



COLEGIO DE POSTGRADUADOS
INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO
HIDROCIENCIAS

**Modelación estocástica de los escurrimientos
de la cuenca del Río Amajac, Hidalgo, México**

PEDRO ALVARADO MEDELLIN

T E S I S

PRESENTADA COMO REQUISITO PARCIAL

PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO

2007

La presente tesis titulada: **Modelación estocástica de los escurrimientos de la cuenca del Río Amajac, Hidalgo, México** realizada por el alumno: PEDRO ALVARADO MEDELLIN, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTOR EN CIENCIAS
HIDROCIENCIAS

CONSEJO PARTICULAR

CONSEJERO

Dr. JUAN ENRIQUE RUBIÑOS PANTA

ASESOR

Dr. JESÚS CHÁVEZ MORALES

ASESOR

Dr. JOSÉ LUIS OROPEZA MOTA

ASESOR

Dr. CARLOS RAMÍREZ AYALA

ASESOR

Dra. ELIZABEHT HERNANDEZ ACOSTA

ASESOR

Dr. ROBERTO QUINTERO LIZAOLA

Dedicatoria

Atziry

Sin tu amor y apoyo no hubiera sido posible alcanzar esta meta

Isaac

y

Atziryta

Los más grandes e importantes proyectos de mi vida

Agradecimientos

A Dios por brindarme salud y entendimiento.

A mis padres, por su ejemplo y amor.

A la Universidad Autónoma de Zacatecas, por su invaluable apoyo.

Al pueblo de México, que a través del CONACYT financia nuestros estudios de postgrado.

A los profesores del Colegio de Postgraduados, en especial a los de Hidrociencias, por compartir sus conocimientos.

A los Drs. Juan Enrique Rubiños Panta y Jesús Chávez Morales, por su dedicación y apoyo incondicional a lo largo de mis estudios doctorales.

A los Drs.:

José Luis Oropeza Mota

Carlos Ramírez Ayala

Roberto Quintero Lizaola

Elizabeth Hernández Acosta

Abel Nolasco Quevedo

Por el apoyo, sugerencias y observaciones en el desarrollo del trabajo de Tesis.

A mis hermanos, familia Alvarado Medellín por su apoyo incondicional.

A todos los amigos del CP, por lo momentos que vivimos juntos, no los olvidaré, no escribo el nombre de alguno en particular para no excluir a nadie.

Esta tesis fue financiada con el apoyo del proyecto Fondos Mixtos Conacyt-Hidalgo 8848, titulado “*Ordenación de Recursos Hídricos del Río Amajac (Subuenca del Panuco) relación socioeconómica y ambientales en el Estado de Hidalgo*”, cuyo responsable es el Dr. Juan Enrique Rubiños Panta, Profesor-Investigador del Colegio de Postgraduados.

HACE ALGUNAS DÉCADAS, el Ing. Thomas King resumió el problema del agua con las siguientes palabras:

“De todas las substancias necesarias para la vida, tal y como las conocemos en la TIERRA, el agua es la más importante, la más conocida y la más maravillosa; y, sin embargo, la mayoría de la gente sabe muy poco acerca de ella.”

En términos más poéticos, Byron trata el mismo tema en su Don Juan: “Hasta que su falta lo hace padecer, el hombre reconoce del agua el valer.”

La historia está llena de leyendas de batallas libradas por el agua, de monarcas y de sacerdotes que la veneraban, de taumaturgos que la han bendecido, de cultura que decayeron después de perder o desperdiciar este elemento, de gente que pereció por su causa.

El agua tiene muchas y muy curiosas facetas. Está universalmente presente y se ha mantenido inmutable, tanto en cantidad como en naturaleza, en el curso de millones de años. Es servidora a la vez que ama y señora del hombre. Sin embargo, la gente sólo se percata de ello cuando le falta o cuando pone en peligro su vida. Entonces su dramático efecto nos llega con fuerza abrumadora... acompañado, con demasiada frecuencia, de falsas impresiones, conclusiones equivocadas y soluciones erradas.

Abel Wolman (1979)

CONTENIDO

I	<i>Introducción</i>	1
II	<i>Justificación</i>	5
III	<i>Objetivos e Hipótesis</i>	7
3.1	Objetivo general	8
3.1.1	Objetivos específicos	8
3.2	Hipótesis	9
IV	Revisión de Literatura	10
4.1	Modelos de series de tiempo	11
4.1.1	Definiciones	11
4.1.2	La variable normal estandarizada	14
4.2	El proceso estocástico y las series de tiempo hidrológicas	15
4.3	Modelos lineales Auto Regresivos univariados	15
4.3.1	Modelos Auto Regresivos AR(p)	16
4.3.2	Estimación de los parámetros Auto Regresivos AR(p)	17
4.3.3	Modelos de Promedios Móviles MA(q)	19
4.3.4	Modelos Auto Regresivo y de Promedio Móviles ARMA(p,q)	20
4.4	Selección de modelos de series de tiempo	21
4.4.1	Etapa 1: Identificación	22
4.4.2	Etapa 2: Estimación de parámetros	26
4.4.3	Etapa 3: Diagnóstico	27
4.4.4	Etapa 4: Pronóstico	29
4.5	Modelos de series de tiempo multivariados	31

4.5.1	Correlograma cruzado	31
4.5.2	Descripción del modelo de series de tiempo multivariado	33
4.5.3	Modelos multivariados generales	34
4.5.4	Propiedades de los modelos multivariados	36
4.6	Modelo multivariado Auto Regresivo anual AR(1)	37
4.6.1	Estimación de los parámetros del modelo	38
4.6.2	Modelación para generación de escurrimientos sintéticos	41
4.6.3	Pruebas de bondad de ajuste de modelos multivariados anuales	42
4.7	Modelos de disgregación	43
4.7.1	Descripción de los modelos de disgregación	44
4.7.2	Modelo de disgregación para un sitio	45
4.7.3	Modelo de disgregación para múltiples sitios	46
4.8	Planeación y manejo de sistemas hídricos con series de tiempo	
4.8.1	Series de tiempo para el pronóstico y generación de escurrimientos	48
4.8.2	Modelos de disgregación	54
4.8.3	Conclusiones de la revisión de literatura	53
4.9	Conclusiones de la revisión de literatura	56
V	<i>Materiales y Métodos</i>	58
5.1	Descripción de la zona de estudio	60
5.1.1	Cuenca del Río Amajac	60
5.2	Clima	62
5.3	Hidrometría de la cuenca del Río Amajac	63
5.4	Modelos en series de tiempo para la cuenca del Río Amajac	68
5.4.1	Modelo univariado para la estación hidrométrica Venados	68
5.4.2	Modelo AR(1) multivariado para cuatro estaciones	

	hidrométricas en la cuenca del Río Amajac	71
	5.4.3 Modelo de disgregación temporal para un sitio	74
	5.4.4 Modelo de disgregación temporal para sitios múltiples	76
	5.4.5 Pruebas de bondad de ajuste para series sintéticas generadas con modelos de disgregación	78
VI	Resultados y discusión	79
6.1	Modelación de series de tiempo univariadas	81
	6.1.1 Identificación del modelo	83
	6.1.2 Estimación de los parámetros	84
	6.1.3 Diagnóstico	85
	6.1.4 Pronóstico	88
6.2	Modelación de series de tiempo multivariadas	90
	6.2.1 Modelo de series de tiempo multivariado temporal AR(1)	91
	6.2.2 Análisis de residuos	94
6.3	Modelos de disgregación para un sitio	97
6.4	Modelo de disgregación para sitios múltiples	101
6.5	Pruebas de bondad de ajuste para modelos de disgregación	107
VII	Conclusiones y recomendaciones	113
7.1	Recomendaciones	115
	Bibliografía	116
	ANEXO I Valores históricos y sintéticos para la modelación de la cuenca del Río Amajac	121
	ANEXO II Descripción del uso de las hojas de cálculo para la generación de series sintéticas	133

A1.1	Modelos Box-Jenkins para la estación Venados	134
A1.2	Modelo de disgregación	139

ÍNDICE DE CUADROS

<i>Cuadro</i>	<i>Descripción</i>	<i>Pág.</i>
5.1	Volúmenes medios anuales registrados en la estación hidrométrica “Presa la Esperanza.	64
5.2	Volúmenes medios anuales registrados en la estación hidrométrica “Venados”.	65
5.3	Volúmenes medios anuales registrados en la estación hidrométrica “San Agustín”.	66
5.4	Volúmenes medios anuales registrados en la estación hidrométrica “Temamatla”.	67
5.5	Esquema de obtención de modelos AR(1) elaborados por mes y año para las estaciones Venados y Temamatla en los periodos de 1964 a 2004.	75
5.6	Esquema de obtención de modelos anuales AR(1) para las estaciones hidrométricas de la cuenca del Río Amajac.	77
6.1	Escurrimientos medios anuales históricos y estandarizados registrados de 1937 a 2004 en la estación hidrométrica Venados.	82
6.2	Valores de la función de auto correlación de la series Venados.	83
6.3	Valores de la función de auto correlación parcial de la serie Venados.	84
6.4	Estimación del parámetro $\hat{\phi}_1$ del modelo ARMA(1,0) con un valor inicial de 0.75 con los valores históricos estandarizados.	85
6.5	Parámetros estadísticos estimados para los modelos ARMA(1,0), ARMA(0,1) y ARMA(1,1).	86
6.6	Modelos ARMA(p,q) para la estación hidrométrica Venados.	86
6.7	Comparación de las propiedades estadísticas de la serie histórica y la generada sintéticamente con el modelo ARMA(1,0).	89

6.8	Datos generados sintéticamente con el modelo AR(1) para la estación Venados.	89
6.9	Series sintéticas anuales generadas con el modelo AR(1) multivariado para las estaciones de la cuenca del Río Amajac.	94
6.10	Comparación de los parámetros estadísticos de las series históricas con las series generadas con el modelo AR(1) multivariado.	95
6.11	Valores medios históricos y sintéticos obtenidos con el modelo de disgregación para la estación Venados.	99
6.12	Valores medios históricos y sintéticos obtenidos con el modelo de disgregación para la estación Temamatla.	101
6.13	Valores medios históricos y sintéticos obtenidos con el modelo de disgregación multivariado AR(1) para la estación Temamatla.	102
6.14	Valores medios históricos y sintéticos obtenidos con el modelo de disgregación multivariado AR(1) para la estación Venados.	104
6.15	Valores medios históricos y sintéticos obtenidos con el modelo de disgregación multivariado AR(1) para la estación San Agustín.	105
6.16	Valores medios históricos y sintéticos obtenidos con el modelo de disgregación multivariado AR(1) para la estación Presa La Esperanza.	106
6.17	Estadístico KS para bondad de ajuste entre las series históricas y sintéticas generadas con modelos de disgregación en las estaciones hidrométricas del Río Amajac.	108
A1.1	Volúmenes medios históricos registrados en la estación Temamatla.	122
A1.2	Volúmenes medios históricos registrados en la estación Venados.	124
A1.3	Volúmenes medios históricos registrados en la estación San Agustín.	127
A1.4	Volúmenes medios históricos registrados en la estación Presa La Esperanza.	129

A1.5	Valores históricos y sintéticos usados para la obtención del estadístico de Kolmogorov-Smirnov, estaciones Temamatla y Venados.	131
A1.6	Valores históricos y sintéticos usados para la obtención del estadístico de Kolmogorov-Smirnov, estaciones San Agustín y Presa La Esperanza.	132

ÍNDICE DE FIGURAS

Figura	Descripción	Pág.
4.1	Etapas para la selección de un modelo Box-Jenkins.	22
4.2	Representación esquemática de la estructura de correlación de un modelo multivariado AR(1).	35
5.1	Diagrama de flujo para la generación de series de tiempo sintéticas.	59
5.2	Ubicación de la cuenca del Río Amajac respecto al estado de Hidalgo y el país.	61
5.3	Estaciones hidrométricas en la cuenca del Río Amajac.	63
6.1	Hidrograma históricos en la estación Venados.	81
6.2	Función de autocorrelación estación Venados.	83
6.3	Función de auto correlación parcial de la estación Venados.	84
6.4	Series histórica y estimada con el modelo ARMA(1,0).	87
6.5	Serie sintética de 68 años de escurrimientos generada con el modelo AR(1) para la estación Venados,	89
6.6	Datos históricos y estimados, estación Temamatla.	96
6.7	Comparativo en la estación Venados.	96
6.8	Comparativo en la estación San Agustín o Río Mezquitlán.	97
6.9	Comparativo en la estación hidrométrica de la Presa La Esperanza.	97
6.10	Correlación entre los promedios mensuales de los datos observados y sintéticos para la estación Venados.	100
6.11	Prueba de correlación entre las medias de los datos observados y sintéticos para la estación Temamatla.	100
6.12	Prueba de correlación entre las medias de los datos observados y sintéticos mensuales y anuales para la estación Temamatla.	103
6.13	Función de Distribución Acumulada (FDA) Empírica para las	

	series de Enero, estación Temamatla.	109
6.14	FDA Empírica para las series de Septiembre, estación Temamatla.	109
6.15	FDA Empírica para las series de Marzo, estación Venados.	110
6.16	FDA Empírica para las series de Septiembre, estación Venados.	110
6.17	FDA Empírica para las series de Enero, estación San Agustín.	111
6.18	FDA Empírica para las series de Septiembre, estación San Agustín.	111
6.19	FDA Empírica para las series de Enero, estación P. Esperanza.	112
6.20	FDA Empírica para las series de Septiembre, estación Presa La Esperanza	112
A2.1	Esquema de obtención del parámetro ϕ_1	137
A2.2	Invocación de la rutina del solver.	137
A2.3	Parámetros de solver.	138

RESUMEN

En México, la Comisión Nacional Del Agua (CNA), es el organismo responsable de registrar y publicar los resultados de la medición de variables que se obtienen de la red de estaciones meteorológicas e hidrométricas. En los últimos años, las redes se han reducido. En algunos ríos, la medición de los escurrimientos, la variable hidrológica más importante, se ha suspendido. En estos casos, la falta de datos históricos provoca que no se pueda realizar una planeación adecuada de los recursos hídricos a nivel local y regional.

Por lo anterior, la modelación matemática, en particular con modelos de series de tiempo, se puede desarrollar para generar series sintéticas con la capacidad de reproducir las propiedades estadísticas de los registros históricos para complementarlos y extenderlos.

Se presenta una metodología basada en hojas de cálculo para generar series sintéticas a través de modelos ARMA(p,q) univariados y multivariados y modelos de disgregación. Los modelos se desarrollaron, aplicaron y validaron a los registros históricos de 4 estaciones (Temamatla, Vendos, San Agustín y Presa La Esperanza) localizadas en cuenca del Río Amajac en el Estado de Hidalgo, México. Se encontró que los modelos anuales AR(1) univariado y multivariado tienen la capacidad de reproducir las propiedades estadísticas de las series históricas. Los modelos de disgregación se desarrollaron para dividir las series sintéticas anuales y obtener series sintéticas mensuales con las características estadísticas de las series históricas en ambos niveles, anual y mensual. Las pruebas de bondad de ajuste que se aplicaron a las series generadas muestran un buen ajuste en comparación con los datos históricos de las cuatro estaciones hidrométricas.

Palabras clave: Modelos AR, modelos de disgregación, procesos estocásticos, series de tiempo hidrológicas, Río Amajac.

ABSTRACT

In México, the Government National Water Commission (CNA, as its initials in Spanish), is the responsible of the meteorological and hydrologic variables measurement, recording and publishing, which were continuously obtained in former established nets of meteorological and hydrometric stations. In the last years those nets have been reduced. In some rivers the measurement of water stream flow, one of the most important hydrologic variables, has been suspended. In those cases, lack of updated historical series of data recordings can profoundly affect the water resources planning at local and regional level.

Therefore, mathematical modeling, specifically time series models, can be developed to generate synthetic data series, which resemble the statistical characteristics of the recorded historical data, to properly complement and extend the historical ones.

A methodology, spreadsheet based, to generate synthetic stream flows with unvaried and multivariate ARMA(p,q) and disaggregation time series models, is presented. The models were developed, validated and applied to water flow historical data series recorded in four gaging stations (Venados, Temamatla, San Agustín y Pres La Esperanza), located in the Amajac River Basin, in the State of Hidalgo, Mexico. It was found that univariate (Venados gaging station) and multivariate (the four gaging stations) AR(1) models were able to resemble well the Annual recorded stream flows data. Disaggregation Models, based in the AR(1) models, were developed to generate monthly synthetic stream flows from the annual generated ones. A statistical test of goodness of fit of synthetically generated monthly data versus the observed ones was found to be good enough for the four gaging stations.

Key words: Amajac River, ARMA models, disaggregation models, hydrologic time series, stochastic processes.

Capítulo I

Introducción

*“La ciencia más útil es aquella cuyo
fruto es el más comunicable”*

Leonardo Da Vinci

¿Por qué se debe poner atención una y otra vez al manejo de los recursos hídricos a nivel de cuenca? Cuando el agua es el tema del día en la escena internacional, dada la situación crítica en varios países, se debe recordar que existen más de 300 grandes ríos y que las cuencas en donde drenan cubren más de la mitad de la tierra que emerge sobre el planeta. Más de 200 ríos son internacionales, es decir, estos cruzan los límites entre países, esta misma situación aplica dentro de los países debido a las divisiones políticas e institucionales. Por lo tanto, la cuenca, es tema de nuestro interés como un sistema que está compuesto por recursos naturales y que las comunidades humanas dependen de éstos, Burton (2003).

El agua es un recurso natural que se requiere para muchos usos en la sociedad en nuestros días. De la misma manera en que la demanda del agua crece, decrece la disponibilidad per cápita. En la actualidad existe el interés en el desarrollo de metodologías y técnicas que ayuden a mejorar el proceso de toma de decisión en problemas relacionados con el manejo del agua a corto, medio y largo plazo, lo que se denomina ordenación de los recursos hídricos. Más aún, la mayoría de las naciones tienen la necesidad de producir substancialmente la cantidad de productos agrícolas para cubrir los requerimientos de su propia población, además realizar tratados con otros países, haciendo imperativo la utilización óptima de los recursos de agua y tierra, Chávez (1985).

El territorio mexicano, igual que en el resto del mundo, está dividido naturalmente en cuencas. Las cuencas más importantes del país corresponden a grandes Ríos como el Lerma, el Santiago, el Balsas, el Bravo, el Panuco, el Papaloapan, el Coatzacoalcos, el Grijalva, el Usumacinta, el Mayo, el Yaqui y otros.

La cuenca del Río Panuco en extensión, es la cuarta más grande del país, y si se clasifica por el volumen medio anual escurrido es la quinta, por lo que se considera una cuenca importante. Pertenece a la Región Hidrológica 26, Río Panuco, que

frecuentemente es afectada por desastres naturales relacionados con ciclones generados en el Golfo de México y del Mar Caribe, Programa Regional Hidráulico Comisión Nacional del Agua, (2003). La cuenca del Río Amajac se ubica en la cuenca alta del Río Panuco.

En el presente trabajo se desarrolla una metodología basada en modelos de series de tiempo como una herramienta para auxiliar en la toma de decisiones en el manejo de los recursos hídricos en cuencas hidrológicas a nivel local, y regional. Se expone una metodología para formular modelos estocásticos que permitan seguir paso a paso la construcción de modelos hidrológicos en los que se involucran variables aleatorias que simulan el comportamiento de los fenómenos naturales. En el caso de este estudio, las variables que se analizan son los volúmenes medios anuales y mensuales escurridos en la cuenca del Río Amajac en el Estado de Hidalgo, afluente del Río Panuco.

En el capítulo 4 se presenta la teoría básica en la que se sustenta el análisis de series de tiempo para la generación de series sintéticas con modelos Auto Regresivos y de Promedios Móviles, ARMA(p,q) por sus siglas en inglés, y modelos de disgregación para modelos univariados y multivariados. Y se presenta también, la revisión sobre estudios que se han realizado utilizando técnicas y modelos de análisis de series de tiempo.

En el capítulo 5 se detalla la metodología para generar series de tiempo sintéticas, para los casos de modelos univariados y multivariados. Es importante mencionar que los modelos de series de tiempo en este estudio se han aplicado al caso de volúmenes de escurrimiento, pero se pueden usar en general en fenómenos climatológicos e hidrológicos que se comportan en forma aleatoria en el tiempo y en el espacio, como la lluvia, la temperatura, la evaporación, la contaminación en ríos o cuerpos de agua, entre otros.

En el capítulo 6 en primer lugar, se presentan los resultados de la aplicación de los modelos de series de tiempo a una estación hidrométrica (caso univariado). Se describe la metodología para identificar, estimar, y definir un modelo de series de tiempo tipo ARMA a los datos de escurrimientos históricos de la estación hidrométrica “Venados”. Se realiza un análisis estadístico de los residuos que se obtienen de la comparación de los registros históricos con los generados por el modelo.

Posteriormente, se describe la construcción y aplicación de un modelo multivariado para series de tiempo considerando los datos históricos de escurrimiento en cuatro estaciones hidrométricas (caso multivariado). El modelo es del tipo Auto Regresivo AR(1). Las estaciones hidrométricas que se usaron para construir el modelo multivariado son: Temamatla, Venados, San Agustín o Río Mezquititlán, y Presa La Esperanza, pertenecientes todas a la cuenca del Río Amajac en el Estado de Hidalgo.

Para los dos modelos se analizan los parámetros estadísticos representativos de las funciones de probabilidad que describen los volúmenes escurridos. Estos parámetros son la media \bar{z} , desviación estándar s_z y el coeficiente de auto correlación r_k . Para ambos casos se describe como se construyeron los modelos para generar series sintéticas que preserven las propiedades estadísticas de las series originales.

Se presenta la metodología de elaboración de modelos de *disgregación* para generar series de datos sintéticos mensuales a partir de datos anuales que fueron generados previamente. La metodología de generación de datos a escalas menores con modelos de disgregación ha demostrado que la generación y simulación de series hidrológicas sintéticas reproducen satisfactoriamente las propiedades estadísticas de las series originales, lo que permite tener la posibilidad de analizar el manejo de los recursos ante diversos escenarios de ocurrencia.

Capítulo II

Justificación

*“La imaginación es más importante
que el saber”*

Albert Einstein

En la cuenca del Río Amajac, en algunas ocasiones existe abundancia de escurrimientos provocados principalmente por fenómenos naturales provenientes del Golfo de México y el mar Caribe. El 75% de las lluvias se concentran en los meses de Junio a Octubre. A lo anterior se une la orografía que causa que la cuenca sea de rápida respuesta. Por lo descrito anteriormente, es necesario contar con herramientas de pronóstico y/o simulación que permitan elaborar planes de manejo de los recursos hídricos en la cuenca con la finalidad de disminuir los efectos causados por fenómenos extremos y llevar a cabo un ordenamiento adecuado de los recursos hídricos en conjunto con otros recursos naturales.

De la cuenca del Río Amajac, en promedio salen $1,600 \cdot 10^6 \text{m}^3$ lo que indica que el aprovechamiento del recursos se puede mejorar bajo un ordenamiento del mismo, sin embargo, esto requiere la utilización de herramientas metodológicas que permitan analizar el comportamiento de los volúmenes escurridos.

Se han desarrollado diferentes técnicas para analizar los escurrimientos como una variable que aparece aleatoriamente en el espacio y en el tiempo. Uno de los métodos que ha tenido aceptación en la comunidad científica en las últimas décadas es la modelación de series de tiempo. Esto se debe principalmente, a que estos modelos permiten hacer un análisis de las variables hidrológicas que se presentan aleatoriamente en el espacio y en el tiempo; esta variabilidad espacial y temporal la pueden reflejar este tipo de modelos estocásticos.

Capítulo III

Objetivos e Hipótesis

*"La fuerza no proviene de la
capacidad física sino de la voluntad
indomable."*

Mahatma Gandhi

3.1 Objetivo general

Desarrollar una metodología para la generación de series de escurrimientos sintéticos que auxilie en la toma de decisiones a los responsables del manejo de los recursos hídricos de las cuencas hidrológicas tanto en el ámbito local como regional.

3.1.1 Objetivo específicos

1. Desarrollar modelos estocásticos basados en la teoría de series de tiempo que ayude en el pronóstico de los escurrimientos de la cuenca del Río Amajac a corto, mediano y largo plazo con una alta confiabilidad.
2. Desarrollar la metodología para construir modelos en series de tiempo que consideren una y varias estaciones hidrométricas.
3. Desarrollar la metodología para la generación de datos sintéticos mensuales a partir de datos anuales a través del uso de modelos de disgregación.

3.2 Hipótesis

El comportamiento de las variables hidrológicas, como el escurrimiento superficial se puede reproducir sintéticamente aplicando las teorías de análisis y modelación de series de tiempo.

La identificación y selección de los modelos de series de tiempo se mejora con el uso de técnicas de estimación de errores y la definición de sus parámetros con técnicas de optimización.

A través del uso de modelos de disgregación, es posible generar series de tiempo sintéticas que conserven las propiedades estadísticas de las series originales a escalas de tiempo menor.

Capítulo IV

Revisión de Literatura

“La ciencia es como la tierra, sólo se puede poseer un poco de ella”

Francios Voltaire-Marie Arouet

4.1 Modelos de series de tiempo

El desarrollo de los modelos matemáticos de este capítulo se refiere principalmente, a las técnicas para la generación de escurrimientos y otros procesos estocásticos que se utilizan en estudios de simulación de sistemas de recursos hidráulicos. Aunque, las técnicas tratadas aquí, salvo algunas modificaciones, se pueden utilizar para modelar otros fenómenos hidrológicos o meteorológicos, la discusión se centrará directamente sobre el uso de los modelos para generar series de escurrimientos.

4.1.1 Definiciones

Los escurrimientos generados se llamarán *sintéticos* para distinguirlos de las observaciones históricas. El campo de estudio en el que se estudian los procesos hidrológicos a través de las series de tiempo ha sido llamado *hidrología operacional* o más recientemente *modelación estocástica hidrológica* o simplemente *hidrología estocástica*.

La causa principal del desarrollo de la simulación con series de tiempo se debe a que en los estudios de simulación de cuencas, embalses, etc., se usan en conjunto muchas series de tiempo como escurrimientos, lluvia, evaporación y/o temperatura para evaluar las propiedades estadísticas de diferentes alternativas de funcionamiento de sistemas hidráulicos. Con este propósito, los escurrimientos y otros valores sintéticos deben tener la capacidad de representar y preservar las propiedades estadísticas de los registros históricos.

Serie de tiempo. Conjunto de observaciones en el que se mide la variación en el tiempo de algún fenómeno, como el escurrimiento en un río, el nivel de un embalse, entre otros. Los valores tomados en el tiempo se llaman *estados*. El

tiempo y el estado pueden ser variables continuas o discretas. En general, las series de tiempo hidrológicas pertenecen a la categoría de *estados continuos y tiempo discreto* en el que las observaciones son hechas a un fenómeno que cambia continuamente a intervalos de tiempo generalmente iguales.

Variable aleatoria estocástica. Los eventos cuantificados o registrados cuando su *resultado u ocurrencia* es incierta. Evento es el fenómeno en estudio (por ejemplo los 12 escurrimientos mensuales de un río en un año), es el resultado de valores reales registrados. Proceso es un conjunto de resultados. Proceso estocástico, si los resultados son inciertos.

Lag o retraso. Es la relación que existe en una variable y su valor registrado k periodos de tiempo antes. Dicho de otras manera la relación que existe entre z_t y z_{t-k} . En este sentido la palabra lag o retraso será usada indistintamente en el presente trabajo.

En general, una serie de tiempo se puede dividir en una componente determinística la cual se usa con propósitos predictivos y una componente estocástica que consiste en el efecto de la probabilidad.

La componente estocástica está constituida por oscilaciones irregulares y efectos aleatorios y para describirlo se requieren conceptos de probabilidad. Una característica importante que poseen las series de tiempo discretas es la *estacionalidad*. Para definir matemáticamente esta propiedad se denotan las funciones de probabilidad conjunta de funciones tomadas en los tiempo $t = 1, 2, \dots, N$ y nuevamente en los tiempos $t+1, t+2, \dots, t+N$, como $f(z_1, z_2, \dots, z_N)$ y $f(z_{t+1}, z_{t+2}, \dots, z_{t+N})$, respectivamente. Entonces, para un proceso estocástico estacionario, los dos conjuntos de funciones de densidad de probabilidad de N dimensiones es idéntico para todos los valores de N y t .

En general existen dos clases de modelo en series de tiempo, los paramétricos, que son el tipo de modelos que se usan en esta investigación, llamados así debido a que se requiere el cálculo de parámetros para conservar la estructura de varianza y covarianza en el espacio y en el tiempo. Los modelos no paramétricos no requieren o requieren un mínimo de parámetros para la representación de las propiedades estadísticas de las series históricas; para esto, se generan funciones de distribuciones empíricas, a través de bloques de remuestreo (bootstrap) de los datos originales.

Es importante resaltar los conceptos y las aplicaciones en que los modelos estocásticos se basan. Estas se describen a continuación:

- (i) En primera instancia, si la secuencia de los registros no representa adecuadamente una serie, no se puede desarrollar un análisis confiable. Se asume también que una muestra representa una población con suficiencia, esto también implica que la condición de estacionalidad es invocada; de otra manera, se pueden utilizar modelos no estacionarios.
- (ii) En segundo lugar, debe ser posible identificar y rectificar errores de medida y de registro, si estos existieran, en un primer análisis de los datos. También, cualquier efecto forzado en las series de tiempo hidrológicas causado por los diferentes usos del río o por la construcción de una represa se debe tomar en cuenta, ya que las series de tiempo representan los procesos naturales tal como ocurren.
- (iii) Por último, el modelo debe ser manejable, este deberá mantener las propiedades estadísticas e hidrológicas del proceso, así que los diferentes tipos de eventos y secuencias que ocurren de igual manera se generan de forma realista.

En la práctica, estas condiciones no se pueden alcanzar completamente, y con frecuencia hay ciertas suposiciones. Con tales consideraciones en mente un

modelo que sea lo suficientemente flexible y al mismo tiempo parsimonioso con respecto a sus parámetros debe satisfacer los propósitos generales para lo que se construye. El término parsimonia se refiere al hecho de que si para un mismo fenómeno existen dos modelos que lo pueden representar, se deberá seleccionar el más sencillo, es decir, el que requiera menos parámetros para su ajuste.

4.1.2 La variable normal estandarizada

Una de las funciones de distribución de probabilidad (PDF) que se utiliza en hidrología es la Normal, definida como una función de probabilidades de una variable aleatoria continua z :

$$f(z) = \frac{1}{\sigma_z \sqrt{2\pi}} e^{-(z-\mu_z)^2 / 2\sigma^2} \quad (4.1)$$

En la que μ_z y σ_z^2 son la media y la varianza de z respectivamente, comúnmente se denota como $N(\mu, \sigma)$. Los parámetros se estiman con la muestra disponible. Si se hace un cambio de variable: $t = (z-\mu)/\sigma$, la ecuación (4.1) queda como sigue:

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad (4.2)$$

Que se conoce como normal estándar, donde (4.2) representa una función sin parámetros, es decir, $\mu = 0$, y $\sigma^2 = 1$, y que se denota como $N(0,1)$. Esta función se utiliza comúnmente en hidrología, particularmente en modelos de series de tiempo como función de densidad de probabilidades para el análisis de errores aleatorios. Generalmente, para el caso de volúmenes medios anuales, la función

de densidad de probabilidad con transformación logarítmica log-normal, se aplica para estandarizar esta variable, Yevjevich (1972).

4.2 El proceso estocástico y las series de tiempo hidrológicas

Muchas variables aleatorias importantes en los recursos hidráulicos están en función de valores que cambian con el tiempo. Registros históricos de lluvias o escurrimientos en un lugar en particular son una secuencia de datos observados llamados *series de tiempo*. En las series de tiempo los valores se ordenan en función del tiempo, y generalmente es el caso de que el valor observado de la variable aleatoria en un tiempo dado influye en la distribución de la variable aleatoria en tiempos posteriores; en tal sentido, las observaciones no son independientes.

Las propiedades de un proceso estocástico se deben determinar por una serie de tiempo simple o realización. Para hacer esto, primeramente se hacen ciertas suposiciones. Primero, se asume que el proceso es estacionario, esto significa que la función de distribución de probabilidad del fenómeno no cambia a través del tiempo. Dicho de otra manera: si $X(t)$ es un proceso estocástico estacionario, entonces:

$$F_{X(t)}[X(t)] = F_X[X(t)] \quad (4.3)$$

4.3 Modelos lineales Auto Regresivos univariados

En un modelo auto regresivo lineal, el valor real de una variable se iguala a una suma ponderada de un número asignado previamente de valores ponderados y una variable que es completamente aleatoria, en este caso, la palabra lineal,

significa meramente que el valor real es dependiente aditivamente sobre valores pasados y no se refiere a sus cuadrados o raíces cuadradas.

Este tipo de modelos es la mejor aproximación a muchos procesos naturales, es en particular es aplicable, por ejemplo, a gastos en un río el cual se complementa con cantidades de agua subterránea dependientes en el tiempo, escurrimiento superficial, retención en una cuenca y otros problemas parecidos. El proceso de precipitación por ejemplo, esta fuertemente ligado al sistema de circulación atmosférico y temperaturas oceánicas que pueden persistir por largos periodos de tiempo Loucks, *et al.* (1981).

El método para la solución de series de tiempo proviene del trabajo clásico de Markov (1907). La teoría básica de Yule (1921, 1927) y Walker (1931); Yule fue quien evidentemente trabajó con estos modelos, fue quien usó por primera vez el término de correlación. Wold (1954) hizo un análisis teórico de estos modelos y más recientemente Box y Jenkins (1976) fueron quienes presentaron una metodología sistemática para el desarrollo de estos modelos Kottegoda (1980).

El modelo es aplicado a un componente estocástico z_t (como escurrimiento mensual, anual, etc.) el cual se analiza como una variable aleatoria. Para esto se asume que la tendencia y la temporalidad se removieron con anterioridad. Se propone que el valor de z_t en el tiempo t se conforma de la suma ponderada de los p valores en los tiempo $t-1, t-2, \dots, t-p$ y un número aleatorio a_t ; los a_t valores $t=0, \pm 1, \pm 2, \dots$, son mutuamente independientes, siguen una distribución normal y están idénticamente distribuidos.

4.3.1 Modelos Auto Regresivos $AR(p)$

El modelo lineal auto regresivo de orden p o $AR(p)$ tiene la forma:

$$z_t = \phi_{p,1}z_{t-1} + \phi_{p,2}z_{t-2} + \phi_{p,3}z_{t-3} + \dots + \phi_{p,p}z_{t-p} + a_t \quad (4.4a)$$

$$= \sum_{i=1}^p \phi_{p,i}z_{t-i} + a_t \quad (4.4b)$$

Donde $\phi_{p,i}, i = 1, 2, 3, \dots, p$ son los parámetros auto regresivos o ponderaciones. Las propiedades de z_t y a_t son $E(z_t) = E(a_t) = 0$, $\text{var}(z_t) = E(z_t^2) = \sigma_a^2$, $\text{var}(a_t) = E(a_t^2) = \sigma_a^2$, $\rho_k = E(z_t z_{t-k}) / \sigma_z^2$ y para $k=1, 2, 3, \dots$, $E(a_t a_{t-k}) = E(a_t z_{t-k}) = 0$ donde la última igualdad significa que el número aleatorio actual es independiente de valores ponderados en el mismo proceso. E es la media esperada o valor medio del término entre paréntesis y var es la varianza, ρ_k es el coeficiente de auto correlación teórico y z_t es la variable en estudio (escurrimientos, niveles de embalses, precipitación etc.).

De ahora en adelante se asume, sin perder la generalidad, que la varianza σ_a^2 del componente estocástico a_t es igual a uno. Esto significa que en las aplicaciones, los números generados con la ecuación (4.4) se deben multiplicar por la desviación estándar de la variable modelada y la media se debe sumar a cada número. Estos y otros parámetros se estiman a partir de secuencias de observaciones.

4.3.2 Estimación de los parámetros Auto Regresivos AR(p)

Si la ecuación (4.4) se multiplica por z_{t-1} y se toman esperanzas, se tiene que:

$$E(z_t z_{t-1}) = \phi_{p,1}E(z_{t-1} z_{t-1}) + \phi_{p,2}E(z_{t-2} z_{t-1}) + \phi_{p,3}E(z_{t-3} z_{t-1}) + \dots + \phi_{p,p}E(z_{t-p} z_{t-1}) + E(a_t z_{t-1}) \quad (4.5)$$

Dado que $E(a_t z_{t-1}) = 0$ y tomando en cuenta otras propiedades que se dieron anteriormente, se obtiene la siguiente ecuación:

$$\rho_1 = \phi_{p,1} + \phi_{p,2}\rho_1 + \phi_{p,3}\rho_2 + \dots + \phi_{p,p}\rho_{p-1} \quad (4.6)$$

Donde $\phi_{p,p}$ es el parámetro auto regresivo, ρ_p es el coeficiente de auto correlación teórico. Además si la ecuación (4.4) se multiplica por $z_{t-2}, z_{t-3}, \dots, z_{t-p}$ y se toman esperanzas después de cada multiplicación, se obtienen un número de p relaciones llamadas Yule-Walker, citado por Kottogoda (1980). Estas se pueden representar en forma matricial de la siguiente manera:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \cdot \\ \cdot \\ \cdot \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdot & \cdot & \cdot & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdot & \cdot & \cdot & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \cdot & \cdot & \cdot & \rho_{p-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdot & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \phi_{p,1} \\ \phi_{p,2} \\ \phi_{p,3} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{p,p} \end{bmatrix} \quad (4.7a)$$

O abreviando se puede escribir:

$$\rho_p = P_p \phi_p \quad (4.7b)$$

En donde una condición necesaria para que sea un proceso estacionario es que la matriz de auto correlación P_p sea positiva, esto es, el determinante y todos sus determinantes menores principales deben ser mayores o iguales a cero. Por lo tanto:

$$\phi_p = P_p^{-1} \rho_p \quad (4.8)$$

Para resolver la ecuación (4.8), los estimadores de los coeficientes de auto correlación $r_1, r_2, r_3, \dots, r_p$ se sustituyen en P_p y ρ_p y con las ecuaciones de Yule-Walker para estimar los parámetros auto regresivos $\phi_{p,1}, \phi_{p,2}, \dots, \phi_{p,p}$. Para un proceso normal lineal, esto da una buena aproximación con los métodos de máxima verosimilitud o de mínimos cuadrados.

4.3.3 Modelos de Promedios Móviles MA(q)

En el proceso de promedios móviles MA (q) de orden q, el valor de la variable aleatoria es una suma ponderada de q+1 números aleatorios. Aquí, la ponderación asignada a la suma de los números aleatorios es la unidad, y los números aleatorios generados en tiempos anteriores se multiplican por $-\theta_{q,1}, -\theta_{q,2}, \dots, -\theta_{q,q}$ respectivamente. De acuerdo a esto, el modelo tiene la forma:

$$z_t = a_t - \theta_{q,1}a_{t-1} - \theta_{q,2}a_{t-2} - \dots - \theta_{q,q}a_{t-q} \quad (4.9a)$$

$$= a_t - \sum_{j=1}^q \theta_{q,j}a_{t-j} \quad (4.9b)$$

Donde $\theta_{q,1}$ es el parámetro de promedios móviles. El componente estocástico a_t es equivalente en este caso al resultado de un filtro lineal con “ruido blanco” debido a que el pronóstico de z_t se hace a partir de los errores aleatorios a_t . Tal como anteriormente $E(z_t) = 0$ y $E(z_t^2) = 1$; también, $E(a_t a_{t-k}) = 0$ para $k \neq 0$. Si se eleva al cuadrado la ecuación (4.9) y se toman esperanzas, la varianza y el componente aleatorio se estiman con:

$$\sigma_a^2 = \frac{1}{1 + \theta_{q,1}^2 + \theta_{q,2}^2 + \dots + \theta_{q,q}^2} \quad (4.10)$$

Si se multiplica el lado izquierdo de la ecuación (4.10) por z_{t-k} y el lado derecho por $a_{t-k} - \theta_{q,1}a_{t-1-k} - \theta_{q,2}a_{t-2-k} - \dots - \theta_{q,q}a_{t-q-k}$, se toman esperanzas y se sustituyen por σ_a^2 en la ecuación (4.8) se tiene:

$$\rho_k = \frac{(-\theta_{q,k} + \theta_{q,1}\theta_{q,k+1} + \theta_{q,2}\theta_{q,k+2} + \dots + \theta_{q,q-k}\theta_{q,q})}{(1 + \theta_{q,1}^2 + \theta_{q,2}^2 + \dots + \theta_{q,q}^2)} \quad \text{para } k=1,2,3,\dots,q \quad (4.11)$$

Y $\rho_k = 0$, $k > q$ donde ρ_k es el coeficiente teórico de auto correlación del retraso k .

4.3.4 Modelos Auto Regresivos y de Promedios Móviles ARMA (p,q)

Los modelos mixtos auto regresivos y de promedios móviles se generan combinando las ecuaciones (4.4) y (4.9).

$$z_t = \phi_{p,1}z_{t-1} + \phi_{p,2}z_{t-2} + \dots + \phi_{p,p}z_{t-p} + a_t - \theta_{q,1}a_{t-1} - \theta_{q,2}a_{t-2} - \dots - \theta_{q,q}a_{t-q} \quad (4.12)$$

El efecto lineal en los acuíferos y las cantidades independientes de lluvia justifica la aproximación representada por el proceso de escurrimiento en un río a través de los modelos ARMA. Estos también son análogos a los modelos conceptuales en la hidrología paramétrica en la que se basan algunos estudios de embalses.

La ecuación (4.12) puede escribirse como:

$$\phi_p(B)z_t = \theta_q(B)a_t \quad (4.13)$$

En donde B es un operador de retraso para el caso de series de tiempo univariadas, la ecuación (4.13) se puede expresar como:

$$\phi_p(B) = 1 - \phi_{p,1}B - \phi_{p,2}B^2 - \dots - \phi_{p,p}B^p \quad (4.14)$$

y

$$\theta_q(B) = 1 - \theta_{q,1}B - \theta_{q,2}B^2 - \dots - \theta_{q,q}B^q \quad (4.15)$$

Las ecuaciones (4.14) y (4.15) representan dos polinomios de orden p y q respectivamente, donde las condiciones para que un modelo sea estacionario e invertible respectivamente, son que las raíces de las ecuaciones $\phi_p(B) = 0$ y $\theta_q(B) = 0$ ó que estas dos raíces, también llamadas las raíces de B , deberán estar fuera del círculo unitario (estas condiciones son aplicables a los modelos univariados AR(p) y MA(q) respectivamente).

4.4 Selección de modelos de series de tiempo

La selección del modelo ARMA (p,q) se obtienen por un algoritmo como el de Box y Jenkins (1994); este es un procedimiento iterativo, en la figura 4.1 se resume esquemáticamente esta metodología.

La metodología de Box-Jenkins es muy útil en el pronóstico de valores futuros de una serie de tiempo que tiene valores que son estadísticamente dependientes o relacionados unos de los otros, como es el caso de los escurrimientos en una estación hidrométrica (series univariadas). La metodología de Box-Jenkins para el desarrollo del modelo de pronóstico de una serie consiste en cuatro etapas básicas. Tales etapas son llamadas *identificación, estimación, diagnóstico y pronóstico*, figura 4.1.

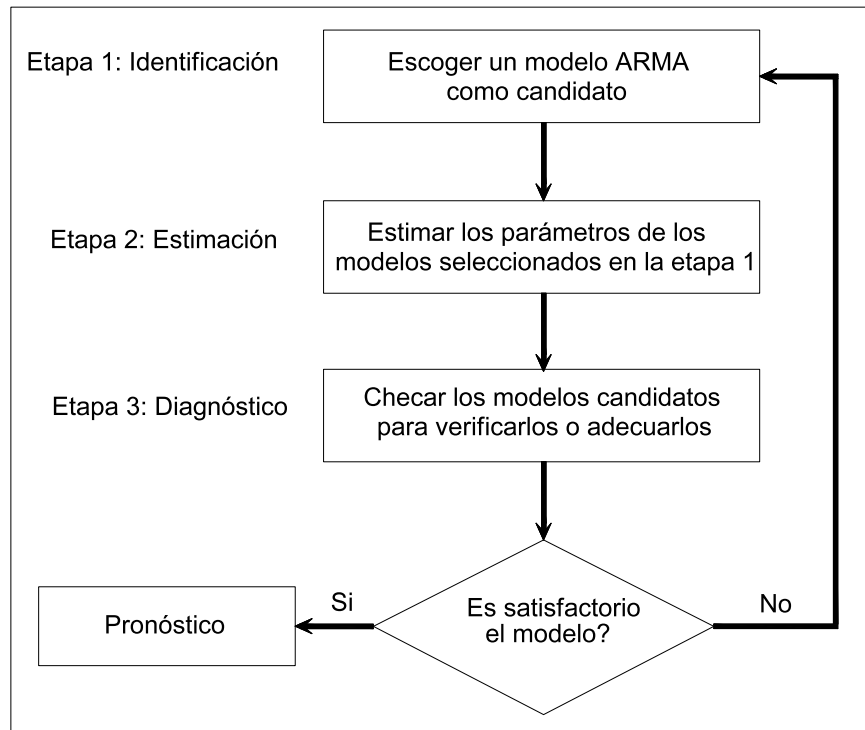


Figura 4.1. Esquema de la metodología para la selección de un modelo de series de tiempo, Box-Jenkins (1994).

Las etapas del método de Box-Jenkins para la construcción de los modelos de series de tiempo univariados, son las siguientes:

4.4.1 Etapa 1: Identificación

La identificación de un modelo tentativo para hacer el pronóstico se selecciona a través del análisis de datos históricos, las herramientas que permite hacer la selección del modelo son la Función de Auto Correlación, (FAC) y Auto Correlación Parcial, (FACP).

Para obtener los coeficientes de la función de auto correlación de la serie de tiempo se utiliza la siguiente ecuación:

$$r_k = \frac{\sum_{t=1}^{n-k} (z_{t+k} - \bar{z})(z_t - \bar{z})}{\sum_{t=1}^n (z_t - \bar{z})^2} \quad (4.16)$$

Donde r_k es el coeficiente de auto correlación de la serie histórica retrasada k periodos de tiempo; \bar{z} promedio de los datos históricos; n número total de observaciones, z_t es la observación en el tiempo t .

Las funciones de Auto Correlación y Auto Correlación Parcial en conjunto son la herramienta más útil para el análisis e identificación de un modelo que sea valido para los datos de una serie en particular.

Con la Función de Auto Correlación Parcial r_{kk} de la FACP se estima la interacción que existe entre cada uno de los retrasos k con la ecuación:

$$r_{kk} = \begin{cases} r_1 & \text{si } k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} \end{cases} \quad (4.17)$$

Donde $r_{kk} = r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j}$, ($k = 2,3,\dots; j = 1,2,\dots, k-1$)

Pankratz (1983), propuso una guía para identificar modelos AR(p) o MA(q) o una combinación de estos. Para ello se basa en las FAC y FACP teóricas que cada modelo ARMA tienen asociadas. En seguida se resumen:

1. Los procesos estacionarios Auto Regresivos AR tienen FAC teóricas que cortan repentinamente el cero después de un número corto de lags y FACP teóricas que cortan repentinamente el cero después de unos pocos retrasos y enseguida los valores de los lags se alternan en signo en lags posteriores. La longitud del lag del último valor antes de que la FACP corte el cero es al orden (p) del proceso AR.
2. El proceso de promedios móviles MA tiene una FAC que corta a cero después de un cierto número de lags. La longitud del último valor de la FAC antes de que la función corte a cero es el orden de (q) del proceso MA. Su FACP teórica cae más allá del cero y posteriormente sus valores se alternan en signo.

Inferencia estadística en la etapa de identificación

Como se sabe, en las estadísticas clásicas, por lo general se quiere inferir sobre el comportamiento de una población, sin embargo, para tener toda la información relevante a esa población sería casi imposible o a un costo muy elevado, por eso se utiliza la inferencia con base en una muestra, en conjunto con algunos conceptos de probabilidad y fórmulas de las matemáticas estadísticas. La muestra se puede usar para estimar las características de la población o hacer algunas pruebas de hipótesis sobre la población, este tipo de razonamiento inductivo se le llama inferencia estadística.

Es importante recalcar lo que se resume en el párrafo anterior ya que las pruebas que se realizan a las series de tiempo siempre se calculan bajo supuestos y por ende se obtienen estimadores de los valores teóricos.

Un coeficiente de auto correlación estimado (r_k) es un estimador de su correspondiente coeficiente de correlación (ρ_k) teórico (desconocido). No se puede esperar que cada r_k sea igual a ρ_k dado el error de muestreo.

Para poder calcular este error, se usa la ecuación de Bartlett citado por Pankratz (1983), que permite calcular el error estándar de una distribución de valores de r_k . La ecuación es la siguiente:

$$S(r_k) = \left(1 + 2 \sum_{j=1}^{k-1} r_j^2 \right)^{1/2} n^{-1/2} \quad (4.18)$$

Donde $S(r_k)$ es el error estándar entre el estimador (r_k) y su teórico (ρ_k), n el número de observaciones de la serie original, j es el número de retrasos que se consideran para el cálculo de FAC y FACP, que generalmente se usa $n/4$. Esta aproximación es apropiada para procesos estacionarios con errores aleatorios que tienen una distribución normal.

Una prueba estadística que se aplica para saber si los valores estimados de los coeficientes de auto correlación ρ_k son significativos estadísticamente es la prueba t , aunque la función no tiene una distribución como la t de student, es asintóticamente normal y comúnmente se utiliza como una t estadística en análisis de regresión. Como una regla para descartar la significancia se debe concluir lo siguiente:

$$\rho_k = 0 \quad \text{si} \quad \left| t_{r_k} \right| = \left| \frac{r_k}{S_{r_k}} \right| \leq 2 \quad (4.19)$$

Se usará como regla que si $\left| t_{r_k} \right| \geq 2$ es razonable concluir que $\rho_k \neq 0$. Así, por ejemplo si $\left| t_{r_k} \right| > 2$ para los retrasos 1, 2, 3, ..., q , y si $\left| t_{r_k} \right| \leq 2$ para los lags $q+1$, $q+2, \dots$, es razonable concluir que la función de auto correlación teórica corta el cero después del retraso q .

Con los antecedentes descritos anteriormente, es posible hacer las pruebas necesarias para el ajuste de los diferentes modelos ARMA(p,q) a los datos hidrológicos.

4.4.2 Etapa 2: Estimación de parámetros

Para estimar los parámetros $\phi_{p,p}$ y $\theta_{q,q}$ de los modelos ARMA se han desarrollado diferentes métodos como el de mínimos cuadrados no lineales, momentos, máxima verosimilitud, algoritmo de Marquardt, entre otros. Con el desarrollo que han tenido las computadoras personales se facilita la creación de nuevos algoritmos (entre los que se tiene los de optimización) que permiten contar con otras herramientas para hacer la estimación de los valores de estos parámetros.

Estimación de parámetros por medio de optimización

Para estimar los valores de los parámetros $\hat{\phi}_p$ y $\hat{\theta}_q$ del modelo de series de tiempo se utiliza un método de optimización que se detalla a continuación:

a). Se proponen valores iniciales de $\hat{\phi}_p$ y/o $\hat{\theta}_q$ que se sustituyen en la siguiente ecuación:

$$\hat{z}_t = \hat{\phi}_1 z_{t-1} \dots - \hat{\phi}_p z_{t-p} + \hat{\theta}_1 a_{t-1} \dots - \hat{\theta}_q a_{t-q} \quad (4.20)$$

Con la ecuación (4.20) se genera una nueva serie:

b). Se calculan las diferencias o residuos entre la serie histórica y la serie generada:

$$a_t = (z_t - \hat{z}_t) \quad (4.21)$$

c). Se obtienen los cuadrados de las diferencias a_t^2 . La sumatoria de las diferencias al cuadrado entre la serie histórica y la generada es la función objetivo a minimizar como se describe:

$$\min \sum_{t=1}^N a_t^2 = \min \sum_{t=1}^N (z_t - \hat{\phi}_1 z_{t-1} - \dots - \hat{\phi}_p z_{t-p} + \hat{\theta}_1 a_{t-1} + \dots + \hat{\theta}_q a_{t-q})^2 \quad (4.22)$$

Sujeta a las condiciones de:

$$|\phi_p| < 1; \text{ estacionalidad} \quad (4.23)$$

$$|\theta_q| < 1; \text{ invertibilidad} \quad (4.24)$$

4.4.3 Etapa 3: Diagnóstico

El modelo estadístico adecuado es aquel cuyos valores de los errores aleatorios o proceso de ruido blanco son independientes entre sí, es decir, que no están auto correlacionados. En la práctica no es posible observar los errores aleatorios (a_t), pero se tienen sus estimadores, los residuos (\hat{a}_t) calculados en la etapa de estimación. En la etapa de verificación se usan los residuos para probar la hipótesis de independencia de los errores aleatorios.

Existe una razón muy importante para tener interés en que los errores aleatorios sean independientes. Los errores aleatorios son una componente de la variable que se analiza. Así, si los errores aleatorios están correlacionados en la serie de datos, entonces hay un patrón de auto correlación en la variable en estudio que no se ha encontrado y que se deberá tomar en cuenta en los términos AR o MA

del modelo en análisis. Si los errores están correlacionados se tendrá que reformular el modelo.

La herramienta básica para el análisis en la etapa de diagnóstico es la *función de auto correlación* aplicada a los errores \hat{a}_t . La ecuación con la que se obtienen los coeficientes es:

$$\hat{a}_t = \frac{\sum_{t=1}^{n-k} (\hat{a}_t - \bar{a})(\hat{a}_{t+k} - \bar{a})}{\sum_{t=1}^n (\hat{a}_t - \bar{a})^2} \quad (4.25)$$

Una vez que se calcularon los coeficientes de correlación de los residuos es importante determinar la hipótesis de la significancia estadística diferente de cero de cada uno de los coeficientes, para ello, se usa la ecuación (4.18) de Bartlett (1946), citado por Pankratz (1983), para calcular el error estándar de las Función de Auto Correlación residual.

La prueba más importante para determinar de la independencia de los errores aleatorios es la prueba χ^2 . Ljung-Box (1978), citado por Pankratz (1983), desarrollaron una prueba en la que se analizan todos los coeficientes como un conjunto, para ello se prueba la hipótesis nula de correlación en la variable aleatoria \hat{a}_t .

$$H_0 : \rho_1(a) = \rho_2(a) = \dots = \rho_k(a) = 0 \quad (4.26)$$

Una prueba para determinar la auto correlación en el proceso de ruido blanco, visto como un conjunto, es la prueba estadística Q^* de Ljung-Box;

$$Q^* = n(n+2) \sum_{k=1}^K (n-k)^{-1} r_k^2(\hat{a}) \quad (4.27)$$

En donde n es el número total de observaciones, r_k es el número de retrasos considerados para la obtención de las funciones de auto correlación y auto correlación parcial. La prueba Q^* tiene una distribución aproximada a la prueba de Chi cuadrada χ^2 con $(k-m)$ grados de libertad, siendo m el número de parámetros estimados en el modelo ARMA(p,q), el valor de Q^* se obtiene de una tabla de valores críticos para la función χ^2 .

4.4.4 Etapa 4: Pronóstico

Sea t el periodo de tiempo actual, para el pronóstico, el principal interés es en periodos de tiempo futuro para la variable en series de tiempo, denotadas como z_{t+l} , donde $l \geq 1$. El periodo de tiempo t es llamado el origen del pronóstico, y l es llamado el pronóstico hacia delante o *futuro*. En el análisis de los modelos ARMA, el pronóstico depende de las observación de una variable realizadas hasta el periodo t . Sea la información contenida en un conjunto de observaciones (z_t, z_{t-1}, \dots) , designado como I_t . Entonces el pronóstico de z_{t+l} , será designado como $\hat{z}_t(l)$ es la esperanza matemática de z_{t+l} dado I_t . Esto es:

$$\hat{z}_t(l) = E(z_{t+l} | I_t) \quad (4.28)$$

Refiriéndose a un origen de tiempo t es deseable hacer un pronóstico con el error medio cuadrático mínimo de z_{t+l} , donde $l \geq 1$ es el intervalo de tiempo futuro. El pronóstico de un modelo ARMA(p,q) se puede expresar como una suma infinita ponderada de valores aleatorios previos al tiempo t , Guerrero (2003):

$$\hat{z}_t = \mu + \psi_0 a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \quad (4.29)$$

Donde ψ_j son los coeficientes de los valores aleatorios anteriores al tiempo t , μ es la media de la serie histórica.

El cálculo de ψ_i es:

$$\psi_0 = 1$$

$$\psi_1 = \phi_1 - \theta_1$$

$$\psi_2 = \phi_1(\phi_1 - \theta_1)$$

y en general $\psi_j = \phi_1^{j-1}(\phi_1 - \theta_1)$

Así la varianza del error del pronóstico (con $\psi_0 = 1$) es:

$$\text{var}[e_t(l)] = E[e_t^2(l)] = \sum_{j=0}^{l-1} \psi_j^2 \sigma_a \quad (4.30)$$

Donde $e_t(l)$ es el error en el origen t para el pronóstico en el periodo l , Var es la varianza del error y σ_a es la varianza de las diferencias \hat{a}_t vistas como un proceso aleatorio.

El intervalo de confianza $(1-\alpha)$ para un pronóstico con error medio cuadrático mínimo $z_t(l)$ es:

$$z_{t+l} = z_t(l) \pm u_{\alpha/2} \left\{ 1 + \sum_{j=1}^{l-1} \psi_j^2 \right\}^{1/2} \sigma_a \quad (4.31)$$

Donde $u_{\alpha/2}$ es el nivel de probabilidad para una distribución normal estándar.

4.5 Modelos de series de tiempo multivariados

Desde los años 60 del siglo pasado se han centrado esfuerzos en desarrollar métodos para analizar las propiedades estocásticas de series hidrológicas y así brindar esquemas para el desarrollo de series hidrológicas multivariadas. Debido a que en la planeación, diseño y operación de sistemas de recursos hidráulicos generalmente se involucran varias fuentes superficiales y usos del agua, el análisis estocástico y la modelación multivariada se vuelven relevantes. En hidrología estocástica, un modelo multivariado se refiere a una misma variable en diferentes sitios, ejemplo, escurrimientos medios anuales registrados en diferentes estaciones hidrométricas.

Las series de tiempo hidrológicas periódicas, como mensuales, semanales y diarias tienen unos ciertos periodos y propiedades estocásticas. Las propiedades periódicas se dan generalmente en forma de medias periódicas desviaciones estándar periódicas y coeficientes de sesgo periódicos, mientras que las propiedades pueden estar representadas por coeficientes de auto correlación periódicos o constantes. Para el caso de las series de tiempo anuales el coeficiente de auto correlación se considera constante, sin embargo, en general este no es el caso para series de tiempo con periodos menores dado que los coeficientes varían con el tiempo. Cuando se trabaja con series periódicas multivariadas, los coeficientes de correlación cruzada lag-0, lag-1, (o cualquier otro lag) también pueden ser periódicos llevando a matrices de parámetros periódicos del modelo multivariado.

4.5.1 Correlograma cruzado

La función de covarianza C_k que mide la dependencia lineal entre variables con k número de unidades de tiempo retrazadas en las series de tiempo $Z(t)$, $t > 0$ con

media cero se define por un proceso de covarianza estacionario expresado como:

$$C_k = E \{ Z_{(t+k)} Y_{(t)} \} \quad (4.32)$$

Donde E denota el valor esperado, $Z_{(t+k)}$ y $Y_{(t)}$ son las variables que se analizan.

La ecuación (4.32) es útil para la modelación de series de tiempo multivariadas y para la determinación de estructuras de interdependencia entre dos series de tiempo $\{Z(t), t \in T\}$ y $\{Y(t), t \in T\}$ no solamente en el mismo periodo de tiempo sino en series de tiempo con k periodos de tiempo retrazados. Un ejemplo de esto es la dependencia lineal entre los escurrimientos de un río en varios puntos de aforo, o en una estación aguas abajo y los antecedentes de lluvia en la cuenca del río. Así, la función de covarianza cruzada se define como:

$$C_{ZY}(k) = E \{ Z_{(t)} Y_{(t+k)} \} \quad (4.33)$$

Y también:

$$C_{YZ}(k) = C_{ZY}(-k) = E \{ Z_{(t+k)} Y_{(t)} \} \quad (4.34)$$

Donde las dos series tienen media cero, C_{ZY} es el coeficiente de correlación múltiple. Nótese que $C_{ZY}(k) \neq C_{ZY}(-k)$ porque no es una función par, en contraste con la igualdad $C_k = C_{-k}$, en concordancia con la ecuación (4.33). La función de correlación cruzada entre las dos series $Z_{(t)}$ y $Y_{(t)}$ con medias $E(Z)$ y $E(Y)$ y varianzas $var(Z)$ y $var(Y)$, está dada por:

$$\rho_{ZY}(k) = E \left[\frac{\{Z(t) - E(Z)\} \{Y(t+k) - E(Y)\}}{\sqrt{var(Z) var(Y)}} \right] \quad (4.35)$$

Similarmente $\rho_{yz}(k)$ corresponde a la ecuación (4.34). Estas se estiman a partir de dos conjuntos de observaciones $(z_t, y_t; t=1, 2, 3, \dots, N)$, con medias estimadas \bar{z} y \bar{y} con los coeficientes de correlación cruzada:

$$r_{zy}(k) = \frac{\left\{ \sum_{t=1}^{N-k} (z_t - \bar{z})(y_{t+k} - \bar{y}) \right\}}{\left[\left\{ \sum_{t=1}^N (z_t - \bar{z})^2 \right\} \left\{ \sum_{t=1}^N (y_t - \bar{y})^2 \right\} \right]} \quad (4.36)$$

Para los lags $k = 0, 1, 2, \dots, N$. Los coeficientes de correlación cruzados para lags negativos son iguales a los coeficientes $r_{yz}(k)$, para lags positivos, se obtienen cambiando el numerador de la ecuación (4.36). Así se obtiene un correlograma cruzado para $k = \pm 1, \pm 2, \dots$

4.5.2 Descripción del modelo de series de tiempo multivariado

El modelo lag-1 univariado Auto Regresivo se denomina modelo AR(1); en general, el modelo lag-p se presenta como el modelo AR(p). En el caso de modelos multivariados se usa la misma notación. AR(p) significa modelo multivariado de orden p , o modelo multivariado auto regresivo de orden p . En general, la notación es similar a la que se usa para modelos univariados. La diferencia es el uso de letras mayúsculas para la representación de vectores o matrices de variables aleatorias y parámetros. Se utilizan letras griegas para definir variables aleatorias, si están subrayadas significa que son vectores o matrices.

4.5.3 Modelos multivariados generales

Considérese una serie estacionaria multivariada $z_t^{(i)}$, $i = 1, \dots, n$, donde n es el número de series de tiempo o variables. Sin perder la generalidad, se asume que las $z_t^{(i)}$ son todas variables estandarizadas con media cero y desviación estándar uno, es decir que, $E[z_t^{(i)}] = 0$ y $\text{Var}[z_t^{(i)}] = 1$, para $i = 1, \dots, n$. El modelo AR(1) (modelo multivariado lag-1) de las series $z_t^{(i)}$ $i = 1, \dots, n$ se puede representar en forma matricial de la siguiente manera:

$$Z_t = A_1 Z_{t-1} + B \underline{\varepsilon}_t \quad (4.37)$$

Donde Z_t es un vector de $(n \times 1)$ elementos, $z_t^{(i)}$, $i = 1 \dots n$ sitios o variables; A_1 y B son matrices de $(n \times n)$ parámetros, $\underline{\varepsilon}_t$ es un vector de $(n \times 1)$ variables aleatorias independientes, distribuidas normalmente con media cero y desviación estándar uno. Se asume que el vector $\underline{\varepsilon}_t$ no tiene correlación en el espacio y en el tiempo, esto es, $E[\varepsilon_t^{(i)} \varepsilon_t^{(j)}] = 0$ para $i \neq j$ o $E[\varepsilon_t \varepsilon_t^T] = I$, T es la matriz transpuesta. La ecuación (4.37) en forma expansiva se puede expresar como:

$$\begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \\ \cdot \\ \cdot \\ z_t^{(n)} \end{bmatrix} = \begin{bmatrix} a^{11} & a^{12} & \cdot & \cdot & \cdot & a^{1n} \\ a^{21} & a^{22} & \cdot & \cdot & \cdot & a^{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a^{n1} & a^{n2} & \cdot & \cdot & \cdot & a^{nn} \end{bmatrix} \begin{bmatrix} z_{t-1}^{(1)} \\ z_{t-1}^{(2)} \\ \cdot \\ \cdot \\ z_{t-1}^{(n)} \end{bmatrix} + \begin{bmatrix} b^{11} & b^{12} & \cdot & \cdot & \cdot & b^{1n} \\ b^{21} & b^{22} & \cdot & \cdot & \cdot & b^{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ b^{n1} & b^{n2} & \cdot & \cdot & \cdot & b^{nn} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \cdot \\ \cdot \\ \varepsilon_t^{(n)} \end{bmatrix} \quad (4.38)$$

La estructura de correlación de z_t de la ecuación (4.37) o (4.38) indica una correlación cruzada lag-0 o lag-1 en el espacio y una correlación serial lag-1 en el tiempo. Este tipo de estructura de correlación se muestra gráficamente en la figura 4.3, donde las flechas horizontales representan el coeficiente de

correlación serial individual con lag-1 ρ_1^{ii} y ρ_1^{jj} , las flechas en diagonal representan la correlación cruzada lag-1 ρ_1^{ij} y ρ_1^{ji} , y finalmente las flechas verticales representan las correlaciones cruzadas lag-0 ρ_0^{ij} . Cuando esta correlación permanece constante, entonces los coeficientes ρ son iguales para los tiempos $(t, t+1)$, $(t+1, t+2)$, etc.

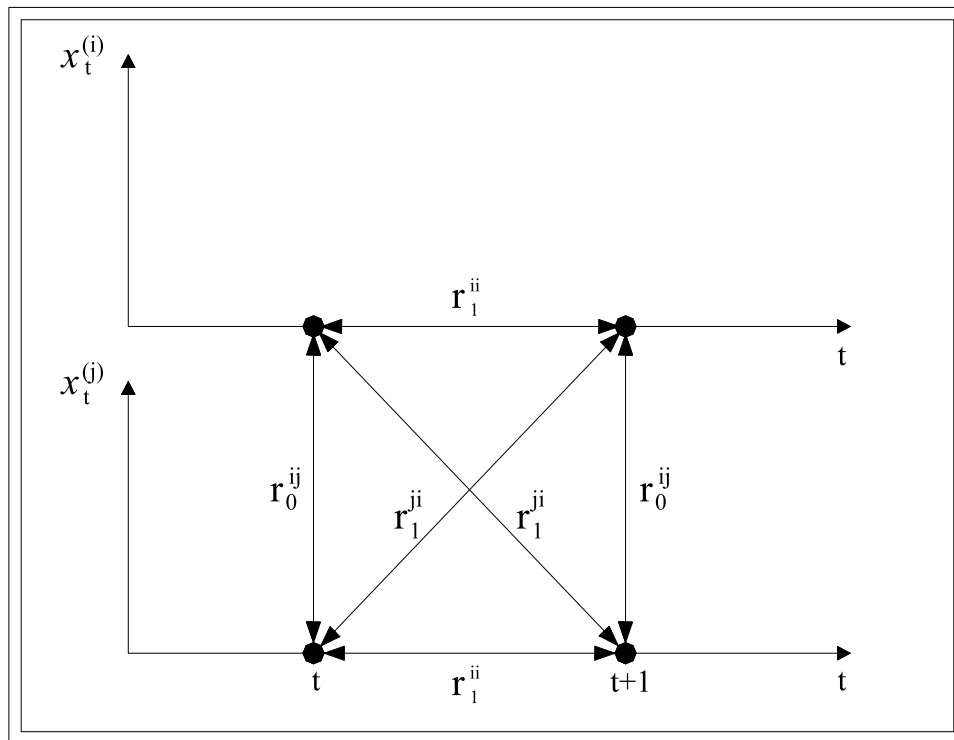


Figura 4.2. Representación esquemática de la estructura de correlación de un modelo multivariado AR(1).

Si se supone el caso de dos variables $n = 2$, el modelo de la ecuación (4.37) o (4.38) es:

$$z_t^{(1)} = a_{11}z_{t-1}^{(1)} + a_{12}z_{t-1}^{(2)} + b_{11}\varepsilon_t^{(1)} + b_{12}\varepsilon_t^{(2)} \quad (4.39)$$

$$z_t^{(2)} = a_{21}z_{t-1}^{(1)} + a_{22}z_{t-1}^{(2)} + b_{21}\varepsilon_t^{(1)} + b_{22}\varepsilon_t^{(2)} \quad (4.40)$$

La ecuación (4.31) para la variable 1 muestra la dependencia serial lag-1 entre $z_t^{(1)}$ y $z_{t-1}^{(1)}$ y la dependencia cruzada lag-1 entre $z_t^{(1)}$ y $z_{t-1}^{(2)}$

4.5.4 Propiedades de los modelos multivariados AR(p)

Las propiedades principales de los modelos multivariados AR(p) son la media, la varianza y la estructura de covarianza. En general, un modelo multivariado tiene el vector $\underline{\mu}$ de medias con elementos μ^1, \dots, μ^n ; el vector σ^2 de varianzas con elementos $\sigma^{2(1)}, \dots, \sigma^{2(n)}$ y una estructura de covarianzas la cual se puede representar por la matriz M de (nxn) coeficientes. Para una fácil presentación y sin perder la generalidad, se asume que las medias son iguales a cero y las varianzas igual a uno para cada una de las series del modelo.

Considerando la matriz de correlación:

$$M_k = E[Z_t Z_{t-k}^T] = \begin{bmatrix} \rho_k^{11} & \rho_k^{12} & \cdot & \cdot & \cdot & \rho_k^{1n} \\ \rho_k^{21} & \rho_k^{22} & \cdot & \cdot & \cdot & \rho_k^{2n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \rho_k^{n1} & \rho_k^{n2} & \cdot & \cdot & \cdot & \rho_k^{nn} \end{bmatrix} \quad (4.41)$$

Donde T es la transpuesta de la matriz, $\rho_k^{ij} = \text{corr}[z_t^{(i)}, z_{t-k}^{(j)}]$ es el coeficiente de correlación lag-k entre $z_t^{(i)}$ y $z_{t-k}^{(j)}$ para $i \neq j$, y es también el coeficiente de auto correlación de orden k para $i = j$.

4.6 Modelo multivariado Auto Regresivo AR(1) anual

Considérese las series hidrológicas anuales $z_t^{(i)}$ en una región donde existen datos de escurrimientos disponibles en n sitios, $i = 1, \dots, n$. Las series anuales Z_t pueden ser representadas o no por la Función de Distribución Normal. Sí no es así, otras Funciones de Probabilidad las pueden representar como la log-normal, gamma, entre otras y dichas series se deben transformar a la normal o por lo menos su coeficiente de sesgo deberá ser cercano a cero. Considérese que la matriz de la función de transformación Y_t se representa por \underline{g} :

$$Y_t = \underline{g}(x_t) \quad (4.42a)$$

O en forma matricial:

$$\begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ \cdot \\ \cdot \\ y_t^{(n)} \end{bmatrix} = \begin{bmatrix} g^{(1)}, & 0, & \cdot & \cdot & \cdot, & 0 \\ 0, & g^{(2)}, & \cdot & \cdot & \cdot, & 0 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ 0, & 0, & \cdot & \cdot & \cdot & g^{(n)} \end{bmatrix} \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ \cdot \\ \cdot \\ x_t^{(n)} \end{bmatrix} \quad (4.42b)$$

Donde las g 's pueden ser diferentes funciones para cada uno de los sitios i . Si los gastos históricos Z son normales, entonces la matriz de transformación $g = I$, es una matriz identidad. Sí un $z_t^{(j)}$ dado es normal, entonces $g^{(j)} = 1$.

El modelo general para las series normalizadas Y_t (transformadas) es:

$$\begin{aligned}
y_t^{(1)} &= \mu^{(1)} + \sigma^{(1)} z_t^{(1)} \\
&\cdot \quad \cdot \quad \cdot \\
&\cdot \quad \cdot \quad \cdot \\
&\cdot \quad \cdot \quad \cdot \\
y_t^{(n)} &= \mu^{(n)} + \sigma^{(n)} z_t^{(n)}
\end{aligned}
\tag{4.43a}$$

Donde $\mu^{(i)}$ y $\sigma^{(i)}$ son la media y la desviación estándar de $y_t^{(i)}$, y $z_t^{(i)}$ es la variable estandarizada correlacionada en el tiempo y con una correlación cruzada en el espacio. La ecuación (4.43) se puede escribir en forma matricial de la siguiente manera:

$$Y_t = \underline{\mu} + \underline{\sigma} Z_t \tag{4.43b}$$

Donde $\underline{\mu}$ es una matriz de $(n \times 1)$, $\underline{\sigma}$ es una matriz diagonal de $(n \times n)$ y Y_t y Z_t son dos matrices de $(n \times 1)$.

4.6.1 Estimación de los parámetros del modelo

Se asume que existen datos de un número N de años disponibles para cada sitio i , cada uno con observaciones $z_1^{(i)}, z_2^{(i)}, \dots, z_N^{(i)}$. Estos datos son normales o es posible, mediante una transformación convertirlos en normales. La serie normal o normalizada es:

$$y_t^{(i)} = [y_1^{(i)}, y_2^{(i)}, \dots, y_N^{(i)}] \quad i = 1, 2, \dots, n \tag{4.44}$$

Las n series deben tener el mismo tamaño de la muestra N , lo que implica que si existen datos faltantes en una serie se deben completar con algún método de complementación de datos hidrológicos, y las series “cortas” se deberán extender, por ejemplo, como modelos ARMA u otro método, para igualar el

tamaño de las otras series. Esta condición es necesaria para la estimación de los parámetros de las matrices del modelo. Los estimadores de los momentos $\underline{\mu}$ y $\underline{\sigma}$ de la ecuación (4.43b) se calculan con las ecuaciones:

$$\hat{\mu}^{(i)} = (1/N) \sum_{t=1}^N y_t^{(i)} \quad i = 1 \dots n \quad (4.45)$$

y

$$\hat{\sigma}^{(i)} = [(1/N) \sum_{t=1}^N \{y_t^{(i)} - \hat{\mu}^{(i)}\}^2]^{1/2} \quad i = 1 \dots n \quad (4.46)$$

De la ecuación (4.46) la variable estandarizada $z_t^{(i)}$ es:

$$z_t^{(i)} = [y_t^{(i)} - \hat{\mu}^{(i)}] / \hat{\sigma}^{(i)} \quad i = 1 \dots n \quad (4.47)$$

Para el modelo AR(1) de la ecuación (4.47) los estimadores de los momentos de los parámetros A_1 y B se pueden obtener con:

$$\hat{A}_1 = \hat{M}_1 \hat{M}_0^{-1} \quad (4.48)$$

y

$$\hat{B} \hat{B}^T = \hat{M}_0 - \hat{A}_1 \hat{M}_1^T = D \quad (4.49)$$

Respectivamente, donde \hat{M}_0 y \hat{M}_1 son las matrices de correlación lag-0 y lag-1 de Z_t , que se obtienen con el procedimiento de correlograma cruzado ecuaciones (4.32) a la (4.36), Kottegoda (1980).

La matriz \hat{B} de la ecuación (4.49) se determina con el procedimiento que se describe en seguida. Para la obtención de los parámetros de la matriz \hat{B} se requiere solucionar la ecuación de matrices $\hat{B} \hat{B}^T = D$. Esto es, dado que los elementos de la matriz D son conocidos, es necesario encontrar elementos en la matriz B tales que el producto de B con su transpuesta B^T sea igual a D .

Cualquier solución para B que produzca $BB^T = D$ es una solución válida. En general existen un número infinito de soluciones que producirán D . La solución para que $BB^T = D$ se puede obtener por análisis del componente principal, pero para identificar B es un procedimiento bastante complejo, Salas *et al.* (1980). Sin embargo, si se asume que B es una matriz triangular inferior, entonces se puede encontrar una solución única por el método de raíces cuadradas, u otros métodos.

Por lo tanto, si B es una matriz triangular inferior y D es una matriz positiva definida, entonces los elementos no cero de B se pueden determinar por Graybill (1969), citado por Kottegoda (1980) como:

$$b^{ij} = d^{ji} / b^{jj} \quad \text{para } j=1 \quad i=1 \dots n \quad (4.50)$$

$$b^{ij} = \left[d^{ij} - \sum_{k=1}^{j-1} (b^{jk})^2 \right]^{1/2} \quad \text{para } j = 2, \dots, n, \quad i=j \quad (4.51)$$

$$b^{ij} = \left[d^{ij} - \sum_{k=1}^{j-1} b^{jk} b^{ik} \right] / b^{jj} \quad \text{para } j = 2, \dots, n-1, \quad i=j+1, \dots, n \quad (4.52)$$

Donde b^{ij} son elementos de la matriz B , d^{ij} son elementos de la matriz D y n es la dimensión de las matrices B y D .

De otra forma, si B es una matriz triangular inferior pero D es una matriz definida positiva o positiva semidefinida, entonces los elementos de B se pueden determinar por:

$$b^{ki} = 0 \quad \text{para todas las } k < i \quad (4.53)$$

$$b^{ki} = 0 \quad \text{para todas las } k \geq i \quad (4.54)$$

$$d^{ii} - \sum_{j>i} (b^{ij})^2 \leq 0 \quad (4.55)$$

Y

$$b^{ki} = \frac{d^{ki} - \sum_{j<i} b^{ij} b^{kj}}{\left[d^{ii} - \sum_{j<i} (b^{ij})^2 \right]^{1/2}}, \text{ para todas las } k \geq i \text{ cuando } d^{ii} - \sum_{j<i} (b^{ij})^2 > 0 \quad (4.56)$$

4.6.2 Modelación para generación de escurrimientos sintéticos

El modelo para la generación sintética de series anuales multivariadas se puede obtener sustituyendo de los parámetros estimados con las ecuaciones (4.37), (4.42) y (4.43). Así, las series anuales $z_t^{(i)}$, $i = 1, \dots, n$ se generan con:

$$\hat{z}_t^{(i)} = \hat{g}^{-1} \hat{y}_t^{(i)} \quad i = 1, \dots, n \quad (4.57)$$

Y

$$\hat{y}_t^{(i)} = \hat{\mu}^{(i)} + \hat{\sigma}^{(i)} \hat{z}_t^{(i)} \quad i = 1, \dots, n \quad (4.58)$$

En donde \hat{g}^{-1} es la función inversa de la transformación a la normal, $\hat{\mu}$ y $\hat{\sigma}$ son estimadores de la media μ y la desviación estándar σ , i es la variable hidrológica observada en un sitio y n es el número de variables en el modelo multivariado. Las series $\hat{z}_t^{(i)}$ se pueden representar por el modelo AR(1) de la ecuación (4.37):

$$\hat{Z}_t = \hat{A}_1 \hat{Z}_{t-1} + \hat{B} \hat{\epsilon}_t \quad (4.59)$$

El cual es la forma matricial del modelo de generación de datos sintéticos. Usando la representación de la ecuación (4.48), el modelo de generación AR(1) se puede escribir de la siguiente manera:

$$\begin{aligned}
 \hat{z}_t^{(1)} &= \hat{a}^{11} \hat{z}_{t-1}^{(1)} + \dots + \hat{a}^{1n} \hat{z}_{t-1}^{(n)} + \hat{b}^{11} \varepsilon_t^{(1)} + \dots + \hat{b}^{1n} \varepsilon_t^{(n)} \\
 \hat{z}_t^{(2)} &= \hat{a}^{21} \hat{z}_{t-1}^{(1)} + \dots + \hat{a}^{2n} \hat{z}_{t-1}^{(n)} + \hat{b}^{21} \varepsilon_t^{(1)} + \dots + \hat{b}^{2n} \varepsilon_t^{(n)} \\
 &\vdots \\
 \hat{z}_t^{(n)} &= \hat{a}^{n1} \hat{z}_{t-1}^{(1)} + \dots + \hat{a}^{nn} \hat{z}_{t-1}^{(n)} + \hat{b}^{n1} \varepsilon_t^{(1)} + \dots + \hat{b}^{nn} \varepsilon_t^{(n)}
 \end{aligned} \tag{4.60}$$

Donde las \hat{a} 's y las \hat{b} 's son los elementos de la matriz de estimación de parámetros A_1 y B , respectivamente, y $\varepsilon_t^{(i)}$ son variables independientes normales e idénticamente distribuidas con media cero y varianza uno.

4.6.3 Pruebas de bondad de ajuste de modelos multivariados anuales

Las pruebas de bondad de ajuste de modelos multivariados no han sido desarrolladas ampliamente en la estadística hidrológica. La mayoría de la literatura se enfoca solo en el planteamiento de los problemas, descripción de los modelos, estimación de los parámetros y en la generación de nuevas series sintéticas. La prueba del ajuste del modelo se hace por comparación de las auto correlaciones de las series históricas con las generadas.

En la prueba de ajuste de un modelo multivariado AR(1) se sugiere el siguiente procedimiento; 1) Comparar las matrices \hat{M}_0 y \hat{M}_1 de correlación históricas con las generadas y los correlogramas históricos con los generados; 2) Probar la independencia en el espacio y en el tiempo de los residuos de las series ε_t ; y 3) probar la hipótesis de normalidad de ε_t .

En el primer procedimiento se compara las matrices de correlación de las series históricas con las generadas así como también los correlogramas individuales de cada una de las series generadas con las históricas. El segundo procedimiento es la verificación de los residuos para probar su independencia en el tiempo y en el espacio.

Los residuos se obtienen con la ecuación:

$$\varepsilon_t = \hat{B}^{-1}(Z_t - \hat{A}_1 Z_{t-1}) \quad (4.61)$$

4.7 Modelos de disgregación

La disgregación de datos de escurrimiento ha sido y continúa siendo un reto en la hidrología estocástica. Las décadas pasadas fueron testigos de numerosos estudios con el objetivo de disgregar datos de escurrimientos, en consecuencia se desarrollaron un gran número de modelos matemáticos, por ejemplo Harms y Campbell, (1967); Valencia y Schaake, (1972); Salas *et al.*, (1980); Stedinger y Vogel, (1984); Bras y Rodríguez-Iturbide, (1985); Grygier y Stedinger, (1988); Lin y Lee, (1990); Santos y Salas, (1992) citados por Maheepala y Perera, (1996).

Los modelos de disgregación se han convertido en la técnica más utilizada para la modelación de series de tiempo hidrológicas. El modelo de Harms y Campbell, (1967), fue la primera propuesta para la disgregación de datos de escurrimientos, Maheepala y Perera, (1996).

El objetivo principal de cualquier modelo de disgregación es permitir la conservación de propiedades estadísticas de las series históricas en las series generadas. Para explicar esto, considérese el caso donde se generan datos mensuales a partir de datos anuales para el estudio de operación de un

embalse. Lo que se desea es conservar en los datos mensuales sintéticos ciertas propiedades estadísticas de los registros históricos.

Adicionalmente también es deseable conservar las propiedades estadísticas de los registros anuales. Así, lo importante es conservar las propiedades estadísticas en ambos niveles, mensual y anual, estas propiedades estadísticas son la media, la desviación estándar, las funciones de distribución, y algunas covarianzas, o equivalentemente correlaciones. Los modelos de disgregación permiten conservar estas estadísticas a más de un nivel. Estos objetivos se aplican a los modelos de disgregación, temporal y espacial.

4.7.1 Descripción de los modelos de disgregación

La generación de datos con modelos de disgregación es un proceso por el cual las series de tiempo se generan dependientes de series de tiempo que ya están disponibles. Generalmente, las series de tiempo independientes se generan previamente. Esto se puede realizar aplicando algún método modelo estocástico. Las series originales son disgregadas o divididas en subseries. Comúnmente, como es el caso de la disgregación temporal, en las subseries (datos semestrales, bimestrales, mensuales, etc.) suma debe ser igual a la serie original, por ejemplo, datos anuales.

La generación de subseries dependientes de las series originales se lleva a cabo con modelos lineales que conservan las propiedades estadísticas importantes entre las subseries y las series originales. De esta manera, las propiedades estadísticas de las series originales y en las subseries se conservan.

Las dos formas básicas de los modelos de disgregación son la temporal y la espacial. Un ejemplo de disgregación temporal es la disgregación de series anuales en series mensuales. Otros ejemplos de disgregación temporal son

series anuales (original) a series semestrales, series semestrales a series trimestrales, series trimestrales a series mensuales, series mensuales a series semanales, etc. Un ejemplo de disgregación espacial es el caso de la división del escurrimiento total de una cuenca en escurrimientos por tributarios. En consecuencia los modelos de disgregación se pueden usar con otros propósitos, tal como disgregar las series de sólidos totales disueltos o la cantidad de contaminantes en un río.

4.7.2 Modelo de disgregación para un sitio

La modelación para una variable, en un sitio o una estación hidrométrica o estación pluviográfica, los registros que se disgregarán corresponden a series de tiempo de escurrimientos, lluvias u otras variables climatológicas. Se seleccionan series de tiempo estacionarias, que deberán ser normales o mediante algún procedimiento válido ser normalizadas. A continuación se describe el proceso de la obtención del modelo de disgregación para cada serie histórica.

1. Con los registros históricos anuales de una estación hidrométrica, se construye un modelo anual que permita generar series de tiempo futuras que conserven los parámetros estadísticos de los registros históricos.
2. Para los registros históricos mensuales se desarrolla un modelo de series de tiempo para cada mes. El modelo de generación de series mensuales debe conservar las propiedades estadísticas de las series históricas mensuales.

Por lo que para la modelación del escurrimiento en una estación hidrométrica, se tendrán 13 modelos (uno para cada mes del año y el modelo anual). Para cualquiera de los casos los volúmenes mensuales en un año en particular deberán sumar el escurrimiento total para un año dado j , es decir:

$$\sum_{i=1}^{12} q_i = Q_j \quad (4.62)$$

Donde q_i es el escurrimiento para cada mes en el año j , i es el número de meses 1 = Enero,..., 12 = Diciembre, Q_j es el escurrimiento total en el año j .

Los meses disgregados q_i para un año dado j , se seleccionan con base en la cercanía que tenga con el valor del total del escurrimiento que se generó con el modelo AR(1) anual con base en el valor mínimo que tenga la siguiente ecuación.

$$\left(\frac{\sum_{i=1}^{12} q_i - Q_j}{\hat{\sigma}_Q} \right)^2 \quad (4.63)$$

Donde $\hat{\sigma}_Q$ es la desviación estándar de los registros históricos anuales.

4.7.3 Modelo de disgregación para sitios múltiples

Al igual que en los modelos de disgregación en una variable, los registros históricos de escurrimientos mensuales y anuales deben estar disponibles para cada una de las estaciones hidrométricas para las que se vayan a generar series de datos sintéticos.

A continuación se describe la generación de datos de escurrimientos medios mensuales:

1. Generar series mensuales para cada sitio aplicando un modelo de generación mensual que conserve satisfactoriamente los parámetros estadísticos mensuales, incluyendo la correlación entre las estaciones hidrométricas. Las series generadas deben tener la longitud suficiente para que los periodos húmedos y secos sean bien representados por el modelo. Estos meses q_i se utilizarán para obtener los volúmenes anuales Q_j , como en el caso de la disgregación de un sitio.

2. Para la generación de series anuales se debe utilizar un modelo que conserve las propiedades estadísticas de los datos generados, las estructuras de correlación entre las variables y la estructura de auto correlación entre la misma variable.

Disgregar las series generadas en el paso 2 utilizando los valores mensuales que se obtuvieron en el paso 1. El criterio de selección de los escurrimientos mensuales q_i será igual que en el modelo de disgregación para un sitio, ecuación (4.63) aplicado a un modelo con variables múltiples.

$$\left(\frac{\sum_{k=1}^{12} q_k^i - Q_j^i}{\sigma_Q^i} \right)^2 \quad (4.64)$$

Donde Q_j^i es el escurrimiento total en el año j en el sitio i , q_i es el escurrimiento para cada mes en el año j en el sitio i , σ_Q^i es la desviación estándar de los registros históricos anuales para el sitio i .

4.8 Planeación y manejo de sistemas hídricos con series de tiempo

La planeación y manejo de sistemas hidráulicos en el contexto de esta investigación se refiere al desarrollo de una metodología que pueda ayudar en el proceso de la toma de decisiones, como una guía para operar esos sistemas en periodos de tiempo futuros en función del análisis de sus registros históricos. En el estudio se analizan las variables escurrimientos medios anuales en una cuenca.

Se revisó la literatura sobre procesos estocásticos, variables aleatorias y simulación hidrológica con series de tiempo, para los casos de modelos AMRA(p,q) (Auto Regressive Moving Average) a nivel anual y mensual univariados y multivariados. Otro tipo de modelos que se analizaron fueron los modelos de disgregación de datos. Estos últimos se consideran debido a que en años recientes los modelos de disgregación han sido usados con éxito en la generación de datos a periodos menores a partir de datos de periodos mayores, ejemplo, mensuales a partir de anuales que pueden auxiliar en la toma de decisiones en las políticas de operación en sistemas hidráulicos con varios embalses y en el manejo de agua a nivel de cuenca, Kumar *et al.* (2000).

4.8.1 Series de tiempo para el pronóstico y generación de escurrimientos

Enayetur y Hayat, (2000) desarrollaron un modelo de series de tiempo para generar series de tiempo sintéticas y hacer una simulación hacia el futuro con la finalidad de determinar la capacidad óptima de un embalse. El propósito del trabajo fue combinar la probabilidad de que el embalse se vaciara con la probabilidad de que el embalse vertiera por exceso. La simulación la hicieron para 100 años en el futuro, para esto seleccionaron un modelo del tipo cadenas de Markov en conjunto con la ecuación de continuidad aplicada al Río Mitta Mitta en Australia. Se analizaron las características estadísticas de los registros

históricos y se determinó que un modelo anual AR(1) log-normal se ajustaba bien a los registros de escurrimientos. Se utilizó un periodo de registro de datos de 34 años. Concluyeron que, para una vida económica del embalse de 34 años para la capacidad que se obtuvo asegura que la presa verterá por primera vez en los primeros 7 años y que en un periodo de vida útil de 34 años se vaciará únicamente dos veces y en total verterá 4 veces. Se simularon 1000 series de 50 hasta 100 años para obtener las diferentes probabilidades.

Voguel *et al.* (1996), diseñaron un modelo no paramétrico para el diseño de la capacidad de almacenamiento de un embalse. El modelo se usa como un modelo de series de tiempo univariado y multivariado. Utilizan un método de remuestreo (bootstrap) o muestro con reemplazo que consiste en generar grandes cantidades de valores de un registro histórico, lo que lleva a una función de probabilidad empírica de fenómenos meteorológicos e hidrológicos; esta es una técnica que genera funciones de probabilidad empíricas sin la necesidad de calcular parámetros ni hacer transformaciones a la normal u otras funciones, ya que permite adaptar una función de probabilidad empírica al fenómeno en estudio, la técnica es fácil de implementar en una hoja de cálculo. El reto fue muestrear con reemplazo varias veces los datos históricos de tal manera que se asegura que la estructura de covarianzas en el tiempo y en el espacio se conserve. Una ventaja de este método es que no se requieren registros históricos largos como es el caso de las series de tiempo paramétricas tradicionales.

La principal desventaja de la metodología es que no se pueden generar valores mayores o menores que los del registro original, el método es incapaz de generar por ejemplo sequías o periodos húmedos menores o mayores a las que existen en los registros. El método fue creado dada la creciente necesidad de operar sistemas complejos y un aumento en el uso de las series de tiempo para tal fin.

Finalmente, se comparan los dos métodos el paramétrico y el no paramétrico. se monto un experimento Monte Carlo para generar 1000 estimaciones con cada uno de los modelos y se calcularon los percentiles 5, 25, 50 75 y 95%.

Los resultados del experimento muestran que el coeficiente de sesgo y el error medio cuadrático obtenido con el modelo no paramétrico, son menores que el modelo paramétrico; pero cuando los registros tienen una mayor correlación, el modelo no paramétrico muestra un coeficiente de sesgo menor que el modelo paramétrico, pero aún el modelo no paramétrico muestra un error medio cuadrático muy parecido al modelo paramétrico.

Rieu *et al.* (1999), utilizaron un modelo llamado modelo híbrido multivariado PAR(1)NT-hybrid para generar escurrimientos sintéticos en la cuenca del Río Han en Corea. Donde la legislación coreana establece que el balance hidrológico se tiene que realizar con modelos estocásticos cada 10 años para estimar la disponibilidad de agua en las cuencas. El objetivo del estudio fue comparar el modelo híbrido con un modelo PAR(1) para disgregación de datos.

La selección del modelo se hizo base en el modelo que mejor conserve las estructuras de correlación en el tiempo y en el espacio. Para ello utilizaron una combinación de modelos paramétricos y no paramétricos. El modelo paramétrico es un modelo tradicional ARMA(p,q) y el otro modelo usa una técnica para generar funciones de probabilidad empíricas. Los autores comentan que una de las ventajas de usar esta técnica es que no es necesaria la transformación para cumplir con la normalidad de los registros históricos, ya que el modelo analiza los residuos generados y adapta una función de probabilidad a estos.

El modelo se mueve a través de bloques “bootstrap” como el componente no paramétrico. El modelo está compuesto por la primera parte de un modelo tradicional, sin embargo la parte de los residuos es en donde incluye la parte de

un modelo no paramétrico, es aquí en donde se usa la técnica “bootstrap” para genera una función de probabilidad para los errores.

Finalmente se comparan los resultados obtenidos con tres modelos el modelo periódico multivariado MPAR(1), por sus siglas en inglés, el modelo MCPAR(1)-hybrid y un modelo para disgregación de datos. Para esto se generaron 300 series de datos, cada serie contiene 30. Los resultados muestran que el modelo híbrido conserva las estructuras de varianza y el coeficiente de sesgo mejor que los otros modelos, además también conserva los coeficientes de correlación cruzada a varios lags, mientras que los otros dos solamente conservan las estructuras lag-0 y lag-1.

Loucks *et al.* (1981) discuten diferentes modelos para la planeación de sistemas hidráulicos en los que abarca optimización y simulación para problemas de un embalse y sistemas con varios embalses con métodos determinísticos y estocásticos. La optimización es principalmente para la maximización de los ingresos netos en sistemas y en los modelos involucra almacenamientos y acuíferos. Los modelos en su mayoría son resueltos con programación lineal y programación dinámica. La optimización se realiza bajo restricciones de uso y tipo de suelo, agua superficial y subterránea y otros como semillas, fertilizantes, maquinaria y mano de obra.

En su texto, los autores hacen una breve introducción a los procesos estocásticos en la que se abordan diferentes métodos para plantear y resolver modelos de series de tiempo para una variable y proponen un modelo de disgregación de datos para modelos anuales. Es importante resaltar que no se hace ninguna propuesta para el análisis de residuos.

Watkins *et al.* (2000) desarrollaron un modelo para el manejo de los lagos de la cuenca alta del Río Colorado en el centro de Texas, USA. Los objetivos principales del modelo son dos: maximizar los ingresos netos esperados de la

venta de agua entregada de manera ininterrumpida para el consumo urbano; maximizar los beneficios de la recreación en los lagos.

El modelo lineal multi etapas es resuelto con el método de programación estocástica, éste ayuda en la toma de decisiones de cuánta agua vender a la ciudad de Austin para el próximo año en función del agua almacenada en los lagos en el año actual y los niveles esperados para el próximo año. Una de las principales ventajas de la programación estocástica es el hecho de que se pueden tomar decisiones sin el conocimiento de variables aleatorias en etapas futuras. Una desventaja del método de programación estocástica es que el modelo matemático puede resultar muy grande o puede requerir un algoritmo de solución especial.

Para la optimización se utilizaron intervalos de tiempo de un mes, dado que una de las características de esta técnica es que mantiene la correlación histórica espacial de los gastos de entrada. La correlación serial mensual es también conservada pero la correlación serial anual no, dado que las secuencias en cada etapa son muestreadas independientemente de las que las preceden en etapas anteriores.

El modelo resultante comprende más de 25,000 columnas 14,000 renglones y más de 70,000 elementos diferentes de cero. El problema fue resuelto utilizando los softwares GAMS/OSL.

Para evaluar el efecto de rechazar la correlación de los gastos de entrada anuales los resultados del modelo son comparados con una nueva técnica para la generación de escurrimientos, esta técnica intenta mantener la correlación de los gastos a través de las etapas en problemas multi etapas. Este nuevo algoritmo es desarrollado y aplicado para series de tiempo hidrológicas.

Draper (2001) desarrolló un modelo de optimización en donde modifica los modelos tradicionales de optimización estocástica implícita e incorpora la consideración del riesgo en la operación de embalses con reglas previamente establecidas.

Aplica la técnica para la optimización de cuatro embalses en el estado de California, USA, el lago Barryessa, el embalse Pine Flat, el embalse New Don Pedro y el lago McClure, aunque hace la optimización, primeramente, para cada embalse por separado con tres escenarios diferentes; *previsión perfecta*, *previsión limitada*, y *operación miope*, se hace el estudio para la proyección del funcionamiento de los almacenamientos durante 20 años. En el estudio de caso se realizó en los almacenamientos anteriores, con el objetivo de minimizar los costos causados por las sequías, es por eso que el estudio tiene un enfoque central en encontrar una función que se adapte mejor para pronosticar los gastos de entrada a cada embalse, ya que en tiempo de secas los volúmenes almacenados tendrán que estar en función de los escurrimientos de los próximos años, es así que se usan los enfoques de previsión mencionados anteriormente.

El autor utiliza multiplicadores de Lagrange para encontrar el valor del agua almacenada en cada uno de los embalses y así poder evaluar los daños económicos provocados por las sequías. Se usa un modelo de series de tiempo de Thomas-Fiering y Markov para representar la correlación serial entre los gastos de entrada mensuales; para tener mayor aproximación se discretizaron los gastos de entrada y de entrega.

Finalmente el autor concluyó que el escenario que mejor representa la serie de datos hidrológicos es la *previsión limitada*, esto dado que el modelo de optimización implícitamente estocástico se adapta mejor al periodo de análisis en que se desea realizar la optimización.

4.8.2 Modelos de disgregación

Tarboton *et al.* (1998), desarrollaron un modelo de disgregación el cual hace la división de escurrimientos en el espacio y en el tiempo. La principal ventaja del modelo, es la capacidad de conservar las características estadísticas de las series originales en más de una escala, datos semestrales, bimestrales, mensuales, etc. Los autores usan una técnica de modelación no paramétrica y argumentan que el modelo tiene la capacidad de generar datos sintéticos sin la necesidad de asumir o transformar los datos en una función de probabilidad normal. El modelo funciona con base en una estimación de una función de densidad no paramétrica empírica.

El método conserva una de las propiedades más importantes de los modelos de disgregación que es la aditividad, esto es, que la suma de los meses generados debe ser igual al escurrimiento generado anualmente.

El modelo es aplicado al Río San Juan en el estado de Nuevo México y muestran como el procedimiento de simulación condicionado conserva una variedad de atributos estadísticos.

La principal desventaja que muestra la estimación de densidad no paramétrica es que es aplicable solo a una variable, es decir, que no se aplica como modelo multivariado, lo que implica que si se deseará hacer la simulación a nivel regional, se tendría que analizar una a una los escurrimientos de cada río por separado.

Kumar *et al.* (2000), usaron un modelo de disgregación en combinación con un modelo de optimización para hacer la división de escurrimientos mensuales a diarios. Con este tipo de modelación se conservan los atributos estadísticos de las series de tiempo en múltiples sitios y en más de una escala.

Una de las principales contribuciones de su investigación es la habilidad del modelo para producir escurrimientos diarios con las propiedades estadísticas de los datos históricos a nivel mensual y diario debido a que la dimensionalidad de los modelos aumenta conforme disminuye la escala de tiempo. Es decir que, la cantidad de datos que se requiere generar para obtener escurrimientos a nivel diario es de una magnitud muy grande y esto aumenta si se desea hacer al mismo tiempo una disgregación espacial y temporal como es el caso de esta investigación.

El modelo de optimización encuentra los escurrimientos a nivel diario y se obtienen mes por mes debido a la dimensionalidad del modelo. El modelo es no paramétrico, es decir que, no se hacen transformaciones a la normal ni se supone a priori una función de distribución de probabilidad de los escurrimientos históricos.

La investigación se aplicó en el Río San Juan en el estado de Colorado, USA y se hizo con la finalidad de hacer una planeación diaria a largo plazo. Se compararon los resultados con un modelo paramétrico y con las propiedades estadísticas de los registros históricos a nivel diario y mensual.

Maheepala y Perera (1996) desarrollaron un modelo de disgregación temporal en el cual utilizaron “fragmentos mensuales”. Los fragmentos mensuales son una relación del escurrimiento mensual dividido en el total anual. De esta manera los autores generan escurrimientos sintéticos mensuales y después anuales, los primeros se utilizan como los “fragmentos” que en total suman el escurrimiento anual. Los fragmentos utilizados como válidos son aquellos que su suma anual sea cercana a un año dado generado previamente con un modelo anual.

El modelo se construyó para la simulación de los Ríos Murray, Mitta, Loddon Hunter, Goulburn en una primera fase de análisis como modelo univariado, después se utilizó como modelo multivariado en los Ríos Kiewa, Dartmouth y Hume en Australia.

Se compararon los datos con los registros históricos, y se encontró que el modelo conserva las propiedades estadísticas de los registros históricos como son la media, desviación estándar y el coeficiente de auto correlación en ambas modalidades univariado y multivariado. El modelo de disgregación fue utilizado en el dominio del tiempo.

4.9 Conclusiones de la revisión de literatura

De la revisión de literatura que se realizó se desprende que, los modelos de series de tiempo hidrológicos, han tomado una gran importancia, esto se debe principalmente a que, con estos modelos se puede estudiar la variabilidad que existe en la naturaleza. En algunos países del mundo es por ley que los balances hidrológicos se deban realizar con análisis de series de tiempo (modelos ARMA y/o modelos no paramétricos). En los últimos años en el mundo se han desarrollado nuevas propuestas de modelos de series de tiempo; estos presentan algunas variantes de la metodología para resolver un modelo ARMA(p,q) con respecto a los otros como es el caso de los modelos no paramétricos, que obtienen funciones de probabilidad empíricas que se adaptan a la cantidad de datos existentes.

Aunque aún no se han desarrollado una metodología general para el caso de modelos multivariados no paramétricos, parece que tienen un futuro promisorio ya que estos modelos no requieren grandes cantidades de registros históricos, como es el caso de los modelos paramétricos.

Una de las causas que han coadyuvado a que los modelos para series sintéticas tomen un importante auge es debido al desarrollo de las computadoras personales, lo que ha permitido hacer simulaciones de grandes cantidades de información en intervalos de tiempo “cortos”.

Capítulo V

Materiales y Métodos

*“Los hombres se distinguen menos
por sus cualidades que por la cultura
que ellos mismos se proporcionan”*

Confucio

En el presente capítulo se desarrolla la metodología para la generación de datos de escurrimientos sintéticos a través de modelos de series de tiempo en la cuenca del Río Amajac, Hidalgo, México. Se detallan los métodos en lo referente al análisis de datos y la forma de generar series sintéticas a partir de registros históricos, con la finalidad de auxiliar en la planeación de los recursos hidráulicos a nivel de cuenca, operación de embalses y cuerpos de agua. En la figura 5.1 se muestra un diagrama de flujo del desarrollo de la metodología.

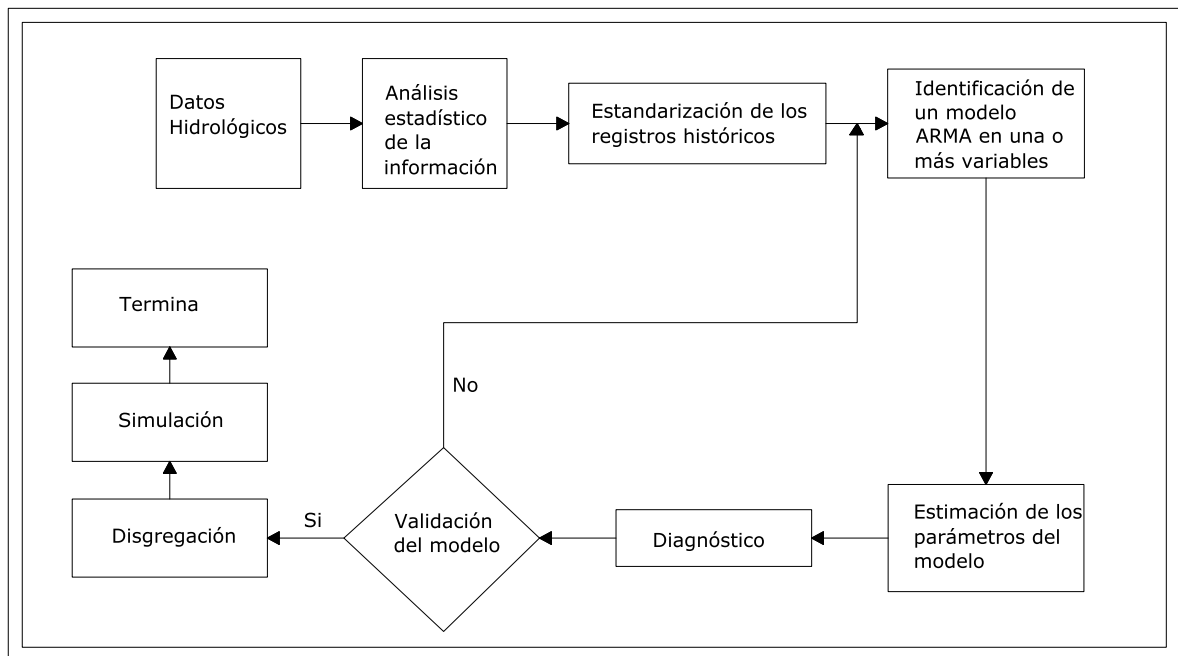


Figura 5.1. Diagrama de flujo de la metodología para la generación de series de tiempo sintéticas.

5.1 Descripción de la zona de estudio

El territorio mexicano, igual que en el resto del mundo, está dividido en múltiples cuencas hidrográficas, algunas de las más importantes cuencas corresponden a los grandes ríos nacionales como el Lerma, el Santiago, el Balsas, el Bravo, el Panuco, el Papaloapan, el Coatzacoalcos, el Grijalva, el Usumacinta, el Mayo, el Yaqui y otros de menor tamaño.

5.1.1 Cuenca del Río Amajac

La Región Hidrológica Río del Panuco corresponde a la vertiente del Golfo de México y se considera como una de las más importantes del país, tanto por su superficie, que la ubica en el cuarto lugar nacional, como por el volumen de sus escurrimientos, que le otorgan el quinto lugar. Debido a su gran superficie, se dividió esta región en dos: Alto Panuco y Bajo Panuco.

La zona del Alto Panuco comprende las cuencas de los Ríos Tula y San Juan del Río, que son afluentes del Río Moctezuma; las cuencas Meztlán y Amajac, que originan el Río Amajac. La zona del Bajo Panuco comprende las cuencas de los Ríos Extoraz, Bajo Amajac, Tempoal, Moctezuma, Tampaón y Panuco.

El estado de Hidalgo se encuentra casi en su totalidad dentro de la cuenca del Río Amajac, con una superficie aproximada de 19,793 Km². Sólo una pequeña extensión de la porción oriental forma parte de la región Tuxpan-Nautla, la cual cubre un área de 1,111.52 Km². Las coordenadas geográficas que delimitan esta cuenca van de los 19° 46' a 21° 25' de Latitud Norte y 98° 01' a 99° 26' de Longitud Oeste, figura 5.2.

El Río Amajac tiene sus orígenes en la parte sureste de la cuenca del Alto Panuco, en los límites de los Estados de Hidalgo y Puebla, en la zona del

parteaguas con el Río Tecolutla de la región hidrológica No. 27, a una elevación de 2950 m.s.n.m., en el cerro Tlacholoya. En el estado de Puebla donde nace con el nombre de Río San Lorenzo, se alimenta por pequeños gastos provenientes principalmente de manantiales.

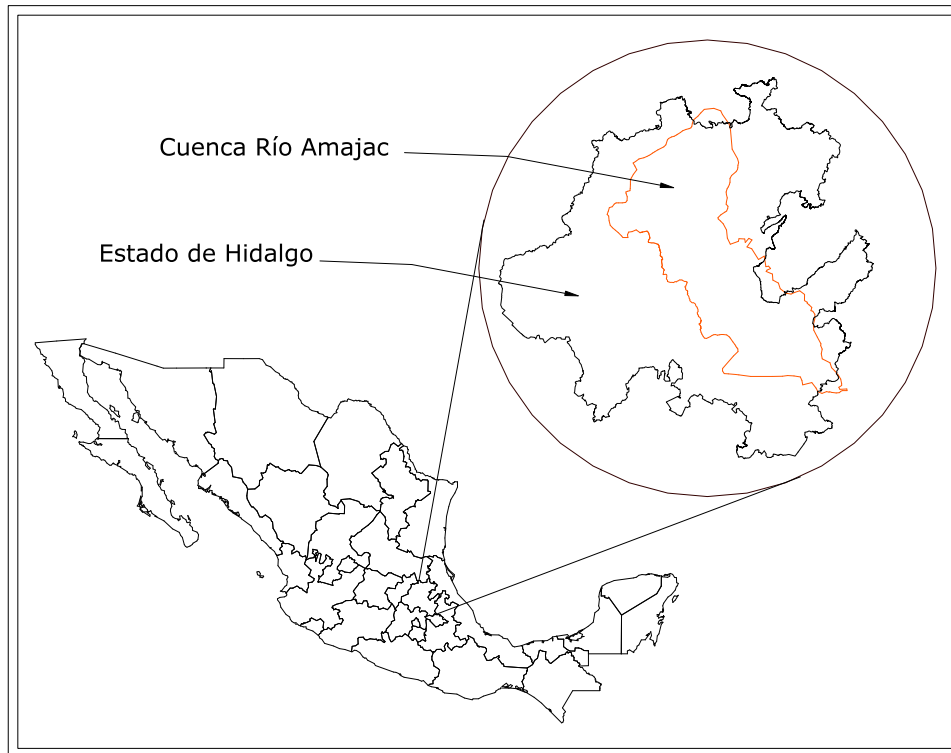


Figura 5.2. Ubicación de la cuenca del Río Amajac.

En sus orígenes sigue un rumbo norponiente con el nombre de Río Tlacholoya, cambiando al de arroyo Casa Blanca; pasa al poniente de la ciudad de Tulancingo Hidalgo, en donde recibe por margen derecha, a los 2200 m.s.n.m., las aportaciones del Río Tulancingo. Aguas abajo de esta confluencia se le conoce como Río Grande de Tulancingo.

A 17 Km. aguas abajo de la ciudad de Tulancingo, a 2000 m. de elevación, la corriente penetra a un estrecho y profundo cañón, siendo el resto de la cuenca de una topografía sumamente accidentada. Pasa por las poblaciones de San

Sebastián, Venados, Mezquititlán y Meztitlán, Hgo., y descarga sus aguas en la laguna de Meztitlán a una elevación de 1100 m.s.n.m.

La laguna de Meztitlán se encuentra aproximadamente 15 Km. al noroeste de la cabecera municipal de Meztitlán. Sus coordenadas geográficas son: Latitud Norte: 20° 41' 00", Longitud Oeste: 98° 52' 00", y una Altitud de 1,300 msnm.

Los excedentes de la laguna son descargados al Río Almolón a través de dos túneles; este Río, a su vez forma un solo caudal con las descargas del Río Amajaque, que desembocan en el Río Amajac, y continua con dirección norte hacia los límites de los Estados de Hidalgo y San Luis Potosí, en donde cambia su nombre a Río Moctezuma.

5.2 Clima

El tipo de clima de la región en la Subcuenca de Tulancingo, según la clasificación de Copen, citado por García, (1987), es una zona fría a semifría. La temperatura media anual es de 13°C a 17°C y la precipitación pluvial media anual es de 702 a 1,102 mm.

La Subcuenca del Amajac, según la misma clasificación, es una zona templada, con una temperatura media anual de 17°C y precipitación pluvial media anual de 514 mm. La Subcuenca del Amajaque se clasifica como una zona cálida, con una temperatura media anual de 22°C y precipitación pluvial media anual de 1,852 mm.

5.3 Hidrometría de la cuenca el Río Amajac

Desde principios de los años 30's se establecieron diferentes estaciones hidrométricas en la cuenca, en la figura 5.3 se muestra su localización.

En el extremo sur de la cuenca, en los límites con el estado de Puebla donde nace el Río San Lorenzo y se une con el Río Grande de Tulancingo, está la estación "Presa La esperanza", que se ubica aguas abajo de la Presa en donde se unen las descargas de la presa y el Río San Lorenzo; la estación tiene registros hidrométricos que datan desde el año de 1963 a 2004, ver cuadro 5.1.

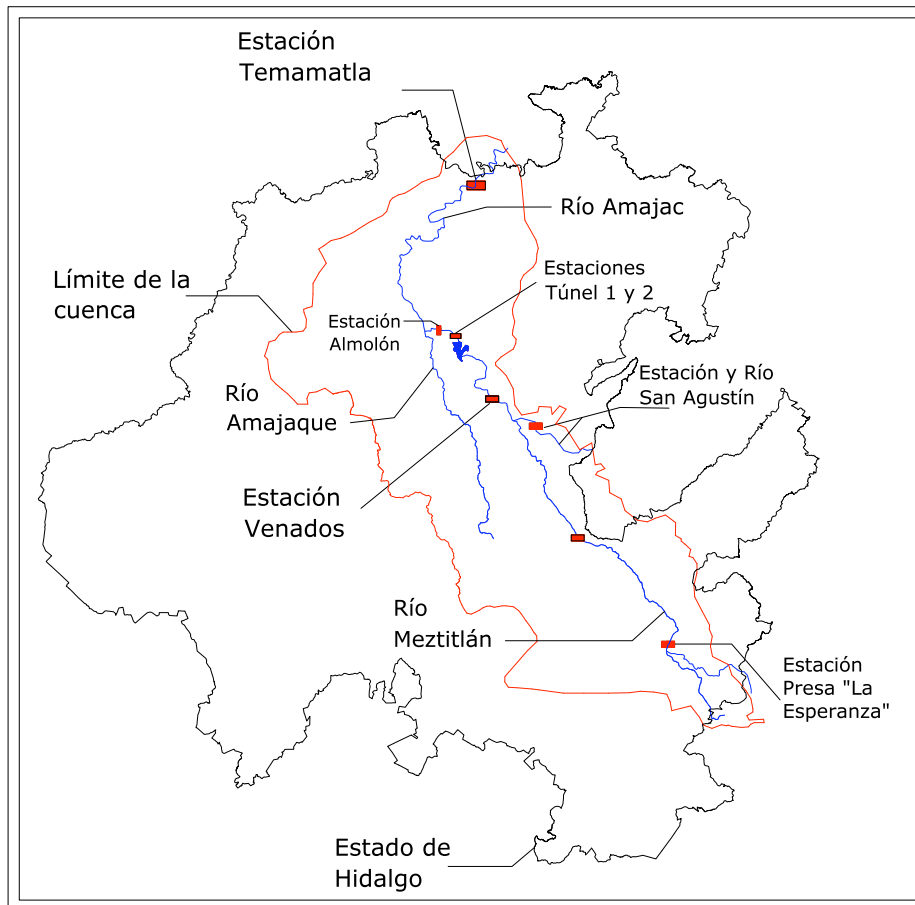


Figura 5.3. Estaciones hidrométricas en la cuenca del Río Amajac.

Cuadro 5.1. Volúmenes medios anuales registrados en la estación hidrométrica “Presa la Esperanza.

Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)
1964	2,616	1975	15,170	1986	5,650	1997	5,884
1965	7,420	1976	16,792	1987	4,031	1998	6,902
1966	5,416	1977	6,872	1988	7,348	1999	29,041
1967	5,700	1978	4,439	1989	8,401	2000	49,201
1968	4,744	1979	9,631	1990	6,520	2001	5,895
1969	15,476	1980	5,935	1991	6,736	2002	6,156
1970	4,848	1981	14,586	1992	11,554	2003	3,766
1971	5,318	1982	4,064	1993	13,357	2004	9,006
1972	9,267	1983	824	1994	7,460	$\bar{V} =$	9,445
1973	13,293	1984	10,511	1995	8,867	$S_V =$	8,134
1974	16,007	1985	6,892	1996	5,681	$\hat{r}_k =$	0.303

Con rumbo norte, el Río pasa por la población de Alcholoaya, a partir del poblado el Vite, se le conoce propiamente como Río Meztlán. En esta zona recibe por la margen derecha al arroyo Las Granadas y por la margen izquierda al arroyo de Regla, donde aproximadamente a 25 kilómetros aguas abajo cruza por debajo del puente Venados en la carretera Federal 105 Pachuca-Huejutla, en este puente, la Gerencia de Aguas Superficiales e Ingeniería de Ríos de la CNA, tiene instalada la estación hidrométrica “Venados” donde se miden también azolves. En el cuadro 5.2 se muestran los registros históricos medios anuales de la estación Venados.

Cuadro 5.2. Volúmenes medios anuales registrados en la estación hidrométrica “Venados”.

Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)
1937	143,522	1955	665,130	1973	204,001	1991	107,910
1938	107,608	1956	239,782	1974	327,325	1992	133,952
1939	151,964	1957	81,375	1975	276,189	1993	150,153
1940	76,004	1958	293,175	1976	182,573	1994	124,212
1941	242,297	1959	169,404	1977	75,755	1995	94,489
1942	231,323	1960	95,529	1978	154,641	1996	80,463
1943	98,323	1961	104,136	1979	159,131	1997	83,772
1944	537,773	1962	66,224	1980	110,511	1998	175,597
1945	165,370	1963	142,989	1981	279,660	1999	406,913
1946	162,406	1964	70,053	1982	90,667	2000	532,439
1947	121,500	1965	118,386	1983	83,558	2001	118,567
1948	99,589	1966	141,115	1984	248,507	2002	120,391
1949	86,715	1967	121,894	1985	169,614	2003	81,489
1950	84,387	1968	132,084	1986	80,004	2004	128,606
1951	87,186	1969	276,421	1987	72,411	$\bar{V} =$	169,537
1952	266,992	1970	166,319	1988	163,561	$S_V =$	115,507
1953	121,046	1971	126,790	1989	127,289	$\hat{r}_k =$	0.239
1954	331,421	1972	118,441	1990	172,331		

Antes de descargar a la laguna de Meztitlán, se une con las aguas del Río San Agustín o Mezquititlán en donde se ubica la estación hidrométrica que tiene el mismo nombre. Los volúmenes medios anuales del Río San Agustín se muestran en el cuadro 5.3 Finalmente el Río descarga en la Laguna de Meztitlán aguas abajo del poblado Vega de Meztitlán.

Cuadro 5.3. Volúmenes medios anuales registrados en la estación hidrométrica “San Agustín”.

Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)
1961	22,547	1973	34,597	1985	25,345	1997	15,520
1962	18,022	1974	81,419	1986	12,228	1998	24,154
1963	16,951	1975	52,747	1987	19,861	1999	101,891
1964	17,609	1976	40,157	1988	54,494	2000	122,591
1965	31,667	1977	16,774	1989	17,388	2001	34,365
1966	20,078	1978	18,890	1990	57,411	2002	31,548
1967	31,137	1979	27,442	1991	36,123	2003	16,094
1968	23,244	1980	13,399	1992	44,617	2004	27,728
1969	32,330	1981	76,242	1993	67,667	$\bar{V} =$	33,946
1970	30,188	1982	19,940	1994	33,653	$S_V =$	24,221
1971	17,114	1983	13,811	1995	16,451	$\hat{r}_k =$	0.299
1972	21,763	1984	45,591	1996	10,847		

Los excedentes de la laguna de Meztlán son descargados por dos túneles, el primero se localiza a una altura de 1,232.7 y el segundo a una elevación de 1244.7 m.s.n.m. Ambos tienen una sección tipo herradura.

Con las descargas de la laguna, el Río cambia de nombre a Río Almolón, este a su vez, se une con los escurrimientos provenientes de la cuenca del Río Amajaque, que nace en la ciudad de Pachuca.

Las descargas del Río Almolón y Amajaque reciben el nombre de Río Amajac, que continua su camino con dirección Norte hacia los límites de los estados de Hidalgo y San Luis Potosí, en donde se ubica la estación hidrométrica “Temamatla”, cuyos volúmenes medios anuales se muestran en el cuadro 5.4.

Cuadro 5.4. Volúmenes medios anuales registrados en la estación hidrométrica “Temamatla”.

Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)	Año	Volumen (10 ³ m ³)
1961	1,630,454	1973	1,705,268	1985	1,594,821	1997	744,945
1962	956,760	1974	2,603,078	1986	1,006,747	1998	2,196,262
1963	911,181	1975	2,749,546	1987	1,211,072	1999	2,244,254
1964	658,456	1976	2,371,100	1988	1,691,033	2000	1,231,918
1965	954,642	1977	741,867	1989	1,428,778	2001	1,173,091
1966	1,153,738	1978	1,909,952	1990	2,175,859	2002*	1,151,076
1967	1,396,813	1979	1,687,121	1991	1,924,470	2003*	759,467
1968	1,458,660	1980	1,128,568	1992	2,054,028	2004*	1,338,810
1969	2,062,820	1981	3,277,267	1993	2,592,879	$\bar{V} =$	156,9340
1970	1,590,402	1982	751,928	1994	1,132,198	$S_V =$	658,704
1971	1,314,588	1983	1,479,814	1995	1,261,740	$\hat{r}_k =$	0.107
1972	1,238,522	1984	3,263,748	1996	1,141,213		

* Valores generados con un modelo AR(1) anual.

Aunque existen más estaciones hidrométricas en la cuenca, no son consideradas en el estudio debido a que sus registros no están actualizados y/o tienen demasiados registros incompletos.

Los registros históricos mensuales correspondientes a los datos de las estaciones mostrados en los cuadros 5.1 al 5.4 se muestran en los cuadros A1.1 al A1.4 del Anexo I.

5.4 Modelos en series de tiempo para la cuenca del Río Amajac

En el presente capítulo se discute sobre el trabajo estadístico para la generación de datos en la hidrología estocástica. El tema central serán los métodos para generar datos sintéticos los cuales son necesarios para el análisis de políticas de operación y manejo de sistemas hidráulicos en el corto, mediano y largo plazo. En el capítulo se presenta la metodología para la generación de datos a partir de modelos ARMA y modelos de disgregación para una variable y para variables múltiples.

5.4.1 Modelo univariado para la estación hidrométrica Venados

Se desarrolló el método de Box y Jenkins para la construcción de los modelos de series de tiempo en una variable para los volúmenes medios anuales escurridos en la estación “Venados” en la cuenca del Río Amajac. Esta estación fue seleccionada principalmente por ser la estación cuya longitud de registros es mayor, de esta manera se tiene mayor certeza en la estimación de los diferentes parámetros de los modelos.

En el desarrollo del método de Box y Jenkins, se selecciona un modelo en particular a partir de datos históricos de una variable aleatoria, en este caso, una estación hidrométrica. El modelo será seleccionado a partir de una clase de modelos para series de tiempo estacionarias conocidos como ARMA(p,q).

En el caso de la modelación de series de tiempo anuales, es necesario que los registros hidrométricos estén estandarizados a media cero y desviación estándar uno, para esto se usa la ecuación (4.2) aplicada a los volúmenes medios anuales históricos en la estación hidrométrica.

La identificación de un modelo ARMA(p,q) tentativo que represente los datos históricos es seleccionado a través del análisis de las serie históricas, las herramientas que permite hacer la selección del (los) modelo (s) son las Funciones de Auto Correlación (FAC) y Auto Correlación Parcial (FACP).

Para el cálculo de la FAC se utiliza la ecuación:

$$r_k = \frac{\sum_{t=1}^{n-k} (z_{t+k} - \bar{z})(z_t - \bar{z})}{\sum_{t=1}^n (z_t - \bar{z})^2} \quad (5.1)$$

Donde r_k es el coeficiente de auto correlación de la serie histórica retrasada k periodos de tiempo; \bar{z} promedio de los datos históricos; n número total de observaciones, z_t es la observación en el tiempo t .

Para calcular la FACP se utiliza la siguiente ecuación:

$$r_{kk} = \begin{cases} r_1 & \text{si } k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} \end{cases} \quad (5.2)$$

Donde el coeficiente de auto correlación parcial $r_{kk} = r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j}$,
($k = 2,3,\dots; j = 1,2,\dots, k-1$)

Obtenidas las FAC y FACP se estiman los parámetros de los modelos que tentativamente se seleccionaron como candidatos para ser el modelo definitivo que represente los registros de la estación. En este caso, y con base en el

análisis de las FAC y FACP, como posibles candidatos, quedaron los modelos ARMA(1,0), ARMA(0,1) y ARMA(1,1).

Para estimar los valores de los parámetros $\hat{\phi}_p$ y $\hat{\theta}_q$ de los 3 modelos de series de tiempo se utilizó una metodología de optimización que se detalla a continuación:

a). Se proponen valores iniciales de $\hat{\phi}_p$ y/o $\hat{\theta}_q$ que se sustituyen en la siguiente ecuación:

$$\hat{z}_t = \hat{\phi}_1 z_{t-1} \dots - \hat{\phi}_p z_{t-p} + \hat{\theta}_1 a_{t-1} \dots - \hat{\theta}_q a_{t-q} \quad (5.3)$$

Con la ecuación (5.3) se genera una nueva serie para cada uno de los modelos tentativos;

b). Se calculan las diferencias o residuos entre la serie histórica y la serie generada con cada uno de los diferentes modelos:

$$a_t = (z_t - \hat{z}_t) \quad (5.4)$$

c). La función a minimizar es la sumatoria de las diferencias al cuadrado entre la serie histórica y la generada como se muestra:

$$\min \sum_{t=1}^N a_t^2 = \min \sum_{t=1}^N (z_t - \hat{\phi}_1 z_{t-1} - \dots - \hat{\phi}_p z_{t-p} + \hat{\theta}_1 a_{t-1} + \dots + \hat{\theta}_q a_{t-q})^2 \quad (5.5)$$

Sujeta a las condiciones:

$$|\phi_p| < 1; \text{ estacionalidad} \quad (5.6)$$

$$|\theta_q| < 1; \text{invertibilidad} \quad (5.7)$$

Se aplican las ecuaciones (4.18) y (4.19) junto con las ecuaciones (4.25) y (4.27) para hacer el diagnóstico de cada modelo candidato seleccionado en la etapa de identificación. En esta etapa se prueba la independencia estadística, o dicho de otra manera, se debe verificar que no exista correlación entre las diferencias o errores aleatorios de los valores de la serie histórica y las series generadas con los modelos ARMA(1,1), ARMA(1,0) y ARMA(0,1).

Finalmente, con el modelo que mejor representa los datos históricos de la estación hidrométrica Venados, se realiza el pronóstico. Este se realiza para corto, mediano y largo plazo con el objetivo de inferir el comportamiento de la variable en el futuro. En lo que refiere al presente estudio, el objetivo es inferir en el comportamiento de los volúmenes medios anuales escurridos en la estación hidrométrica Venados.

Para hacer el pronóstico a corto plazo se utilizan las ecuaciones (4.28) a la (4.31) con un intervalo de confianza previamente establecido, y en caso de que se quiera simular el comportamiento de una variable en el mediano y largo plazo, se generan series cuya longitud de tiempo será igual al periodo que se desea estudiar, y de esta manera se tienen diferentes escenarios del comportamiento de la variable.

5.4.2 Modelo AR(1) multivariado para las estaciones hidrométricas en la cuenca del Río Amajac

En el desarrollo del modelo de series de tiempo multivariado existen condiciones que se deben cumplir, tales condiciones son:

Los n sitios o variables deberán tener el mismo tamaño de la muestra, lo que implica que si existen datos faltantes en alguna de las series se deben completar con algún método.

Las series “cortas” se deberán extender para igualar el tamaño de las otras. Esta condición es necesaria para la estimación de los parámetros de las matrices del modelo. Además, se asume que los datos de cada una de las series están estandarizados, es decir, tienen media cero y desviación estándar uno.

El modelo AR(p) multivariado en general se escribe como:

$$Z_t = A_1 Z_{t-1} + B \underline{\varepsilon}_t \quad (5.8a)$$

Donde Z_{t-1} es un vector de $(n \times 1)$ elementos de $z_t^{(i)}$, donde $i = 1 \dots n$ variables que corresponden a diferentes sitios de la cuales A_1 y B son matrices de $(n \times n)$ parámetros; $\underline{\varepsilon}_t$ es un vector de $(n \times 1)$ variables aleatorias independientes, distribuidas normalmente con media cero y desviación estándar uno.

En forma matricial el modelo AR(1) mutivariado se presenta de la siguiente manera, Matalas (1967), citado por Salas *et al.* (1985):

$$\begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ z_t^{(n)} \end{bmatrix} = \begin{bmatrix} a^{11} & a^{12} & \cdot & \cdot & \cdot & a^{1n} \\ a^{21} & a^{22} & \cdot & \cdot & \cdot & a^{2n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ a^{n1} & a^{n2} & \cdot & \cdot & \cdot & a^{nn} \end{bmatrix} \begin{bmatrix} z_{t-1}^{(1)} \\ z_{t-1}^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ z_{t-1}^{(n)} \end{bmatrix} + \begin{bmatrix} b^{11} & b^{12} & \cdot & \cdot & \cdot & b^{1n} \\ b^{21} & b^{22} & \cdot & \cdot & \cdot & b^{2n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ b^{n1} & b^{n2} & \cdot & \cdot & \cdot & b^{nn} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_t^{(n)} \end{bmatrix} \quad (5.8b)$$

La estructura de correlación de z_t de la matriz A_1 en la ecuación (5.8) indica una correlación cruzada lag-0 o lag-1 en el espacio y una correlación lag-1 en el

tiempo. Esta correlación cruzada se obtiene a partir de las ecuaciones (4.32) a la (4.36).

Los parámetros del modelo se obtienen con las ecuaciones (4.46) y (4.47), para el proceso de estandarización se tiene que:

$$z_t^{(i)} = [y_t^{(i)} - \hat{\mu}^{(i)}] / \hat{\sigma}^{(i)} \quad i = 1, \dots, n \quad (5.9)$$

Para el modelo multivariado AR(1) de la ecuación (5.8) los estimadores de los momentos de los parámetros A_1 y B se obtienen con:

$$\hat{A}_1 = \hat{M}_1 \hat{M}_0^{-1} \quad (5.10)$$

y

$$\hat{B}\hat{B}^T = \hat{M}_0 - \hat{A}_1 \hat{M}_1^T = D \quad (5.11)$$

Donde \hat{M}_0 y \hat{M}_1 son las matrices de correlación lag-0 y lag-1 de Z_t , que se obtienen con el procedimiento de correlograma cruzado ecuaciones (4.32) a la (4.36) Kottegoda, (1980).

Para el cálculo de los coeficientes de la matriz B de la ecuación (5.9) se usa el procedimiento de Graybill (1969), citado por Kottegoda, (1980), ecuaciones (4.44) a la (4.49).

Para la prueba de bondad de ajuste del modelo multivariado se usa la ecuación (4.61) aplicada al análisis de los residuos de cada una de las series que componen el modelo.

5.4.3 Modelo de disgregación temporal para un sitio

Para el desarrollo de los modelos de disgregación de series de tiempo en una variable se seleccionaron las series de tiempo de los escurrimientos superficiales registrados en las estaciones hidrométricas Temamatla y Venados. El proceso de la obtención del modelo de disgregación para cada estación hidrométrica se describe a continuación.

1a. Con los registros que se muestran en los cuadros 5.3 y 5.4 para las estaciones Venados y Temamatla respectivamente, se construye un modelo anual AR(1) para cada estación siguiendo la metodología que se describe en apartado 5.4.1. Posteriormente se generan datos sintéticos para cada estación.

2a. De igual manera, con los registros históricos mensuales de los cuadros A1.1 y A1.2 se construye un modelo AR(1) para cada mes del año y para cada estación. Por lo que se tendrán 13 modelos AR(1) por estación hidrométrica, uno para cada mes y el modelo anual, como se muestra en el esquema del cuadro 5.5. Al igual que en el paso 1a, se deberán generar datos sintéticos mensuales con longitud igual al periodo que se desea simular para cada estación.

3a. Las series generadas en el paso 1a se disgregan utilizando los valores mensuales que se obtuvieron en el paso 2a. El criterio de selección de los escurrimientos mensuales q_i será con base en la siguiente ecuación:

$$\sum_{i=1}^{12} q_i = Q_j \quad (5.12)$$

Donde q_i es el escurrimiento para cada mes en el año j , i es el número de meses 1 = Enero, ..., 12 = Diciembre, Q_j es el escurrimiento total en el año j .

Para cualquiera de los casos de disgregación, en una o múltiples variables, la longitud de las series sintéticas deberá ser de tal tamaño que en ellas se reflejen los periodos húmedos y secos.

Cuadro 5.5. Esquema de obtención de modelos AR(1) elaborados por mes y año para las estaciones Venados y Temamatla en los periodos de 1964 a 2004.

Volúmenes medios mensuales 10^3 m^3					
Año	Enero	Febrero	...	Diciembre	Anual
1964	8,319	6,163	...	8,318	143,664
1965	4,441	3,575	...	7,804	107,703
1966	5,182	3,937	...	7,614	152,068
⋮	⋮	⋮	⋮	⋮	⋮
2004	3,733	2,388	...	4,716	128,606
	AR(1) Enero	AR(1) Febrero	...	AR(1) Diciembre	AR(1) anual

Los volúmenes mensuales generados sintéticamente en un año en particular deberán cumplir con la siguiente condición:

Los escurrimientos mensuales sintéticos q_i obtenidos en el paso 2a para un año en particular j , se seleccionan en función de la cercanía que tengan con el valor del total del escurrimiento sintético generado con el modelo AR(1) anual del paso 1a y con base en el valor mínimo que tenga la siguiente ecuación.

$$\left(\frac{\sum_{i=1}^{12} q_i - Q_j}{\sigma_Q} \right)^2 \quad (5.13)$$

Donde σ_Q es la desviación estándar de los registros históricos anuales

5.4.4 Modelo de disgregación temporal para sitios múltiples

Al igual que en los modelos de disgregación para una variable, los registros históricos de escurrimientos mensuales y anuales deberán estar disponibles para cada una de las estaciones para las que se vayan a generar datos sintéticos.

La obtención de volúmenes medios mensuales sintéticos con modelos de disgregación se describe a continuación:

1b. Para la generación de datos sintéticos con modelos de disgregación para variables múltiples, se deberá usar un modelo que conserve las estructuras de correlación en el espacio y en el tiempo. Para el caso de las estaciones hidrométricas de la cuenca del Río Amajac se utilizó el modelo de Matalas (1967). Con este modelo se deberán generar series de datos sintéticos anuales con las propiedades estadísticas de las series históricas, como se muestra en el apartado 5.4.2.

2b. Para generar series sintéticas mensuales para cada sitio, se deberá utilizar un modelo que permita conservar las propiedades estadísticas de los registros históricos en las series generadas sintéticamente. Para el caso de la cuenca del Río Amajac, se utilizó la metodología mostrada en el apartado 5.4.1.

Los valores sintéticos mensuales q_i se usarán para obtener los volúmenes anuales Q_j como en el caso de la disgregación para un sitio. En el cuadro 5.6 se muestra un esquema para la obtención de modelos anuales AR(1) para las estaciones de la cuenca del Río Amajac.

Cuadro 5.6. Esquema de obtención de modelos anuales AR(1) para las estaciones hidrométricas de la cuenca del Río Amajac.

Volúmenes medios anuales 10^3 m^3				
Año	Temamatla	Venados	San Agustín	P. Esperanza
1964	1,432,420	153,553	37,052	7,681
1965	2,685,853	224,593	36,850	10,027
1966	1,791,526	140,700	10,305	3,321
⋮	⋮	⋮	⋮	⋮
2004	1,434,372	137,405	23,826	11,668
	AR(1) ⁱ	AR(1) ⁱ	AR(1) ⁱ San	AR(1) ⁱ P.
	Temamatla	Venados	Agustín	Esperanza

3b. Las series generadas en el paso 1b se disgregan utilizando los valores mensuales que se obtuvieron en el paso 2b. El criterio de selección de los escurrimientos mensuales q_i será igual que en el modelo de disgregación para un sitio, ecuación (5.13), ahora aplicado a sitios múltiples.

$$\left(\frac{\sum_{k=1}^{12} q_k^i - Q_j^i}{\sigma_Q^i} \right)^2 \quad (5.14)$$

Donde Q_j^i es el escurrimiento total en el año j para el sitio i , q_i es el escurrimiento para los meses del año j para el sitio i , σ_Q^i es la desviación estándar de los registros históricos anuales para el sitio i .

5.4.5 Pruebas de bondad de ajuste para series sintéticas generadas con modelos de disgregación

Para determinar la relación que existe entre las series históricas y sintéticas generados con modelos de disgregación en una y múltiples variables, se utilizará la prueba de Kolmogorov-Smirnov (KS) para dos muestras independientes.

La prueba KS, permite detectar si existen discrepancias entre las frecuencias relativas acumuladas de las dos muestras objeto de estudio. Lo anterior propicia que este estadístico pueda advertir diferencias no sólo en los promedios, si no, que estas sean debido a la dispersión o la simetría.

Para aceptar las hipótesis de esta prueba, se debe concluir que sí:

$H = 0$; Las distribuciones de las poblaciones son iguales.

$H = 1$; Las distribuciones de las poblaciones son distintas.

Para las cuatro estaciones hidrométricas, se aplicará el estadístico de KS en las series históricas en los meses de Enero y Septiembre para el periodo de 1964 a 2004 en conjunto con series sintéticas generadas con modelos de disgregación para los mismos meses. Las series sintéticas deben tener la misma longitud de tiempo que las series históricas. Los meses de Enero y Septiembre se analizan por considerarse representativos de los periodos secos y húmedos en el año, respectivamente. En los cuadros A1.5 y A1.6 del Anexo 1, se muestran los conjuntos de series históricas y sintéticas a las que se aplicó el estadístico de KS.

Capítulo VI

Resultados y Discusión

*“El hombre encuentra a Dios detrás
de cada puerta que la ciencia logra
abrir”*

Albert Einstein

En este capítulo se presentan los resultados que se obtuvieron de la aplicación de las metodologías propuestas para generar datos sintéticos a través de modelos de series de tiempo y modelos de disgregación. Los modelos se aplicaron a los registros históricos anuales y mensuales de las estaciones hidrométricas Presa la Esperanza, Venados, San Agustín y Temamatla localizadas en Río Amajac, en el Estado de Hidalgo, México.

En el apartado 6.1, se presentan los resultados de la aplicación de la metodología para la construcción de un modelo univariado. El modelo de series de tiempo tipo ARMA(1,0) univariado se aplicó a los escurrimientos medios anuales observados en la estación hidrométrica Venados del Río Amajac.

Posteriormente, en el apartado 6.2, se muestran los resultados de la generación de datos sintéticos anuales con un modelo de series de tiempo multivariado, se construyó el modelo de Matalas, (1967), para cuatro estaciones hidrométricas de la cuenca del Río Amajac.

Los resultados de la aplicación de un modelo de generación de datos sintéticos a través de un modelo de disgregación para una variable se presentan en el apartado 6.3. El modelo de disgregación univariado se construyó para las estaciones Venados y Temamatla, la primera se consideró por tener el mayor periodo de registro, y la segunda por ser la estación más importante, pues se localiza en la salida de la cuenca.

La generación de datos sintéticos anuales y mensuales por medio de modelos de disgregación para sitios múltiples se presenta en el apartado 6.4. Se muestran los resultados de la aplicación de un modelo AR(1) para variables múltiples en conjunto con un modelo de disgregación.

6.1 Modelación de series de tiempo univariadas

Para el caso de la construcción de un modelo ARMA(p,q) univariado, el modelo se desarrolló para la estación hidrométrica Venados, ubicada aguas arriba de la laguna de Mezquitlán con coordenadas 20°28'00" N y 98°40'45" W, en dicha estación se tienen escurrimientos mensuales registrados desde el año 1937 a 2004, el hidrograma histórico de los volúmenes anuales para el mismo periodo se muestran en la figura 6.1.

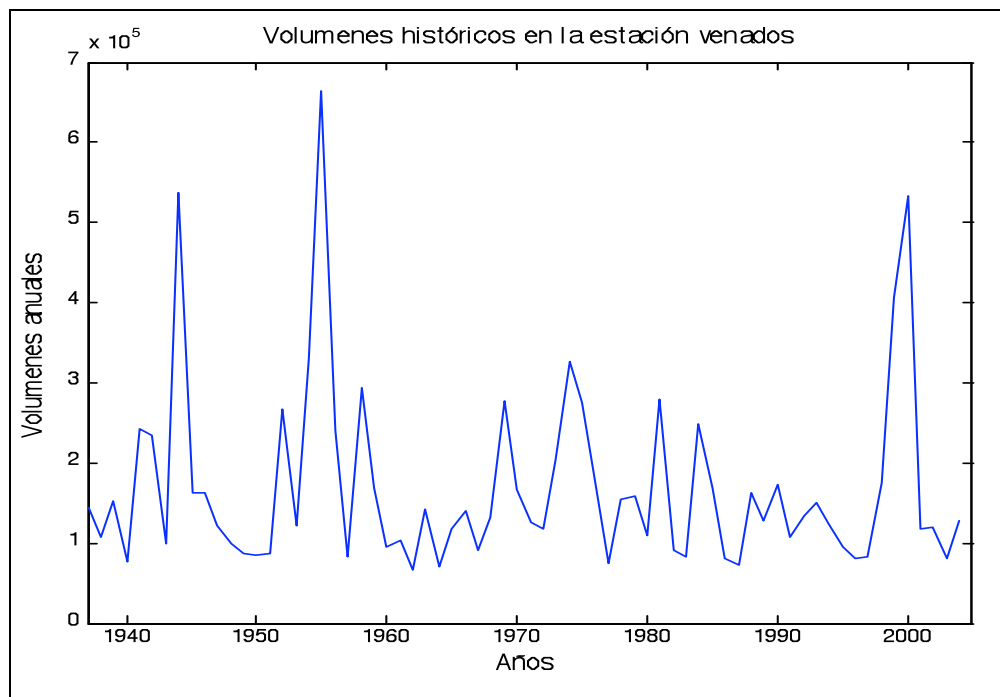


Figura 6.1 Hidrograma histórico de la información registrada en la estación hidrométrica Venados, Hidalgo, México.

En el cuadro 6.1 se muestran los datos de los volúmenes medios anuales históricos que se usaron para la construcción del modelo AR(1) en la estación hidrométrica Venados.

Cuadro 6.1. Escurremientos medios anuales históricos y estandarizados registrados de 1937 a 2004 en la estación hidrométrica Venados.

Año	Históricos (V_t) 10^3 m^3	Estandariz. (z_t)	Año	Históricos (V_t) 10^3 m^3	Estandariz. (z_t)	Año	Históricos (V_t) 10^3 m^3	Estandariz. (z_t)
1937	143,556	-0.2249	1960	95,529	-0.6408	1983	83,474	-0.7451
1938	107,608	-0.5362	1961	104,144	-0.5662	1984	248,408	0.6829
1939	151,964	-0.1521	1962	66,224	-0.8945	1985	169,506	-0.0003
1940	76,004	-0.8098	1963	142,989	-0.2299	1986	79,916	-0.7759
1941	242,301	0.6300	1964	70,138	-0.8606	1987	72,328	-0.8416
1942	234,323	0.5609	1965	118,386	-0.4429	1988	163,484	-0.0524
1943	98,323	-0.6166	1966	141,115	-0.2461	1989	127,205	-0.3665
1944	538,043	3.1905	1967	91,534	-0.6754	1990	172,235	0.0234
1945	162,370	-0.0621	1968	132,084	-0.3243	1991	107,825	-0.5343
1946	162,406	-0.0617	1969	276,421	0.9254	1992	133,850	-0.3090
1947	121,500	-0.4159	1970	166,319	-0.0279	1993	150,054	-0.1687
1948	99,589	-0.6056	1971	126,790	-0.3701	1994	124,118	-0.3932
1949	86,715	-0.7171	1972	118,441	-0.4424	1995	94,404	-0.6505
1950	84,387	-0.7372	1973	203,897	0.2975	1996	80,381	-0.7719
1951	87,186	-0.7130	1974	327,218	1.3652	1997	83,688	-0.7433
1952	266,992	0.8438	1975	276,078	0.9224	1998	175,532	0.0519
1953	121,046	-0.4198	1976	182,465	0.1119	1999	406,807	2.0543
1954	331,421	1.4016	1977	75,662	-0.8128	2000	532,322	3.1410
1955	665,130	4.2908	1978	154,540	-0.1298	2001	118,474	-0.4421
1956	239,782	0.6082	1979 \square	159,029	-0.0910	2002	120,303	-0.4263
1957	82,005	-0.7579	1980	110,412	-0.5119	2003	81,408	-0.7630
1958	292,875	1.0679	1981 \square	279,546	0.9525	2004	128,510	-0.3552
1959	169,224	-0.0027	1982	90,575	-0.6837			

\square Datos faltantes generados con un modelo AR(1) anual.

Para el desarrollo del modelo de series de tiempo de la estación hidrométrica se utilizó la metodología de Box-Jenkins como se describe a continuación:

6.1.1 Identificación del modelo

Para la identificación de un modelo se usaron las ecuaciones (5.1) y (5.2) se obtuvieron las Funciones de Auto Correlación (FAC) y Auto Correlación Parcial (FACP), Figuras 6.2 y 6.3, sus valores respectivos se muestran en los cuadros 6.2 y 6.3.

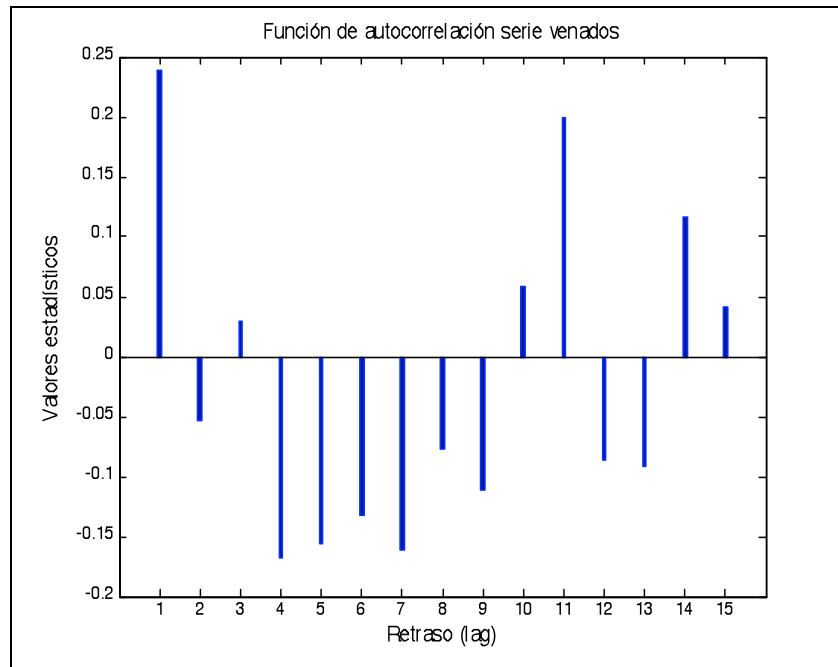


Figura 6.2. FAC de la estación Venados.

Cuadro 6.2. Valores de la FAC de la series Venados.

Retraso	1	2	3	4	5	6	7	8	9	10
r_k	0.24	-0.05	0.03	-0.17	-0.16	-0.13	-0.16	-0.08	-0.11	0.06

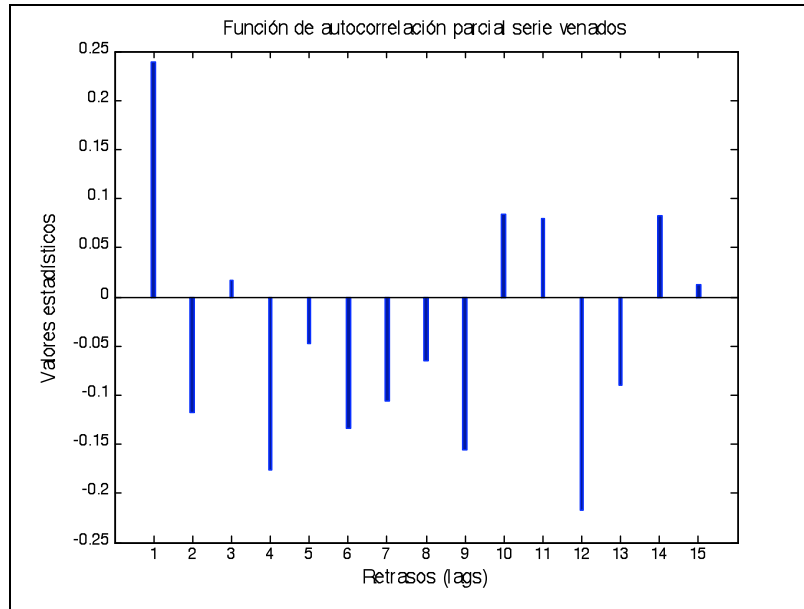


Figura 6.3. FACP de la estación Venados.

Cuadro 6.3. Valores de la FACP de la serie Venados.

Retraso	1	2	3	4	5	6	7	8	9	10
r_{kk}	0.24	-0.12	0.016	-0.18	-0.05	-0.13	-0.11	-0.06	-0.16	0.084

En el análisis de la FAC y FACP se puede observar que los valores de los primeros retrasos de las funciones caen repentinamente, cortan el cero y posteriormente se alternan en signo en diferentes retrasos. Este es el comportamiento típico de un proceso estacionario, Pankratz (1983) que se puede representar por un modelo ARMA(1,0), ARMA(1,1) o ARMA(0,1).

6.1.2 Estimación de los parámetros

El paso siguiente es estimar los parámetros $\hat{\phi}_{p,p}$ y $\hat{\theta}_{q,q}$ de los modelos identificados. Para la estimación de los parámetros de cada modelo se utilizó la

metodología de optimización propuesta en la sección 4.4.2, ecuaciones (4.20) a la (4.24). En el cuadro 6.4 se muestra la metodología para estimar el parámetro de un modelo AR(1) o ARMA(1,0). Los parámetros de los modelos ARMA(1,1) y ARMA(0,1) se estimaron de igual manera.

Cuadro 6.4. Estimación del parámetro $\hat{\phi}_1$ del modelo ARMA(1,0) con un valor inicial de 0.75 con los valores históricos estandarizados.

t Año	z_t	$\hat{z}_t = \hat{\phi}_1 z_{t-1}$	$\hat{a}_t = z_t - \hat{z}_t$	\hat{a}^2
	---	---	---	---
1937	-0.2249	---	-0.2249	0.0506
1938	-0.5362	-0.1687	-0.3675	0.1350
1939	-0.1521	-0.4021	0.2500	0.0625
⋮	⋮	⋮	⋮	⋮
2004	-0.3552	-0.5723	0.2171	0.0471

6.1.3 Diagnóstico

Se analizaron los residuos que se obtuvieron de la comparación entre los modelos y la serie histórica de la estación Venados. El análisis que se hizo a los residuos es principalmente para probar la independencia estadística de los mismos. La ecuación (4.25) permite probar la hipótesis de independencia. Con la ecuación (4.27) se hizo un análisis de correlación aplicado a los residuos, vistos como un conjunto.

En el cuadro 6.5 se muestran los resultados de las diferentes pruebas hechas a los modelos, los valores mínimos de significancia y los valores máximos que se permiten en las pruebas, en el cuadro 6.6 se muestran los modelos con sus respectivos parámetros.

Cuadro 6.5. Parámetros estadísticos, estimados para los modelos ARMA(1,0), ARMA(0,1) y ARMA(1,1) en la estación Venados.

Concepto	Modelo					
	ARMA(1,0)		ARMA(0,1)		ARMA(1,1)	
	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{\phi}_1$	$\hat{\theta}_1$
Parámetro estimado	0.239	0.0	0.0	-0.321	-0.386	-0.682
Error Estándar	0.12	0.0	0.0	0.12	0.24	0.27
Estadístico t	2.0	0.0	0.0	-2.75	-2.86	-1.30
Valores mínimos de t	1.96	0.0	0.0	1.96	1.96	1.96
Estadístico Q^*	4.69		4.32		3.72	
Valores máximos de Q^*	13.4		14.3		12.0	

Cuadro 6.6. Modelos ARMA(p,q) para la estación hidrométrica Venados

Modelo	Forma del modelo	$\hat{\phi}_p$	$\hat{\theta}_q$
ARMA(1,0)	$\hat{z}_t = \bar{z}(1 - \hat{\phi}_1) + \hat{\phi}_1 z_{t-1} + a_t$	0.239	---
ARMA(0,1)	$\hat{z}_t = \bar{z}(1 - \hat{\phi}_1) + a_t - \hat{\theta}_1 a_{t-1}$	---	- 0.321
ARMA(1,1)	$\hat{z}_t = \bar{z}(1 - \hat{\phi}_1) + \hat{\phi}_1 z_{t-1} - \hat{\theta}_1 a_{t-1} + a_t$	- 0.386	- 0.682

Los tres modelos pasan las pruebas de independencia estadísticas (Prueba Q^*), cuadro 6.5, pareciera que cualquiera de los tres modelos puede representar los datos históricos de la estación Venados, sin embargo, se tiene que seleccionar solamente uno de ellos. Según el principio de parsimonia, se selecciona el modelo que brinde buenos resultados y que tenga el mínimo de parámetros para estimar. Para esto se usa el estadístico t , que indica la significancia estadística de los parámetros. Para este fin, los valores de los parámetros de los modelos AMRA(1,1), ARMA(1,0) y ARMA(0,1) deberán ser significativos al 95% de probabilidad, o dicho de otra forma, los valores de los parámetros deberán tener un valor absoluto mayor a 1.96. Por lo anterior, el modelo ARMA(1,1) queda descartado, puesto que el valor del parámetro $\hat{\theta}_1$ no es significativo al 95% de probabilidad como se muestra en el cuadro 6.5.

El valor de la prueba para Q^* para los modelos ARMA(1,0) y ARMA(0,1) es menor para el modelo ARMA(1,0), por lo tanto el modelo seleccionado para representar los datos de volúmenes medios anuales de la estación hidrométrica Venados es el modelo ARMA(1,0) o AR(1). En la figura 6.4 se muestran los valores históricos y los estimados por el modelo AR(1) o ARMA(1,0).

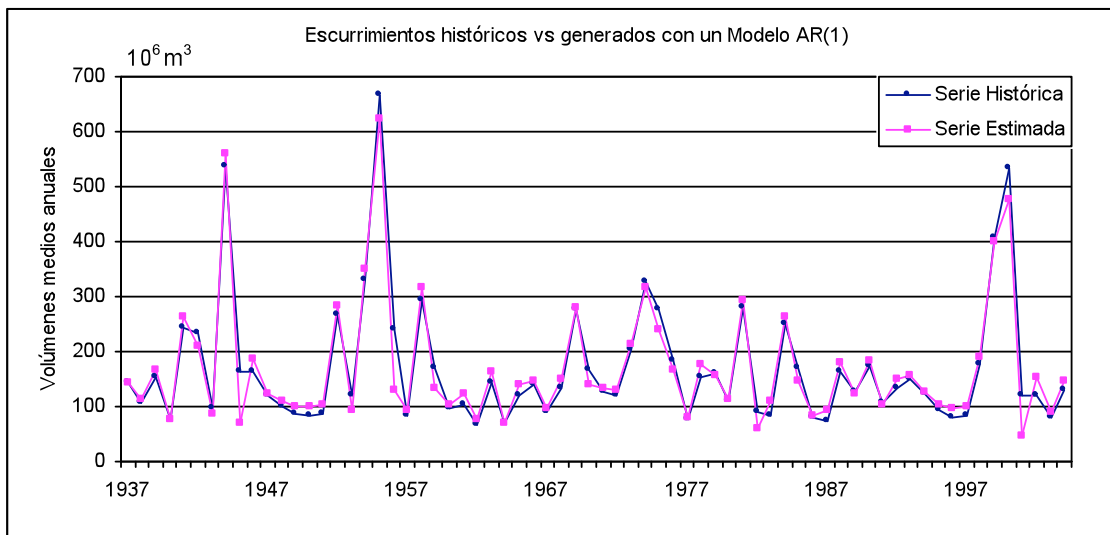


Figura 6.4. Series histórica y estimada con el modelo ARMA(1,0).

6.1.4 Pronóstico

En la última etapa de la modelación de series de tiempo se usa el modelo para hacer pronósticos a corto plazo, éste se hace para el periodo inmediato al último registro disponible. Esta etapa es la más importante para generar series de tiempo sintéticas con límites de confianza dentro de un rango de probabilidad previamente establecido. Esto ofrece información importante para el planificador, Salas *et al.* (1985).

El modelo tipo ARMA(1,0) para la estación hidrométrica Venados queda establecido de la siguiente manera y se especifica como:

$\hat{z}_t = \hat{\phi}_1 z_{t-1} + a_t$	(6.1)
ó	
$\hat{z}_t = 0.2397 z_{t-1} + a_t$	(6.2)

A partir de las ecuaciones (4.29), (4.30) y (4.31) y con un intervalo de confianza del 25% se pronostica para el siguiente periodo de tiempo un volumen medio anual dentro de los límites:

Inferior $127,145 \cdot 10^3 \text{ m}^3$

Superior $189,109 \cdot 10^3 \text{ m}^3$

Con la finalidad de comparar las propiedades estadísticas de los datos históricos contra los generados con el modelo ARMA(1,0) que se obtuvo para la estación Venados, se generaron series sintéticas con la misma longitud de la serie original (68 años). Los resultados de esta comparación se muestran en el cuadro 6.7 y en la figura 6.5 se presenta la gráfica de la serie sintética, en el cuadro 6.8

se muestran los datos anuales generados con el modelo AR(1) de la estación venados.

Cuadro 6.7. Comparación de las propiedades estadísticas de la serie histórica y la generada sintéticamente con el modelo ARMA(1,0).

Parámetro	Serie histórica	Serie sintética
Media (10^3 m^3)	169,537	169,120
Desviación Estándar (10^3 m^3)	115,500	93,713
Coefficiente de auto correlación	0.2397	0.377
Años	68	68

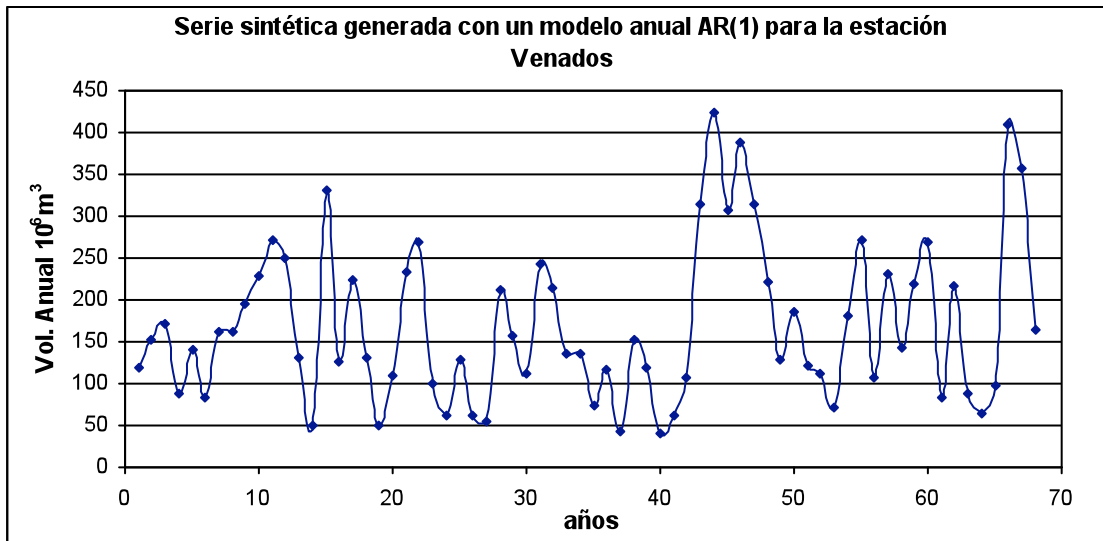


Figura 6.5. Serie sintética de 68 años de escurrimientos generada con el modelo anual AR(1) para la estación Venados.

Cuadro 6.8. Datos generados sintéticamente con el modelo AR(1) para la estación Venados.

Periodo t (años)	Generados (10 ³ m ³)	Periodo t (años)	Generados (10 ³ m ³)	Periodo t (años)	Generados (10 ³ m ³)	Periodo t (años)	Generados (10 ³ m ³)
1	118,499	19	50,855	37	42,109	55	272,161
2	152,912	20	108,545	38	152,859	56	107,019
3	171,037	21	232,331	39	119,653	57	230,358
4	89,272	22	268,157	40	40,303	58	142,398
5	140,965	23	100,256	41	62,317	59	219,580
6	84,240	24	61,203	42	107,106	60	268,461
7	162,171	25	128,547	43	313,780	61	82,467
8	162,263	26	62,801	44	423,417	62	217,243
9	195,226	27	55,257	45	306,687	63	87,239
10	228,256	28	212,187	46	387,894	64	64,937
11	270,881	29	156,259	47	314,184	65	98,703
12	249,728	30	112,114	48	221,132	66	410,627
13	131,968	31	242,910	49	128,875	67	358,144
14	49,306	32	214,027	50	185,358	68	163,280
15	330,231	33	134,689	51	120,695	$\bar{V} =$	169,120
16	126,030	34	136,238	52	111,887	$S_V =$	93,713
17	223,359	35	74,541	53	71,461	$\hat{r}_k =$	0.377
18	132,066	36	117,056	54	181,411		

6.2 Modelación de series de tiempo multivariadas

Las cuatros estaciones cuyos registros se utilizaron en este análisis fueron Temamatla, Venados, Río San Agustín y Presa la esperanza. Los datos fueron tomados de la gerencia del distrito de riego 028 Presa la esperanza y 008 Laguna de Meztlán. En los cuadros 5.1 al 5.4 se muestran los volúmenes medios anuales históricos en las estaciones, respectivamente.

Una de las condiciones para la construcción de un modelo de series de tiempo Multivariado es que las series en los n sitios sean iguales, para el caso de las cuatros estaciones hidrométricas en la cuenca del Río Amajac el periodo de tiempo en que coinciden es de los años de 1964 a 2004, excepto en la estación Temamatla en donde faltan los registro de los años 2003 y 2004, para esta estación en particular, se utilizó un modelo AR(1) anual univariado para extender su registro y un modelo de disgregación para complementar las series mensuales al 2004 y así uniformizar las series con las otras estaciones.

Otra condición, es que, cada una de las series históricas deberá estar estandarizada a media cero y desviación estándar uno. Para esto se realizó un proceso de normalización en las cuatro estaciones hidrométricas mediante el uso de la ecuación (5.9) aplicada a los logaritmos naturales de los registros históricos.

6.2.1 Modelo de series de tiempo multivariado temporal AR(1)

Para la obtención de las matrices de correlación M_0 y M_1 de las series históricas de los volúmenes de escurrimiento en las cuatro estaciones hidrométricas, se aplicaron las ecuaciones (4.32) a (4.36), y se obtuvieron las matrices que se muestran a continuación.

Para la matriz de correlación lag-0 se tiene:

$$M_0 = \begin{bmatrix} 1 & 0.6673 & 0.6234 & 0.4463 \\ 0.6673 & 1 & 0.8229 & 0.7848 \\ 0.6234 & 0.8229 & 1 & 0.7362 \\ 0.4463 & 0.7848 & 0.7362 & 1 \end{bmatrix}$$

Y para la matriz de correlación lag-1:

$$M_1 = \begin{bmatrix} 0.0492 & -0.0441 & -0.0967 & -0.2007 \\ 0.2173 & 0.2857 & 0.1480 & 0.1018 \\ 0.2672 & 0.2718 & 0.2222 & 0.1603 \\ 0.3811 & 0.3172 & 0.2226 & 0.2030 \end{bmatrix}$$

La matriz A de la ecuación (4.48) queda de la siguiente manera:

$$\hat{A} = \hat{M}_1 \hat{M}_0^{-1} = \begin{bmatrix} 0.1012 & 0.2767 & -0.1018 & -0.3881 \\ 0.0389 & 0.6347 & -0.2052 & -0.2626 \\ 0.1405 & 0.2683 & -0.0066 & -0.1081 \\ 0.3302 & 0.2616 & 0.1928 & -0.0077 \end{bmatrix}$$

La matriz de D de la ecuación (4.49) queda así:

$$D = \begin{bmatrix} 0.9195 & 0.6208 & 0.6060 & 0.4214 \\ 0.6208 & 0.8673 & 0.7277 & 0.6676 \\ 0.6060 & 0.7277 & 0.9083 & 0.6210 \\ 0.4214 & 0.6676 & 0.6210 & 0.8357 \end{bmatrix}$$

Con el desarrollo de las ecuaciones (4.50) a la (4.56), la matriz B queda de la siguiente manera:

$$B = \begin{bmatrix} 0.9589 & 0 & 0 & 0 \\ 0.6474 & 0.6995 & 0 & 0 \\ 0.6320 & 0.4758 & 0.8575 & 0 \\ 0.4395 & 0.5722 & 0.7178 & 0.6743 \end{bmatrix}$$

El modelo de la ecuación (5.8) para la cuenca del Río Amajac queda:

$$\begin{aligned}
 z_t^{(1)} &= a^{11} z_{t-1}^{(1)} + a^{12} z_{t-1}^{(2)} + a^{13} z_{t-1}^{(3)} + a^{14} z_{t-1}^{(4)} + b^{11} \varepsilon_t^{(1)} \\
 z_t^{(2)} &= a^{21} z_{t-1}^{(1)} + a^{22} z_{t-1}^{(2)} + a^{23} z_{t-1}^{(3)} + a^{24} z_{t-1}^{(4)} + b^{21} \varepsilon_t^{(1)} + b^{22} \varepsilon_t^{(2)} \\
 z_t^{(3)} &= a^{31} z_{t-1}^{(1)} + a^{32} z_{t-1}^{(2)} + a^{33} z_{t-1}^{(3)} + a^{34} z_{t-1}^{(4)} + b^{31} \varepsilon_t^{(1)} + b^{32} \varepsilon_t^{(2)} + b^{33} \varepsilon_t^{(3)} \\
 z_t^{(4)} &= a^{41} z_{t-1}^{(1)} + a^{42} z_{t-1}^{(2)} + a^{43} z_{t-1}^{(3)} + a^{44} z_{t-1}^{(4)} + b^{41} \varepsilon_t^{(1)} + b^{42} \varepsilon_t^{(2)} + b^{43} \varepsilon_t^{(3)} + b^{44} \varepsilon_t^{(4)}
 \end{aligned}
 \tag{6.3}$$

Sustituyendo los valores de las matrices A y B en el modelo AR(1) multivariado de la ecuación (5.8), el modelo multivariado queda como:

$$\begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \\ z_t^{(3)} \\ z_t^{(4)} \end{bmatrix} = \begin{bmatrix} 0.1012 & 0.2767 & -0.1018 & -0.3818 \\ 0.0389 & 0.6347 & -0.2052 & -0.2626 \\ 0.1405 & 0.2683 & -0.0066 & -0.1081 \\ 0.3302 & 0.2616 & 0.1928 & -0.0077 \end{bmatrix} \begin{bmatrix} z_{t-1}^{(1)} \\ z_{t-1}^{(2)} \\ z_{t-1}^{(3)} \\ z_{t-1}^{(4)} \end{bmatrix} + \begin{bmatrix} 0.9589 & 0 & 0 & 0 \\ 0.6474 & 0.6995 & 0 & 0 \\ 0.6320 & 0.4758 & 0.8574 & 0 \\ 0.4395 & 0.5722 & 0.7178 & 0.6743 \end{bmatrix} \begin{bmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \varepsilon_t^{(3)} \\ \varepsilon_t^{(4)} \end{bmatrix}
 \tag{6.4}$$

Con el modelo desarrollado para las cuatro series de tiempo de la cuenca del Río Amajac es posible generar datos de escurrimientos sintéticos para las cuatro estaciones hidrométricas con las propiedades estadísticas de las series originales en periodos de tiempo futuro que permitan hacer una planeación integral de los recursos hídricos de la cuenca del Río Amajac, en el cuadro 6.9 se muestran las series generadas con el modelo de la ecuación (6.3) para un periodo de tiempo de 41 años.

Cuadro 6.9. Series sintéticas anuales generadas con el modelo AR(1) multivariado para las estaciones de la cuenca del Río Amajac.

Periodo de tiempo	Estación hidrométrica			
	Temamatla	Venados	San Agustín	P. Esperanza
1	1,432,420	153,553	37,052	7,681
2	2,685,853	224,593	36,850	10,027
3	1,791,526	140,700	10,305	3,321
4	1,363,681	175,243	34,531	17,038
5	1,619,059	130,775	35,597	5,614
6	1,575,399	134,749	28,913	6,735
7	2,348,571	107,686	32,216	9,368
8	1,862,492	173,212	45,074	19,012
9	1,582,328	138,933	28,152	7,777
10	864,388	81,651	16,645	5,510
11	2,448,766	158,407	47,630	10,357
12	1,003,882	77,156	31,119	10,105
13	1,399,647	74,419	28,525	7,256
14	964,113	67,060	8,666	850
15	3,200,050	183,607	37,961	4,983
16	1,894,587	129,884	18,079	4,914
17	1,108,291	77,805	31,173	7,337
18	811,528	63,728	7,914	1,133
19	1,823,355	184,719	30,066	5,818
20	1,118,708	105,319	16,790	2,825
21	1,087,392	136,316	21,783	3,730
22	2,373,562	225,768	25,070	9,189
23	1,905,381	338,294	110,417	26,896
24	791,186	101,592	18,206	6,034
25	1,733,512	236,675	32,378	4,527
26	1,760,723	244,468	26,486	4,660
27	1,876,915	211,112	45,233	8,926
28	1,500,137	137,046	27,351	16,798
29	1,000,189	64,145	18,401	3,032
30	1,571,508	88,754	25,066	5,402
31	1,385,980	55,436	14,426	5,919
32	1,988,653	75,287	19,428	2,298
33	1,240,324	62,524	15,662	2,824
34	1,653,430	100,536	17,090	3,694
35	2,286,724	355,271	34,015	5,726
36	3,052,166	323,197	28,844	9,522
37	2,302,062	352,961	175,679	73,858
38	2,072,218	198,017	34,861	15,364
39	2,230,850	151,932	50,960	10,035
40	1,785,089	107,946	29,484	3,530
41	1,434,372	137,405	23,826	11,668
Promedio	1,705,635	153,363	33,120	9,300
Desv. Est.	581,174	81,593	28,184	11,583
Coef. r_k	0.042	0.465	-0.035	0.109

6.2.2 Análisis de residuos

Para realizar el análisis de bondad de ajuste del modelo multivariado se compararon los valores estimados de las series individuales con los datos de las series históricas.

Los residuos se obtienen al resolver el modelo de la ecuación (5.8) simultáneamente para las cuatro estaciones.

En forma matricial el modelo queda de la siguiente forma:

$$\begin{bmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \varepsilon_t^{(3)} \\ \varepsilon_t^{(4)} \end{bmatrix} = \begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \\ z_t^{(3)} \\ z_t^{(4)} \end{bmatrix} - \begin{bmatrix} 0.1012 & 0.2767 & -0.1018 & -0.3881 \\ 0.0389 & 0.6345 & -0.2052 & -0.2626 \\ 0.1405 & 0.2683 & -0.0066 & -0.1081 \\ 0.3302 & 0.2616 & 0.1928 & -0.0077 \end{bmatrix} \begin{bmatrix} z_{t-1}^{(1)} \\ z_{t-1}^{(2)} \\ z_{t-1}^{(3)} \\ z_{t-1}^{(4)} \end{bmatrix} \begin{bmatrix} 1.0429 & 0 & 0 & 0 \\ -0.0084 & 1.4937 & 0 & 0 \\ -0.2091 & -0.8288 & 1.1662 & 0 \\ 0.3986 & -0.3852 & -1.2414 & 1.483 \end{bmatrix}$$

En las figuras 6.6 a la 6.9 se muestran las gráficas de las series históricas y las estimadas con el modelo AR(1) multivariado. En el cuadro 6.10 se hace una comparación de las propiedades estadísticas de las series.

Cuadro 6.10. Comparación de los parámetros estadísticos de las series históricas y las series generadas con el modelo AR(1) multivariado.

Estación/parámetros	Temamatla	Venados	San Agustín	P. La esperanza
<i>Históricos</i>				
Media (10^6 m^3)	1,637,717	159,875	35,027	9,446
Desv. Est. (10^6 m^3)	647,626	95,556	24,752	8,134
r_k (Adimensional)	0.107	0.240	0.299	0.302
<i>Sintéticos</i>				
Media (10^6 m^3)	1,705,635	153,363	33,120	9,300
Desv. Est. (10^6 m^3)	581,174	81,593	28,184	11,583
r_k (Adimensional)	0.042	0.465	-0.035	0.109

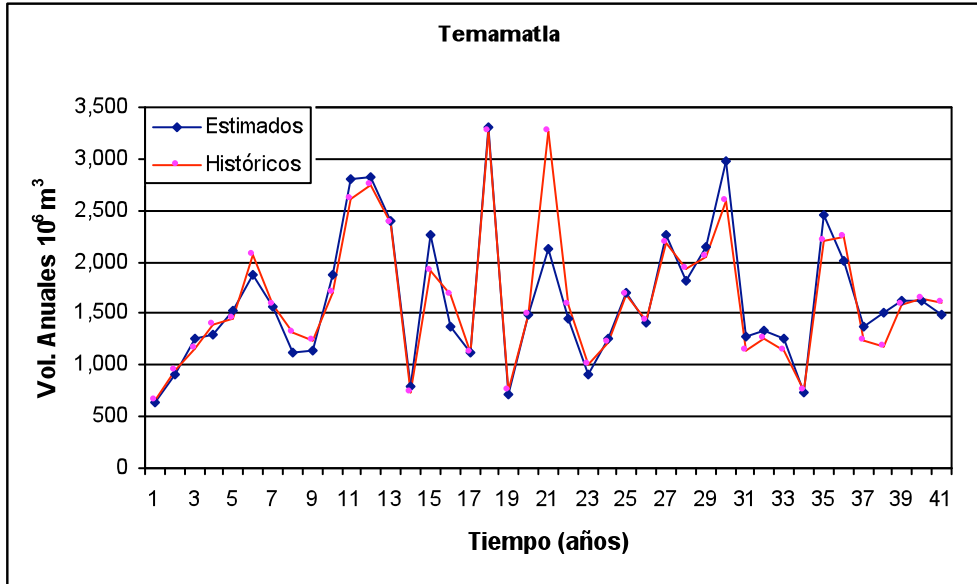


Figura 6.6. Ajuste de los datos históricos con los estimados estación Temamatla.

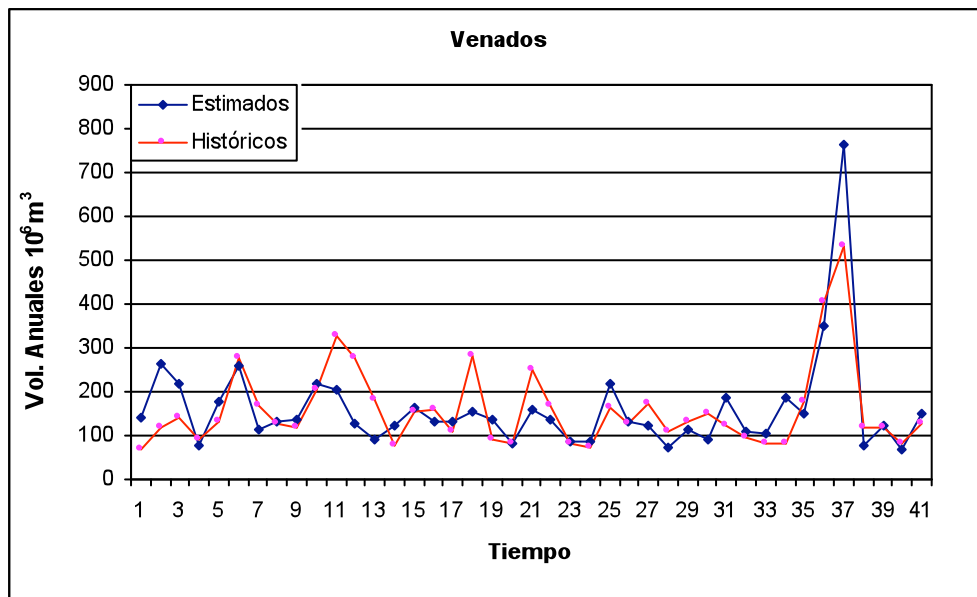


Figura 6.7. Comparativo en la estación Venados.

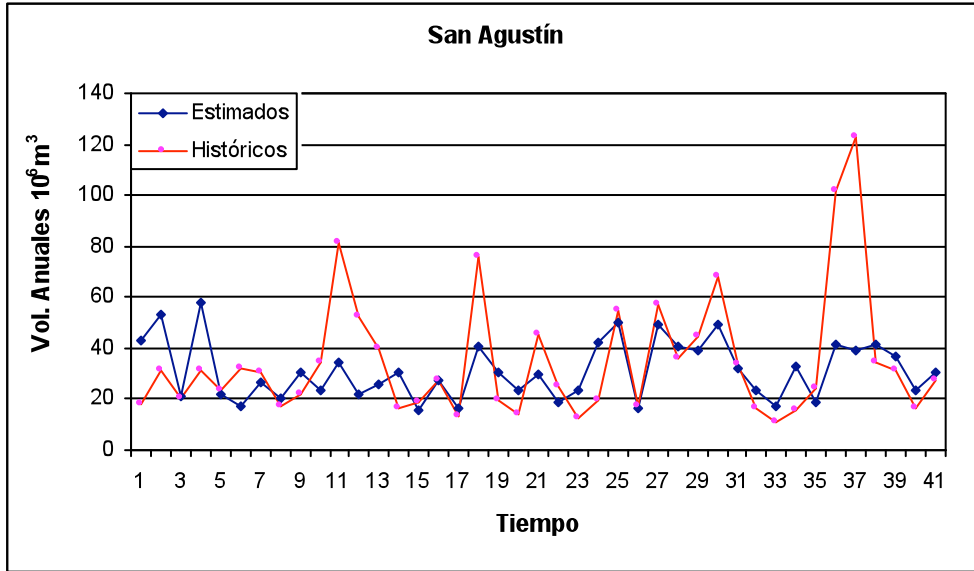


Figura 6.8. Comparativo en la estación San Agustín o Río Mezquitlán.

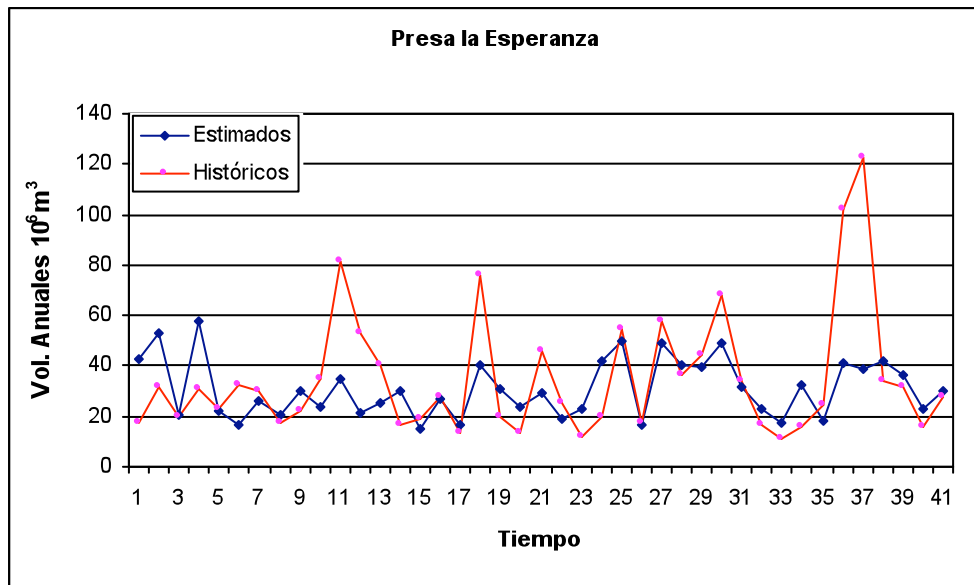


Figura 6.9. Comparativo en la estación hidrométrica de la Presa La Esperanza.

Modelos de disgregación temporal en la cuenca del Río Amajac

Para ilustrar el uso de los modelos de disgregación en la cuenca del Río Amajac se desarrollaron los modelos en series de tiempo univariadas y para variables múltiples. Para que sea posible la disgregación, los registros históricos de volúmenes anuales y mensuales deberán estar disponibles para los sitios que se quieran disgregar, en este caso, para las cuatro estaciones hidrométricas consideradas en este estudio.

6.3 Modelo de disgregación para un sitio

La disgregación o división de datos se realiza a partir de series de tiempo sintéticas que han sido previamente generadas, tales series se pueden generar mediante el uso de diferentes modelos. Para el caso del estudio de la cuenca del Río Amajac, los datos fueron generados mediante el uso de modelos ARMA(p,q) para el caso de una variable, y AR(1) para el caso de variables múltiples.

En el apartado 5.4.1 se detalló la metodología para generar series de tiempo sintéticas en una variable. Los datos mostrados en el cuadro 6.8 se generaron con un modelo anual AR(1) para la estación Venados, estos datos se utilizarán para ilustrar la construcción de un modelo de disgregación en una variable.

El objetivo principal de los modelos de disgregación es conservar las propiedades estadísticas de los registros históricos en las series generadas a ambos niveles, anual y mensual.

Al usar la serie anual generada sintéticamente mostrada en el cuadro 6.8 y utilizando los modelos AR(1) mensuales de la estación venados, se generaron datos sintéticos mensuales. El criterio de selección para un año en particular se hizo con base a la ecuación (5.13).

En el cuadro 6.11 se comparan algunas propiedades estadísticas a nivel anual y mensual entre las series históricas y las generadas con el modelo de disgregación para la estación Venados.

Cuadro 6.11. Valores medios históricos y sintéticos obtenidos con el modelo de disgregación para la estación Venados.

Mes	Media (10^6m^3)		Desv. Est. (10^6m^3)		r_k (Adim.)	
	Históricos	Sintéticos	Históricos	Sintéticos	Históricos	Sintéticos
Enero	6,106	6,061	2,459	2,797	0.343	-0.037
Febrero	4,356	4,055	1,818	1,789	0.325	0.119
Marzo	3,994	3,930	1,791	2,222	0.420	-0.046
Abril	4,225	3,963	2,217	2,705	0.145	-0.241
Mayo	5,430	7,017	3,171	6,531	0.202	0.000
Junio	12,102	12,362	11,789	15,362	-0.084	-0.018
Julio	17,654	13,855	16,485	16,292	-0.065	0.060
Agosto	15,882	14,563	13,686	12,710	-0.030	0.037
Septiembre	45,574	43,953	67,017	56,119	0.105	0.339
Octubre	35,652	41,309	56,169	54,210	0.289	0.229
Noviembre	10,826	10,513	6,914	6,041	0.043	-0.137
Diciembre	7,736	7,537	3,544	3,803	0.242	-0.081
Anual	169,537	169,120	115,507	93,713	0.239	0.377

Como se observa en el cuadro 6.11, las propiedades estadísticas de los registros históricos de la estación Venados son conservadas por el modelo de disgregación en una variable. En la Figura 6.10 se muestra la relación que existe entre los promedios mensuales de las series históricas y sintéticas.

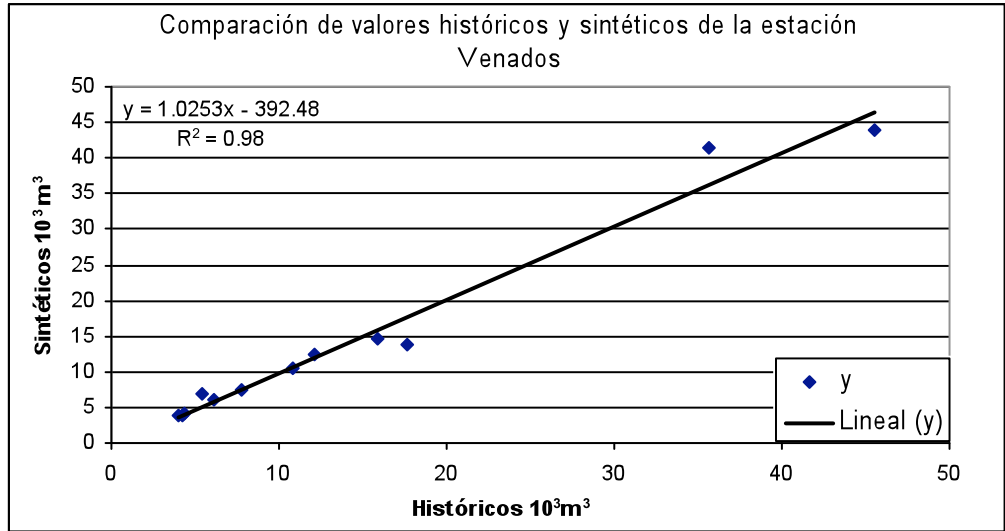


Figura 6.10. Correlación entre los promedios mensuales de los datos observados y sintéticos para la estación Venados.

En el cuadro 6.12 se muestran los promedios a nivel mensual y anual y se comparan las propiedades estadísticas de las series históricas y sintéticas para la estación Temamatla. En la Figura 6.11 se muestra la relación existente entre los promedios de las series mensuales históricas y simuladas.

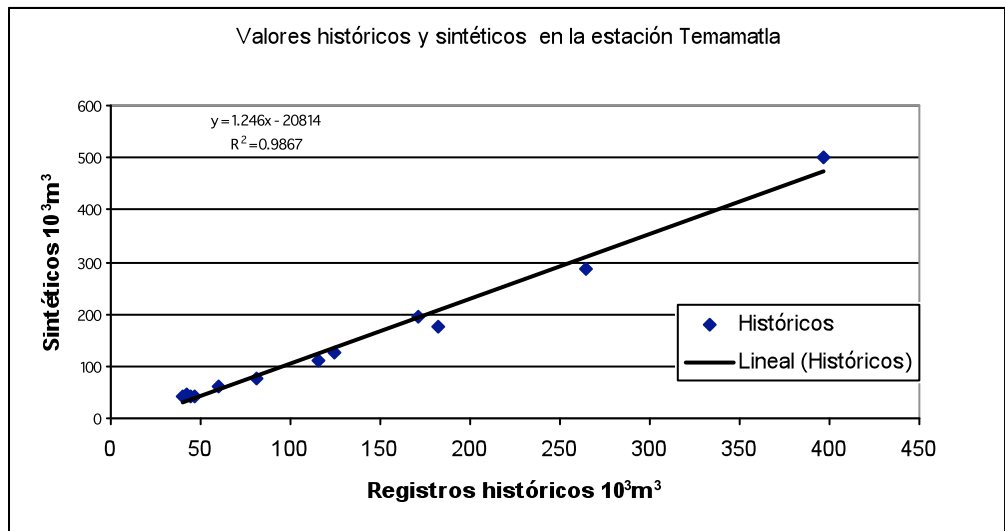


Figura 6.11. Prueba de correlación entre las medias de los datos observados y sintéticos para la estación Temamatla.

Cuadro 6.12. Valores medios históricos y sintéticos obtenidos con el modelo de disgregación para la estación Temamatla.

Mes	Media (10^6m^3)		Desv. Est. (10^6m^3)		r_k (Adim.)	
	Históricos	Sintéticos	Históricos	Sintéticos	Históricos	Sintéticos
Enero	60,231	59,266	23,153	24,915	0.273	-0.020
Febrero	46,152	43,794	16,893	15,717	0.158	-0.035
Marzo	42,767	44,653	10,617	12,285	0.166	0.315
Abril	39,778	41,122	14,133	15,523	0.216	-0.258
Mayo	44,845	42,170	17,503	17,215	0.017	0.023
Junio	115,418	110,991	109,381	72,852	-0.162	0.102
Julio	171,556	194,760	128,369	157,926	-0.115	0.130
Agosto	182,013	176,110	127,734	141,628	-0.265	-0.180
Septiembre	397,110	502,138	351,953	474,885	0.105	0.017
Octubre	264,368	288,523	174,703	199,334	0.256	0.149
Noviembre	123,961	125,070	65,939	66,312	0.172	0.115
Diciembre	81,141	77,038	35,806	31,528	0.120	0.253
Anual	1,569,340	1,705,635	658,704	581,174	0.107	-0.180

6.4 Modelo de disgregación para sitios múltiples

Al igual que los modelos de disgregación para un sitio, los datos que se desean disgregar, deben estar disponibles previamente; estos se debieron generar con un modelo multivariado que preserve las propiedades estadísticas de las series originales en el espacio y en el tiempo.

Los modelos para generar series sintéticas mensuales para cada mes y estación se construyeron como se indica en el apartado 5.4.1. La selección de los volúmenes mensuales para un año y una estación en particular se hizo con base en la ecuación (5.14).

Para la generación de series sintéticas anuales se utilizó el modelo AR(1) multivariado de Matalas (1967), citado por Salas *et al.* (1985), ecuación (5.8), aplicado en conjunto a las cuatro estaciones hidrométricas de la cuenca del Río Amajac.

Para ilustrar la aplicación del modelo de disgregación para sitios múltiples, se utilizaron los datos sintéticos generados previamente en el apartado 5.4.2, tales series anuales se muestran en el cuadro 6.9.

En los cuadros 6.13 al 6.16 se hace una comparación de las propiedades estadísticas de las series históricas y las series que se obtuvieron con el modelo de disgregación para sitios múltiples. Se comparan los promedios a nivel anual y mensual.

Cuadro 6.13. Valores medios históricos y sintéticos obtenidos con el modelo de disgregación multivariado AR(1) para la estación Temamatla.

Mes	Media (10^6m^3)		Desv. Est. (10^6m^3)		r_k (Adim.)	
	Históricos	Sintéticos	Históricos	Sintéticos	Históricos	Sintéticos
Enero	60,834	59,266	23,783	24,915	0.215	0.047
Febrero	46,610	43,794	17,303	15,717	0.091	-0.154
Marzo	43,300	44,653	10,750	12,285	0.112	0.027
Abril	40,285	41,122	14,005	15,523	-0.172	0.044
Mayo	45,727	42,170	17,779	17,215	-0.032	-0.140
Junio	114,480	110,991	108,782	72,852	-0.195	0.181
Julio	169,180	194,760	130,756	157,926	-0.041	0.089
Agosto	186,711	176,110	129,706	141,628	-0.276	-0.121
Septiembre	410,206	502,138	361,074	474,885	0.091	0.011
Octubre	274,700	288,523	176,374	199,334	0.260	-0.059
Noviembre	124,477	125,070	67,675	66,312	0.195	-0.213
Diciembre	82,332	77,038	36,690	31,528	0.126	0.256
Anual	1,598,843	1,705,635	667,260	581,174	0.090	0.042

En la figura 6.12, se muestra la relación que existe entre los promedios de las series mensuales históricas y sintéticas obtenidas con el modelo de disgregación temporal para la estación hidrométrica Temamatla.

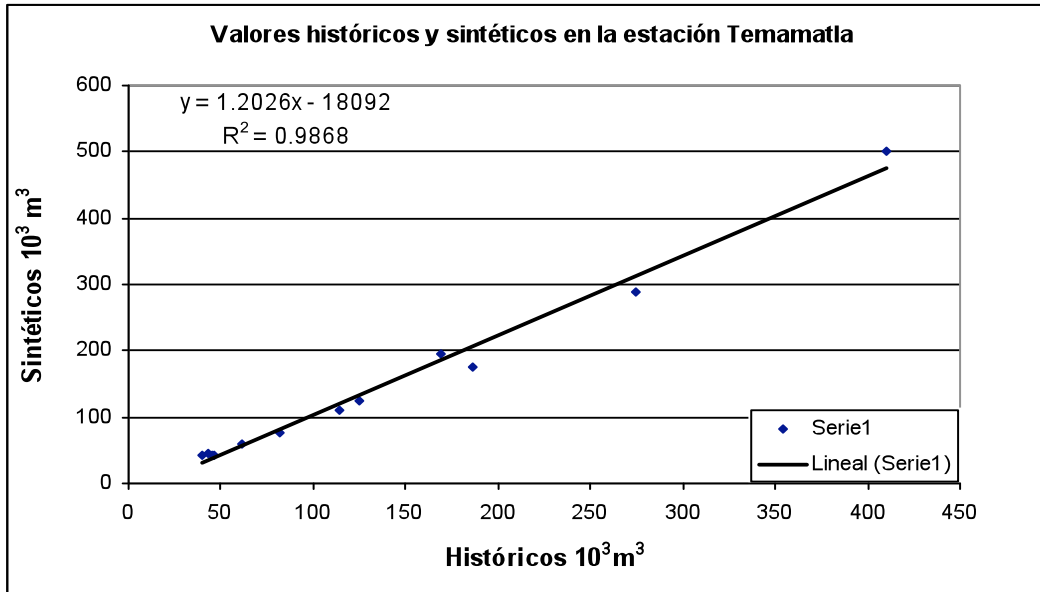


Figura 6.12. Prueba de correlación entre las medias de los datos observados y sintéticos mensuales y anuales para la estación Temamatla.

Cuadro 6.14. Valores medios históricos y sintéticos obtenidos con el modelo de disgregación multivariado AR(1) para la estación Venados.

Mes	Media (10^6m^3)		Desv. Est. (10^6m^3)		r_k (Adim.)	
	Históricos	Sintéticos	Históricos	Sintéticos	Históricos	Sintéticos
Enero	5,374	5,441	1,808	2,648	0.322	0.052
Febrero	3,849	4,057	1,605	1,850	0.323	-0.113
Marzo	3,474	4,374	1,715	2,332	0.510	-0.060
Abril	3,762	3,497	2,180	2,628	0.236	0.064
Mayo	4,647	5,426	3,292	5,965	0.122	0.010
Junio	11,258	11,018	11,885	10,932	-0.031	-0.090
Julio	18,441	17,199	18,635	18,849	-0.053	-0.088
Agosto	17,054	15,504	14,455	14,373	-0.013	-0.046
Septiembre	40,132	32,173	44,849	47,300	0.133	0.508
Octubre	35,643	39,263	64,860	51,622	0.327	0.226
Noviembre	9,423	7,886	5,214	4,199	-0.127	0.069
Diciembre	6,818	7,524	2,790	3,082	0.166	0.165
Anual	159,875	153,363	95,556	81,593	0.342	0.465

Cuadro 6.15. Valores medios históricos y sintéticos obtenidos con el modelo de disgregación multivariado AR(1) para la estación San Agustín.

Mes	Media (10^6m^3)		Desv. Est. (10^6m^3)		r_k (Adim.)	
	Históricos	Sintéticos	Históricos	Sintéticos	Históricos	Sintéticos
Enero	1,564	1,488	852	567	0.068	0.126
Febrero	1,126	1,123	746	663	0.022	-0.326
Marzo	1,071	1,052	659	599	0.229	-0.236
Abril	1,095	1,036	883	674	-0.042	-0.230
Mayo	1,141	1,518	696	2,164	-0.110	-0.123
Junio	1,657	3,700	915	14,415	0.029	0.048
Julio	4,408	2,912	4,619	2,759	-0.029	-0.154
Agosto	3,288	2,640	4,176	1,675	0.000	0.022
Septiembre	7,276	4,229	9,242	3,646	-0.063	0.295
Octubre	7,122	8,397	12,154	21,401	0.213	-0.051
Noviembre	3,444	3,671	5,050	5,137	0.060	-0.093
Diciembre	1,835	1,356	1,349	936	0.217	0.009
Anual	35,027	33,120	24,751	28,184	0.275	-0.035

Cuadro 6.16. Valores medios históricos y sintéticos obtenidos con el modelo de disgregación multivariado AR(1) para la estación Presa La Esperanza.

Mes	Media (10^6m^3)		Desv. Est. (10^6m^3)		r_k (Adim.)	
	Históricos	Sintéticos	Históricos	Sintéticos	Históricos	Sintéticos
Enero	113	438	201	579	0.052	-0.003
Febrero	104	372	141	739	-0.083	-0.046
Marzo	78	357	108	755	-0.221	0.132
Abril	197	260	313	362	-0.286	0.063
Mayo	396	571	578	1,369	0.329	-0.019
Junio	986	688	853	917	-0.134	0.117
Julio	1,012	875	1,467	1,107	0.034	-0.005
Agosto	1,070	1,026	1,107	1,694	-0.185	0.122
Septiembre	2,268	2,068	2,697	2,502	0.029	0.079
Octubre	2,655	1,348	6,764	1,660	0.223	0.343
Noviembre	415	770	541	1,335	0.248	0.030
Diciembre	153	526	195	896	0.195	0.114
Anual	9,446	9,300	8,134	11,583	0.303	0.109

6.5 Pruebas de bondad de ajuste para modelos de disgregación

En el cuadro 6.17 se presentan los resultados de las pruebas de bondad de ajuste de Kolmogorov-Smirnov (KS) realizadas a las series históricas y sintéticas en conjunto. La longitud de los registros fue de 41 años en ambas series y se efectuaron para los meses de Enero y Septiembre, excepto en la estación Venados que por la alta variabilidad se analizó el mes de Marzo y Septiembre.

La prueba más importante para la determinación de la validez de los datos sintéticos es el estadístico H de KS, el objetivo principal de esta prueba es mostrar si dos conjuntos de datos, en este caso, las series histórica y sintética en un mes pertenecen a la misma población. Los resultados de la aplicación de la prueba de KS para dos muestras se presentan en el cuadro 6.17.

Otro indicador importante que se obtiene de la prueba KS son las diferencias máximas entre los datos históricos y los sintéticos, en el cuadro 6.17 se muestra que las series históricas y sintéticas mantienen una estrecha relación. Como se mostró también en las figuras 6.10, 6.11 y 6.12 en donde se puede ver la relación existente entre los promedios de las series históricas y sintéticas.

Como se observa en el cuadro 6.17, se acepta la hipótesis de ajuste a la función Log Normal en todas las series sintéticas, excepto para la estación Presa La Esperanza en el mes de Septiembre, esto se debe a que los datos sintéticos se ajustan a otra función de distribución diferente, por ejemplo, Log Pearson, u otras; lo mismo sucede con las series históricas que son rechazadas.

En las figuras 6.13 a la 6.20 se comparan las Funciones de Distribución Acumulada (FDA) de las series históricas y sintéticas para los meses de Enero y Septiembre en las cuatro estaciones hidrométricas de la cuenca del Río Amajac. Las series históricas y sintéticas a las que se les aplicó la prueba KS se

muestran en los cuadros A1.5 y A1.6. Las pruebas de bondad de ajuste KS se realizaron con el software MATLAB 7.0, (2004).

Cuadro 6.17. Estadístico KS para bondad de ajuste entre las series históricas y sintéticas generadas con modelos de disgregación.

Estaciones/Mes	Estadístico KS	Parámetro			
		Función de Pb que ajusta	Pb	Diferencia Máxima	Pb
Temamatla (Ene)					
Históricos	H = 0	Log Normal	0.28	0.1707	0.548
Sintéticos		Log Normal	0.98		
Temamatla (Sep)					
Históricos	H = 0	Log Normal	0.98	0.1463	0.074
Sintéticos		Log Normal	0.82		
Venados (Ene)					
Históricos	H = 0	Rechazada	---	0.1471	0.424
Sintéticos		Log Normal	0.40		
Venados (Sep)					
Históricos	H = 0	Rechazada	---	0.1324	0.5603
Sintéticos		Log Normal	0.86		
San Agustín (Ene)					
Históricos	H = 0	Log Normal	0.83	0.1463	0.738
Sintéticos		Log Normal	0.56		
San Agustín (Sep)					
Históricos	H = 0	Rechazada	---	0.2195	0.244
Sintéticos		Log Normal	1.0		
P. Esperanza (Ene)					
Históricos	H = 1	Rechazada	---	0.5854	0.560
Sintéticos		Log Normal	0.72		
P. Esperanza (Sep)					
Históricos	H = 0	Log Normal	0.59	0.1951	0.377
Sintéticos		Rechazada	0.21		

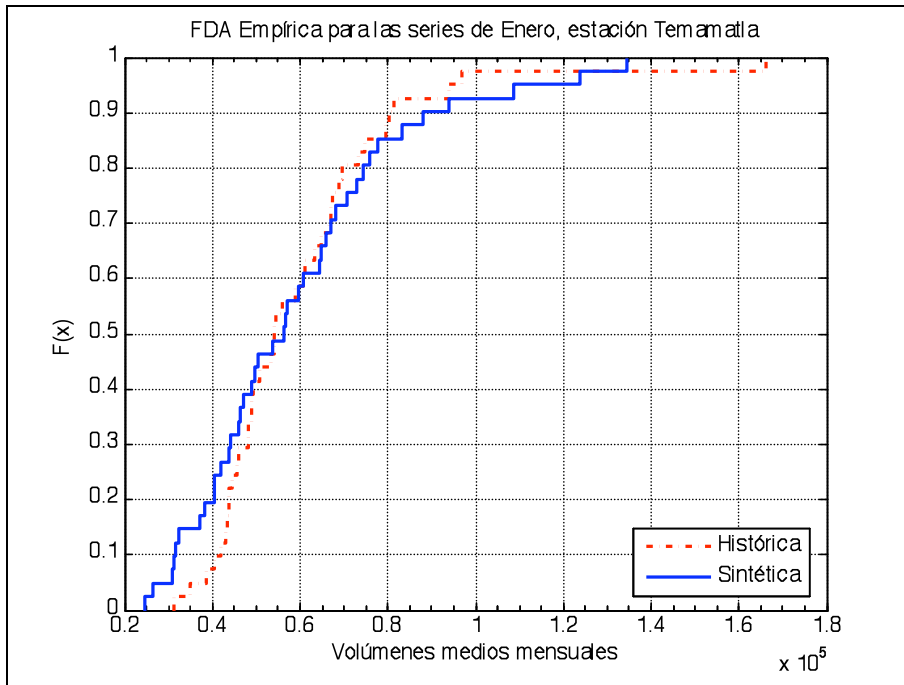


Figura 6.13. FDA Empírica para las series de Enero, estación Temamatla.

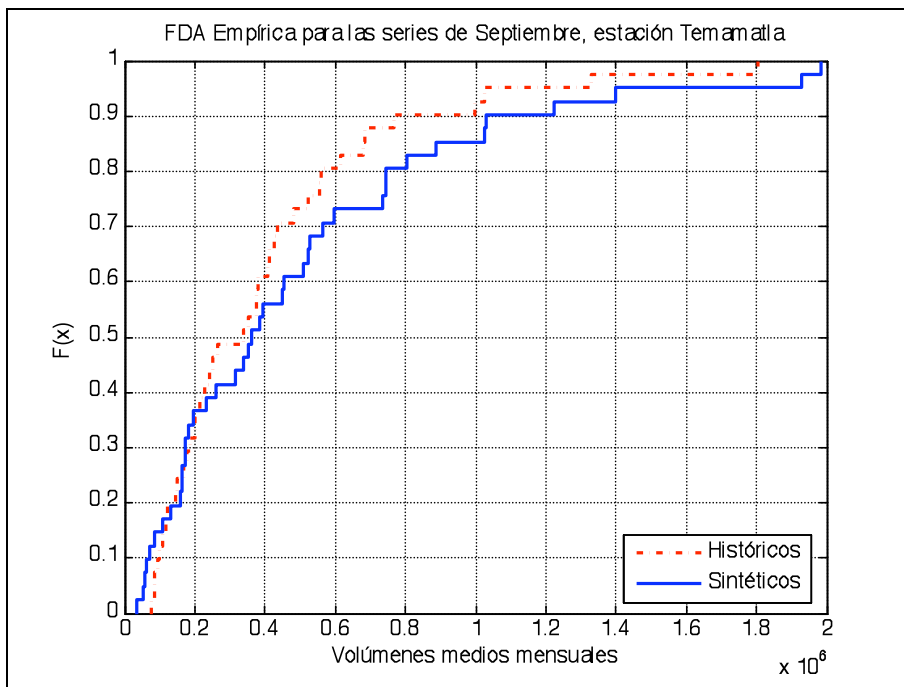


Figura 6.14. FDA Empírica para las series de Septiembre, estación Temamatla.

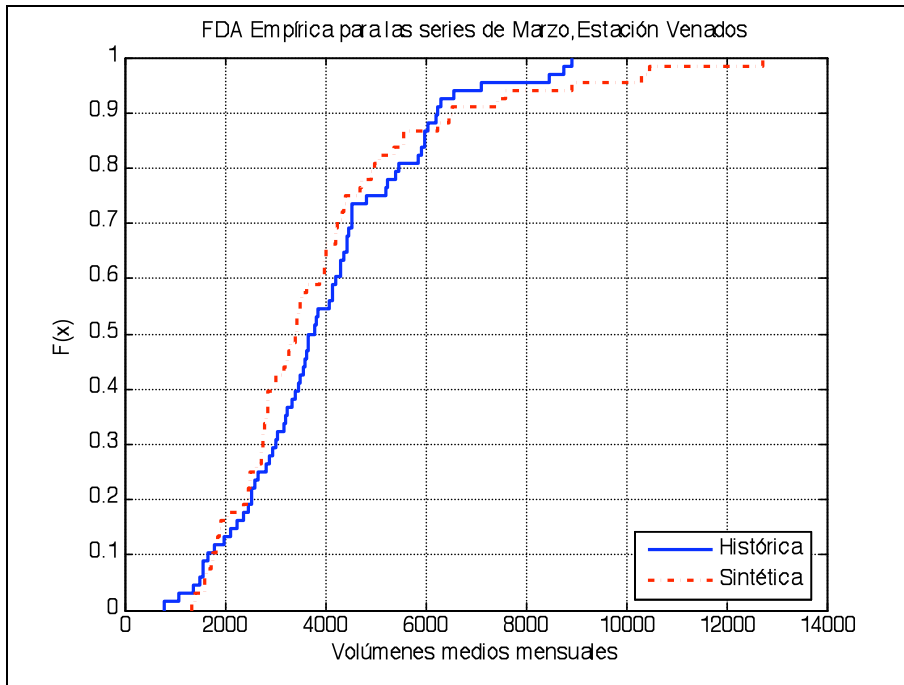


Figura 6.15. FDA Empírica para las series de Marzo, estación Venados.

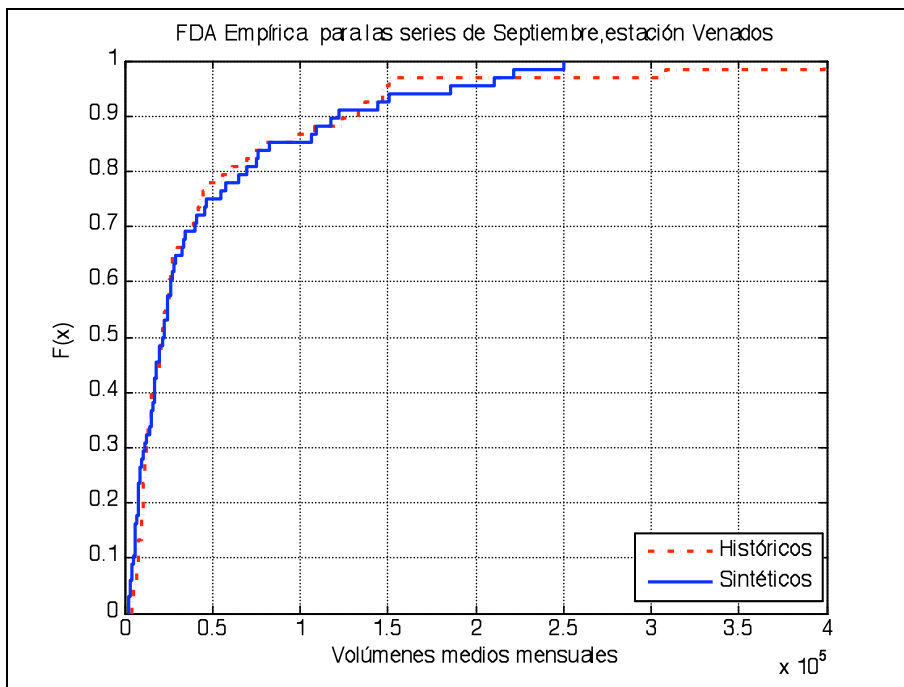


Figura 6.16. FDA Empírica para las series de Septiembre, estación Venados.

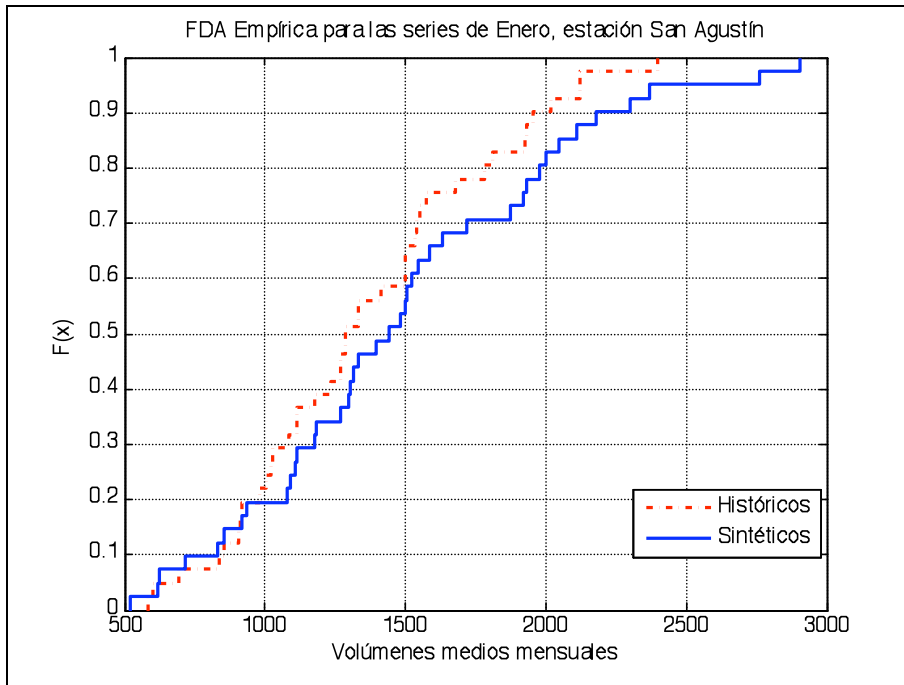


Figura 6.17. FDA Empírica para las series de Enero, estación San Agustín.

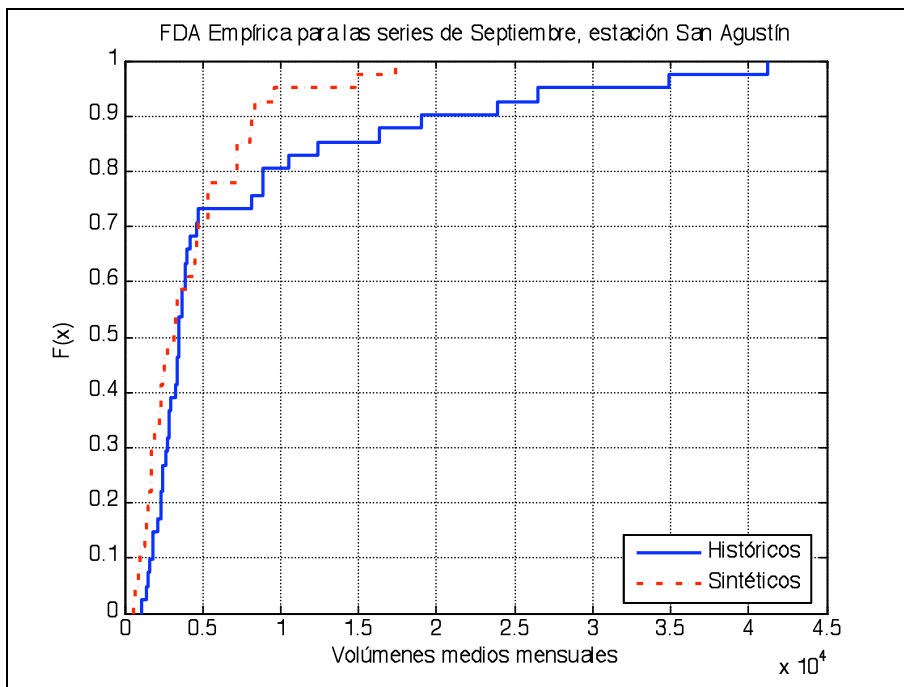


Figura 6.18. FDA Empírica para las series de Septiembre, estación San Agustín.

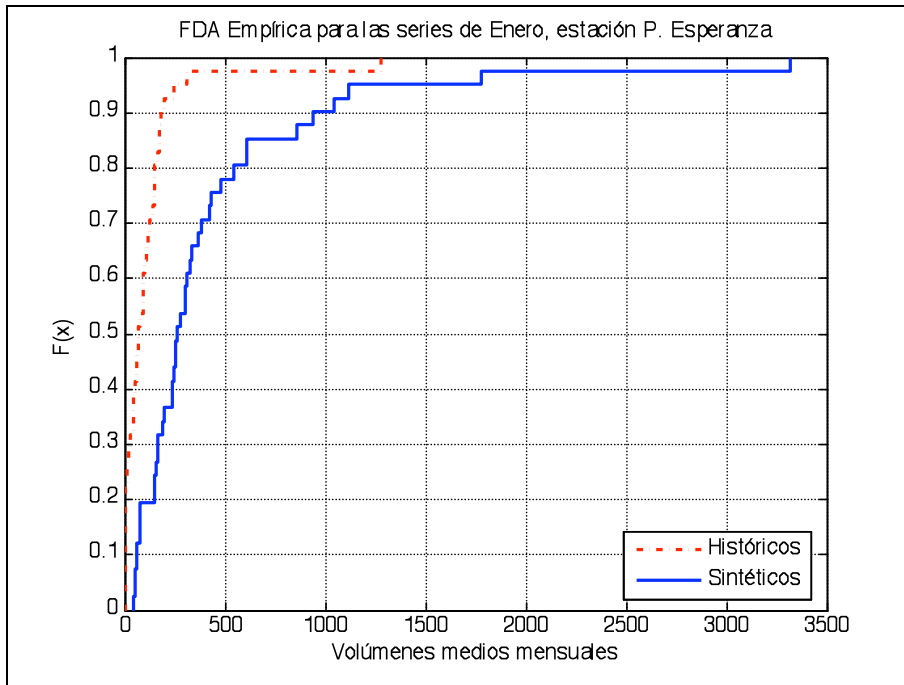


Figura 6.19. FDA Empírica para las series de Enero, estación P. Esperanza.

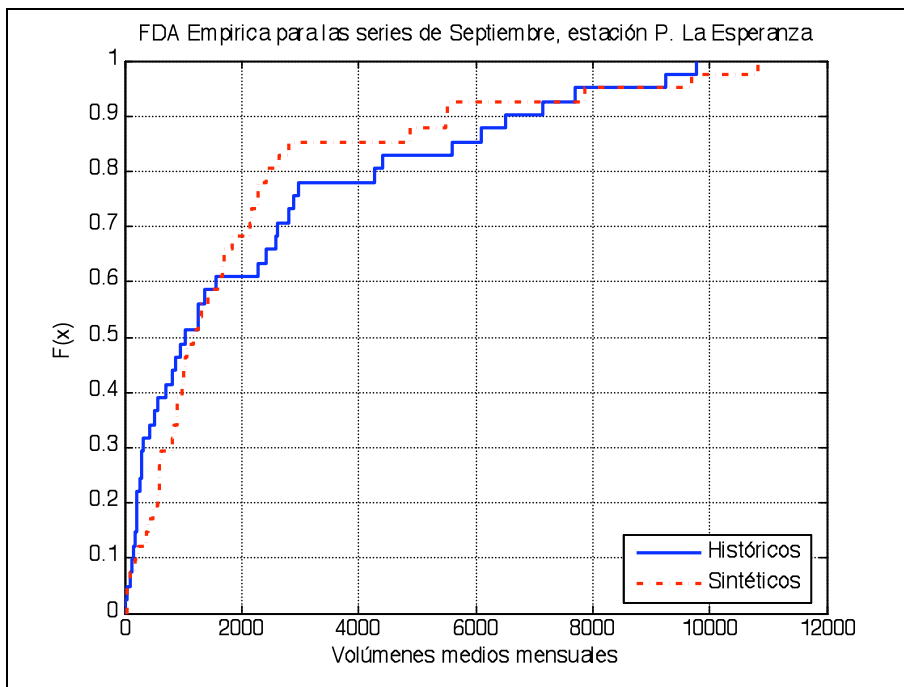


Figura 6.20. FDA Empírica para las series de Septiembre, estación P. Esperanza

Capítulo VII

Conclusiones y Recomendaciones

“La ciencia y la sabiduría, lejos de ser una misma cosa, no tienen entre sí a menudo conexión alguna”

William Cowper

Para generar datos sintéticos, los datos originales son normalizados mediante un proceso de transformación a la función de probabilidades Normal o Log Normal. Cuando se lleva a cabo el proceso inverso, los datos generados pierden las propiedades de dichas funciones de probabilidad, para el caso de los modelos desarrollados en esta investigación, los datos sintéticos mantienen las propiedades de la función Log Normal como lo demuestran las pruebas de bondad de ajuste de Kolmogorov-Smirnov.

La metodología para la estimación de los parámetros $\hat{\phi}_p$ y $\hat{\theta}_q$ representa una solución a un problema estadístico clásico. Comúnmente han sido usados otros métodos para la estimación de parámetros, sin embargo, la aplicación de estos algoritmos se complica conforme crece el número de parámetros por estimar, con la metodología propuesta la simplicidad se mantiene.

El método de estimación de los parámetros de un modelo ARMA(p,q) se desarrolló para la estación Venados como estudio de caso y se resolvió el problema de minimización con el Solver de Excel, sin embargo se pueden utilizar otros software con aplicaciones de optimización. Se presenta con el Solver por la simplicidad en su utilización y el alcance y distribución dicho software.

El análisis de series de tiempo que se realizó a los datos históricos de la estación hidrométrica Venados, muestra que para escala anual, el modelo AR(1) es el que mejor representa la serie histórica.

Las pruebas de bondad de ajuste que se realizaron a las series sintéticas generadas con los modelos AR(1) univariados y multivariados, muestran que las series sintéticas generadas con estos modelos conservan adecuadamente las propiedades estadísticas de las cuatro estaciones de la cuenca del Río Amajac.

Las pruebas de bondad de ajuste que se realizaron a las series históricas y sintéticas en conjunto, muestran que el modelo de disgregación que se

desarrolló para dividir volúmenes anuales a mensuales, permite conservar las propiedades estadísticas (a nivel anual y mensual) de las series históricas.

La metodología mostrada en el presente trabajo permite generar series sintéticas con las propiedades estadísticas de las series históricas. Debido a esto, la metodología puede ser también aplicada para extender registros históricos y/o complementarlos para el caso de variables hidrológicas y más generalmente en registros climatológicos faltantes.

7.1 Recomendaciones

Realizar un programa para simulación y generación de datos sintéticos basado en un lenguaje de programación visual en el que se desarrollen los modelos del presente trabajo y se incluyan además, modelos de disgregación espacial.

Acoplar los modelos de series de tiempo sintéticas en un modelo de simulación/optimización para llevara a cabo la planeación de distritos de riego y el manejo de cuencas.

Los modelos de disgregación propuestos se basan en datos que se generaron previamente, por lo que se recomienda hacer una buena selección del modelo con que se vayan a generar tales datos.

Desarrollar los modelos de disgregación para escalas de tiempo menores, semanas o días (inclusive menores) para la generación de series de tiempo climatológicas.

Capítulo VIII

Bibliografía

“La ciencia es orgullosa por lo mucho que ha aprendido; la sabiduría es humilde porque no sabe más”

William Cowper

Box G.E. P., Jenkins Gwilym M. y Reinsel G.C., 1994, *Time series analysis: forecasting and control*, Prentice-Hall, Englewood Cliffs, NJ, USA.

Burton J., 2003, *Integrated Water Resources Management on a Basin Level (training manual)*, MultiMondes, UNESCO, Canada.

Bowlerman B. L. y O'connell R. T., 1979, *Time series and forecasting*, Duxbury Press, Belmont California, USA.

Chávez M. J., 1985, *An Optimization and Simulation Methodology for Irrigation Planning*, Ph. Dr. Dissertation. University of California at Davis.

CNA (Comisión Nacional del Agua), 2003, *Programa Hidráulico Regional 2002-2006 Región IX Golfo Norte*, CNA, México.

CNA (Comisión Nacional del Agua), 2005, Gerencia del distrito de riego 008 Meztitlán, Comisión Nacional del Agua, Hidalgo, México.

Draper A. J., 2001, *Implicit Stochastic Optimization with Limited Foresight for Reservoir Systems*, Ph. Dr. Dissertation, University Of California at Davis.

García E., 1978, *Modificación al Sistema de Clasificación climática de Köppen; para adaptarlo a las condiciones de la República Mexicana*, Universidad Nacional Autónoma de México, México, DF.

Guerrero G. V. M., 2003, *Análisis estadístico de series de tiempo económicas*, Thomson, México.

Enayetur H. y Hayat K. S., 2000, *Combining Probability of Emptiness and Mean First OverFlow Time of a Dam to Determine its Capacity*, Journal of Spatial Hydrology, Vol. 2, No. 2.

Hillier F. S. Lieberman G. J., 2002, *Investigación de Operaciones*, McGraw-Hill, México.

Jones. D. A., 1983, *Statistical Analysis of Empirical Models Fitted by Optimization*, *Biometrika*, Vol. 70, No. 1, pp. 67-88.

Kottegoda N.T., Lawrance A. J., 1977, *Stochastic Modelling of Riverflow Time Series*, *Journal of the Royal Statistical Society, Series A General*, Vol. 140 , No. 1, pp. 1-47.

Kottegoda, N. T., 1980, *Stochastic Water Resources Technology*, John Wiley and Sons Inc., New York, USA.

Koutsyiannis D., 1998, *Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology*. Department of water resources, National Technical University, Athens, Grece.

Kumar N. D., Lall U. y Petersen M. R., 2000, *Multisite Disaggregation of monthly to Daily Streamflow*. *Water Resources Research*, Vol. 36, No. 7, pp 1823-1833.

Loucks D. P., Stedinger J. R. y Haith D. A., 1981, *Water Resources Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

Maheepala S. y Perera B.J.C., (1996), *Monthly Hydrologic Data Generation by Disaggregation*, *Journal of Hydrology*, Vol. 178, No. 1-4, pp. 277-291.

MatLab 7.0 (2004): MathWorks, Release 14, *Genetic Algorithms and Direct Search Toolbox User's Guide*, Version 1.

Pankratz A., 1983, *Forecasting with univariate Box-Jenkins models, concepts and cases*, Johon Wiley and Sons Inc., New York, USA.

Qin B., Wang W., Ding J. y Fu Q., 2004, *Application of a new Hybrid Model to Predicting Daily Runoffing a Week*, Nature and Science, 2(1).

Voguel R. M. y S. Amy L., 1996, *The moving blocks bootstrap versus parametric time series models*, Water resources research, Vol. 32, No. 6, pp. 1875-1872.

Rieu S., Yup, K. Y. y Lee D. R., 1999, *Streamflow generation using a multivariate hybrid time series model*, School of Civil, Urban, Geosystem Engineering, Soul National University, Seoul, Korea.

Salas, J. D., J. Delleur W., Yevjevich V., y Lane W. L., 1985, *Applied Modeling of hydrologic Time Series*, Water Resources Publications, Littleton, Colorado, USA.

Sharma A., Tarboton D. G. y Lall U., 1997, *Streamflow Simulation; a Nonparametric Approach*, Water Resources Research, Vol. 33, No. 2, pp. 291-308.

Srikanthan R., McMahon T. A. y Sharma A., 2002, *Stochastic Generation of Monthly Rainfall Data*. Thechnical report 02/8. Cooperative Research Centre for Catchment hydrology, Australia.

Stedinger J. R. y Vogel R. M., 1984, *Disaggregation Procedures for Generating Serially Correlated Flow Vectors*, Water Resources Research, Vol. 20, No. 1, pp. 47-56.

Tarboton D. G., Sharma A. y Lall U., 1998, *Disaggregation Procedures for Stochastic Hydrology Based on Nonparametric Density Estimation*, Water Resources Research, Vol. 34, No. 1, pp. 107-119.

Watkins W.D., Lasdon S. L., Nielsen S. S., y Martin W. Q., (2000), *A Scenario-Based Stochastic Programming Model for Water Supplies from the Highland Lakes*, International Transactions in Operational Research, Vol. 7, pp. 211-213.

Yevjevich V., 1972, *Probability and Statistics in Hydrology*, Water Resources Publications, Littleton, Colorado, USA.

Anexo I

*Valores históricos y sintéticos para la
modelación de la cuenca del Río
Amajac*

Cuadro A1.1. Volúmenes medios históricos registrados en la estación Temamatla, (10^3m^3).

Año	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Anual
1964	43,276	32,006	29,552	49,272	31,471	46,403	46,251	32,997	85,733	101,277	60,674	99,544	658,456
1965	43,038	27,135	21,517	28,128	25,059	45,201	82,705	267,899	106,546	192,340	76,459	38,615	954,642
1966	31,428	32,276	29,577	41,628	57,897	199,415	114,252	99,238	185,594	231,985	83,896	46,552	1,153,738
1967	42,041	29,963	37,258	27,397	37,364	37,962	34,474	183,561	523,171	267,016	112,551	64,055	1,396,813
1968	49,157	42,515	40,612	42,367	82,387	112,532	200,241	103,463	406,913	234,940	73,986	69,547	1,458,660
1969	50,898	37,313	34,050	38,384	25,156	25,845	112,563	243,268	999,745	284,561	126,906	84,131	2,062,820
1970	53,502	45,890	40,840	29,937	28,534	228,314	168,417	163,898	480,685	215,230	80,131	55,024	1,590,402
1971	45,602	34,302	34,024	37,644	30,794	62,982	74,114	138,920	177,450	464,926	144,965	68,865	1,314,588
1972	48,914	40,395	39,620	27,269	53,359	199,982	231,938	152,558	93,507	189,626	103,198	58,156	1,238,522
1973	43,657	34,485	32,316	25,226	47,697	170,325	236,762	369,677	212,019	337,223	111,736	84,145	1,705,268
1974	54,192	36,007	52,570	37,203	34,637	161,882	455,235	91,054	1,026,215	368,047	181,143	104,893	2,603,078
1975	74,744	51,719	41,567	34,755	50,163	72,282	145,172	279,655	1,330,534	414,570	158,037	96,348	2,749,546
1976	67,399	48,449	48,231	42,651	62,214	144,041	485,516	288,821	426,349	439,325	195,865	122,239	2,371,100
1977	81,370	62,485	53,410	42,436	40,459	48,768	45,725	50,506	76,649	136,506	59,952	43,601	741,867
1978	34,759	30,017	33,066	28,557	29,325	169,821	61,907	236,889	557,193	494,619	153,996	79,803	1,909,952
1979	59,077	44,786	39,864	43,429	40,260	138,472	65,100	182,605	684,156	166,791	109,456	113,125	1,687,121
1980	69,684	53,325	56,991	42,588	45,403	58,738	45,215	67,733	376,604	175,187	81,019	56,081	1,128,568
1981	80,192	59,148	49,957	46,596	77,914	620,254	386,343	595,232	614,258	428,711	167,964	150,698	3,277,267
1982	79,516	57,297	55,018	49,915	74,692	40,885	43,817	37,201	100,047	107,097	53,288	53,155	751,928
1983	60,725	37,659	33,836	27,694	27,488	30,908	325,801	155,663	376,959	209,687	129,651	63,743	1,479,814
1984	61,273	62,239	41,552	31,066	92,559	132,925	293,259	200,306	1,805,711	321,216	134,166	87,476	3,263,748
1985	66,873	49,918	54,232	105,515	74,201	155,212	436,750	228,453	148,106	145,441	68,846	61,274	1,594,821
1986	54,098	41,550	41,448	39,073	42,326	144,248	109,576	58,755	117,841	153,570	135,913	68,349	1,006,747
1987	54,343	40,209	44,205	34,056	41,724	61,995	304,614	192,653	202,262	125,704	60,909	48,398	1,211,072
1988	38,682	33,495	32,763	42,636	26,681	107,289	161,428	213,694	765,265	129,693	79,045	60,362	1,691,033
1989	48,057	44,997	39,244	40,408	49,865	60,975	94,870	202,973	411,447	216,444	125,467	94,031	1,428,778
1990	66,582	48,187	45,747	43,216	36,461	38,559	226,193	646,826	435,949	312,742	157,041	118,356	2,175,859
1991	64,834	43,244	43,169	39,030	38,593	88,468	332,221	85,665	200,991	699,187	187,449	101,619	1,924,470

1992	96,906	88,786	60,667	50,283	75,545	92,568	145,257	202,628	378,160	544,356	209,496	109,376	2,054,028
1993	68,710	57,352	44,365	39,599	38,707	316,589	365,353	149,459	679,689	435,349	272,389	125,318	2,592,879
1994	73,371	58,767	51,447	54,833	57,089	75,950	51,361	151,573	252,466	184,663	68,448	52,230	1,132,198
1995	46,092	37,984	39,020	29,064	29,063	38,616	142,559	339,824	166,445	161,611	149,421	82,041	1,261,740
1996	56,039	42,613	40,060	36,008	32,496	50,997	59,166	238,813	228,897	219,564	80,391	56,169	1,141,213
1997	44,343	35,684	45,198	63,368	42,790	42,931	63,962	49,485	85,266	150,753	83,111	38,054	744,945
1998	40,475	25,873	27,713	21,792	23,573	22,088	39,319	83,981	560,271	863,611	311,758	175,808	2,196,262
1999	93,823	51,500	45,835	37,593	48,658	58,741	313,585	160,483	351,040	545,178	335,363	202,455	2,244,254
2000	166,099	119,488	78,889	50,562	42,577	159,136	64,413	158,589	124,232	142,255	60,374	65,304	1,231,918
2001	48,199	40,047	36,895	43,087	62,944	48,948	91,203	75,597	340,994	185,931	119,963	79,283	1,173,091
2002*	55,574	61,242	62,027	45,862	43,381	293,996	65,633	57,471	265,403	54,081	91,975	54,431	1,151,076
2003*	37,545	27,977	49,297	17,050	51,588	40,753	81,954	171,964	123,589	75,176	45,087	37,485	759,467
2004*	99,101	62,704	47,642	44,499	21,727	47,684	132,172	245,123	334,091	136,529	62,052	105,486	1,338,810
Promedios	60,834	46,610	43,300	40,285	45,727	114,480	169,180	186,711	410,206	274,700	124,477	82,332	1,598,843
Desv. Est.	23,783	17,303	10,750	14,005	17,779	108,782	130,756	129,706	361,074	176,374	67,675	36,690	667,260
r_k	0.215	0.091	0.112	-0.172	-0.032	-0.195	-0.041	-0.276	0.091	0.260	0.195	0.126	0.090

* Los valores anuales fueron generados con un modelo AR(1) y los valores mensuales con un modelo de disgregación.

Cuadro A1.2. Volúmenes medios históricos registrados en la estación Venados, (10^3m^3).

Año	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC	ANUAL
1937	8,319	6,163	8,455	6,412	9,600	10,748	35,535	14,189	12,387	14,526	8,904	8,318	143,664
1938	4,441	3,575	3,061	2,339	6,235	16,803	17,927	8,722	20,878	7,944	7,879	7,804	107,703
1939	5,182	3,937	4,068	4,693	8,536	23,535	9,852	6,731	27,139	36,245	14,432	7,614	152,068
1940	6,690	5,641	6,239	4,977	4,808	6,367	8,608	7,319	6,722	6,643	6,177	5,813	76,098
1941	5,837	4,443	3,862	7,388	11,734	24,449	14,208	8,640	55,612	80,996	12,791	12,341	242,411
1942	8,460	6,157	4,518	5,373	6,016	13,042	13,544	22,580	108,670	24,867	11,093	10,003	234,433
1943	8,061	6,221	4,476	3,934	6,424	23,499	9,279	6,785	9,492	6,738	6,928	6,486	98,421
1944	5,301	3,753	4,548	2,830	3,915	7,057	6,437	53,733	308,424	100,307	25,214	16,524	538,154
1945	11,756	9,626	6,573	6,692	4,326	11,546	9,315	11,014	22,203	49,506	10,395	9,418	162,480
1946	8,755	6,037	5,861	5,609	6,631	8,416	10,133	9,745	14,434	64,200	12,315	10,270	162,512
1947	6,967	4,553	4,158	5,308	9,747	6,087	3,331	16,269	23,655	20,961	10,280	10,184	121,600
1948	7,948	4,460	3,207	4,503	7,814	6,599	15,934	7,710	19,760	9,796	6,161	5,697	99,687
1949	5,189	3,676	3,577	3,038	3,686	5,528	9,034	6,340	24,970	10,789	5,387	5,501	86,806
1950	5,467	3,789	3,801	5,465	5,158	12,652	11,537	5,070	5,290	17,342	4,228	4,588	84,480
1951	3,704	2,631	3,459	2,294	10,268	9,150	13,613	6,697	11,824	11,665	5,558	6,323	87,277
1952	6,042	4,774	3,608	6,506	10,604	63,023	26,214	11,852	60,832	30,578	32,309	10,650	267,107
1953	9,396	6,015	5,997	4,998	3,518	2,779	9,244	10,432	23,006	21,551	16,033	8,077	121,148
1954	5,711	4,199	3,660	3,555	4,409	4,326	9,448	7,079	77,054	182,363	18,480	11,137	331,525
1955	8,450	5,890	5,483	4,312	5,088	5,963	32,311	54,184	398,711	84,437	37,239	23,062	665,251
1956	17,867	10,471	8,754	7,593	8,975	12,437	18,587	17,811	97,604	16,760	11,784	11,139	239,900
1957	8,390	6,287	6,306	6,362	6,265	7,157	6,213	7,158	7,630	7,803	6,705	5,729	82,102
1958	7,332	4,371	4,291	3,391	8,338	13,165	61,439	20,035	41,015	85,289	28,725	15,484	292,990
1959	9,461	7,356	7,129	12,172	9,581	21,752	18,814	14,668	13,645	32,409	13,493	8,744	169,338
1960	5,920	3,445	3,508	4,017	6,452	5,038	9,975	8,212	25,509	8,663	7,474	7,316	95,623
1961	6,272	4,828	4,815	2,709	1,756	19,315	17,186	11,449	10,244	9,543	12,184	3,843	104,242
1962	3,129	2,441	2,464	4,296	3,190	9,992	6,893	5,789	10,743	5,934	5,320	6,033	66,309
1963	4,830	3,667	3,245	2,312	5,609	10,924	39,807	20,543	16,123	15,135	12,375	8,419	143,090
1964	8,039	5,393	4,220	3,807	5,108	6,544	5,223	4,774	6,154	7,913	6,055	6,908	70,229
1965	4,700	3,480	3,670	4,897	5,783	5,550	8,510	36,456	11,905	12,341	10,703	10,391	118,484

1966	9,595	7,356	8,918	8,638	7,506	12,528	18,952	15,999	15,004	21,477	8,634	6,508	141,225
1967	6,294	5,593	6,035	6,423	5,591	3,341	5,486	9,129	19,457	15,564	4,763	3,858	91,630
1968	5,435	5,142	5,238	8,478	13,638	10,374	14,672	14,870	20,797	16,049	7,839	9,552	132,189
1969	5,948	3,979	4,149	3,551	3,002	4,593	16,767	40,643	137,061	37,519	9,808	9,401	276,527
1970	6,722	5,560	5,216	3,414	3,372	34,556	19,870	18,983	39,579	16,938	6,205	5,904	166,425
1971	5,191	3,026	2,258	2,009	3,520	13,809	10,146	11,918	11,872	37,910	17,688	7,443	126,886
1972	4,849	3,478	3,166	2,494	3,276	11,773	24,794	23,115	12,345	16,232	6,579	6,340	118,538
1973	5,127	3,047	2,382	4,142	2,190	17,290	46,496	37,687	34,692	35,338	8,731	6,775	204,001
1974	4,893	2,854	5,413	2,907	3,418	17,336	65,732	6,566	146,556	51,759	11,403	8,381	327,325
1975	6,849	5,470	5,919	5,573	9,431	11,148	13,408	17,950	146,717	36,136	9,956	7,521	276,189
1976	6,610	6,936	6,214	5,962	5,815	13,307	65,902	27,550	3,929	7,577	19,909	12,754	182,573
1977	7,775	4,974	4,444	3,931	4,151	11,374	5,448	5,033	7,428	9,746	6,762	4,596	75,755
1978	3,881	3,904	4,363	3,935	2,137	25,592	11,538	21,700	29,917	33,330	8,160	6,083	154,641
1979	4,492	3,147	2,582	8,188	4,639	15,355	10,952	31,468	41,562	18,781	9,367	8,496	159,131
1980	6,286	4,610	4,299	4,567	14,463	5,525	6,808	8,930	18,312	21,104	8,323	7,185	110,511
1981	8,202	5,277	3,824	4,125	8,997	70,262	67,905	12,104	13,624	61,600	14,315	9,311	279,660
1982	5,976	4,832	4,532	5,670	7,811	4,324	3,844	4,672	34,868	6,706	4,046	3,294	90,667
1983	3,822	2,742	3,002	2,295	1,123	1,772	18,245	9,373	26,885	6,707	4,046	3,462	83,558
1984	6,727	3,578	2,831	1,147	4,027	8,308	27,200	22,377	150,361	7,670	7,479	6,703	248,507
1985	7,014	3,515	3,634	8,131	7,259	21,213	34,408	21,093	16,899	28,017	9,859	8,464	169,614
1986	4,724	2,514	2,545	4,614	4,441	18,473	7,001	3,279	3,747	18,565	5,023	4,990	80,004
1987	2,502	1,672	1,549	1,268	6,431	5,666	11,300	12,035	7,882	10,478	8,305	3,240	72,411
1988	2,949	2,059	1,647	2,703	201	1,869	5,992	9,040	123,690	6,728	3,328	3,278	163,561
1989	2,786	2,291	1,988	3,522	1,210	2,539	3,264	20,011	69,271	5,934	7,404	6,985	127,289
1990	3,998	3,740	2,949	1,441	5,702	4,748	17,259	82,233	26,490	11,051	6,080	6,543	172,331
1991	4,627	2,994	1,083	1,102	925	3,736	40,893	13,818	9,751	14,489	6,648	7,760	107,910
1992	5,770	7,008	3,348	4,178	6,527	7,662	9,604	11,873	10,451	45,764	14,424	7,239	133,952
1993	4,505	2,610	4,429	3,276	3,887	6,172	26,424	7,222	46,350	20,789	17,404	6,985	150,153
1994	4,021	3,265	2,666	2,929	5,081	13,709	4,410	8,685	15,534	49,955	8,313	5,550	124,212
1995	2,897	2,839	2,531	2,411	1,593	1,181	9,377	32,213	22,005	8,886	5,066	3,405	94,489
1996	4,844	2,152	1,492	1,227	821	3,903	5,693	5,823	10,103	19,686	17,794	6,843	80,463
1997	4,118	2,100	1,801	5,740	5,772	3,259	3,320	2,348	6,536	36,020	7,456	5,218	83,772
1998	4,118	1,557	784	163	57	949	4,915	9,267	133,636	13,742	3,261	3,083	175,597
1999	6,897	4,260	2,897	1,824	1,676	6,376	63,268	19,375	74,632	205,285	8,354	11,963	406,913

2000	10,572	7,719	5,974	5,030	6,364	16,215	6,416	24,160	15,248	388,154	29,864	16,606	532,439
2001	5,275	4,168	3,419	1,584	1,804	13,930	11,073	8,148	42,839	14,456	5,723	6,055	118,567
2002	3,597	2,331	1,553	1,207	2,780	3,840	9,437	6,584	44,751	27,238	12,533	4,454	120,391
2003	3,957	2,266	1,373	1,352	787	5,855	5,598	4,758	25,686	13,066	11,414	5,296	81,489
2004	3,733	2,388	2,113	4,389	8,227	15,609	8,528	15,932	10,898	44,675	7,303	4,716	128,606
Promedios	6,106	4,356	3,994	4,225	5,430	12,102	17,654	15,882	45,574	35,652	10,826	7,736	169,537
Desv. Est.	2,459	1,818	1,791	2,217	3,171	11,789	16,485	13,686	67,017	56,169	6,914	3,544	115,507
r_k	0.3429	0.3248	0.4199	0.1448	0.2022	-0.0840	-0.0649	-0.0304	0.1052	0.2887	0.0427	0.2416	0.2397

Cuadro A1.3. Volúmenes medios históricos registrados en la estación San Agustín, (10³m³).

Año	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Anual
1964	1,553	986	1,869	984	836	887	1,809	1,575	3,420	1,621	1,023	1,046	17,609
1965	2,019	2,319	1,416	1,581	1,112	2,351	2,351	2,351	2,310	6,833	4,438	2,586	31,667
1966	1,178	476	408	376	991	1,433	4,311	2,267	3,864	2,987	1,069	718	20,078
1967	1,573	829	702	556	412	844	10,998	3,785	4,739	3,846	1,683	1,170	31,137
1968	694	379	812	658	863	2,052	3,725	3,721	3,925	3,981	1,692	742	23,244
1969	1,498	712	697	476	366	1,120	11,726	3,856	3,974	4,136	1,983	1,786	32,330
1970	2,119	1,329	1,274	1,306	1,111	2,350	1,564	1,625	8,131	4,448	2,329	2,602	30,188
1971	1,333	1,085	1,272	2,100	1,644	1,346	1,057	1,002	1,613	1,799	1,583	1,280	17,114
1972	581	409	513	297	763	1,952	3,681	3,671	2,925	4,561	1,725	685	21,763
1973	1,498	725	761	482	355	1,170	12,786	4,359	3,322	4,568	2,181	2,390	34,597
1974	2,756	1,568	1,273	1,638	1,235	1,968	8,672	4,256	19,015	19,284	14,368	5,386	81,419
1975	853	467	389	625	456	3,482	2,456	2,867	34,878	3,468	1,864	942	52,747
1976	1,811	2,166	985	912	816	2,142	3,202	3,915	3,425	14,250	4,212	2,321	40,157
1977	1,285	992	1,236	2,064	1,623	1,312	1,025	1,125	1,489	1,870	1,542	1,211	16,774
1978	836	356	323	401	1,539	1,596	3,456	2,769	2,874	2,189	1,823	728	18,890
1979	1,086	943	902	818	1,215	2,487	4,982	3,874	4,178	2,786	2,346	1,825	27,442
1980	1,413	845	786	563	318	586	1,068	1,135	2,710	1,867	1,345	763	13,399
1981	2,897	945	1,659	1,362	1,181	3,946	15,860	8,453	26,538	7,219	4,209	1,973	76,242
1982	918	246	397	465	1,640	1,684	4,238	2,967	2,459	2,392	1,845	689	19,940
1983	1,112	892	861	786	688	653	1,380	1,286	2,803	1,456	1,009	885	13,811
1984	1,953	2,386	1,188	1,419	1,814	2,656	3,516	4,294	3,666	15,301	4,816	2,582	45,591
1985	1,004	416	776	685	1,825	2,835	4,869	3,526	3,630	2,613	2,183	983	25,345
1986	1,028	703	945	683	836	711	1,259	1,067	1,844	1,402	908	842	12,228
1987	601	385	403	381	1,929	1,700	3,390	3,610	2,365	2,410	1,910	777	19,861
1988	983	686	549	901	67	623	1,997	3,013	41,230	2,243	1,109	1,093	54,494
1989	1,286	912	1,225	2,350	1,591	1,280	925	1,016	2,125	1,745	1,612	1,321	17,388
1990	1,332	1,246	983	480	1,901	1,583	5,753	27,411	8,830	3,683	2,028	2,181	57,411
1991	1,542	998	541	368	308	1,245	13,631	4,606	3,250	4,830	2,216	2,587	36,123
1992	1,923	2,336	1,116	1,393	2,176	2,554	3,202	3,958	3,484	15,255	4,808	2,413	44,617
1993	1,502	870	2,215	1,640	1,992	3,180	13,613	3,721	23,877	6,930	5,801	2,328	67,667

1994	2,119	2,329	2,274	2,306	2,111	2,351	1,543	1,635	1,804	8,131	4,448	2,602	33,653
1995	1,535	966	1,869	894	683	788	980	1,557	3,450	1,632	1,033	1,064	16,451
1996	1,232	1,124	981	704	557	1	414	578	2,609	1,557	1	1,089	10,847
1997	1,271	910	1,172	2,054	1,582	1,166	788	787	1,373	1,682	1,532	1,203	15,520
1998	1,267	547	303	246	1	1	1,221	378	16,300	1,855	1,072	963	24,154
1999	1,929	1,270	917	1,197	1,131	1,896	16,757	3,031	8,892	29,350	31,175	4,346	101,891
2000	5,722	3,962	3,794	5,158	3,062	3,066	2,384	4,215	2,474	73,952	7,350	7,452	122,591
2001	1,781	1,718	1,687	1,132	1,156	1,933	2,565	2,455	12,343	3,241	2,307	2,047	34,365
2002	2,398	1,780	932	683	703	668	703	1,239	10,480	4,890	4,374	2,699	31,548
2003	1,024	677	553	308	107	911	377	624	4,600	2,377	3,417	1,120	16,094
2004	1,676	1,280	947	1,454	2,099	1,416	514	1,242	1,082	11,376	2,836	1,808	27,728
Promedios	1,564	1,126	1,071	1,095	1,141	1,657	4,408	3,288	7,276	7,122	3,444	1,835	35,027
Desv. Est.	852	746	659	883	696	915	4,619	4,176	9,242	12,154	5,050	1,349	24,751
r_k	0.068	0.022	0.229	-0.042	-0.110	0.029	-0.029	0.000	-0.063	0.213	0.060	0.217	0.275

Cuadro A1.4. Volúmenes medios históricos registrados en la estación Presa La Esperanza, (10³m³).

Año	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Anual
1964	151	1	90	55	625	711	1	262	326	1	272	121	2,616
1965	1,278	219	1	181	491	837	501	3,452	20	1	438	1	7,420
1966	44	1	187	809	217	1,681	932	239	945	262	39	60	5,416
1967	311	1	1	1	267	239	279	888	2,879	795	13	26	5,700
1968	26	1	1	820	376	1,026	647	521	718	606	1	1	4,744
1969	12	11	1	32	140	709	846	4,028	9,247	408	1	41	15,476
1970	13	1	1	142	31	1	872	985	1,261	1,427	99	15	4,848
1971	71	46	1	32	32	2,346	846	954	440	466	83	1	5,318
1972	80	60	263	1	136	1,188	748	929	285	3,931	1,347	299	9,267
1973	1	51	1	88	146	1,879	4,157	2,967	2,793	857	234	119	13,293
1974	19	1	1	96	597	878	4,049	1	7,705	2,058	381	221	16,007
1975	62	90	1	1	1,501	979	862	736	6,503	3,882	403	150	15,170
1976	96	72	1	273	816	1,556	7,916	1,124	1,263	2,932	508	235	16,792
1977	47	188	1	54	69	500	1	911	254	3,948	580	319	6,872
1978	1	1	386	43	1	2,227	1	856	147	184	539	53	4,439
1979	1	1	1	444	1	269	1	1,752	6,095	889	172	5	9,631
1980	1	1	1	1	2,166	1	373	1,054	1,554	577	104	102	5,935
1981	45	615	419	1	1,186	3,555	2,628	3,552	2,275	132	177	1	14,586
1982	1	1	1	461	1,391	1	1	29	219	1,546	158	255	4,064
1983	97	68	58	1	1	1	53	1	110	200	17	217	824
1984	151	253	125	55	84	1,109	1,326	270	7,135	1	1	1	10,511
1985	1	239	1	1,231	1	3,013	1	848	553	871	132	1	6,892
1986	1	1	1	1	138	2,058	247	872	168	2,119	39	5	5,650
1987	1	1	1	1	302	1,034	1,459	11	31	543	637	10	4,031
1988	53	78	297	383	217	57	1,619	1,718	2,593	201	49	83	7,348
1989	2	62	1	26	1	1,789	1,276	432	4,272	430	1	108	8,400
1990	94	193	5	7	64	744	691	2,743	818	714	136	311	6,520
1991	69	31	1	23	36	609	1,837	1,297	279	1,128	1,202	224	6,736
1992	116	643	146	204	289	1,492	391	1,188	1,366	4,514	1,095	110	11,554

1993	163	149	79	52	86	260	1,618	358	4,420	3,923	1,979	270	13,357
1994	60	66	72	62	109	970	186	332	860	4,480	206	57	7,460
1995	106	103	54	47	114	214	655	4,060	2,969	464	46	35	8,867
1996	173	92	126	76	76	215	307	869	212	1,704	1,368	463	5,681
1997	152	68	166	394	298	465	443	267	115	3,143	217	156	5,884
1998	92	71	41	27	27	55	151	163	5,586	457	155	77	6,902
1999	197	164	107	117	141	293	1,696	299	9,765	13,912	1,926	424	29,041
2000	244	179	224	284	347	1,367	261	922	511	42,178	1,581	1,103	49,201
2001	141	94	107	54	169	1,223	467	777	2,433	214	125	91	5,895
2002	179	116	109	127	843	103	149	409	2,615	992	273	241	6,156
2003	172	99	72	103	190	1,056	142	250	1,027	337	175	143	3,766
2004	121	125	45	1,269	2,494	1,700	836	561	216	1,419	94	126	9,006
Promedio	113	104	78	197	396	986	1,012	1,070	2,268	2,655	415	153	9,446
Desv. Est.	201	141	108	313	578	853	1,467	1,107	2,697	6,764	541	195	8,134
Coef. r_k	0.051	-0.083	-0.221	-0.283	0.323	-0.149	0.034	-0.185	0.029	0.223	0.247	0.194	0.302

Cuadro A1.5. Valores históricos y sintéticos usados para la obtención del estadístico de Kolmogorov-Smirnov, estaciones Temamatla y Venados.

Periodo de tiempo	Temamatla				Venados			
	Enero		Septiembre		Enero		Septiembre	
	Históricos	Sintéticos	Históricos	Sintéticos	Históricos	Sintéticos	Históricos	Sintéticos
1	63,327	74,545	243,420	260,638	8,029	7,687	6,146	69,157
2	48,910	40,348	266,201	733,533	4,696	8,515	11,895	9,494
3	43,745	108,757	144,759	450,035	9,588	5,449	14,992	24,194
4	43,276	123,691	85,733	174,255	6,287	10,508	19,437	75,032
5	43,038	46,508	106,546	802,810	5,431	5,236	20,780	75,670
6	31,428	46,943	185,594	35,474	5,946	3,800	137,009	2,080
7	42,041	68,131	523,171	1,025,330	6,718	7,130	39,554	32,936
8	49,157	49,776	406,913	744,943	5,187	3,183	11,863	11,015
9	50,898	50,592	999,745	525,999	4,845	7,525	12,335	17,515
10	53,502	31,676	480,685	130,089	5,124	1,893	34,674	6,179
11	45,602	49,088	177,450	1,400,921	4,891	2,795	146,508	27,928
12	48,914	46,003	93,507	69,595	6,846	5,742	146,658	20,025
13	43,657	83,130	212,019	163,145	6,606	3,491	3,927	5,567
14	54,192	70,630	1,026,215	87,475	7,765	5,587	7,419	12,128
15	74,744	60,828	1,330,534	1,982,908	3,878	3,196	29,897	4,888
16	67,399	26,445	426,349	385,651	4,489	4,636	41,535	11,746
17	81,370	41,762	76,649	108,984	6,280	16,298	18,296	250,092
18	34,759	32,307	557,193	53,022	8,199	7,213	13,618	117,668
19	59,077	64,566	684,156	173,438	5,970	6,598	34,833	210,953
20	69,684	24,437	376,604	314,620	3,818	5,089	26,858	221,600
21	80,192	72,819	614,258	58,772	6,724	6,924	150,301	34,977
22	79,516	37,096	100,047	507,051	7,010	7,813	16,888	29,277
23	60,725	64,735	376,959	1,032,084	4,719	8,987	3,743	57,983
24	61,273	43,604	1,805,711	160,672	2,499	7,103	7,873	18,274
25	66,873	59,786	148,106	888,156	2,948	4,469	123,632	24,024
26	54,098	53,732	117,841	566,452	2,784	4,029	69,225	7,803
27	54,343	56,465	202,262	60,320	3,996	4,386	26,475	14,365
28	38,682	88,169	765,265	196,842	4,623	7,093	9,743	186,175
29	48,057	38,315	411,447	183,379	5,765	6,748	10,443	17,458
30	66,582	65,855	435,949	165,645	4,502	3,264	46,319	20,106
31	64,834	44,144	200,991	352,240	4,018	6,382	15,522	24,133
32	96,906	40,630	378,160	392,585	2,894	5,694	21,985	39,684
33	68,710	77,764	679,689	230,523	4,839	5,725	10,093	21,497
34	73,371	31,156	252,466	454,911	4,114	7,551	6,529	4,471
35	46,092	30,978	166,445	1,223,014	4,116	6,399	133,587	7,346
36	56,039	93,740	228,897	1,928,523	6,895	9,418	74,613	6,295
37	44,343	56,830	85,266	595,644	10,570	5,114	15,245	5,932
38	40,475	66,863	560,271	359,456	5,271	3,944	42,805	17,000
39	93,823	56,899	351,040	745,786	3,594	9,443	44,718	17,509
40	166,099	75,777	124,232	525,106	3,953	6,308	25,661	26,764
41	48,199	134,373	340,994	337,612	3,730	8,942	10,890	25,929
Promedios	59,950	59,266	408,530	502,138	5,370	6,276	40,110	44,216
Desv. Est.	22,913	24,915	361,419	474,885	1,807	2,590	44,833	62,545

Cuadro A1.6. Valores históricos y sintéticos usados para la obtención del estadístico de Kolmogorov-Smirnov, estaciones San Agustín y P. La Esperanza.

Periodo de tiempo	San Agustín				Presa La Esperanza			
	Enero		Septiembre		Enero		Septiembre	
	Históricos	Sintéticos	Históricos	Sintéticos	Históricos	Sintéticos	Históricos	Sintéticos
1	1,553	1,444	3,420	14,823	151	423	326	971
2	2,019	523	2,310	2,796	1278	241	20	2182
3	1,178	2,369	3,864	683	44	74	945	1057
4	1,573	1,108	4,739	914	311	938	2879	7880
5	694	1,588	3,925	4,420	26	152	718	602
6	1,498	833	3,974	7,257	12	383	9247	631
7	2,119	1,543	8,131	7,220	13	194	1261	1324
8	1,333	2,905	1,613	3,221	71	1119	440	9695
9	581	2,177	2,925	1,401	80	146	285	4881
10	1,498	1,523	3,322	3,192	1	305	2793	2638
11	2,756	1,300	19,015	4,126	19	236	7705	5525
12	853	2,000	34,878	5,389	62	545	6503	899
13	1,811	1,268	3,425	975	96	429	1263	245
14	1,285	625	1,489	1,935	47	55	254	46
15	836	716	2,874	4,746	1	41	147	2275
16	1,086	1,113	4,178	1,246	1	252	6095	1033
17	1,413	2,302	2,710	2,189	1	169	1554	2132
18	2,897	1,179	26,538	970	45	58	2275	391
19	918	617	2,459	7,195	1	308	219	1008
20	1,112	918	2,803	4,658	97	334	110	93
21	1,953	1,180	3,666	8,303	151	63	7135	1704
22	1,004	854	3,630	4,549	1	156	553	1586
23	1,028	1,721	1,844	1,921	1	1781	168	2419
24	601	1,498	2,365	1,751	1	368	31	363
25	983	1,920	41,230	1,714	53	80	2593	579
26	1,286	1,485	2,125	2,377	2	329	4272	602
27	1,332	1,088	8,830	2,481	94	856	818	1259
28	1,542	2,046	3,250	5,332	69	609	279	5475
29	1,923	1,302	3,484	1,609	116	165	1366	184
30	1,502	1,330	23,877	1,547	163	263	4420	1166
31	2,119	1,397	1,804	2,520	60	256	860	1828
32	1,535	934	3,450	3,382	106	56	2969	813
33	1,232	1,508	2,609	5,355	173	80	212	44
34	1,271	1,933	1,373	604	152	193	115	893
35	1,267	1,876	16,300	1,729	92	244	5586	846
36	1,929	1,314	8,892	9,565	197	612	9765	1425
37	5,722	1,977	2,474	8,139	244	3320	511	10837
38	1,781	2,761	12,343	17,347	141	281	2433	2807
39	2,398	1,633	10,480	8,078	179	482	2615	1680
40	1,024	2,109	4,600	2,306	172	306	1027	488
41	1,676	1,078	1,082	3,403	121	1046	216	2271
Promedios	1,564	1,488	7,276	4,229	113	438	2,268	2,068
Desv. Est.	852	567	9,242	3,646	201	579	2,697	2,502

Anexo II

Descripción del uso de las hojas de cálculo para la generación de series sintéticas

En el presente anexo se describe la construcción de series de tiempo sintéticas anuales y mensuales a partir de registros históricos de cuatro estaciones hidrométricas de la cuenca del Río Amajac. Para hacer la descripción de los libros de Excel mencionados en los párrafos siguientes, se hace referencia a las ecuaciones y cuadros descritos en el presente documento en conjunto con los libros de Excel que acompañan la tesis. Para ubicación, se hace referencia a Columnas, Renglones y celdas y se menciona la hoja de cálculo en la que se encuentran los desarrollos de las ecuaciones.

El libro “Box-Jenkins estación venados” contiene varias hojas de cálculo en las cuales se desarrollan las etapas de la construcción de un modelo ARMA(p,q) o modelos Box-Jenkins: dichas etapas se describen a detalle en el apartado 4.4.

El libro “Generación de series sintética con el modelo multivariado AR(1)” muestra la generación de series de tiempo sintéticas multivariadas en el uso del modelo de Matalas, (1967). Para la generación de las series sintéticas multivariadas se utiliza la información de las cuatro estaciones hidrométricas de la cuenca del Río Amajac.

En el libro “Modelo de disgregación univariado de la estación Venados” se describe la generación de series mensuales a partir de series anuales, para ello se utilizan las series sintéticas que fueron previamente y generadas a partir de registros históricos.

A1.1 Modelos Box-Jenkins para la estación Venados

Para la construcción del modelo univariado de la estación Venados se utiliza el libro “Box-Jenkins estación Venados” hasta que se indique lo contrario. Este libro contiene diferentes hojas de cálculo a las cuales se hará referencia a lo largo de este apartado.

Para realizar la construcción de modelos de series de tiempo sintéticas con modelos ARMA(p,q) univariados, se debe verificar el ajuste de los datos históricos a la Función de Distribución Normal, o mediante la ecuación (4.2) llevara a cabo la transformación a la dicha función de probabilidad, este procedimiento se realiza como en la Hoja Etapa 1:Identificación, Columna E.

Una vez que se realiza la estandarización se procede a la “*Identificación*” del (los) modelo (s) que pueden representar a los datos de la estación Venados. Para la identificación del modelo se utilizan las FAC y FACP ecuaciones (6.1) y (6.2) respectivamente. El desarrollo de la FAC se muestra en las columnas J a la AG y se muestra el cálculo para r_1, \dots, r_{10} 10 lags, los valores de estos lags conforman la FAC, figura 7.2, los valores de la función se muestra en la fila 80 y en el cuadro 7.2. El cálculo de la FACP, se muestra en la columna AL.

Para el cálculo de la significancia estadística se obtiene primeramente el error estándar entre los valores estimados de la FAC y FACP y sus respectivos teóricos, para esto se utiliza la ecuación (4.18), el error estándar de los valores de la FAC se muestran en las celdas L102 a la L106. Enseguida se obtiene la significancia estadística de los valores de la FAC y FACP, ecuación (4.19), los valores de t para la FAC se muestran en las celdas L110-L114 y para la FACP L117-L121.

Como se observa en las celdas L110-L121, solamente los valores del primer Lag en la FAC y FACP cumplen con la condición de la ecuación (4.19), por lo que los modelos que pueden representar los datos son ARMA(1,0), ARMA(0,1) y la combinación entre estos ARMA(1,1), como se describe en el apartado 7.1.1.

Identificados los modelos, se continúa con la siguiente etapa:

La estimación de los parámetros se realiza para los tres modelos que fueron seleccionados como candidatos. Para esto se realiza la rutina de optimización descrita en el apartado 4.4.2.

La obtención de los parámetros del modelo se ejemplifica con la estimación del parámetro del modelo ARMA(1,0), la hoja de cálculo que obtiene el desarrollo tiene el mismo nombre, los parámetros de los otros modelos se estiman de igual forma, la estimación de los parámetros de los otros modelos se encuentran en hojas de cálculo cuyo nombre hace referencia al modelo.

En el cuadro 7.4 se muestra un esquema de la estimación del parámetro ϕ_1 del modelo ARMA(1,0). Los valores mostrados en la columna C de la hoja de cálculo “Etapa 2; ARMA(1,0)”, son los mismos valores estandarizados que se utilizaron en la etapa de identificación.

En primero lugar se calculan las series con un valor inicial $\phi_1 = 0.75$, y se obtiene una serie de datos alterna a la histórica, que dista de ser igual a original y de cuyas diferencias entre la serie histórica y estimada brinda un valor de la sumatoria de las a_t , figura A2.1. Para estimar el parámetro ϕ_1 se deberá seleccionar el valor ϕ_1 cuya sumatoria de las diferencias a_t sea mínima, ecuación (4.22), cuyas restricciones son las ecuaciones (4.23) y (4.24).

Para obtener la mínima diferencia se utiliza el solver del Excel, figura A2.2 y A2.3.

	D	E	F	G	H	I	J	K	
Box-Jenkins Etapa 2: Estimación									
Esta hoja de cálculo contiene la estimación de los parámetros phi y theta de los valores de los originales de la estación venados, es decir sin ningún lag para el modelo ARMA (1,0)								Phi=	0.2397
t	valores orig.			a_{t-1}	z_t	a_t	a_t^2		
0									
1	-0.225	-0.2248	0	0	-0.225	0.0506			
2	-0.536	-0.5362	-0.2248	-0.0539	-0.482	0.2326			
3	-0.152	-0.1521	-0.4823	-0.1285	-0.024	0.0006			
4	-0.810	-0.8098	-0.0236	-0.0365	-0.773	0.5981			
5	0.630	0.6301	-0.7733	-0.1941	0.824	0.6793			
6	0.561	0.5610	0.8242	0.1510	0.410	0.1681			
7	-0.617	-0.6165	0.4099	0.1345	-0.751	0.5640			
8	3.190	3.1904	-0.7510	-0.1478	3.338	11.1439			
9	-0.062	-0.0620	3.3382	0.7648	-0.827	0.6836			
10	-0.062	-0.0617	-0.8268	-0.0149	-0.047	0.0022			
11	-0.416	-0.4159	-0.0468	-0.0148	-0.401	0.1609			
12	-0.606	-0.6056	-0.4011	-0.0997	-0.506	0.2559			
13	-0.717	-0.7171	-0.5059	-0.1452	-0.572	0.3271			
14	-0.737	-0.7372	-0.5719	-0.1719	-0.565	0.3196			
15	-0.713	-0.7130	-0.5653	-0.1767	-0.536	0.2876			
16	0.844	0.8439	-0.5363	-0.1709	1.015	1.0298			
17	-0.420	-0.4198	1.0148	0.2023	-0.622	0.3870			
18	1.402	1.4016	-0.6221	-0.1006	1.502	2.2566			
19	4.291	4.2908	1.5022	0.3360	3.955	15.6404			
Etapa 2; ARMA(1,1) \ Etapa 2; ARMA(1,0) \ Etapa 2; ARMA(0,1) \ E 3; Diagnóst									

Figura A2.1 Esquema de obtención del parámetro ϕ_1

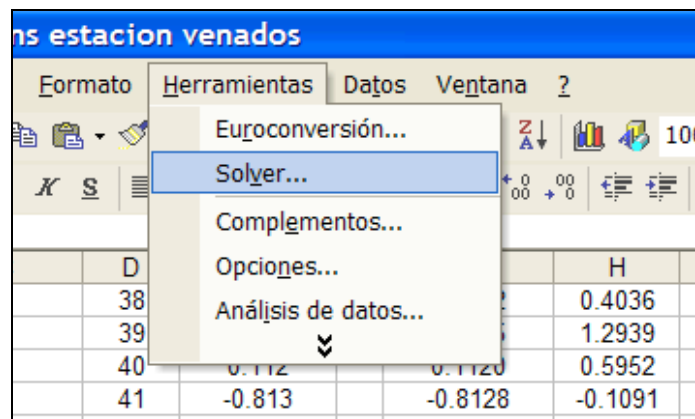


Figura A2.2. Invocación de la rutina del solver.

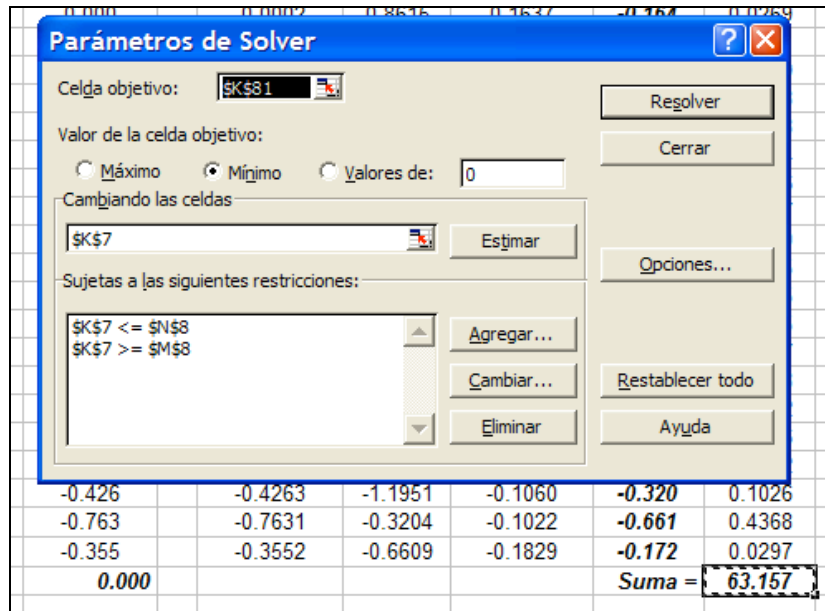


Figura A2.3 Parámetros de solver.

Los parámetros requeridos para optimizar los valores son: *celda objetivo*, en esta celda se deberá tener la sumatoria de las a_t ; *cambiando las celdas*: el valor de esta celda es el del parámetro ϕ_1 ; *sujeta a las restricciones*: en este paso se incluyen las restricciones de las ecuaciones (4.23) y (4.24).

El diagnóstico se realiza a las diferencias o residuos a_t que se obtuvieron para cada modelo ARMA(p,q), para el caso del modelo ARMA(1,0) se encuentran en la columna J.

Para hacer el diagnóstico se obtiene la FAC residual, para ello se aplica la ecuación (4.16) pero ahora a la serie de residuos que resulta de estimar los parámetros de los modelos.

La ecuación (4.27) analiza los valores de la FAC residual como un conjunto, el desarrollo de esta ecuación se encuentra en la celda M89, para el diagnóstico de los tres modelos.

Para realizar el pronóstico, con modelos ARMA se presentan dos alternativas, la primera a corto plazo, esta se realiza con las ecuaciones 4.28 a la 4.31 y la segunda con la generación de series de tiempo sintéticas con longitud igual al periodo que se desea analizar.

A1.2 Modelo de disgregación

Para ejemplificar el desarrollo de un modelo de disgregación se hará referencia al libro “*Modelo de disgregación univariado de la estación Venados*” a menos que se indique lo contrario.

Para la aplicación del modelo de disgregación se utilizan series sintéticas que se generaron previamente. Para ilustrar el uso del modelo, se utilizará la serie que se muestra en el cuadro 7.8, esta serie se obtuvo con el modelo ARMA(1,0) para la estación Venados.

El primer paso para la modelación de series sintéticas es la transformación a la función de probabilidad Normal, en este caso se aplicó a los logaritmos naturales de los valores históricos como se muestra en el libro, el proceso de normalización se muestra en las celdas E5-E72 a la AQ5-AQ72 que están en color azul.

Con la información que se obtuvo de la transformación a la PDF Log Normal, se construyen los modelos AR(1), un modelo anual y uno para cada mes del año, como se muestra en la hoja de cálculo. El modelo anual se muestra en las celdas J86-J185; para el mes de Enero son las celdas R86-R185 y de igual manera para los otros meses del año, hasta Diciembre cuyas celdas son CF86-CF185. La sumatoria de los volúmenes sintéticos mensuales q_i donde $i = 1$ para Enero...12 para Diciembre se muestra en las celdas CI86-CI185.

A continuación se describe como utilizar la hoja para la disgregación de datos sintéticos anuales a mensuales.

1A. Generar de series sintéticas anuales con modelo de las celdas J86-J185. Se deberá seleccionar la serie sintéticas con las propiedades estadísticas que más se acerquen a las de la serie histórica. Una vez seleccionada esta serie se deberá copiar en las celdas CL86-CL185.

En la columna CO se muestra el cálculo de las diferencias entre las series anuales de la columna CL (que es la serie sintética anual encontrada en el paso 1A) y el total anual de la sumatoria de las series sintéticas mensuales, columna CI. El objetivo es encontrar los valores de la de la columna CI iguales a los de la columna CL, para esto se usa el criterio de mínimas diferencias de la ecuación (4.63).

2A. Los valores sintéticos de la columna CL, se obtienen uno por uno, esto es, para el primer valor de la serie anual, la celda CL86, se debe encontrar **el** valor de la serie de la columna CI86-CI186 **con la mínima diferencia** respecto a la celda CL86, ecuación (4.63).

3A. En las celdas CR86-CR185 a la DD86-DD185 se calculan las fracciones de los modelos mensuales con respecto a su total.

Las fracciones del valor encontrado en el paso 2A, se copian en las celdas DG86-DS86 y se multiplican por el valor de la celda CL86, de esta manera se obtienen los fragmentos mensuales de la celda CL86 (celdas DU86-EG86).

Para el siguiente valor de la serie anual, la celda CL87 se sigue el mismo procedimiento. Con la tecla F9 se generan nuevos cálculos de las celdas en toda la hoja, ahora se debe encontrar un valor en la columna CI con la diferencia mínima con respecto a la celda CL87, los fragmentos correspondientes a este

valor se copian ahora en las celdas DG87-DS87. Posteriormente se multiplican por el valor de la celda CL87 y se obtienen los valores mensuales para el segundo valor de la serie de la serie anual.

Este procedimiento se sigue tantas veces como valores sintéticos se desee encontrar.

Para la disgregación de series de modelos multivariados AR(1), es el mismo procedimiento, se disgrega un sitio a la vez.