm)
! NTC: Numero de Dt del tiempo de concentración (o sea, también número de ordenadas del HTA).
! XN es el coeficiente del histograma HTA (1,5 normalmente)
! QSUB1: CAUDAL SUBTERRANEO INICIAL (m3/s, es el caudal de base en el río)
! QSUP1: CAUDAL SUPERFICIAL INICIAL, GENERALMENTE IGUAL A CERO ANTES DE LA LLUVIA
! R1: Es la altura de agua inicial del reservorio R (mm)
! AIMP: PORCENTAJE DE AREAS IMPERMEABLES DE LA CUENCA (x/1)
!
! Si es GR, los parámetros son x1, x2, x3 y x4.
!
!Cuarta fila: Todos los valores de los parámetros que requiere el modelo en modo previsión.
! Están descriptos en el programa.
!
!Quinta fila: Parám y datos que requiere el Algoritmo MOCOM (por eso sólo se usarán en el
! programa si "Opc=3" u "Opc=4"):
! Nip: Número de individuos de la población
! Nic: Número de individuos del complejo
! NiterP: Numero de iteraciones en que son tomados los complejos de la población.
! NiterC: Numero de veces en que es sometido el Complejo al algor de Nelder y Mead.
! Nunif: es el número de veces que el vector aleat unif se automultiplica.
!

```

Figura 7.6b.: Archivo de entrada Ent_Main – parte 2

```

!
!Sexta fila: (Límites inferiores de los todos los parámetros, cualquiera sea el modelo
!elegido (IPH2 o GR).
!
!Septima fila (Límites superiores de los todos los parámetros, cualquiera sea el modelo
!elegido (IPH2 o GR).
!
!Octava fila en adelante: Lista de códigos de las cuencas a leer por cualquiera de las
!opciones, a saber:
! Cgo: Vector de Códigos de la cuenca según algún criterio
! País: Vector de Códigos (Flag) del país o característica del archivo de datos a
!leer (Frh,Frj,Ar,Br,ES,NN)
! NT: Vector de Número de Dt de la serie de cada archivo (cantidad de datos)

```

Figura 7.6c.: Archivo de entrada Ent_Main – parte 3

7.2. ALGORITMO MOCOM-UA - OPTIMIZACIÓN

El método de optimización se utiliza para realizar la calibración automática multi-objetivo, para correrlo con éxito es necesario efectuar algunos cambios de datos del archivo Ent_Main, se muestra a continuación (figura 7.7):

Archivo	Edición	Formato	Ver	Ayuda
	4	1	2	10
	1	300		
	1069.23	2.86	295.40	2.832 12. 16.2
	1	2	3	4 5 6
	100	10	5000000	1 10
	200.	-5.	20.	1.1 10. 6.
	1200.	3.	300.	2.9 13. 7.5
Gualleg00_ArD		2349		
A3422010_FrH		82791		
Y2102010_FrH		83994		
V7135010_FrH		84000		

Figura 7.7.: Archivo de entrada Ent_Main – MOCOM

En la primera fila, la primera columna debe ser el número 4, que corresponde al modo optimización (Mocom-GR4), ver figura 7.6a., la segunda columna debe ser 1 ya que es una única cuenca en estudio.

La segunda fila, la primera columna puede tomar valores de 1 al 6, siendo estos los Dt (diferenciales de tiempo horizonte respecto al horizonte actual).

La sexta y séptima fila especifican los límites mínimos y máximo respectivamente, encolumnados representan de izquierda a derecha los límites de X1, X2, X3 y X4.

La octava fila especifica el nombre del archivo de entrada de datos (ver figura 7.2.) conjuntamente con la cantidad de datos para efectuar

primeramente la calibración automática y determinar los parámetros X1, X2, X3 y X4 asociados a la mejor función objetivo, ya sea el coeficiente de Nash-Sutcliffe o coeficiente de Persistencia. Se utilizo para el caso en estudio la primera mitad de los datos, utilizando la segunda mitad para la modelación.

En síntesis se necesitan los siguientes datos, que están claramente explicados en las figuras 7.6.:

- ✓ Nip: Número de individuos de la población.
- ✓ Nic: Número de individuos de cada complejo.
- ✓ Nis: Número de individuos de cada subcomplejo.
- ✓ NiterP: Número de veces en que la población es dividida en complejos.
- ✓ NiterC: Número de veces en que son elegidos subcomplejo de cada complejo.
- ✓ Niterf: Número de veces que el vector aleatorio uniforme se automultiplica.
- ✓ X1 máx: Límite máximo del parámetro X1, en mm.
- ✓ X1 min: Límite mínimo del parámetro X1, en mm.
- ✓ X2 máx: Límite máximo del parámetro X2, en mm.
- ✓ X2 min: Límite mínimo del parámetro X2, en mm.
- ✓ X3 máx: Límite máximo del parámetro X3, en mm.
- ✓ X3 min: Límite mínimo del parámetro X3, en mm.
- ✓ X4 máx: Límite máximo del parámetro X4.
- ✓ X4 min: Límite mínimo del parámetro X4.

Con:

- ✓ X1: Capacidad del reservorio de producción, en mm.
- ✓ X2: Coeficiente de la funcionintercambio, en mm
- ✓ X3: Capacidad del reservorio de transferencia, en mm.
- ✓ X4: Tiempo de base del hidrograma unitario HU1.

Para efectuar dicha calibración se utilizaron los siguientes límites de los diferentes parámetros:

Parámetros	X1	X2	X3	X4
Unidades	[mm]	[mm]	[mm]	[día]

Mínimo	200	-5	20	1.1
Máximo	1200	3	300	2.9

Además de los siguientes parámetros algorítmicos:

Nip	Nic	NiterP	Niter C	Nunif
100	10	5000000	1	10

Se llevó a cabo la calibración seis veces, modificando entre una y otra el horizonte de pronóstico de calibración. Se comenzó con un horizonte de 1 día y luego se fue incrementando hasta llegar a los 6 días. Como la lógica indica, los mejores resultados de los coeficientes de Nash y Stutcliffe utilizados como función objetivo para calibrar en simulación (FO1), y el coeficiente de persistencia, función objetivo en previsión (FO2), se obtendrían para horizontes más cercanos, pero ocurre lo siguiente:

Coefficiente de Nash-Sutcliffe

Para cada Dt se obtiene un coeficiente de Nash-Sutcliffe (coeficiente que ajusta en modo simulación), o en la programación es la Función Objetivo 1, esto se lo representa a continuación en el gráfico donde se representa Coef de Nash-Sutcliffe vs Dt.

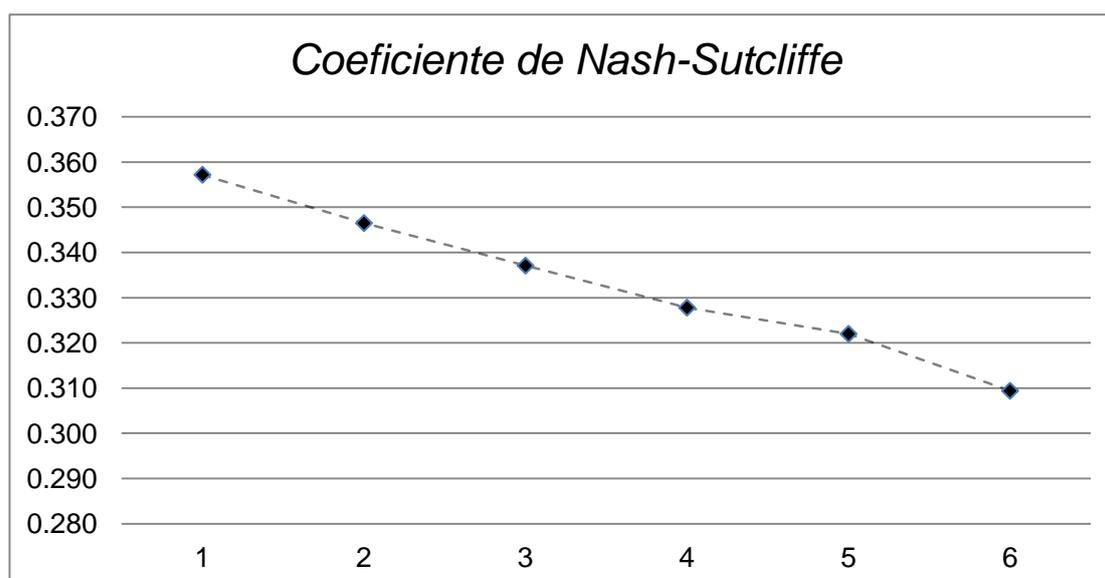


Figura 7.8.: Variación del coeficiente de Nash-Sutcliffe

Se ve claramente como disminuye a medida que aumenta el Dt, esto está relacionado a que hay mayor incertidumbre para pronósticos más alejados del horizonte actual, lo cual es lógico.

Coeficiente de Persistencia

Además del coeficiente anterior se obtiene el propio de persistencia (coeficiente que se utiliza para ajustar en modo pronóstico) que debería responder de la misma forma que el de la figura 7.8, sin embargo debido a causas que no son motivo de análisis de la presente práctica supervisada.

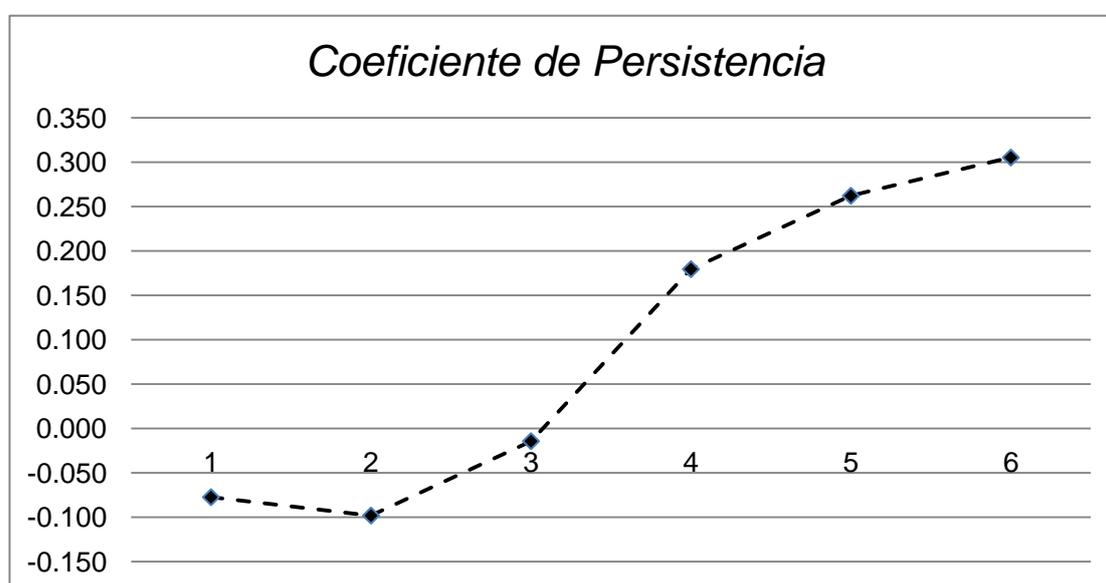


Figura 7.9.: Variación del coeficiente de Persistencia

Resulta inexplicable e ilógica la curva que se ve en gráfico, motivo por el cual sería muy interesante de aquí en adelante hacer un análisis para llegar a una conclusión coherente.

Se obtuvieron los siguientes parámetros:

X1	X2	X3	X4	Coeficiente de Nash-Sutcliffe	Salidas
354.321	1.926	294.675	2.847	0.357	Salida 1
367.907	2.248	293.040	2.888	0.347	Salida 2
356.736	2.754	299.454	2.891	0.337	Salida 3
362.095	2.408	298.312	2.818	0.328	Salida 4
365.404	2.234	299.653	2.884	0.322	Salida 5

381.277	2.524	299.937	2.894	0.309	Salida 6
X1	X2	X3	X4	Coefficiente de Persistencia	Salidas
1117.229	2.723	299.799	2.777	-0.077	Salida1
875.392	2.923	299.557	2.886	-0.098	Salida2
823.840	2.956	299.091	2.851	-0.014	Salida3
1139.874	2.608	299.631	2.782	0.180	Salida4
1143.725	2.974	297.748	2.877	0.262	Salida5
1100.509	2.830	299.400	2.806	0.305	Salida6

Estos parámetros se los utilizarán a posteriori para efectuar la modelación, los primeros para el modo simulación y los restantes para modo pronóstico.

7.3. MODELO GR - SIMULACIÓN Y PRONÓSTICO

La modelación se puede efectuar en modo simulación o en modo pronóstico, el primero ligado al coeficiente de Nash-Sutcliffe, mientras que el segundo al de Persistencia.

Se necesitan los siguientes datos, que están claramente explicados en las figuras 7.6.:

- ✓ NT: Número de datos de la serie ingresada.
- ✓ A: Área de la cuenca.
- ✓ Dt: Intervalo de tiempo, en horas.
- ✓ P: Precipitación en cada intervalo de tiempo, en mm.
- ✓ EP: Evapotranspiración potencial en cada intervalo de tiempo, en mm.
- ✓ Qobs: Caudal observado promedio en cada intervalo de tiempo, en m³/seg.
- ✓ X1: Capacidad del reservorio de producción, en mm.
- ✓ X2: Coeficiente de la función intercambio, en mm
- ✓ X3: Capacidad del reservorio de transferencia, en mm.
- ✓ X4: Tiempo de base del hidrograma unitario HU1.

A continuación se plantea la metodología que se llevo a cabo para cada modo, incluyendo gráficos que muestran las salidas correspondientes para distintos días de horizonte.

Los pasos que se siguieron:

1. Calibración automática con la primera mitad de los datos, se encuentran los juegos de parámetros correspondientes para esta primera mitad.
2. Se efectúa la modelación en modo simulación con la segunda mitad de los datos.
3. Ídem al paso anterior pero en modo previsión.
4. Se realiza nuevamente la calibración automática pero con la segunda mitad de los datos para realizar verificaciones.
5. Se verifican las salidas del paso 2 tras modelar en modo simulación con la primera mitad de los datos.
6. Ídem al paso 5 pero en modo previsión, con vistas en la verificación del paso 3.

Para las salidas con horizonte de 1 día se establecen conclusiones de acuerdo al análisis visual, mientras que las demás se las plantea en el capítulo de anexos (capítulo 10).

7.3.1. Modo Simulación

Para llevar a cabo la modelación en modo simulación es necesario disponer de el juego de parámetros que mejor ajusta la función objetivo para los distintos días de horizonte respecto de la situación actual, tal como en la calibración automática es necesario efectuar nuevos cambios de datos del archivo Ent_Main, se muestra a continuación (figura 7.10), especialmente la inclusión de éstos párametros de ajuste:

Archivo	Edición	Formato	Ver	Ayuda
	2	1	1	10
	1	300		
	1069.23	2.86	295.40	2.832
	1	2	3	4
	100	10	5000000	1
	200.	-5.	20.	1.1
	1200.	3.	300.	2.9
Gualleg01_ArD		2349		
A3422010_FrH		82791		
Y2102010_FrH		83994		
V7135010_FrH		84000		
!				

Figura 7.10.: Archivo de entrada Ent_Main – Modo Simulación

En la primera fila, la primera columna debe ser el número 2, que corresponde al modelo GR4, ver figura 7.6a., la segunda columna debe ser 1 ya que es una única cuenca en estudio, y la tercera y muy importante también tiene que tomar el valor de 1 ya que éste indica el modo simulación.

La segunda fila, la primera columna puede tomar valores de 1 al 6, siendo estos los Dt (diferenciales de tiempo horizonte respecto al horizonte actual).

La tercera fila es aquella donde se colocan los parámetros encontrados con el algoritmo MOCOM en el inciso anterior, de izquierda a derecha se colocan X1, X2, X3 y X4.

La octava fila especifica el nombre del archivo de entrada de datos (ver figura 7.2.) conjuntamente con la cantidad de datos para efectuar la modelación, en este caso se la realiza para la segunda mitad y para la calibración automática con la primera mitad, posteriormente se realiza el mecanismo inverso en busca de verificaciones.

Se llegan a las siguientes conclusiones tras realizar un análisis de las salidas para un horizonte de 1 día ($Dt=1$) en la figura 7.11., en el eje de las abscisas se representa con una unidad de 1 a 2349 los días que representan la primera mitad de los datos que se extienden desde el 01/01/1998 al 06/06/2004:

✓ Se copian bastante bien los picos a pesar de tener coeficientes de Nash-Sutcliffe para nada cercanos a 1.

✓ Se puede observar que el momento de ocurrencia del pico para los caudales simulados presenta un cierto retardo respecto del momento de ocurrencia real

✓ La velocidad con la que disminuyen los caudales observados es superior a la que rige la variación de los caudales simulados. Esto se debe a que la velocidad de vaciado de los reservorios del modelo no se ajusta bien a la realidad.

✓ El primer lapso de tiempo toma valores muy cercanos a cero, ya que es el tiempo que se toma el programa para empezar a modelar.

✓ Los primeros dos picos no llegan a superar los reales, mientras que en el tercero sí.

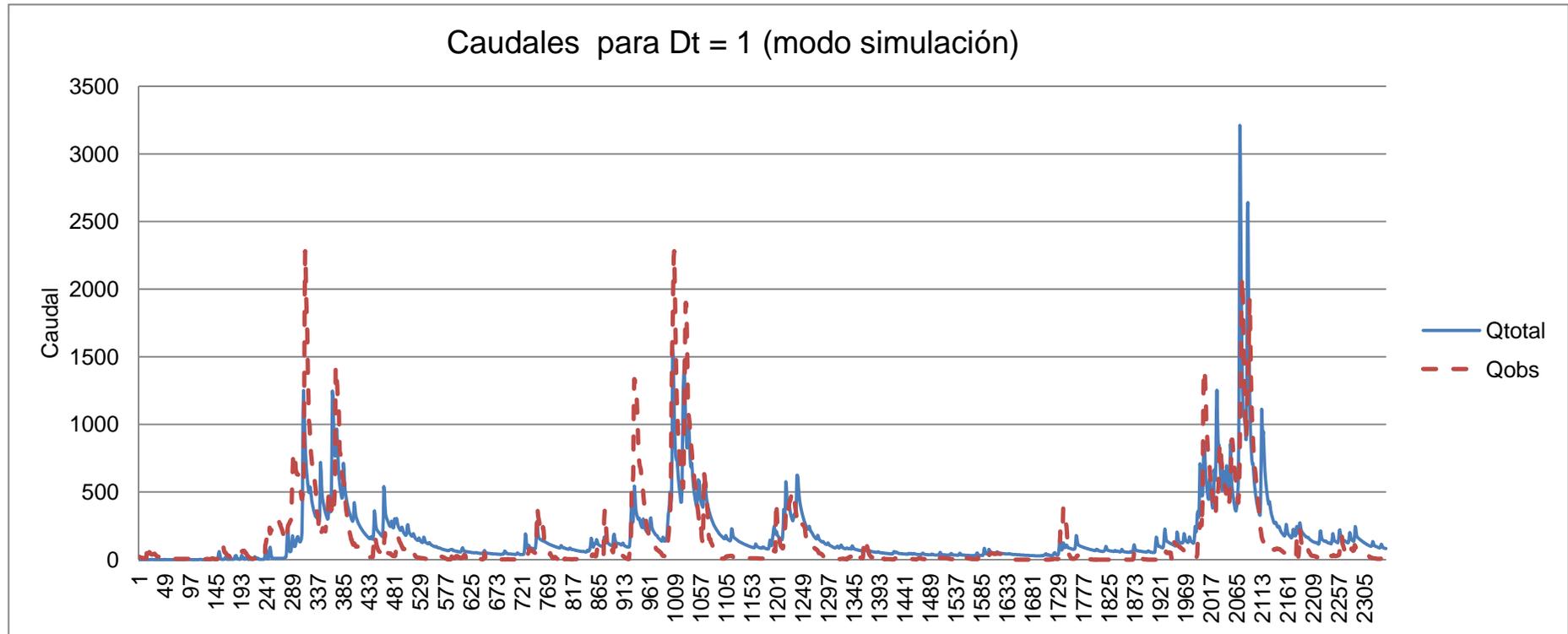


Figura 7.11a.: Salida modo simulación – primera mitad de los datos

Período de tiempo: 1 = 01/01/98 al 2349 = 06/06/04

Qtot=caudales modelados

Qobs=caudales reales

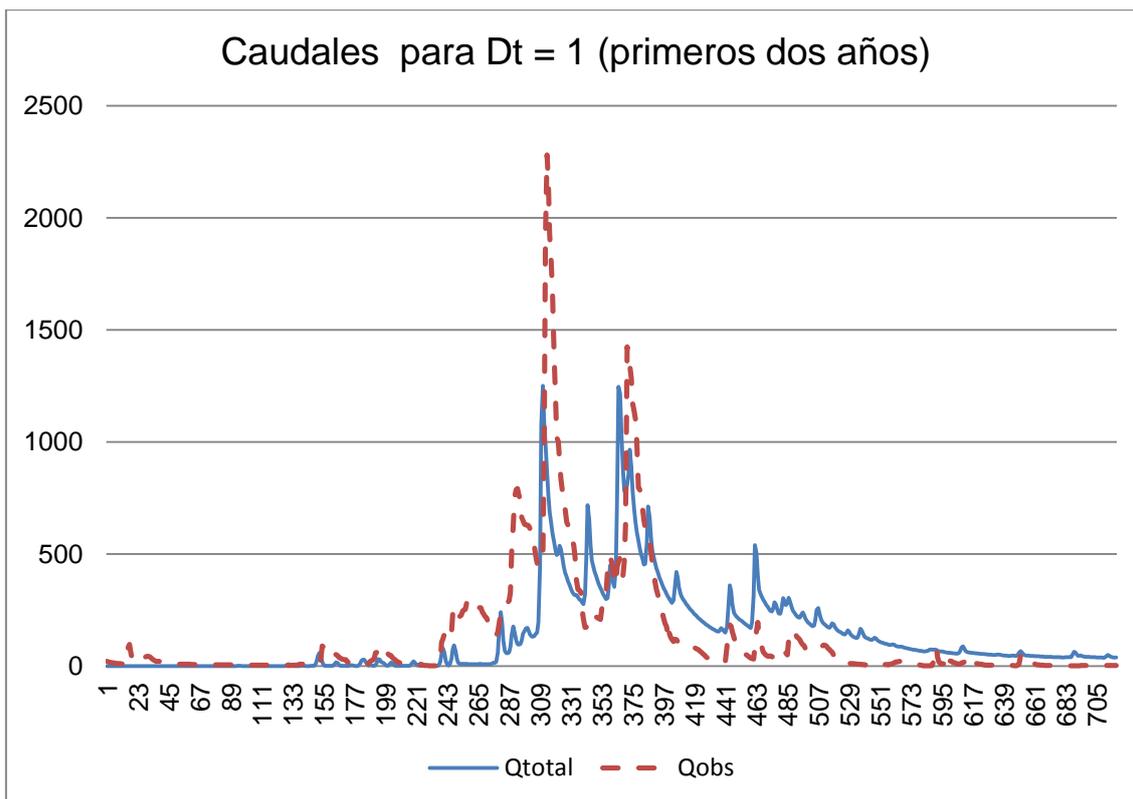


Figura 7.11b.: Salida modo simulación – primeros dos años

Período de tiempo: 1 = 01/01/98 al 705 = 06/12/99

En la figura 7.11b. se ve con claridad un retardo en la parte ascendente del hidrograma modelado respecto al real, y como en parte descendente no vaciar el reservorio, es decir no llega a anularse el caudal como si lo hacen los reales.

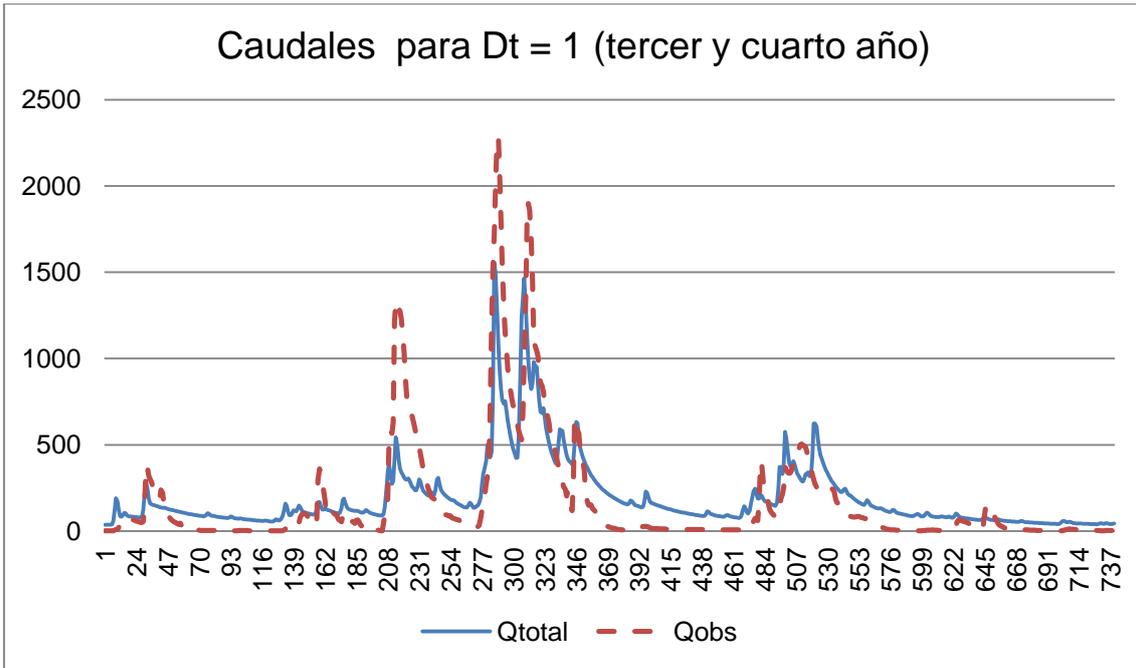


Figura 7.11c.: Salida modo simulación – tercer y cuarto año
 Período de tiempo: 1 = 07/12/99 al 705 = 12/12/01

En la figura 7.11c. se sigue apreciando lo de la anterior, incluso en ambas gráficas no se supera el pico como se había anticipado en una de las conclusiones primeras.

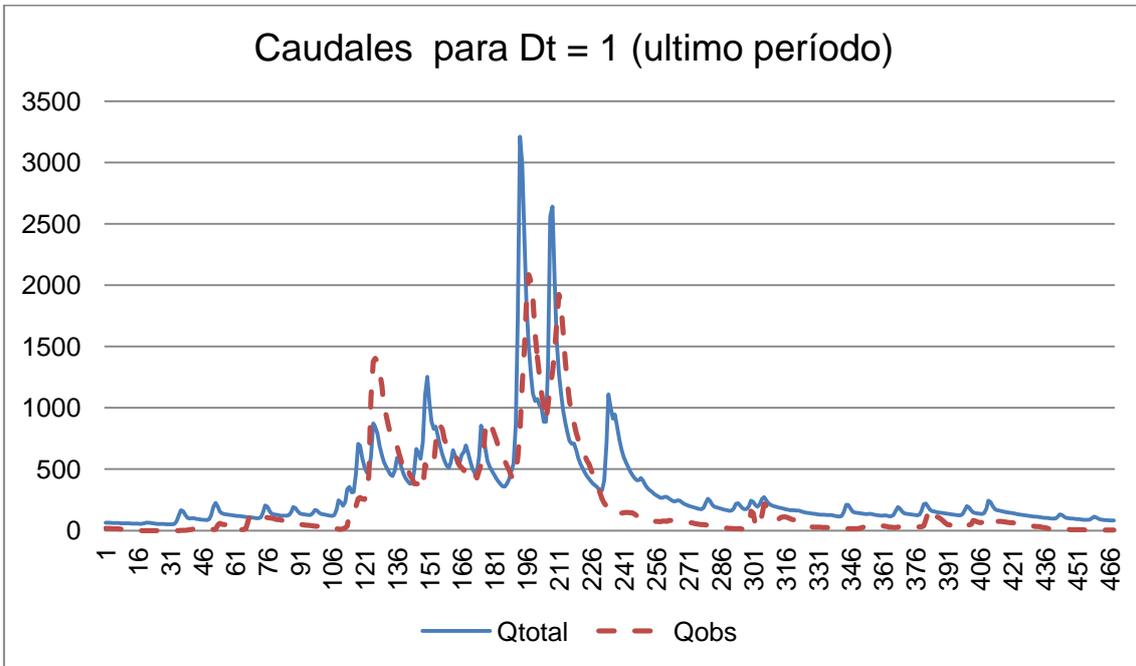


Figura 7.11d.: Salida modo simulación – último período
 Período de tiempo: 1 = 26/02/03 al 2349 = 06/06/04

En la figura 7.11d. cambia sustancialmente la diferencia, el pico de la curva modelada supera la real, y no se observa un retardo en la parte ascendente del hidrograma, pero si se mantiene la característica de falta de llenado del reservorio.

Ahora bien como se había introducido al comienzo del inciso 7.3. se realizó el mismo trabajo a modo de verificación usando para el modelo la primera mitad de los datos, que a diferencia de la primera modelación, estos se habían usado para la calibración automática, de esta labor se llegó al gráfico de la figura 7.12.

Observando este último se ve claramente que la modelación actúa muy parecido al caso anterior, incluso cumpliéndose las mismas conclusiones, especialmente como se mantiene la diferencia del llenado de los reservorios, es decir para el modelo los reservorios no se logran llenar a diferencia de lo real que si lo logra.

La serie para efectuar la modelación se extiende en este caso desde el 07/06/2004 al 10/11/2010, es decir cada unidad del eje de las abscisas representa un día de ese período.

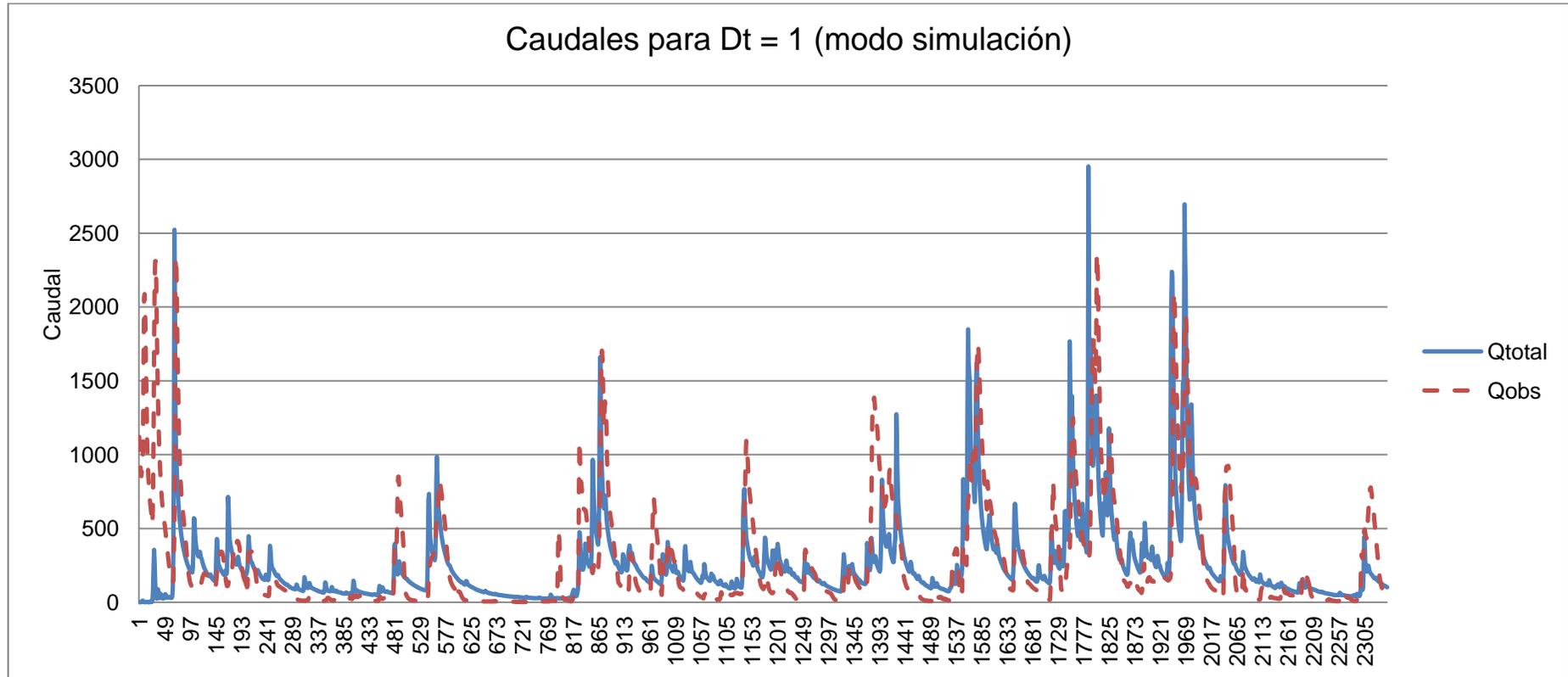


Figura 7.12.: Salida modo simulación – segunda mitad de los datos

Período de tiempo: 1 = 01/01/98 al 2349 = 06/06/04

Qtotal=caudales modelados

Qobs=caudales reales

7.3.2. Modo Pronóstico

Para llevar a cabo la modelación en modo pronóstico es necesario disponer de aquel juego de parámetros que mejor ajusta la función objetivo para los distintos días de horizonte respecto de la situación actual, de esta manera se hace necesario efectuar algunos cambios de datos del archivo Ent_Main, al igual que el caso anterior, se muestra a continuación (figura 7.13):

Archivo	Edición	Formato	Ver	Ayuda		
	2	1	2	10		
	1	300				
	1045.23	2.83	285.40	2.725	12.	16.2
	1	2	3	4	5	6
	100	10	5000000	1	10	
	200.	-5.	20.	1.1	10.	6.
	1200.	3.	300.	2.9	13.	7.5
Gua_leg01_ArD		2349				
A3422010_FrH		82791				
Y2102010_FrH		83994				
V7135010_FrH		84000				
!						
:						

Figura 7.13.: Archivo de entrada Ent_Main – Modo Previsión

En la primera fila, la primera columna debe ser el número 2, que corresponde al modelo GR4, ver figura 7.6a., la segunda columna debe ser 1 ya que es una única cuenca en estudio, y la tercera y muy importante tiene que tomar el valor de 2 ya que éste indica el modo previsión o pronóstico.

Para la segunda fila, la primera columna puede tomar valores de 1 al 6, siendo estos los Dt (diferenciales de tiempo horizonte respecto al horizonte actual).

La tercera fila es aquella donde se colocan los parámetros encontrados con el algoritmo MOCOM en el inciso anterior, de izquierda a derecha se colocan X1, X2, X3 y X4.

La octava fila especifica el nombre del archivo de entrada de datos (ver figura 7.2.) conjuntamente con la cantidad de datos para efectuar la modelación, en este caso se la realiza para la segunda mitad de los datos para el caso de efectuada la calibración automática con la primera mitad y viceversa.

Se llegan a las siguientes conclusiones tras realizar un análisis de las salidas para un horizonte de 1 día (Dt=1) en la figura 7.14a., es de aclarar

que en el eje de las abscisas se representa con una unidad de 1 a 2349 los días que representan la primera mitad de los datos que se extienden desde el 01/01/1998 al 06/06/2004:

- ✓ Los caudales modelados presentan un ajuste casi perfecto respecto a los caudales reales.
- ✓ Se aprecian diferencias en los picos, donde para los tres que se observan, la curva del modelo es inferior a la real.
- ✓ El primer lapso de tiempo toma valores muy cercanos a cero, ya que es el tiempo que se toma el programa para empezar a modelar.
- ✓ A diferencia del gráfico anterior, aquí los reservorios se llenan de la misma forma que los reales.

Se hace un zoom (figura 7.14b.) en uno de los picos para apreciar una de las únicas diferencias que se pueden apreciar al comparar ambos hidrogramas:

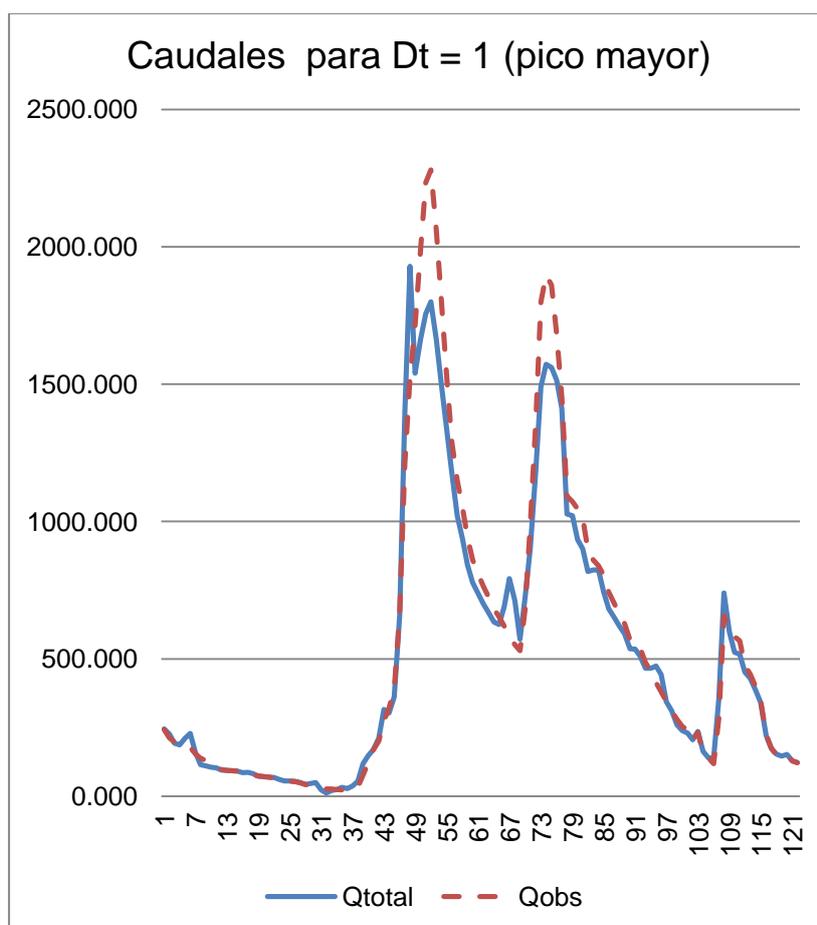


Figura 7.14b.: Salida modo pronóstico– pico central

La modelación copia bastante bien a lo real, solo la único a destacar en diferencias es que los picos reales siempre son sustancialmente mayores.

Nuevamente se hace el mismo análisis de verificación con la segunda mitad de los datos. Se puede ver en la figura 7.15a.

Las conclusiones a las que se aborda son exactamente las mismas al caso anterior, es decir no presenta diferencia el modo de modelar con una u otra serie de datos.

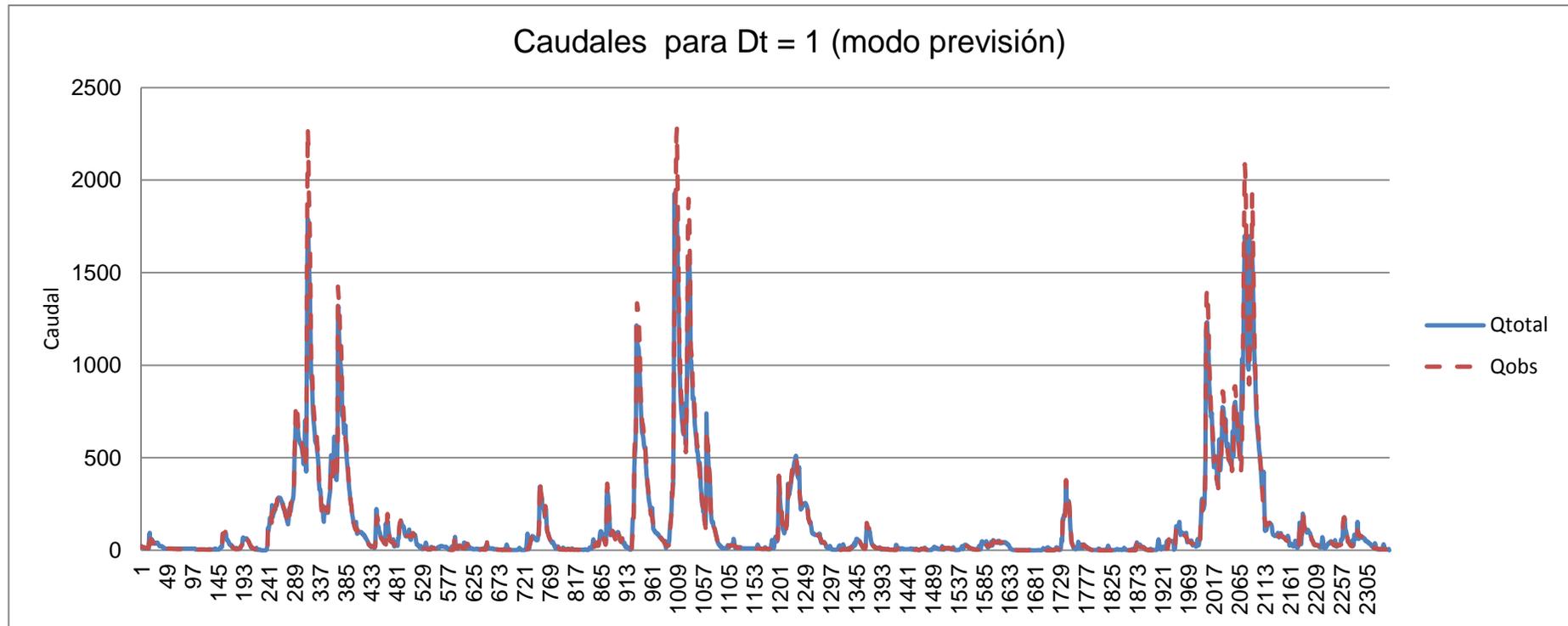


Figura 7.14a.: Salida modo pronóstico– primera mitad de los datos

Período de tiempo: 1 = 01/01/98 al 2349 = 06/06/04

Qtot=caudales modelados

Qobs=caudales reales

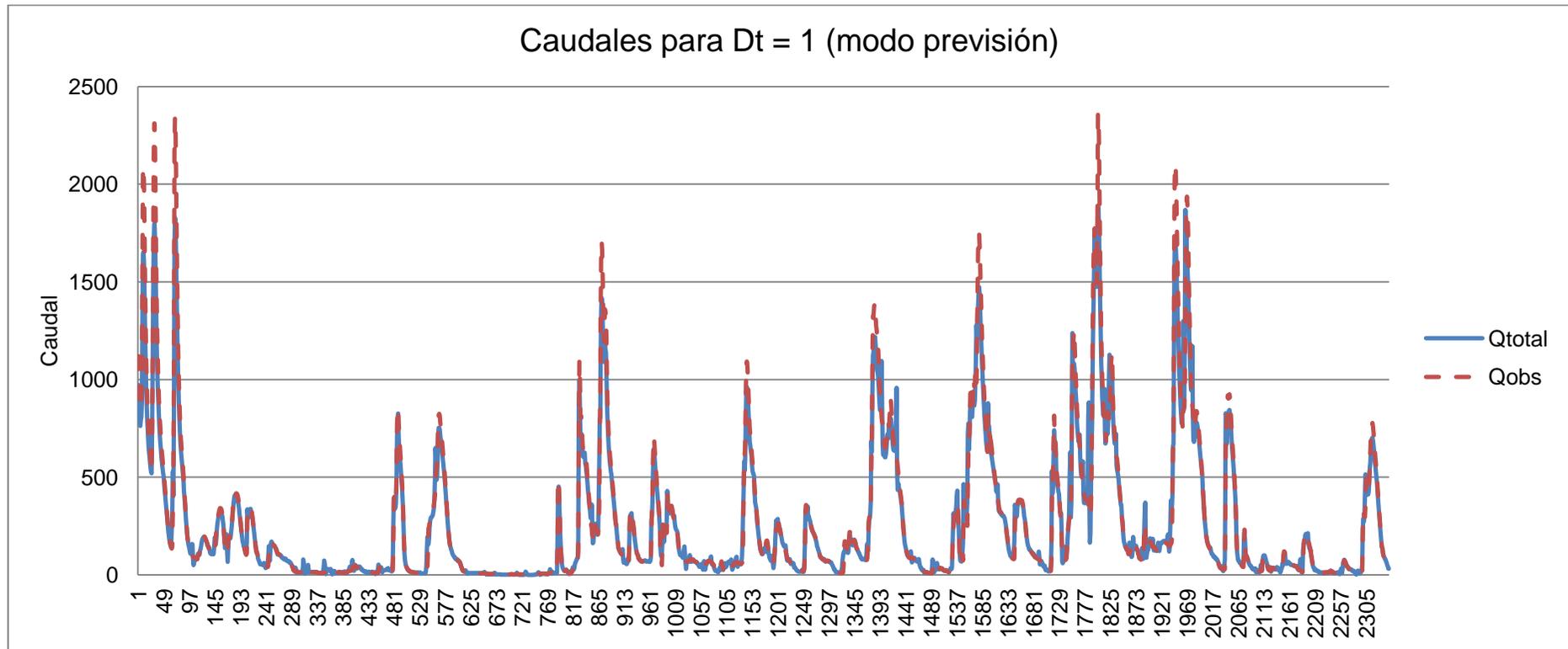


Figura 7.15.: Salida modo pronóstico – segunda mitad de los datos

Período de tiempo: 1 = 01/01/98 al 2349 = 06/06/04

Qttotal=caudales modelados

Qobs=caudales reales

8. CONCLUSIONES

El principal objetivo de esta práctica profesional supervisada consistió en poner a prueba la capacidad de un método de optimización, basado en algoritmos genéticos, para calibrar automáticamente un modelo hidrológico en diferentes tipos de cuencas argentinas. Para ello se eligió el modelo hidrológico Génie Rural.

Como primera medida se puso mucho énfasis en caracterizar de la mejor manera a las cuencas, recabando información de diferentes fuentes, se la reacomodó, incluso se rellenaron espacios vacíos aplicando criterios lógicos. Dicha tarea, planteada para dos cuencas argentinas, fue desarrollada con éxito, salvo para el caso del río Salado donde no se pudieron contar con datos de caudales, motivo por el cual solo se pudo poner a prueba la calibración para el río Gualeguay. A pesar de esto se dedicó mucho más tiempo en el reacomodamiento de los datos del Salado, acciones no menos importantes por conformar un puntapié inicial para que en un futuro cuando se dispongan la totalidad de los datos se pueda terminar el proceso.

Se realizó un barrido estadístico de las series completas para realizar recortes donde no se aprecien comportamientos normales, e incluso caracterizarlas según la dinámica de cambio de los datos y abordar a conclusiones tales que informen si los datos actuales se ven o no influenciados por los anteriores.

Por último se puso a prueba la calibración automática exponiendo sus salidas a través de gráficos, donde tras un análisis visual se abordó a variadas conclusiones netamente perceptivas, sin centrarse demasiado en el comportamiento estrictamente teórico, ya que escapa del objetivo del desarrollo de éste trabajo final.

Se plantean diversas conclusiones de diferentes tipos, las de origen estrictamente técnico y las más importantes que son aquellas que dejaron algún aprendizaje humano, que son la base de la formación profesional, entre ellas se destacan:

✓ Para realizar este tipo de trabajo, especialmente de recopilación de datos es sumamente necesario contar con diversos contactos que faciliten dicha labor, situación con la cual se contó a través del Ing Alonso F.J.

✓ La información recolectada requirió de un trabajo minucioso de procesamiento, es decir que a pesar de contar con la ventaja de poseer datos, situación difícil en nuestro medio, la información requiere de un procesamiento importante.

✓ Fue necesario incursionar en software nunca antes usados como ArcGis y la utilización de macros en Excel.

✓ Se aplicaron conceptos estadísticos ligados a situaciones reales, como es el comportamiento de series de precipitaciones y caudales.

✓ Se comprobó que la hidrología del mundo real tiene muchísima complejidad, que sólo se comprende trabajando en casos reales.

✓ Se verificó la escasa disponibilidad de datos de modelación existentes en Argentina.

✓ Se obtuvo información en múltiples formatos: en papel, escaneados, fotos, en planillas, en documentos de texto, en imágenes satelitales, etc, que condujo a la necesidad de darles distintos tratamientos.

✓ Se observa que el trabajo manual parece insustituible, ya que si bien los programas agilizan la mayor parte de las tareas, la prueba final debe ser la observación de consistencia en los datos.

✓ Se aprendió el manejo de un programa hidrológico de gran difusión en Europa, como es el GR.

✓ Se aprendió el manejo de un método de optimización global multiobjetivo de gran difusión a nivel mundial, como es el MOCOM-UA.

✓ Se obtuvieron coeficientes de ajuste que subestiman la calidad del ajuste, con respecto a la calidad observada al menos en la mayor parte de las series probadas. Esto podría deberse a que en ciertos tramos hay falencias grotescas en la consistencia PQ que fueron dejadas para no extender demasiado la depuración de las series.

✓ Se observó que, aun existiendo deficiencias en ciertos tramos, el método de optimización arribó a juegos de parámetros que explican bastante bien la mayor parte de las series simuladas.

✓ Se destaca que las FO son semejantes en simulación que en pronóstico, siendo sin embargo mejores en este último caso.

✓ Se destaca que las FO son semejantes en el proceso de calibración que en el de verificación.

✓ Se observa que, para ciertos casos, a medida que aumenta el horizonte de pronóstico, se reduce el error del ajuste, lo cual es extraño, por lo cual debería ser objeto de futuros análisis.

✓ Se aplicaron muchos criterios de tipo ingenieriles, aspecto de mucha importancia relacionado al objetivo principal de la PS.

✓ Se trabajó en equipo, de mucha importancia e indispensable para el logro de cualquier objetivo, por la coordinación que se logra, disminución de tiempos de trabajo, entre otros.

9. BIBLIOGRAFÍA

9.1. REFERENCIAS BIBLIOGRÁFICAS

Alonso, F.J. (2008). Calibración automática de un modelo hidrológico basada en algoritmos genéticos, aplicación a dos cuencas típicas argentinas. Tesis de Maestría. FCEFyN. Universidad Nacional de Córdoba.

Zoraida Omaña G. (2006). *Introducción a la Investigación de Operaciones*. Universidad de Carabobo, Venezuela.

Bertoni, J.C. (2003). Hidrología Avanzada. Notas de clases. Curso de Postgrado.

Duan, Q., Sorooshian, S., Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* Vol. 28, Nº 4 pp. 1015-1031.

HEC-HMS. (2000). Hydrologic Modeling System. Technical Reference Manual. U.S.Army, Hydrologic Engineering.

Chocat, B. (1997). Encyclopédie de l'hydrologie urbaine et de l'assainissement. Eurydice 92, Ed. Lavoisier, Paris, 1124 p.

Chow, V.T., Maidment, D.R., Mays, L.W. (1994). Hidrología Aplicada. Mc Graw Hill.

Clarke, R.T. (1973). A review of some mathematical models used in hydrology, with observations on their calibration and use. *Journal of Hydrology*, 19, 1-20.

Castillo, E., Conejo, A., Pedregal, P., García, R. y Alguacil, N. (2002). Formulación y Resolución de Modelos de Programación Matemática en Ingeniería y Ciencia. Universidad de Castilla-La Mancha, España.

Palahí M., Pukkala T. (2004). Métodos de optimización Heurística para la resolución de modelos de planificación forestal. Cuadernos de S.E.C.F. 18:103-108. I Reunión del grupo de trabajo sobre modelización forestal. Palencia, marzo, 2004.

Kuri, A., Galaviz, J. (2002). Algoritmos Genéticos. Fondo de Cultura Económica, UNAM, IPN.

Viveros, E.R.C.; Barán, B.; Almirón, M. (1998). Optimización Basada en Colonias de Hormigas con Heurística de Inicialización. In: IX Panel de Informática, 1998, Asunción. EXPOMÁTICA '98.

Goldberg, D.E. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Pub. Co.

Duan, Q., Sorooshian, S., Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research Vol. 28, Nº 4 pp. 1015-1031.

Yapo, P.O., Gupta, H. V. and Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. Journal of Hydrology. Vol. 204. Pags. 83-97.

Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W. and Sorooshian, S. (2003). Effective and efficient algorithm for multiobjective optimization of hydrologic models. Water Resources Research. Vol. 39. Nº 8. Pags. 5-1 a 5-19.

Vrugt, J.A., Gupta, H.V., Bouten, W. and Sorooshian, S. (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research. Vol. 39. Nº 8. Pags. 1-1 a 5-16.

Herring, C., Palmore, J.I. (1989). Random number generators are chaotic. ACM SIGPLAN Notices archive. V 24, Issue 11, P.76-79

Eckhardt, R. (1987). Stan Ulam, John Von Neumann, and The Monte Carlo Method. Los Alamos Science Special Issue.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002). Molecular Biology of the Cell. New York: Garland Publishing. 4th ed.

Gomez Rocha, H. y Acosta, J., Velez-Langs, O. (2006). Utilización de la Computación Evolutiva y la Teoría de Juegos para la Automatización en Sistemas de Soporte a la Negociación. Cap.3. Trabajo Final de la Carrera de Ingeniero de Sistemas. Universidad de Magdalena (Santa Marta-Colombia).

Collischonn, W. (2001). Simulação hidrológica de grandes bacias. Tesis Doctoral. UFRGS – IPH.

Beven, K., Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological Process.*, 6, 279-298.

Sorooshian, S. and Gupta, V.K. (1985). The Analysis of Structural Identifiability: Theory and Application to Conceptual Rainfall-Runoff Models. *Water Resources Research*. Vol. 21. N° 4, Pags. 487-495.

Beveridge, G., Schechter, R. (1979). *Optimization Theory and Practice*. MacGraw-Hill Book Co.

Ibbitt, R.P. (1970). Systematic parameter fitting for conceptual models of catchment hydrology. Ph.D. dissertation, Imperial College of Science and Technology, University of London.

Johnston, P.R., Pilgrim, D.H. (1976). *Parameter optimization for watershed model*. *Water Resources Research* Vol. 12, N° 3 pp. 477-486.

Brazil, L.E. (1988). *Multilevel calibration strategy for complex hydrologic simulation models*. PhD Dissertation Colorado State University, Fort Collins.

Kuczera, G., (1997). *Efficient subspace probabilistic parameter optimization for catchment models*. *Water Resources Research*. Vol. 33. N° 1. Pags. 177-185.

Tassi, R. ; Bravo, J. M. ; Meller, A. ; Collischonn, W. (2006). *Avaliação do impacto de diferentes funções-objetivo na calibração automática do modelo hidrológico IPH2*. In: XXII Congreso Latinoamericano de Hidráulica, 2006, Ciudad Guayana. Anais del XXII Congreso Latinoamericano de Hidráulica.

Madsen. H. (2000). *Automatic calibration of a conceptual rainfall-runoff model using multiple objectives*. *Journal of Hydrology*. Vol. 235. Pags. 276-288.

Nash, J.E., Sutcliffe, J.V. (1970). *River flow forecasting through conceptual models part I — A discussion of principles*. *Journal of Hydrology*, 10 (3), 282–290.

Nelder, J.A., Mead, R. (1965). *A Simplex method for function minimization*. Computer Journal, 7: 303-313.

Bacchiega J.D.; Bertoni J.C.; Maza. J.A. (2003). Pericia Hidráulica correspondiente al Expediente N° 1341/2003 - Anexo V: Dinámica hídrica de la cuenca del río Salado.

Bianchi J., Goniadzki D. *Modelación hidrológica de la cuenca del río Gualeguay para pronóstico y alerta hidrológico utilizando información de sensores remotos de microondas*. Instituto Nacional del Agua. Ezeiza, Argentina.

Giordano, L., (2011). *Cuenca del río Gualeguay en Rosario del Tala. Proceso de transformación lluvia total-escorrentía total. Desarrollo de herramientas de alerta y pronóstico hidrológico*. Instituto Nacional del Agua. Ezeiza, Argentina.

Goniadzki, D., M. Thibeault, H. Lozza, M. Uriburu Quirno, & D. Dadamia, (2011). *SACD/Aquarius Announcement of Opportunity. Assimilation of Soil Moisture Estimates into Flow-Forecasting Hydrologic Models*. Instituto Nacional del Agua. Ezeiza, Argentina.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andreassian, V., Anctil, F., Loumagne, C. (2005.b). Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 – Towards a simple and efficient potential evapotranspiration model for rainfallrunoff modelling. *Journal of Hydrology* 303(1-4), 290-306.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andreassian, V., Anctil, F., Loumagne, C. (2005.a). Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? *Journal of Hydrology* 303(1-4), 275-289.

9.2. PÁGINAS WEBS

CONICYT. Programa Explora. Divulgación y Valoración de la Ciencia y la Tecnología.

USACH (Universidad de Santiago de Chile). <http://optimos.usach.cl/PH.htm>. Página de Métodos Heurísticos de Optimización. Consultada en Noviembre de 2006.

UROU, Facultad de Ingeniería, Universidad República Oriental del Uruguay. Introducción a la investigación de operaciones. Apuntes de clases. <http://www.fing.edu.uy/inco/cursos/io/>. Uruguay. Consultada en Enero de 2007.

Subsecretaría de Recursos Hídricos. Argentina. <http://www.hidricosargentina.gov.ar/>. Consultada en agosto del 2013.

Enciclopedia Wikipedia. http://en.wikipedia.org/wiki/Latent_heat

Food and Agriculture Organization of the United Nations. Radiación extraterrestre. <http://www.fao.org/docrep/x0490e/x0490e0j.html>

10. ANEXO

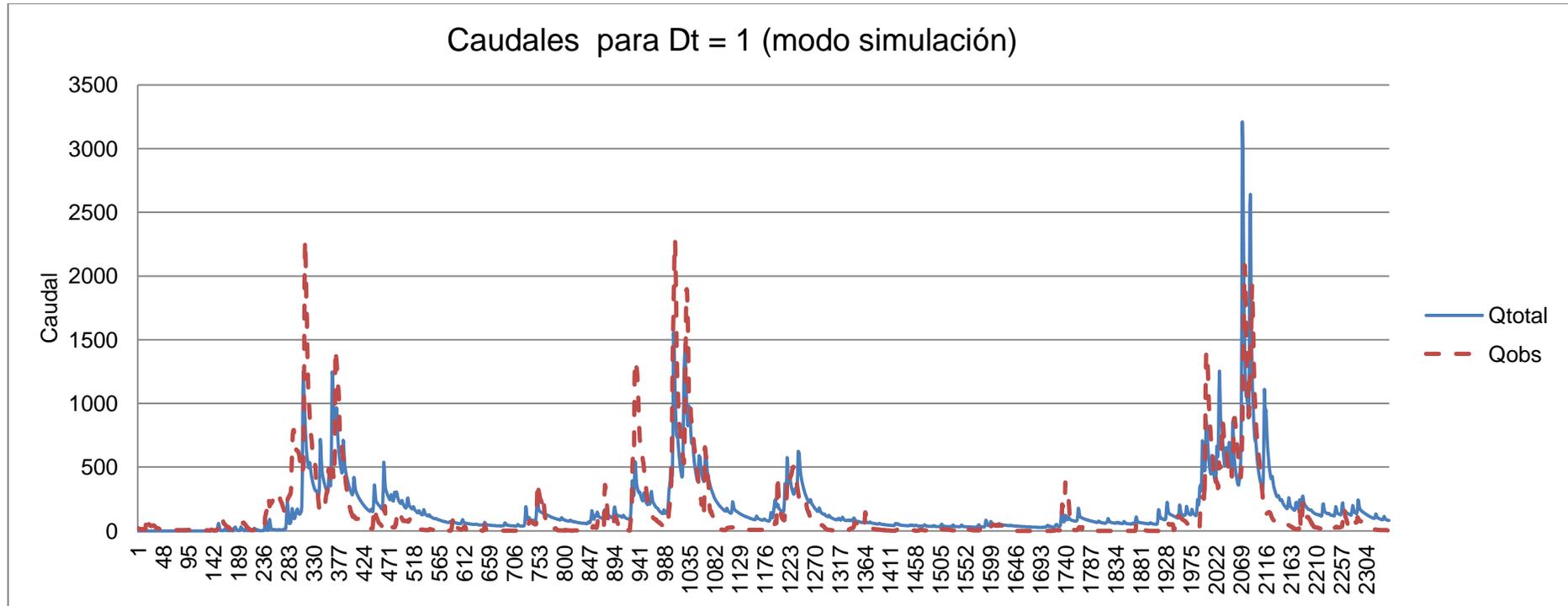


Figura 10.1.: Salida modo simulación – primera mitad de los datos – horizonte=1 día

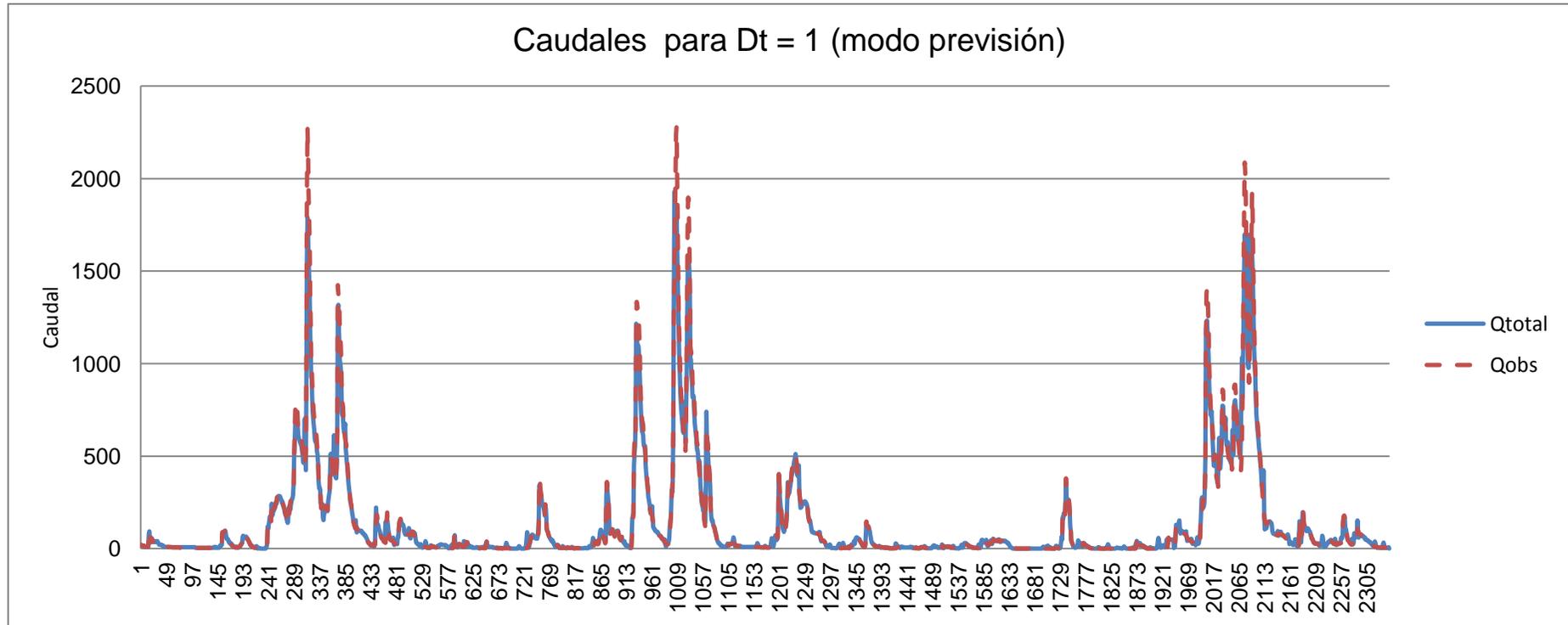


Figura 10.2.: Salida modo previsión – primera mitad de los datos – horizonte=1 día

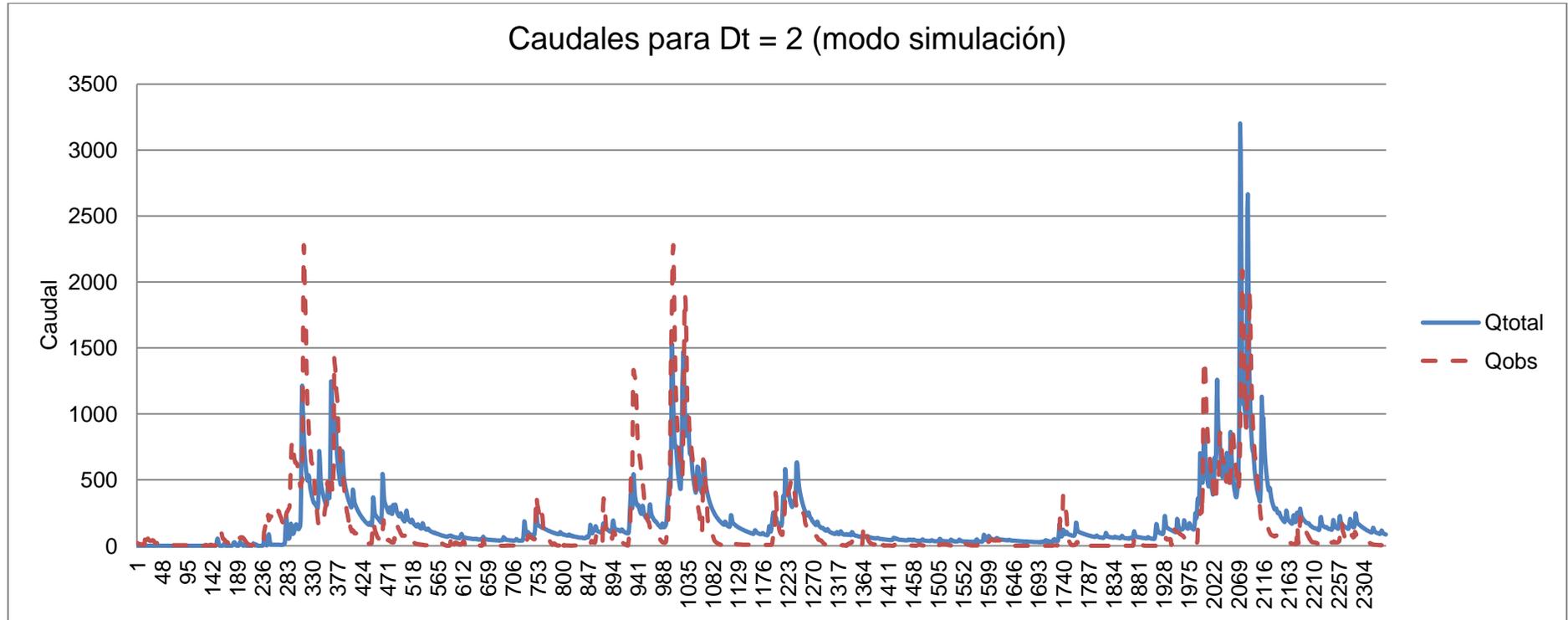


Figura 10.3.: Salida modo simulación – primera mitad de los datos – horizonte=2 días

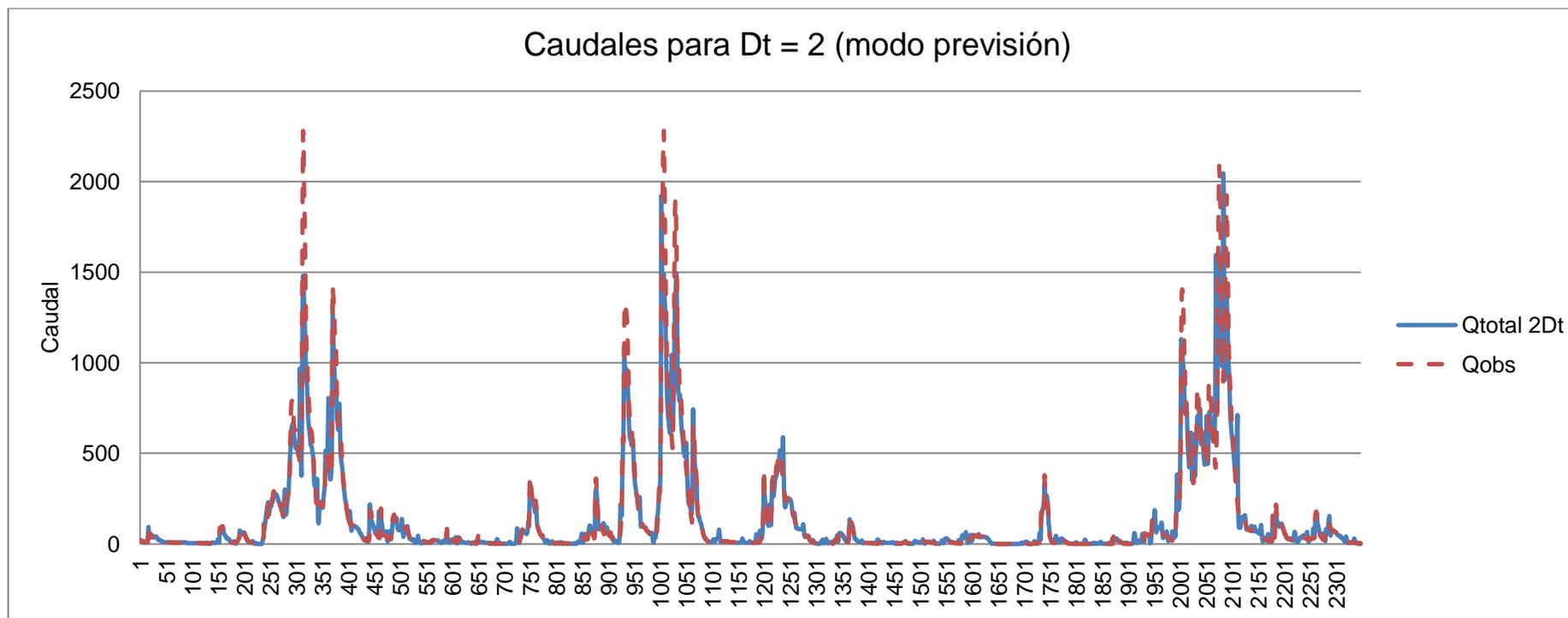


Figura 10.4.: Salida modo previsión – primera mitad de los datos – horizonte=2 días

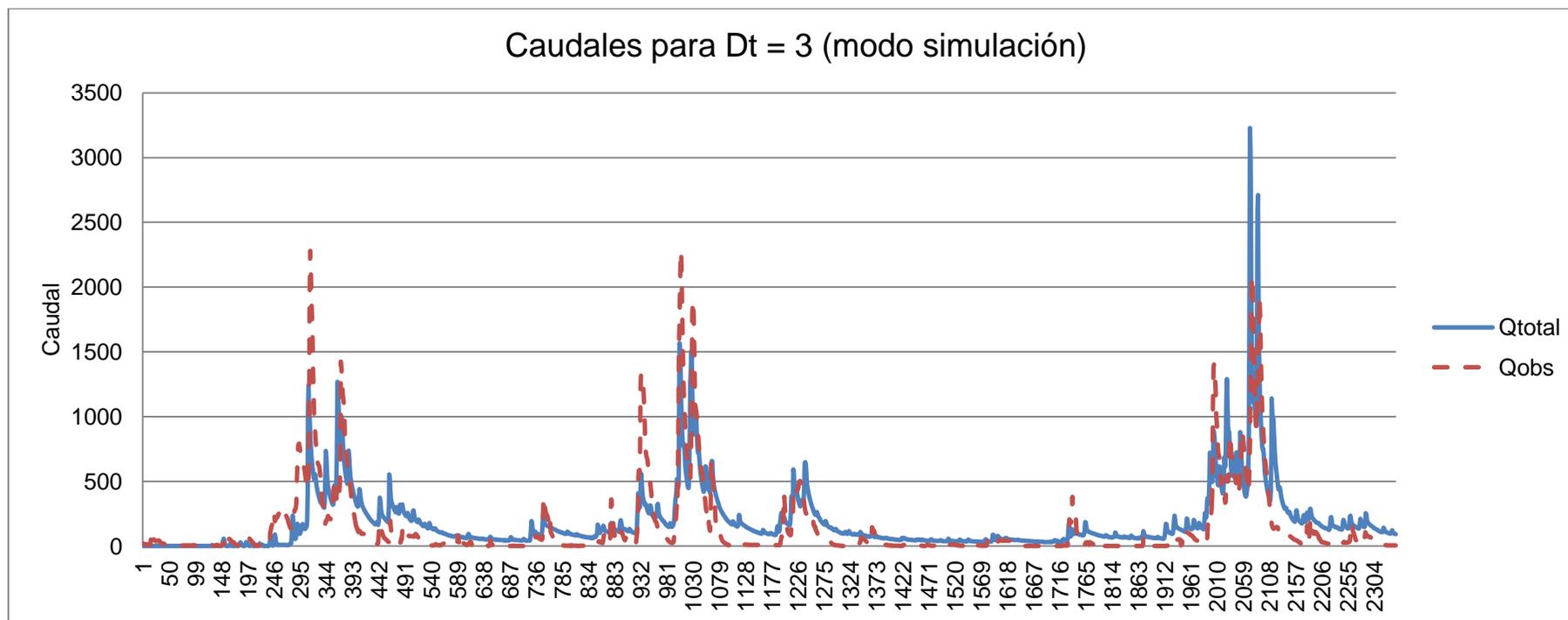


Figura 10.5.: Salida modo simulación – primera mitad de los datos – horizonte=3 días

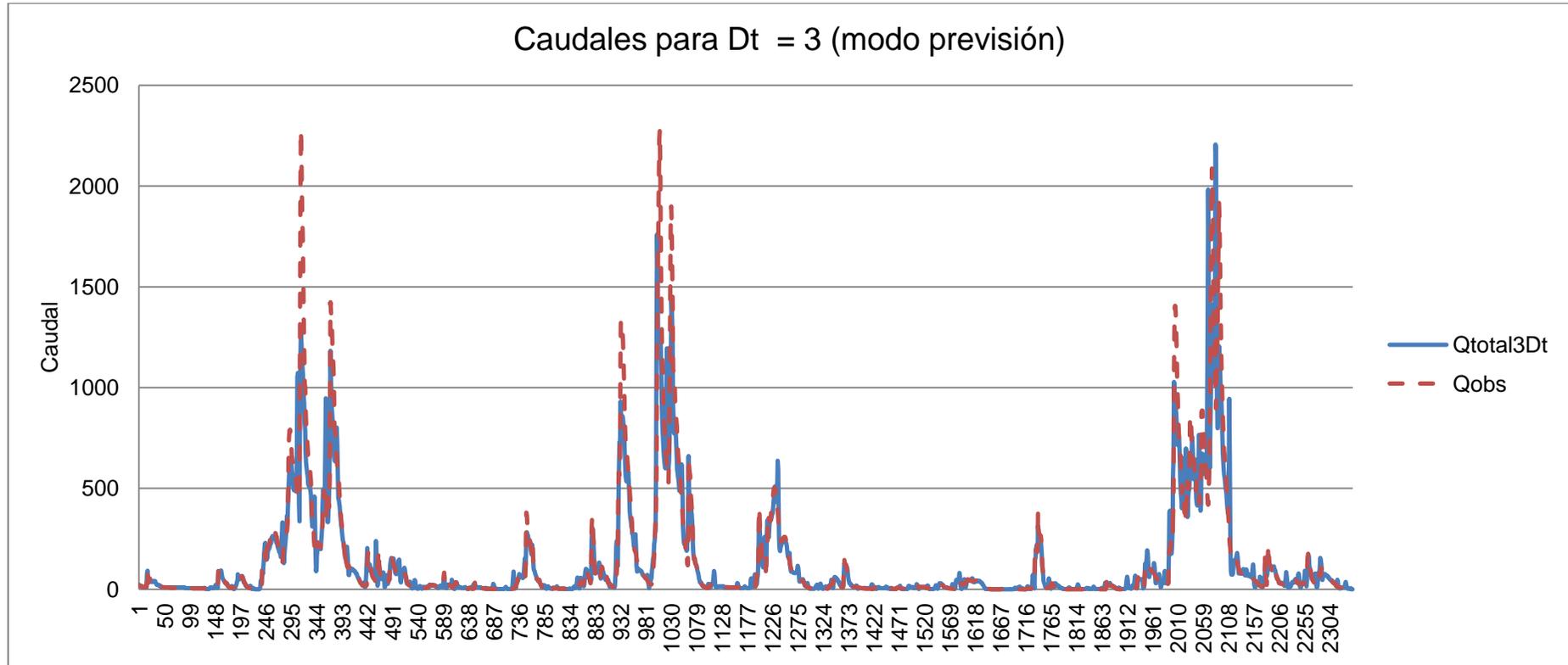


Figura 10.6.: Salida modo previsión – primera mitad de los datos – horizonte=3 días

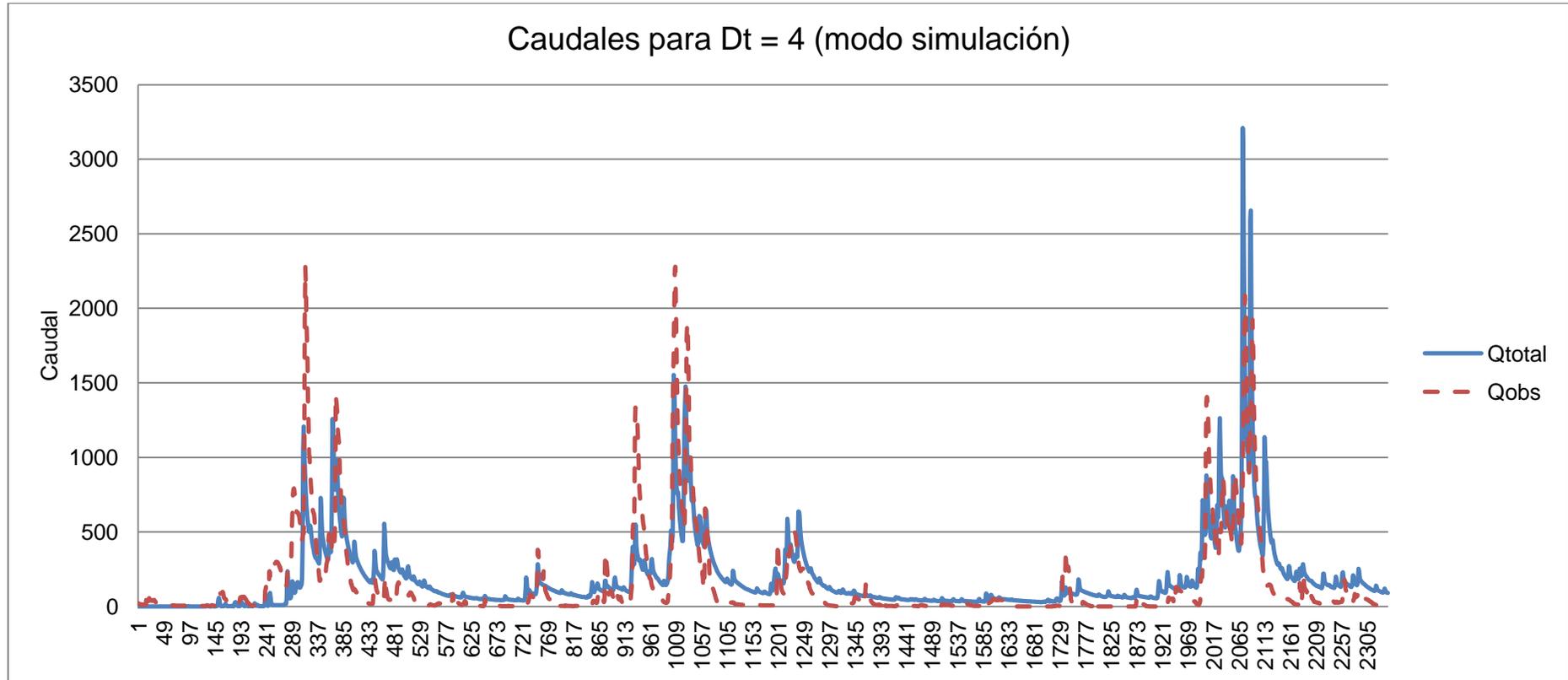


Figura 10.7.: Salida modo simulación – primera mitad de los datos – horizonte=4 días

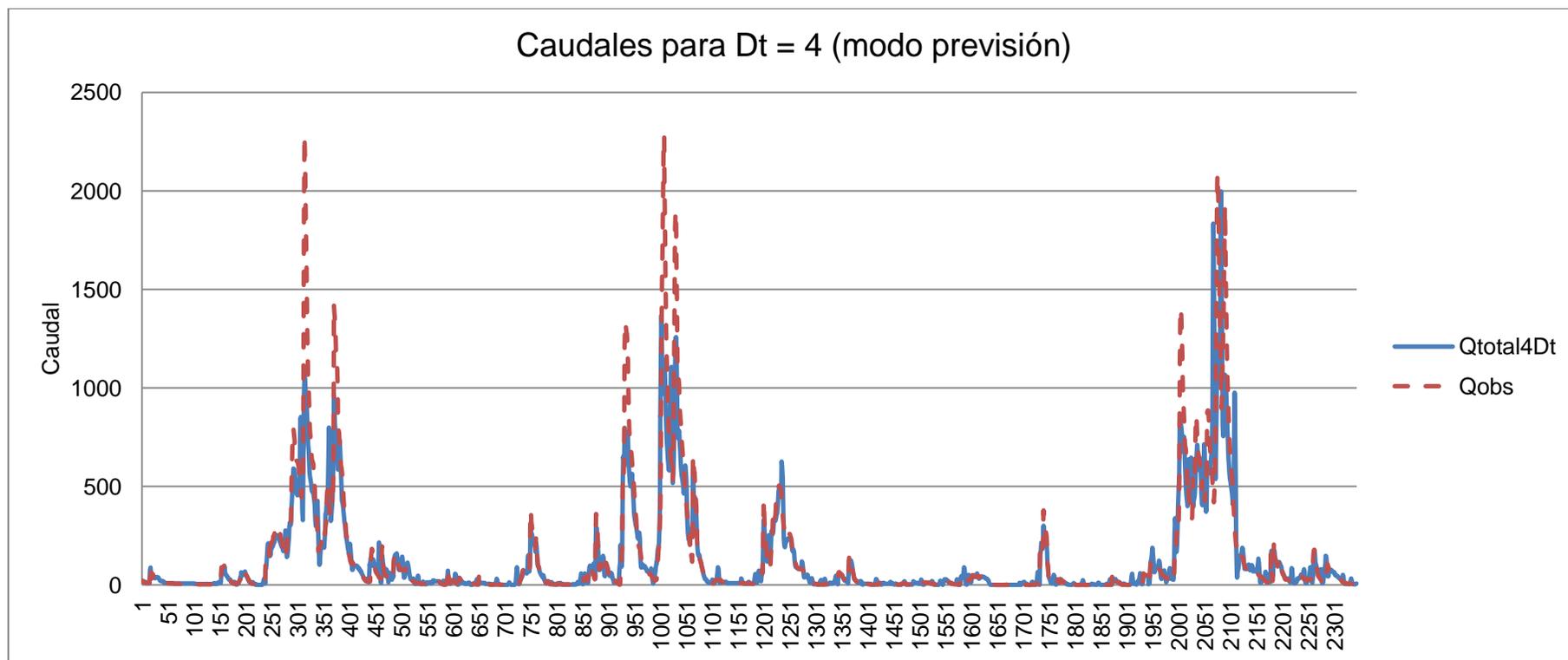


Figura 10.8.: Salida modo previsión – primera mitad de los datos – horizonte=4 días

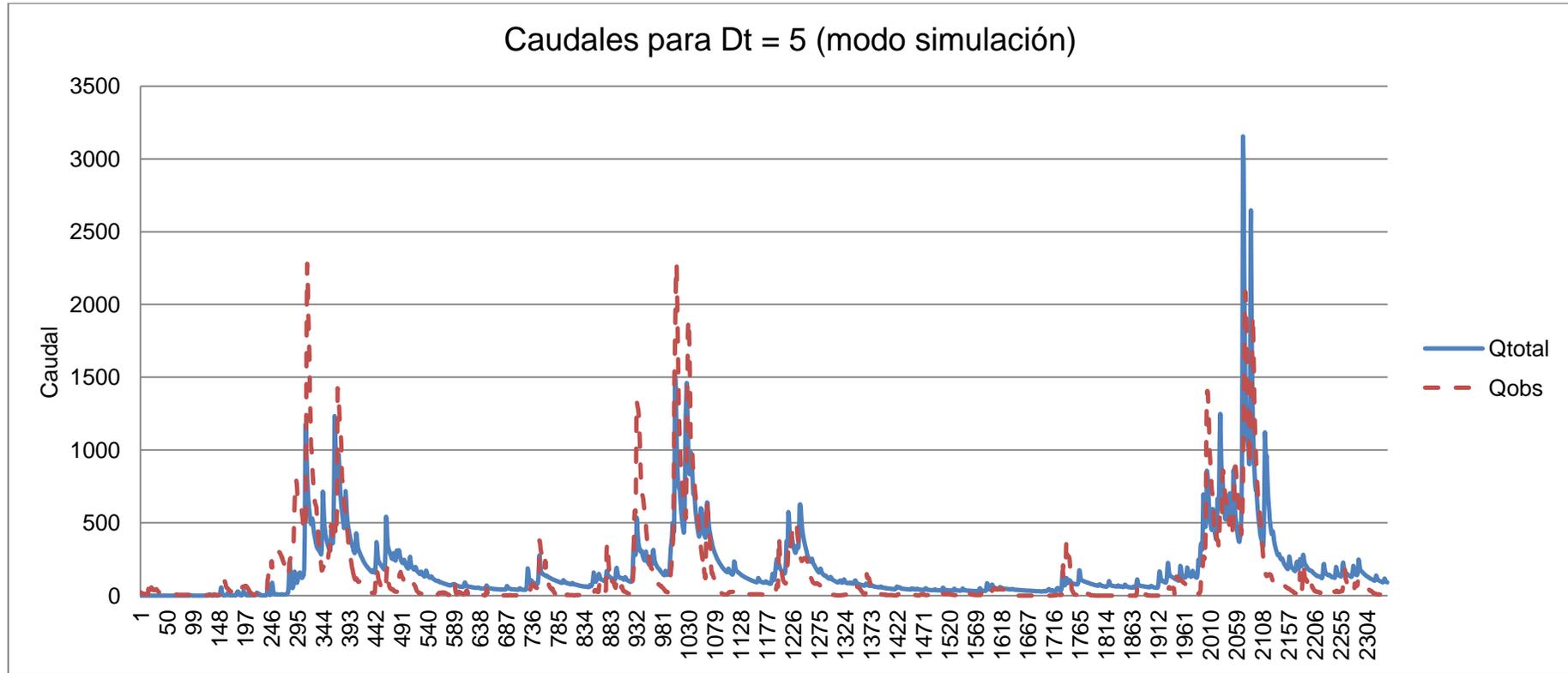


Figura 10.9.: Salida modo simulación – primera mitad de los datos – horizonte=5 días

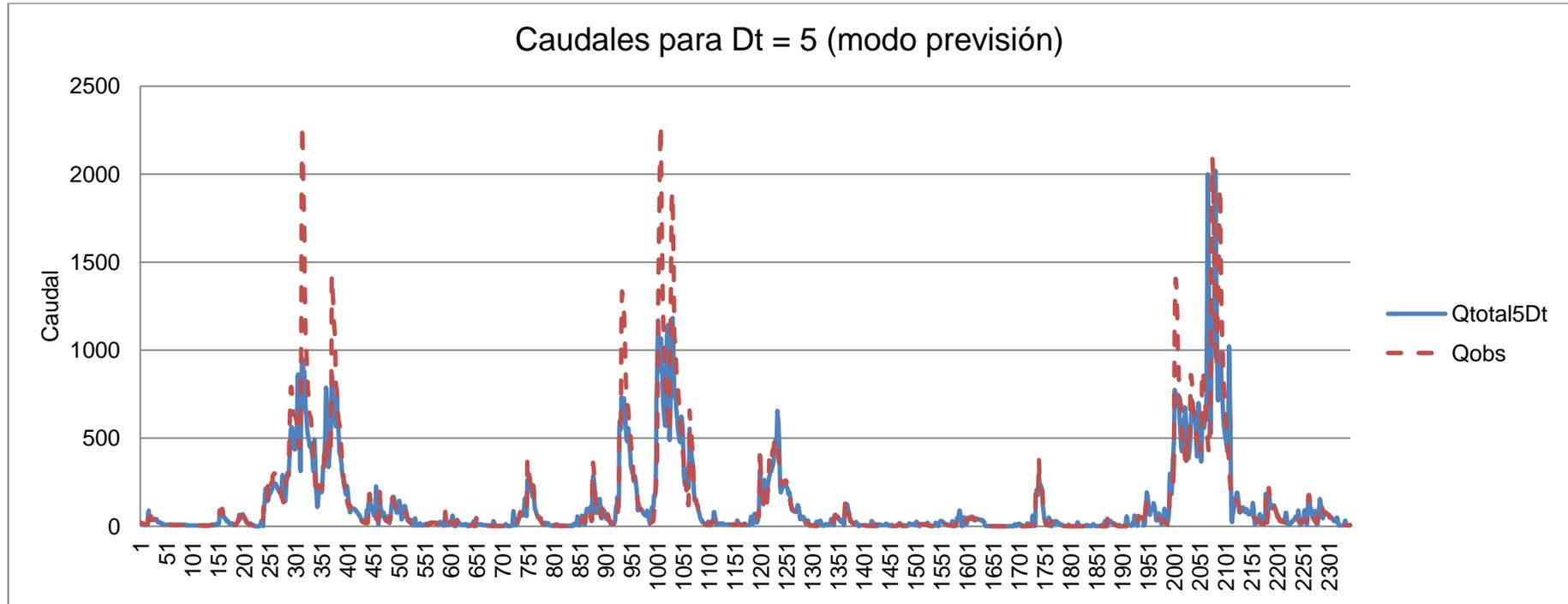


Figura 10.10.: Salida modo previsión – primera mitad de los datos – horizonte=5 días

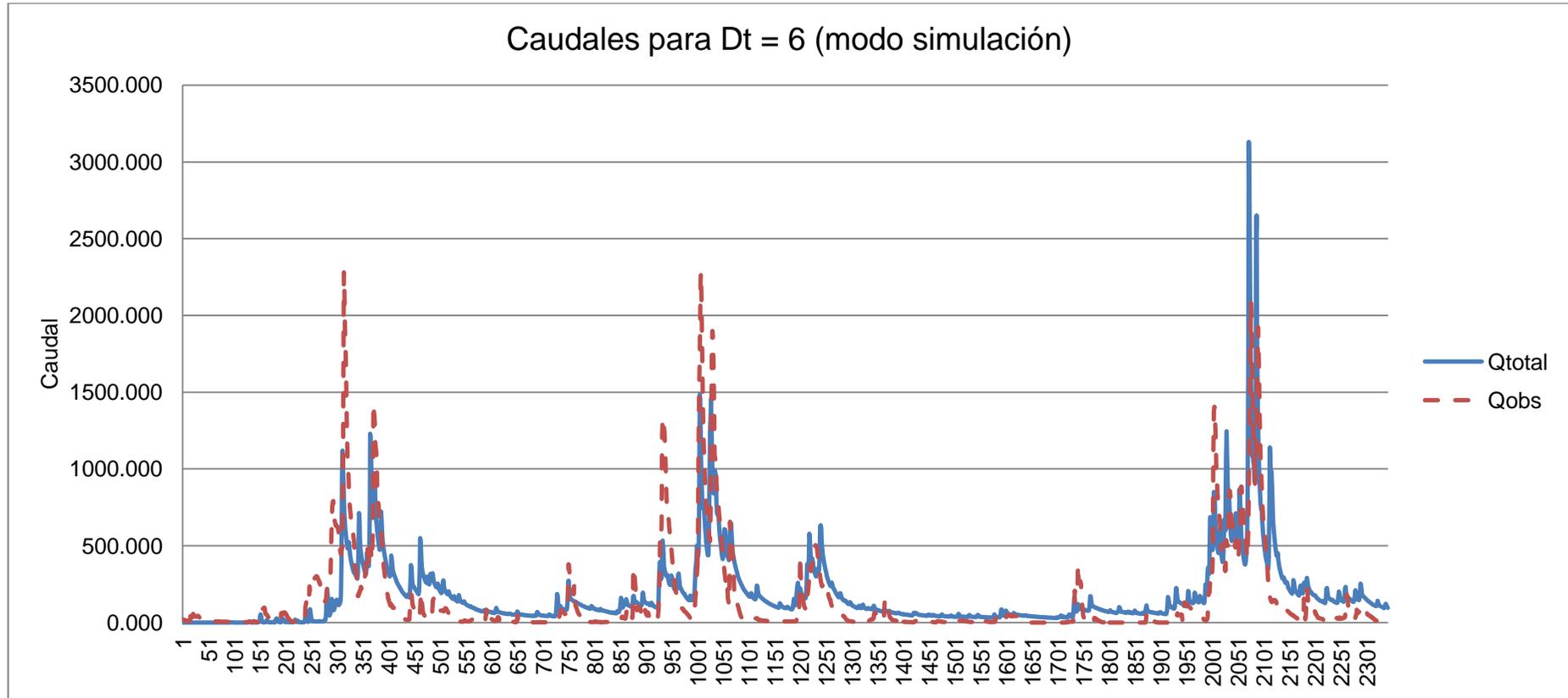


Figura 10.11.: Salida modo simulación – primera mitad de los datos – horizonte=6 días

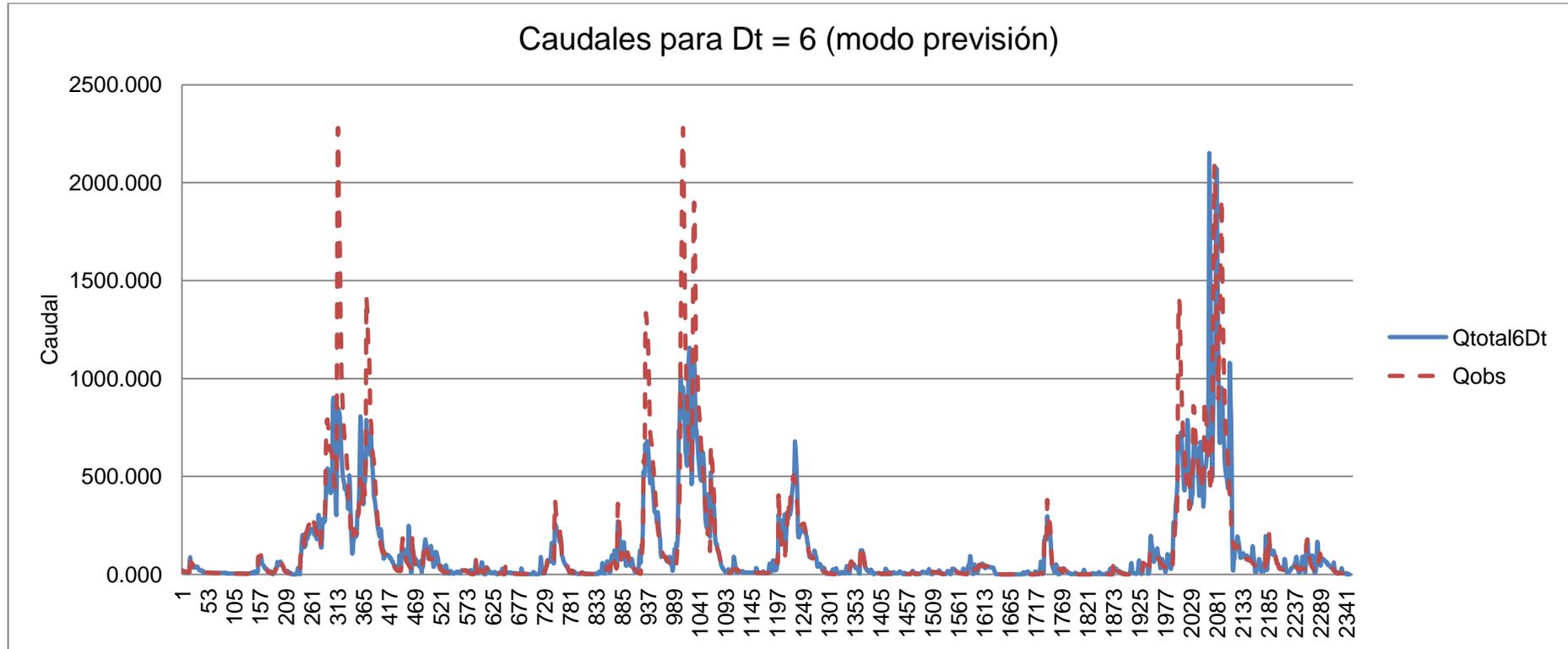


Figura 10.12.: Salida modo previsión – primera mitad de los datos – horizonte=6 días

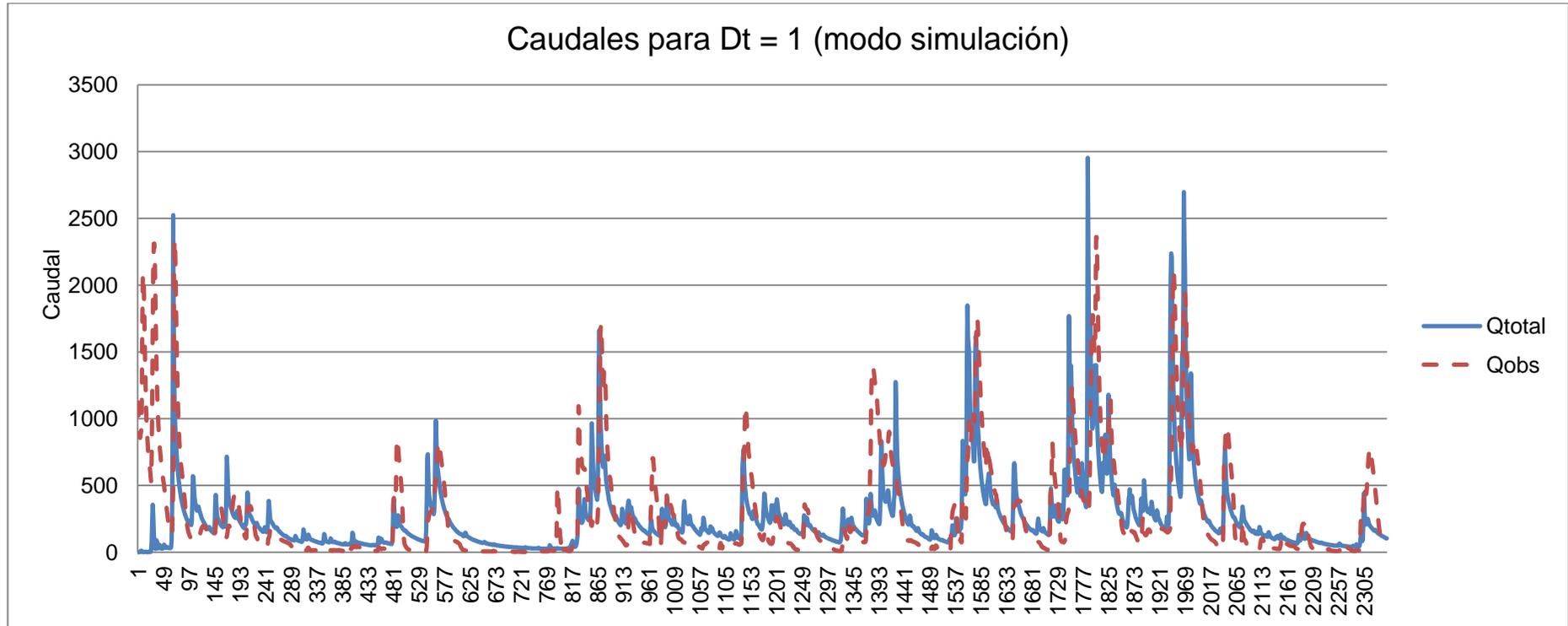


Figura 10.13.: Salida modo simulación – segunda mitad de los datos – horizonte=1 día

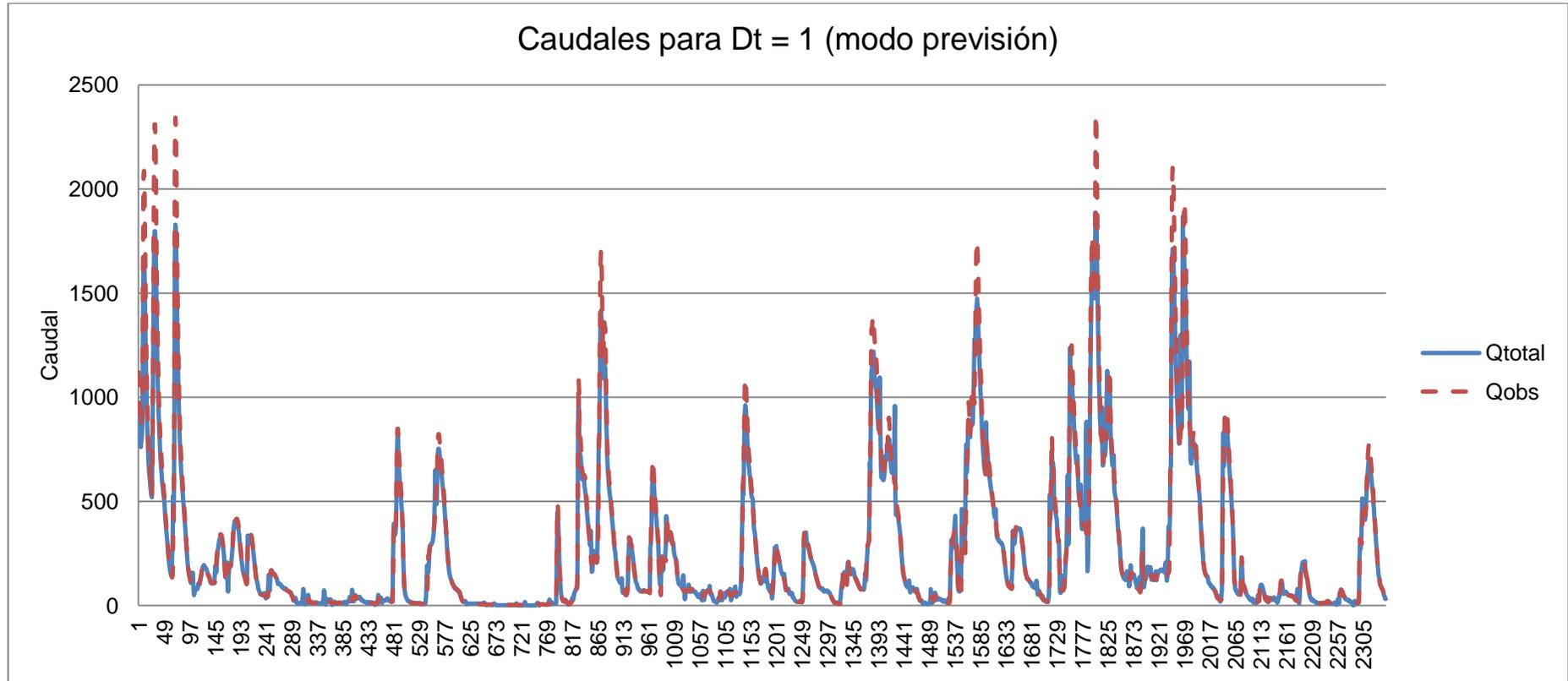


Figura 10.14.: Salida modo previsión – segunda mitad de los datos – horizonte=1 día

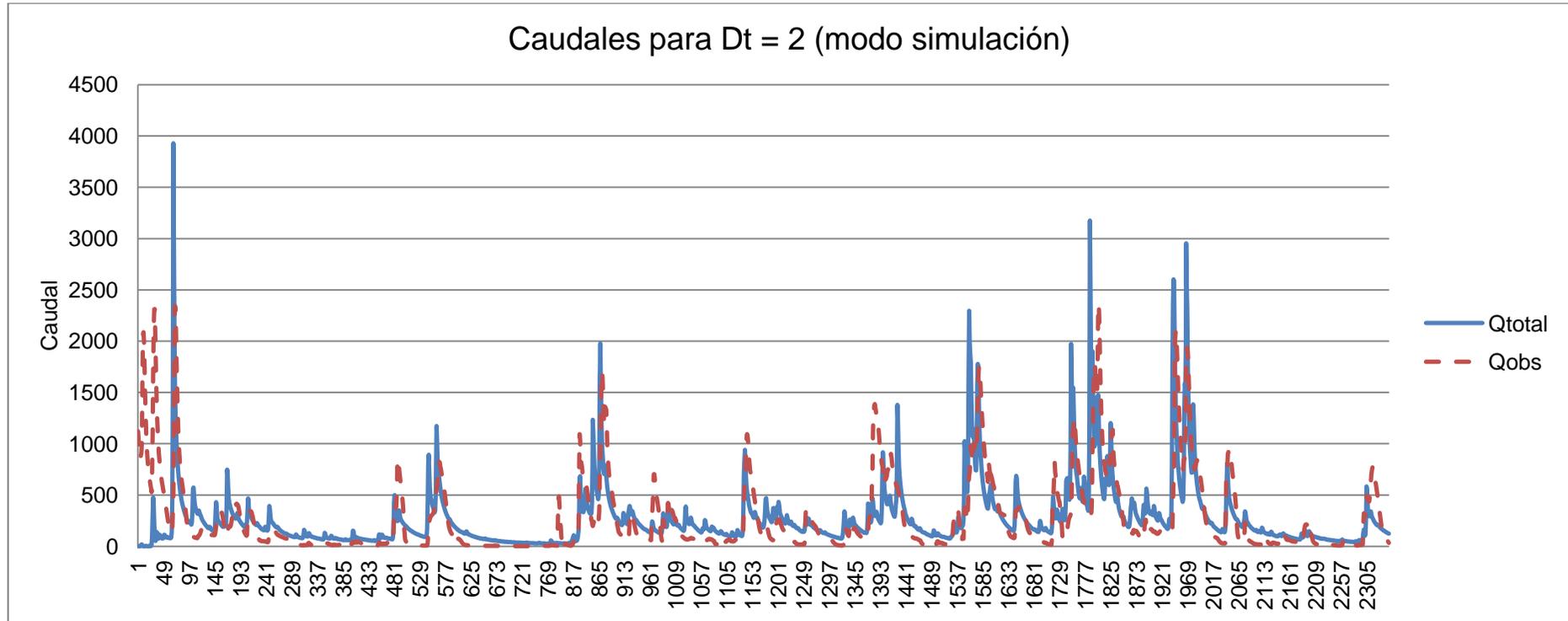


Figura 10.15.: Salida modo simulación – segunda mitad de los datos – horizonte=2 días

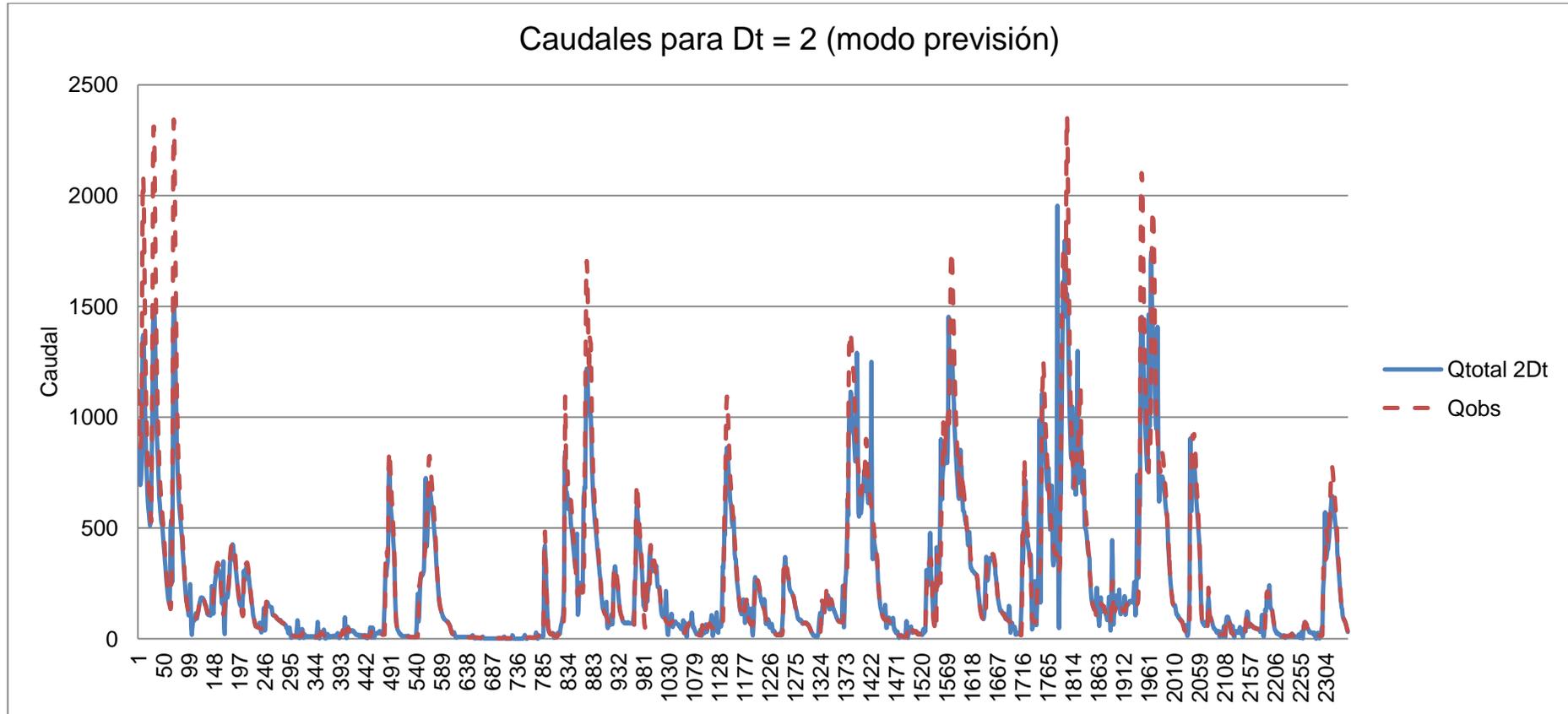


Figura 10.16.: Salida modo previsión – segunda mitad de los datos – horizonte=2 días

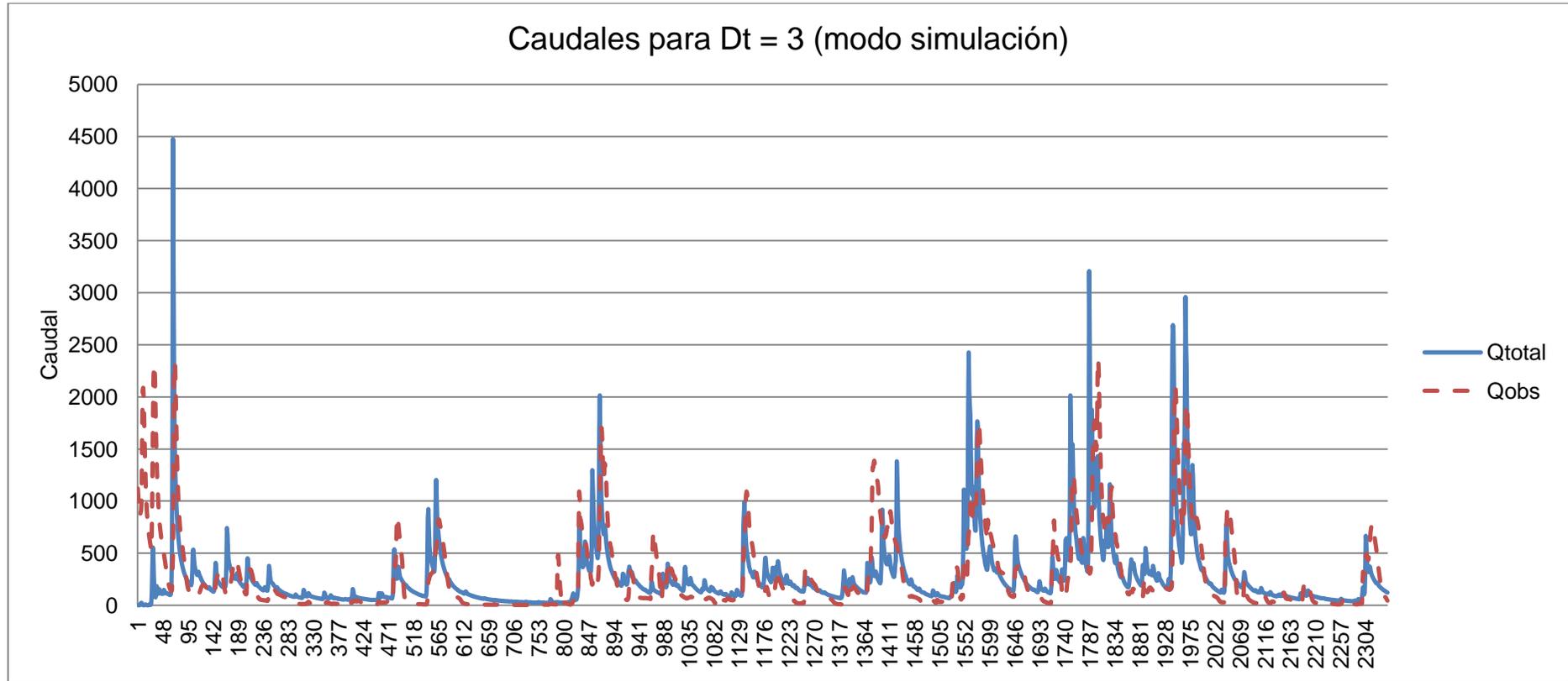


Figura 10.17.: Salida modo simulación – segunda mitad de los datos – horizonte=3 días

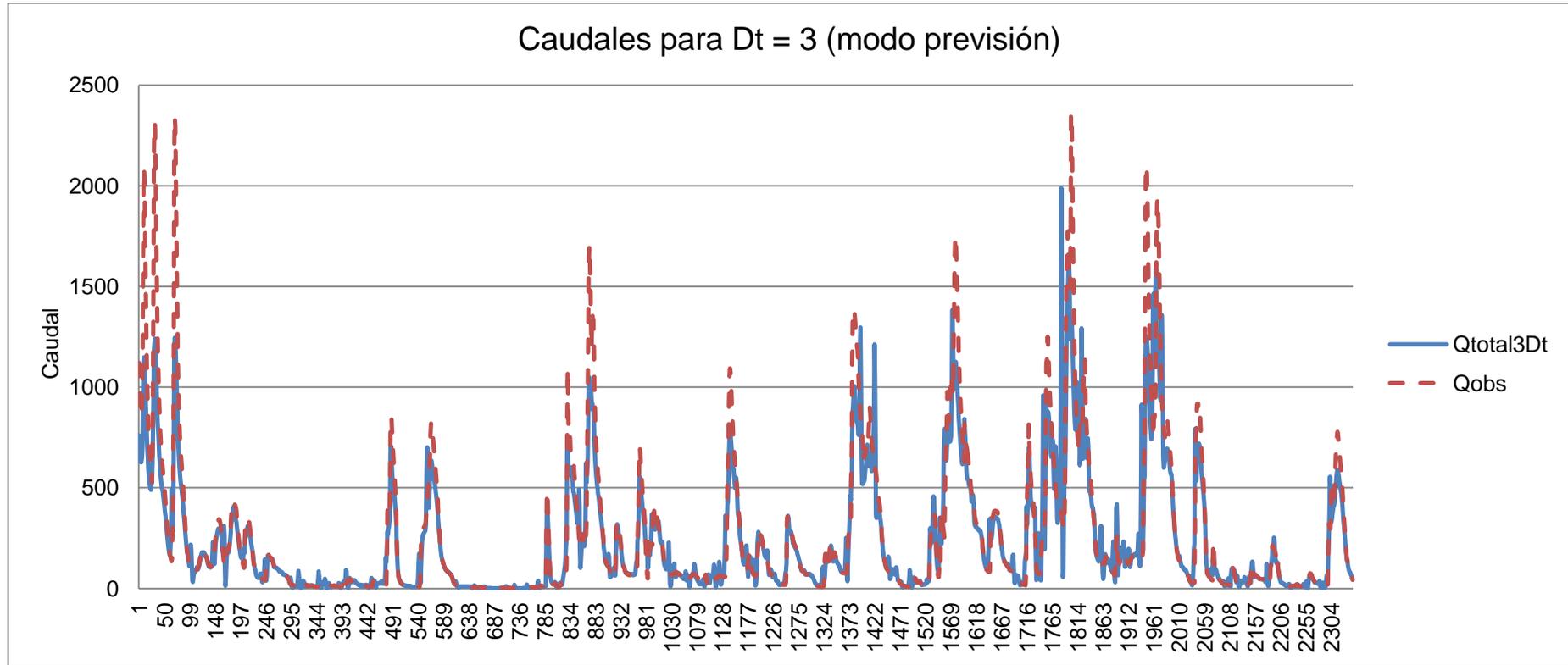


Figura 10.18.: Salida modo predicción – segunda mitad de los datos – horizonte=3 días

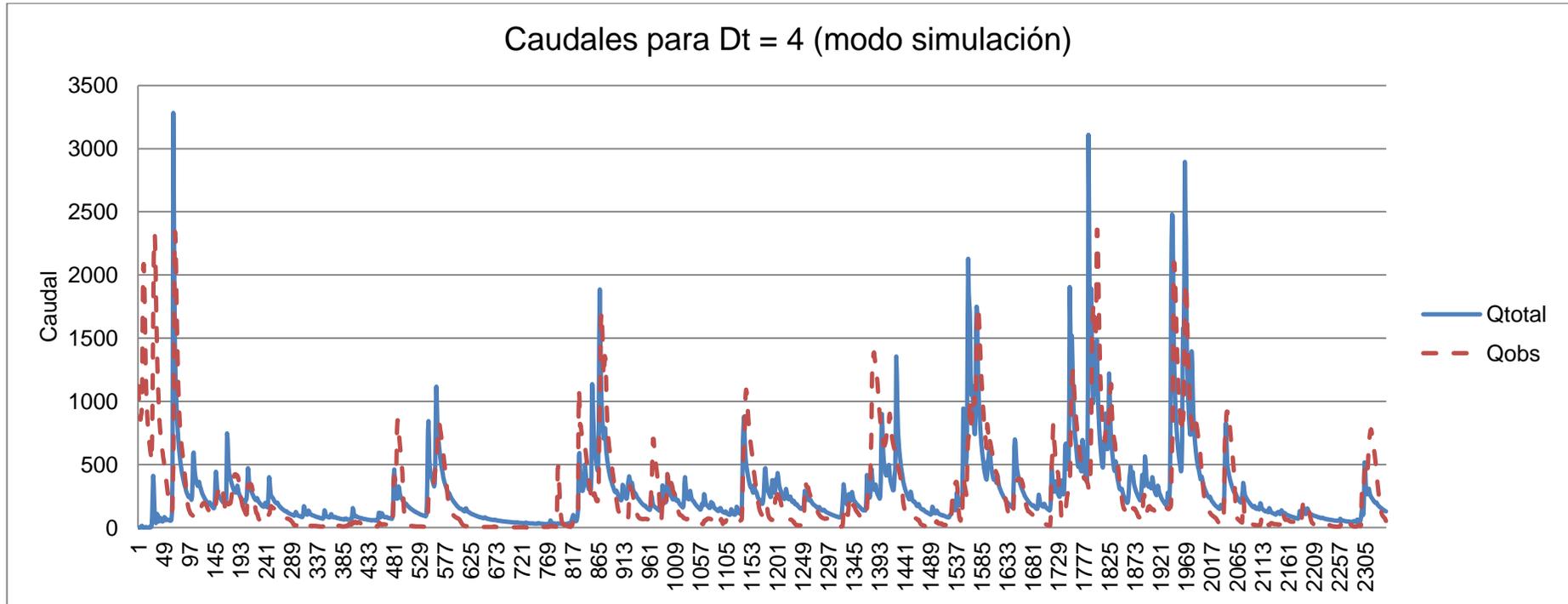


Figura 10.19.: Salida modo simulación – segunda mitad de los datos – horizonte=4 días

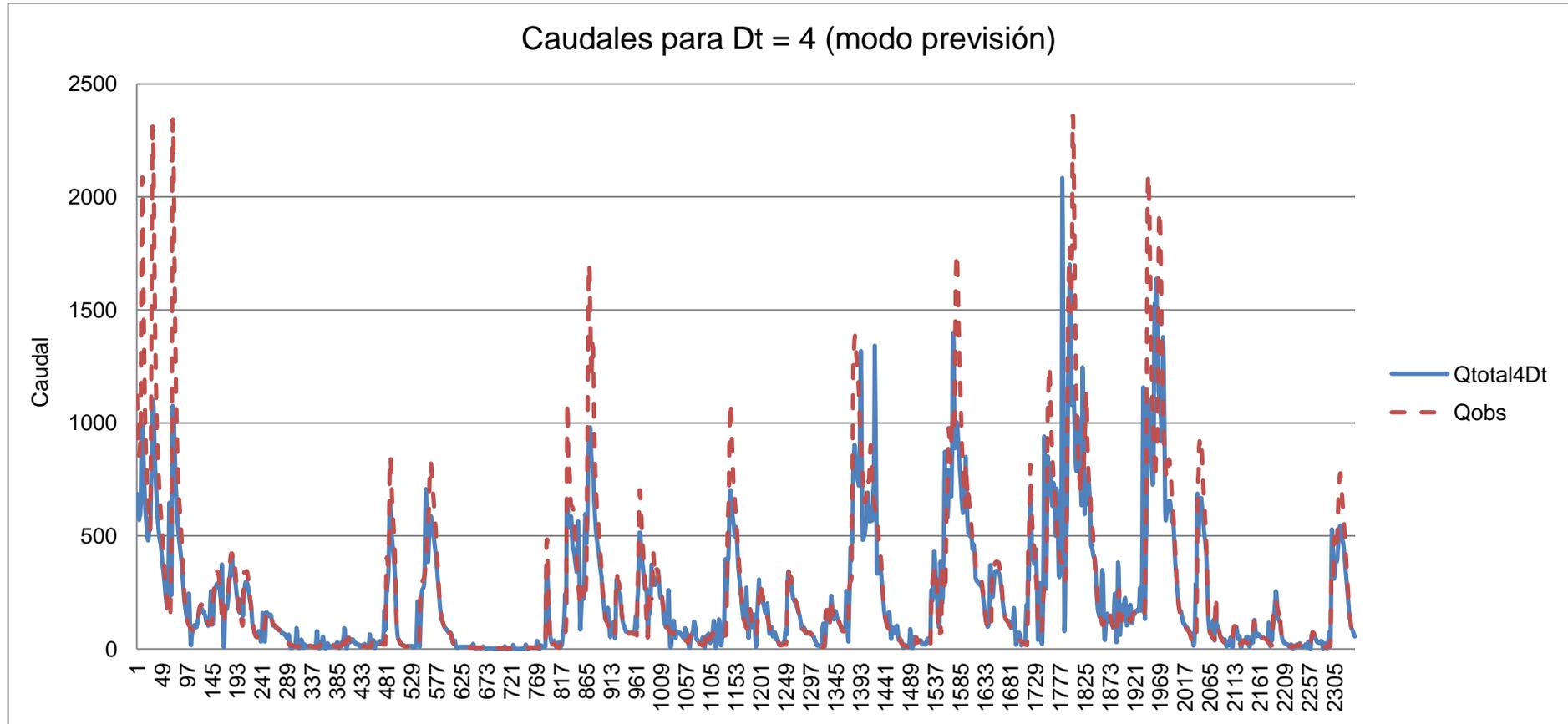


Figura 10.20.: Salida modo previsión – segunda mitad de los datos – horizonte=4 días

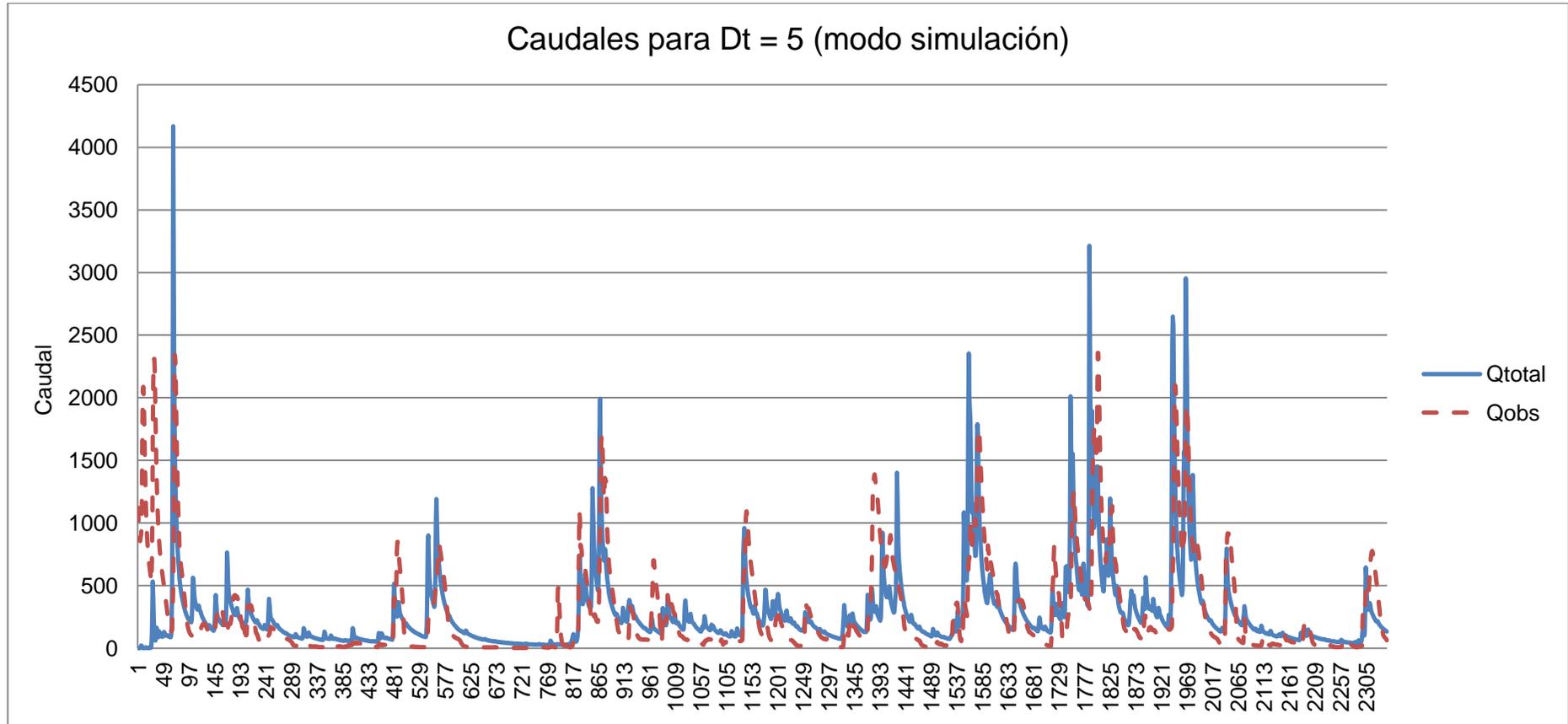


Figura 10.21.: Salida modo simulación – segunda mitad de los datos – horizonte=5 días

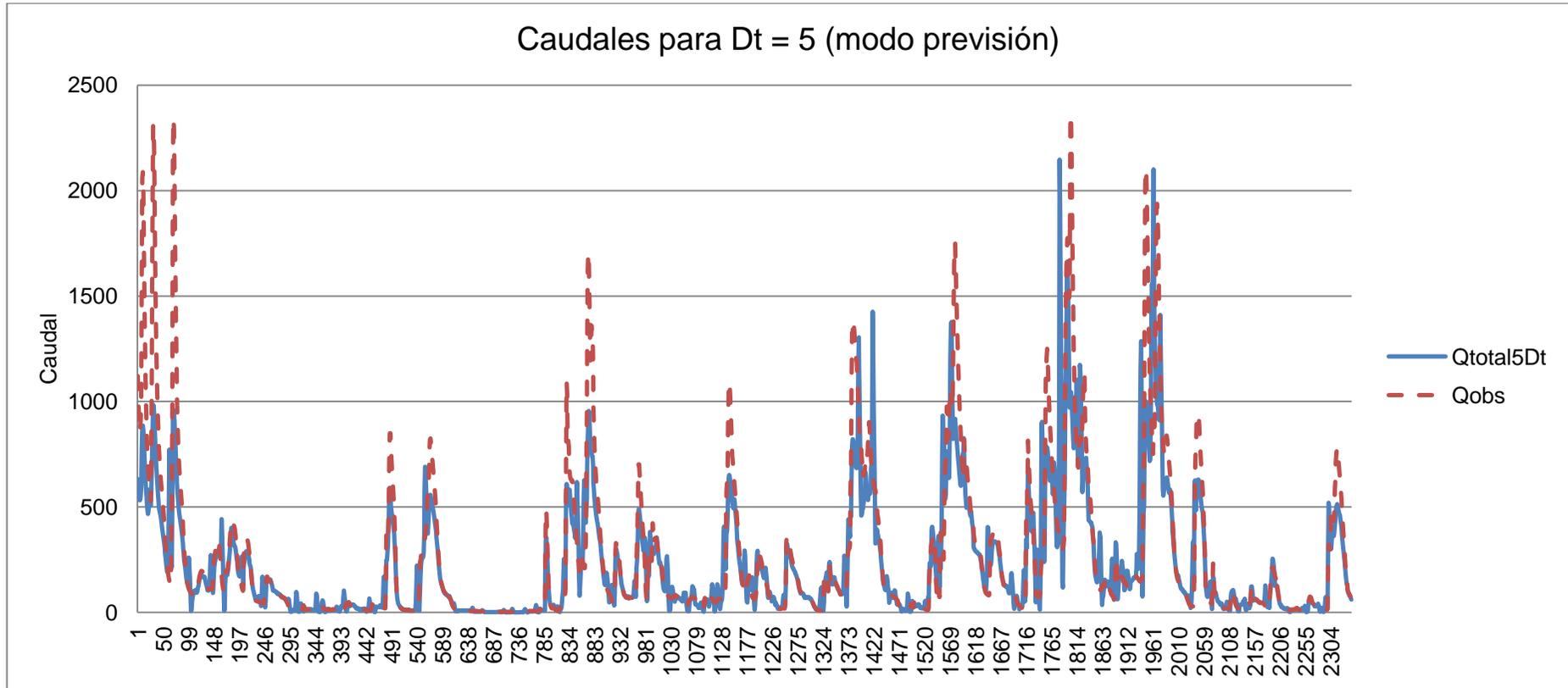


Figura 10.22.: Salida modo previsión – segunda mitad de los datos – horizonte=5 días

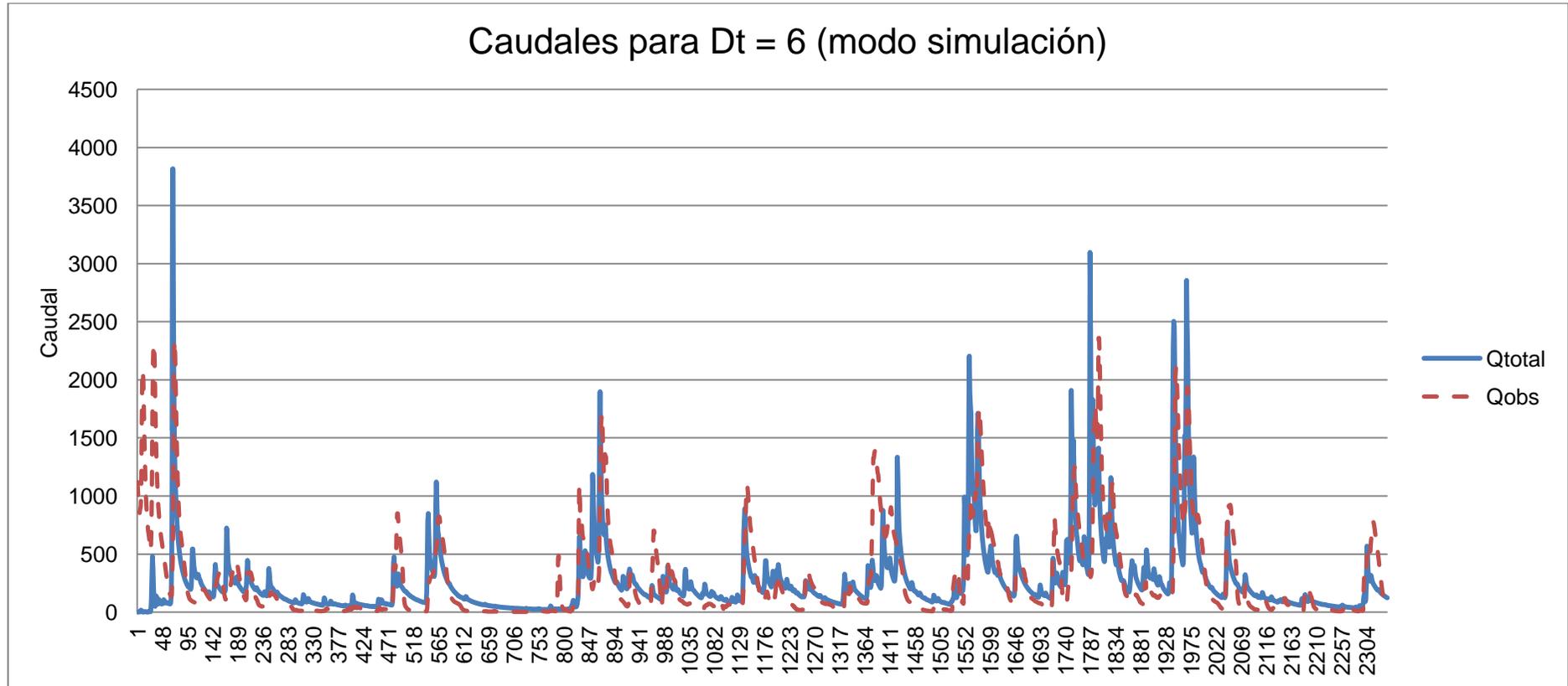


Figura 10.23.: Salida modo simulación – segunda mitad de los datos – horizonte=6 días

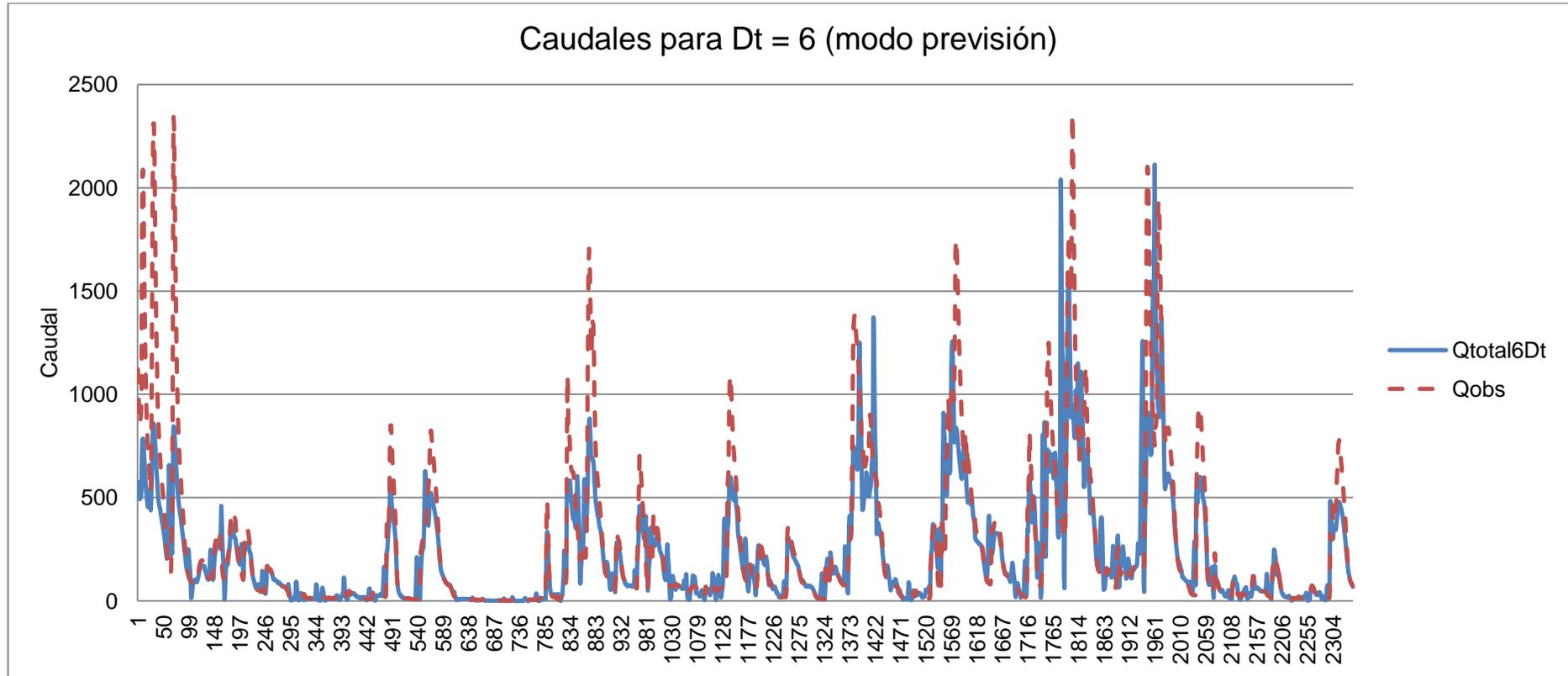


Figura 10.24.: Salida modo previsión – segunda mitad de los datos – horizonte=6 días