



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAestrÍA Y DOCTORADO EN INGENIERÍA

FACULTAD DE INGENIERÍA

MODELOS DE HIDROLOGÍA ESTOCÁSTICA

T E S I S

QUE PARA OPTAR EL GRADO DE:

MAESTRA EN INGENIERÍA

INGENIERÍA DE SISTEMAS – INVESTIGACIÓN DE OPERACIONES

P R E S E N T A:

ACT. GABRIELA VÁZQUEZ ENRIQUEZ

Tutor: DR. SERVIO TULIO GUILLÉN BURGUETE



CIUDAD UNIVERSITARIA

2006



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Jurado Asignado:

Presidente: Dr. Fuentes Maya Sergio
Secretario: Dra. Elizondo Cortes Mayra
Vocal: Dr. Guillén Burguete Servio Tulio
1^{er} Suplente: M.I. Téllez Sánchez Rubén
2^{do} Suplente: Dr. Ordorica Mellado Manuel

Universidad Nacional Autónoma de México
Posgrado de la Facultad de Ingeniería

Tutor de Tesis :

Dr. Guillén Burguete Servio Tulio

FIRMA

Agradecimientos

Este documento es fruto de un trabajo colectivo en el que intervinieron muchas personas que, de una u otra forma, fueron importantes para la conclusión del mismo y a las que agradezco por todo su apoyo y amistad.

A mi familia, por siempre estar a mi lado.

A todos y cada uno de mis amigos

A mis sinodales, Dra. Elizondo Cortes Mayra, M.I. Téllez Sánchez Rubén y al Dr. Ordorica Mellado Manuel y especialmente al Dr. Fuentes Maya Sergio por su apoyo y confianza.

Al Dr. Servio Tulio Guillén, por guiarme y aconsejarme durante este trabajo.

Introducción	I
1. Hidrología en México	1
1.1 Introducción	1
1.2 Marco de Referencia	2
1.2.1 Geografía	2
1.2.1 Demografía	2
1.2.2 Orografía, Hidrología y Vertientes.	3
1.2.3 Climas.	4
1.2.4 Desigual Distribución Espacial y Temporal de las Precipitaciones.	4
1.3 Tipología de Cuencas en México.	5
1.3.2 Cuencas Hidrológicas	8
1.3.3 Principales Ríos	9
1.3.4 Regiones Hidrológicas	10
1.4 Desafíos en Materia de Agua	11
1.5. La Gestión del Agua en México	12
1.5.1 Objetivos Generales de los Consejos de Cuencas	15
1.5.2 Funciones Generales	16
1.5.3 Funciones Específicas	16
1.6 Situación Actual	18
2. Hidrología Básica	22
2.1 Introducción	22
2.1.1 Objetivos de los Estudios Hidrológicos	23
2.2 Ciclo Hidrológico	27
2.3 Hidrología Estocástica	30
2.3.1. Modelos Hidrológicos.	32
2.4 Modelación de la incertidumbre Hidrológica	35
2.4.1 Propiedades de las Series Hidrológicas	36
2.4.2 Características de las Estadísticas Disponibles	37
2.5 Hidrología Aplicada	39
3. Modelos de Planificación de Sistemas Hidrológicos	45
3.1 Introducción	45
3.2 Esquema general de modelación	46
3.2.1 Modelos Típicos Disponibles	48
3.2.2 Propiedades Estadísticas a Modelar	49
3.2.3 Periodicidad y su Representación	51
3.3 Modelación de la Dependencia Temporal	52
3.3.1 Identificación de Modelos Temporales Periódicos	52
3.3.2 Estimación de Parámetros	54
3.4 Construcción del Modelo	56
3.5 Diagnóstico o Verificación del Modelo.	57
3.5.1 Pruebas de Independencia	58
3.5.2 Prueba de Normalidad	59
3.5.3 Promedio Nulo y otras Propiedades	59
3.5.4 Reformulación	60
3.5.5 Sobreajuste	60
3.6 Modelado de la Dependencia Espacial.	60
3.7 Uso de los Modelos	63
3.7.1 Simulación	63
3.7.2 Pronóstico	66
3.8 Otros modelos	69
3.8.1 Modelo GOL y el Paradigma de la PDE	69
3.8.2 Modelo OMSIC	71
3.8.5 Cambio de Paradigma: De la PDE (Programación Dinámica Estocástica) a la PDDE(Programación Dinámica Dual Estocástica)	72
3.8.6 Modelo SDDP	75
3.8.7 Modelo PLP (Programación de Largo Plazo)	77

3.9 Limitaciones de los Modelos	79
3.10 Generación de Escenarios para la Hidrología Futura	82
4. Modelado de Series de Tiempo Hidrológicas	83
4.1 Introducción	83
4.2 Modelos Autoregresivos	84
4.3 Modelos Autoregresivos de Medias Móviles	85
4.3.1 Propiedades de los Procesos ARMA (p,q)	85
4.4 Modelo para Series de Tiempo Periódicas: El modelo PARMA	87
4.4.1 Propiedades de los Procesos PARMA	88
4.5. Modelado de la Dependencia Espacial	90
4.6. Metodología de Ajuste de un Modelo PARMA	95
4.6.1 Identificación	96
4.6.1.1. Función de Autocorrelación Simple (FACS)	96
4.6.1.2 Función de Autocorrelación Parcial (FACP)	97
4.6.1.3 Procedimiento de Identificación	98
4.6.2 Estimación de Parámetros	100
4.6.3 Test de Diagnóstico	101
4.6.4 Transformación de los Datos.	101
5. Modelo para Selección de Caudales Históricos	103
5.1 Introducción	103
5.2 Modelo de Selección, Basado en un Modelo PARMA	104
5.3 Árbol de Escenarios	109
5.4 Aplicación del Modelo	109
5.5 Validez y Limitaciones del Modelo	115
5.6 Perspectivas	116
5.6.1 Espacio Vectorial de las Variables Aleatorias Normales	117
5.7 Número de Aperturas en Cada Etapa	120
5.7.1 Modelo de Sensibilidad de la FCF	121
Conclusiones y Recomendaciones	126
Apéndice I - Series de Tiempo	129
1. Introducción	129
2. Teoría	130
2.1 Series de Tiempo	130
2.2 Tendencia	130
2.3 Variaciones Estacionales.	132
2.3.3 Variaciones Cíclicas	132
2.3.4 Variaciones Residuales	132
3. Análisis de la Tendencia	133
3.1 Método Gráfico	133
3.2 Método de las Medias Móviles	134
4. Aplicación	134
Apéndice II.-Modelos ARIMA	139
1. Introducción	139
2 Definición y Conceptos Básicos de los Modelos ARIMA	140
2.1 Proceso Estocástico y Estacionalidad	140
2.2 Modelos Autoregresivos	143
2.4 Modelo de Medias Móviles	144
2.5 Interpretación de un Modelo de Medias Móviles	144
2.6 Condiciones y Raíces Unitarias para los Modelos AR y MA	145
2.7 La Estacionalidad de las Series Temporales en la Realidad	146
3 Modelo ARIMA(p,d,q) SARIMA(P,D,Q)	150
3.1 Funciones de un Proceso Estocástico Estacionario	150
3.2 Aplicación de estas Funciones a Muestras Concretas	152
4. Proceso Estocástico Estacionario Lineal Discreto	154
4.1 Modelos MA(1)	156
4.2 Modelos AR(1)	159
5. Identificación del Modelo	162

6 Breve Reseña sobre la Estimación de los Modelos ARIMA	163
6.1 Problemas Iniciales: los Valores Iniciales y la No Linealidad.	164
6.2 Predictor Óptimo.	165
6.3 Predicción Puntual	166
7 Características de las Predicciones Realizadas con Modelos ARIMA	168
7.1 Selección de Modelos	168
Referencias	170
Hidrología Básica	170
Hidrología Aplicada.	171
Hidrología Estocástica.	171
Modelos ARIMA	171
Consultas de internet	172

Introducción

Antecedentes

La Hidrología es la ciencia que estudia la distribución, cuantificación y la utilidad de los recursos hídricos que están disponibles en la Tierra. Estos recursos se distribuyen en la atmósfera, la superficie terrestre y las capas del suelo y son estudiados por diferentes ramas de la Hidrología.

La Hidrología Básica estudia los conceptos físicos del ciclo hidrológico, los métodos de recolección de información hidrológica y los procedimientos clásicos de procesamiento de datos estadísticos. Las técnicas que permiten la utilización de los recursos hidráulicos en proyectos de Ingeniería pertenecen al campo de la Hidrología aplicada. Los eventos hidrológicos, tales como aguaceros, caudales, niveles de embalse, etc., son eventos estocásticos y son estudiados por la Hidrología Estocástica, la cual aplica la teoría de los procesos estocásticos y la programación lineal matemática a la solución de problemas de optimización y al estudio de los eventos hidrológicos, que son parte de la ingeniería de sistemas. La Hidrología Estocástica se caracteriza porque, de un lado tienen un patrón medio de comportamiento a largo plazo, y por el otro el pronóstico de sus magnitudes en un momento dado tiene un mayor o menor grado de incertidumbre. El patrón medio corresponde a lo que se denomina la tendencia general o componente determinística y la incertidumbre constituye la componente aleatoria del evento.

Si se desea ser más realista, en la mayoría de las veces, sólo se cuenta con valores estimados o valores promedio y por lo tanto deben considerarse como tales, es decir como observaciones probabilísticas.

Por lo anterior, cuando en un problema de optimización matemática, se tiene una o varias variables aleatorias involucradas con distribuciones conocidas o no, se dice que el modelo corresponde a un problema de optimización estocástica, también llamado de programación estocástica.

La optimización estocástica como parte de la programación matemática ha sido relativamente poco explorada a pesar de su gran utilidad y versatilidad en el modelo de problemas con incertidumbre. Los modelos estocásticos ofrecen una gran variedad de posibilidades de investigación en aspectos relacionados con la optimización y el control de sistemas.

En algunos problemas de optimización, el decisor determina un conjunto de variables de control sujetas a un conjunto de restricciones; en un problema de optimización estocástica, el decisor tiene control sobre las variables de decisión pero no sobre las variables aleatorias que intervienen en el problema.

Problemática

Los extremos hidrológicos y en particular las precipitaciones intensas causantes de inundaciones son eventos naturales que han existido desde la antigüedad. En el presente siglo han sido muchos los eventos de este tipo en todo el mundo.

En los últimos años los daños producidos por los extremos hidrológicos se han incrementado. Estos han producido mayores daños debido al crecimiento poblacional y la urbanización en sitios de potencial peligro. En años recientes se presentaron en México huracanes que afectaron a varios estados de la República Mexicana, especialmente las zonas cercanas a los ríos y a la costa.

Existe la posibilidad de poder prevenir este tipo de catástrofes evitando pérdidas económicas, materiales y sobre todo pérdidas humanas. En la actualidad algunos países realizan estudios y tienen programas de monitoreo hidrológico, estos han sido probados y aplicados con gran éxito.

Al mismo tiempo existen modelos para aprovechar los extremos hidrológicos y transformarlos en ganancias financieras, en algunos países latinos como Argentina, Colombia y Chile entre otros ya son aplicados estos proyectos y actualmente tienen gran éxito, se ha dado entre estos, políticas de compartir información entre ellos, pero también proporcionar información requerida a otros países.

En México la participación y el interés sobre este tema es muy poco, debido a la falta de información y personas dedicadas a este tema.

Objetivo

El propósito central de este trabajo es dar a conocer los modelos que están siendo utilizados en la programación hidrológica estocástica en otros países principalmente Argentina, Chile y Colombia ya que tiene similitud a la situación actual de México, para abarcar el problema de la aleatoriedad en extremos hidrológicos, así como la comparación de éstos para saber en qué casos es conveniente usarlos.

Hipótesis

En muchos sistemas de tiempo con características estocásticas, es posible aproximarse con sistemas discretos y obtener los datos estadísticos a intervalos fijos de tiempo mediante proceso de programación estocástica, ejemplos de ellos se pueden encontrar en líneas de producción, sistemas flexibles de manufactura, sistemas de tráfico. Por lo que puede ser aplicado en sistemas hidrológicos.

Algunos modelos desarrollados tienen características muy específicas, de acuerdo a cada situación, es posible aproximarse a la realidad, aunque existe el riesgo de crear modelos demasiado complejos que son complicados de resolver.

El problema de incertidumbre es resuelto aplicando variables aleatorias al modelo y estos modelos pueden ser aplicados para poder predecir el comportamiento de caudales de ríos y evitar pérdidas materiales y humanas.

Contenido

El trabajo está desarrollado de la siguiente manera:

A lo largo del capítulo 1, se da una visión general de la situación, geográfica y de la manera en que esta estructurados y organizados los sistemas de cuencas y ríos en México, para conocer las condiciones en las que se esta trabajando y las posibles aplicaciones de otros modelos

En el capítulo 2, se presentan algunos antecedentes de la Hidrología Básica, sus componentes, su historia, el papel que desempeña, así como la importancia que tiene en la actualidad y poder entender los conceptos y modelos que se presentan en los siguientes capítulos.

En el capítulo 3, se desarrolla el tema de los Modelos de Planificación de Sistemas Hidrológicos, se describe el estado del arte sobre este tema, se mencionan las principales metodologías y aportaciones más relevantes que se han originado en esta área en países como Argentina, Chile y Colombia, donde ya se han desarrollado algunos modelos.

En el capítulo 4, se da una breve clasificación de los algoritmos más relevantes para el Modelado de Series Hidrológicas, algunos criterios de comparación entre los algoritmos y algunos aspectos relacionados con su eficiencia probada en países antes mencionados

En el capítulo 5, Finalmente se hace un análisis comparativo entre los Modelos para la selección de caudales Históricos, se muestran algunas ventajas y desventajas que se han podido observar en estos tres países y el posible éxito en México.

Se cuenta con dos apéndices, el primero de ellos es relacionado con Series de Tiempo y el segundo con el trabajo de modelos ARIMA, desarrollados por Box y Jenkins, ya que esta información se menciona a lo largo de la tesis y da una visión introductoria acerca del tema

1. Hidrología en México

1.1 Introducción

El agua es un recurso natural finito, que se encuentra desigualmente distribuido en el tiempo y en el espacio, es altamente vulnerable a la contaminación y es imprescindible para sostener la vida, el desarrollo económico y los ecosistemas. Esta condición específica del agua obliga a los gobiernos, a las empresas y a los ciudadanos, a actuar con responsabilidad, alentando la coordinación, la cooperación y la complementariedad en sus ámbitos de actuación.

En materia de agua, México impulsa una reforma profunda de su sector hidráulico para enfrentar con mayores posibilidades de éxito, el reto de mantener el crecimiento de la economía nacional sin afectar la sustentabilidad de los recursos naturales y atender las necesidades de agua y servicios de una población que supera ya los 100 millones de personas y presenta rezagos importantes en los servicios básicos de abastecimiento de agua, drenaje y saneamiento.

El agua por las implicaciones sociales, económicas y políticas de su manejo, y por la incertidumbre de las variables que determinan el comportamiento del ciclo natural que globalmente la reproduce y recicla, debe ser considerada un asunto estratégico y de seguridad nacional por todos los países y gobiernos.

La gestión mexicana actual del agua es más integral; y reconoce a las cuencas hidrográficas y a los acuíferos, como las unidades territoriales más apropiadas para lograr su manejo eficaz.

La generación de energía es una actividad de importantes repercusiones para el desarrollo productivo, la calidad ambiental y el crecimiento económico de la sociedad. La Ingeniería en distintas especialidades, pero particularmente en la Eléctrica, está comprometida, por la magnitud de sus relaciones con el medio ambiente y por el uso de los recursos hidráulicos para generar energía, con el estudio de los fenómenos meteorológicos que originan la lluvia; de las características y propiedades de las cuencas como unidades geográficas de conversión de la lluvia en caudales aprovechables; de las propiedades básicas de los fluidos y su movimiento a través de conductos artificiales y del diseño técnico y económico de estructuras, instalaciones y equipos utilizados para el aprovechamiento de la energía hidráulica.

1.2 Marco de Referencia

1.2.1 Geografía

México se ubica en el hemisferio norte. Su territorio forma parte de las grandes cuencas definidas por los sistemas hidrológicos de escala continental. Su costa este colinda con el océano Pacífico que a su vez forma parte de la llamada cuenca del pacífico; por el oeste se tiene la vertiente del golfo al que desembocan grandes cuencas como la del Río Misisipi que cubre gran parte de los EUA.

En el centro norte se ubican cuencas cerradas. Limita al norte a través de 3,152 km; con los Estados Unidos de América y al sur con los países centroamericanos de Guatemala y Belice con quienes comparte una frontera de 1,149 km; al este con el golfo de México, con 3,294 km; y al oeste con el océano Pacífico, con 7,828 km. Conforme su ubicación geográfica, su territorio suma cerca de 2 millones de km² y forma parte de las grandes cuencas hidrográficas del continente americano. MEXICO: Estructura y organización política.

1.2.1 Demografía

La población mexicana alcanzó a finales del año 2000 los 100 millones de personas y crece a un ritmo promedio del orden de 1.2 por ciento anual. Se encuentra distribuida en 80 ciudades mayores de 100 mil habitantes que concentran el 48 por ciento de la población total; 2,592 localidades con una

población entre 2,500 y 100 mil habitantes que suman otro 27 por ciento de la población nacional y con cerca de 200 mil pequeñas poblaciones de menos de 2,500 habitantes en las que viven el restante 26 por ciento de la población total. La conurbación de la Zona Metropolitana de la Ciudad de México concentra una población de 19.6 millones de habitantes. La distribución demográfica mexicana se caracteriza como de alta concentración urbana y fuerte dispersión rural.

1.2.2 Orografía, Hidrología y Vertientes.

México es un país de montañoso. En su territorio se forman cadenas de montañas que definen su hidrología y sus cuencas. En la vertiente del océano pacífico existe una gran diversidad de pequeñas y medianas cuencas con trayectorias cortas y pendientes muy pronunciadas. En el Golfo de México se forman extensas planicies a partir de una hidrología que se forma en las montañas y desciende hacia el Golfo de México. En ambos casos, la hidrología está definida por la posición geográfica del país que se ubica al paso de los grandes movimientos ciclónicos que se forman tanto en el océano Pacífico como en el Atlántico y en el mar Caribe. Cada año ocurren 30 eventos ciclónicos en promedio. Entre 4 y 5 causan severos daños.

Al analizar el comportamiento de la precipitación en gran parte de la Republica mexicana se establece dos periodos bien marcados; la estación lluviosa y la poco lluviosa o seca. Acumulándose en la primera aproximadamente el 90% de la lluvia anual. En la figura 1 se puede observar la distribución histórica de la precipitación media anual en el territorio nacional. A la vez se señala que en los estados a lo largo de la costa del Golfo de México, centro del país, parte del sur de la costa del Pacífico, así como de la península de Yucatán, las precipitaciones presentan una distribución bimodal dentro del periodo de lluvias con máximas en Junio y Septiembre y mínimas en Julio y Agosto, caracterizándose este ultimo mes por una disminución significativa en los días con lluvia y un aumento de las temperaturas máximas, asociado a una menor nubosidad y a una mayor incidencia de la radiación solar.

1.2.3 Climas.

En las cuencas y subcuencas hidrológicas de México se encuentran todos los Climas. En el norte, predominan los climas secos en grandes zonas desérticas, áridas y semiáridas. En el centro del país, los climas son templados. En el sur, los climas son tropicales y húmedos. Fig. 1.1

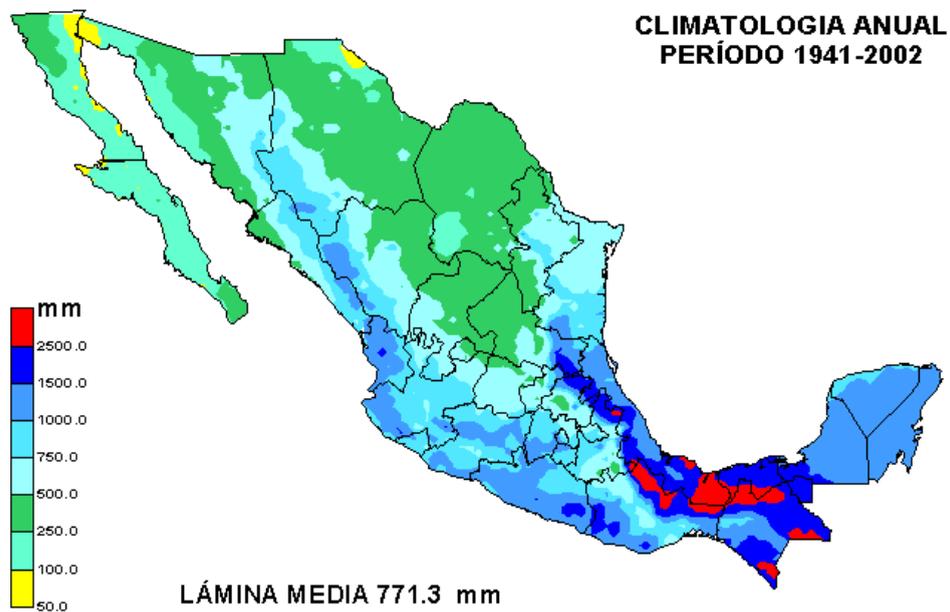


Figura 1.1. Precipitación media anual (1941-2002) CNA
Fuente: Instituto de Astronomía y Meteorología

1.2.4 Desigual Distribución Espacial y Temporal de las Precipitaciones.

Las precipitaciones pluviales están desigualmente distribuidas en el territorio, provocando problemas de escasez en gran parte del país y de abundancia y drenaje en otras zonas. Las zonas áridas y semiáridas del norte, ocupan 40 por ciento del territorio nacional y las precipitaciones son menores a 500 milímetros por año. Otro 30 por ciento del territorio corresponde a las mesetas centrales que reciben precipitaciones entre 500 y 1, 000 milímetros por año. El restante 30 por ciento, está ocupado por zonas tropicales y semi tropicales en donde se producen cada año precipitaciones entre 1,000 y 2,000 milímetros, pero donde también, en algunos años pueden ocurrir precipitaciones de 8 ó 10,000 milímetros.

1.3 Tipología de Cuencas en México.

Por la complejidad de territorio, México cuenta con todo tipo de cuencas, abiertas y cerradas; costeras y de montaña; grandes, medianas y pequeñas. Una de las clasificaciones divide al país en 837 cuencas hidrográficas de diferentes características y tamaños.

Para efectos de planificación, de organización institucional y de participación social, ha sido necesario el agrupamiento de las cuencas en sistemas hidrológicos completos y en lo que hemos llamado sistemas hidrológico – administrativos.

Con el propósito de administrar y regular los usos y aprovechamientos de las aguas superficiales y subterráneas, se dividió el país en 13 Regiones Hidrológico Administrativas. En cada región existe una Gerencia Regional de la Comisión Nacional del Agua que es la entidad federal reconocida por la Ley de Aguas Nacionales como la autoridad en la materia.

Otra división del país realizada a partir de su hidrología es la que define regiones de participación social. Esto se hizo así para tomar en cuenta el tamaño de los territorios regionales y para facilitar la intervención de los usuarios y de otros grupos organizados de la sociedad en el manejo del agua a través de Consejos de Cuenca y otras formas de organización social en los ámbitos de subcuenca y acuífero.

Las principales cuencas hidrográficas de México son las siguientes (Fig. 1.2):



Figura 1.2. México: División por cuenca hidrológica

Fuente: Director General de la Comisión Nacional del Agua, México.

Cuenca del Río Bravo.- Es una cuenca de más de 457 mil km². El 51 por ciento de su territorio se encuentra en los Estados Unidos de Norteamérica y el 49 por ciento es territorio mexicano. El cauce principal sirve de límite fronterizo entre México y los EUA por más de 2 000 km.

En la parte mexicana de la cuenca residen poco más de 8 millones de personas y es una de las zonas de mayor crecimiento demográfico (7 % anual), por lo que el manejo y administración eficaz de los recursos hídricos constituye uno de los grandes desafíos.

En los últimos 10 años se ha presentado una prolongada sequía que ha dificultado cumplir compromisos que se tienen establecidos entre México y los EUA desde 1944 en que se suscribió el Tratado de Límites y Aguas entre los dos países. Por ésta razón, en los últimos años han existido fuertes diferencias entre los dos países, que están siendo encauzadas mediante el incremento de la eficiencia en los usos actuales del agua alentando la modernización y tecnificación de las áreas de riego existentes, así como a través de la reglamentación de la distribución de las aguas superficiales de la parte mexicana de la cuenca.

La Cuenca del Río Colorado también es frontera entre México y los Estados Unidos; sin embargo, en este caso, más del 99 por ciento del territorio de la cuenca se encuentra en los Estados Unidos y solo el uno por ciento en México; esto hace que México resienta los efectos del manejo del agua y de la cuenca que se hace en la parte alta (EUA). Actualmente es una preocupación la disminución de los caudales del río y su repercusión en la ecología de la zona del delta que se encuentra en el extremo norte del golfo de California.

Cuencas de los Ríos Grijalva y Usumacinta.- En el Sur, México limita con las Repúblicas de Guatemala y Belice, y a través de estos países se vincula con la región Centro Americana. En esta zona se localizan los ríos Grijalva y Usumacinta que son los más caudalosos de México y proporcionan la mayor parte de la energía hidroeléctrica del país.

Cuencas peninsulares.- En las penínsulas de Baja California y de Yucatán se depende de las aguas subterráneas.

En Baja California el clima es semidesértico y las precipitaciones son muy escasas la mayor parte del año, por lo que prácticamente no existen corrientes superficiales permanentes.

La Península de Yucatán por su parte, está formada por un macro acuífero regional. Por las características del suelo, los escurrimientos rápidamente se infiltran. Su clima es semitropical y el terreno es de lomeríos muy suaves. Esta es la zona de la cultura maya ampliamente conocida en el mundo.

Cuencas cerradas.- Una parte del territorio mexicano está ocupado por cuencas endorreicas.

El Valle de México hasta el siglo pasado era una cuenca cerrada. Actualmente drena artificialmente parte de sus aguas hacia el Golfo de México a través de la cuenca del Río Pánuco. Para satisfacer las necesidades de agua que demanda la zona metropolitana de la Ciudad de México, en la que residen 19.6 millones de personas según los datos del censo de 2000, se conecta con las cuencas de los Ríos Lerma y Balsas las que transfieren parte de sus disponibilidades.

En el norte centro se ubican las cuencas de los Ríos Nazas y Aguanaval y las cuencas del Altiplano. Es una zona semidesértica en la que se encuentran importantes ciudades mexicanas como San Luis Potosí, Saltillo y Torreón, que son polos de desarrollo agropecuario e industrial.

Cuencas Costeras.- Tanto en la vertiente del pacífico, como en la vertiente del Golfo, México tiene una gran cantidad de ríos, resultado de su orografía.

Las cuencas que se forman por el paralelismo de las cordilleras con las costas, son relativamente pequeñas y con ríos de corta trayectoria y grandes pendientes, por lo que sus caudales son de difícil aprovechamiento, salvo en el pacífico norte, en donde se encuentra las mayores y más prósperas zonas de riego del país.

Macrocuencas.- Los grandes sistemas hidrológicos de México están representados por los Ríos Grijalva y Usumacinta en el sur del país; Coatzacoalcos, Papaloapan y Pánuco en el golfo; Balsas, Lerma y Santiago en el centro y el Bravo en el Norte.

Todas estas cuencas tienen grandes territorios de drenaje. De éstas la más pequeña es Coatzacoalcos con 24,000 km² y la más grande es Bravo con 227,000 km² sólo en la parte mexicana.

Todas las grandes cuencas mexicanas están claramente diferenciadas tanto por sus características hidrográficas como por su desarrollo socio-económico. Por ejemplo, en el sistema Grijalva Usumacinta se genera el 25 por ciento de la energía hidroeléctrica del país. En la del Río Coatzacoalcos se produce el 15 por ciento del petróleo nacional. Las cuencas del Papaloapan y del Río Pánuco son importantes por su producción agropecuaria. La del Río Lerma concentra 13 por ciento de la población del país y el 17 por ciento de la industria nacional.

1.3.2 Cuencas Hidrológicas

El agua que escurre en un río es captada en un área determinada, por lo general por la conformación del relieve. A esta área se le llama cuenca hidrológica. A su vez, las cuencas hidrológicas se agrupan en regiones hidrológicas. En este mapa se presenta la división del país en cuencas hidrológicas, indicando mediante colores la abundancia de agua en cada cuenca. Se ilustran en color rojo, las localidades con más de 500 mil habitantes hasta 1995, y en líneas de color azul los ríos más caudalosos del país. Observe la relación entre los principales ríos y el nivel de escurrimiento superficial de la cuenca que ellos atraviesan. Las zonas representadas en color verde más claro tienen poca agua, y las más oscuras, la tienen en mayor cantidad. Los estados de Veracruz, Tabasco, Oaxaca y Chiapas, ubicados en las cuencas más húmedas, tienen un escurrimiento superficial mayor a los 10 mil litros por persona al año. En cambio, estados como Baja California, Baja California Sur o Coahuila tienen un escurrimiento superficial disponible no mayor a los 200 litros por persona. Fig.1.3



Figura 1.3 Cuencas Hidrológicas

Fuente: <http://mapserver.inegi.gob.mx/geografia/espaol/datosgeograf/basicos/hidrologia/rios/cuencas.gif>

1.3.3 Principales Ríos

Los ríos son las corrientes de agua que fluyen sobre sus cauces. Pueden ser de dos tipos, según su estacionalidad: perenes, con agua todo el año, e intermitentes, con agua sólo en alguna parte del año, por lo general la época de lluvias. Los principales ríos, por su escurrimiento medio anual en millones de metros cúbicos, son los siguientes: Usumacinta, 56 mil; Papaloapan, 47 mil; Grijalva, 25 mil; Coatzacoalcos, 22 500; Balsas, 14 500; Bravo, 13 mil; Pánuco, 12 mil; Hondo, 11 mil; Lerma-Santiago, 8 500; Tecolutla, 5 500; y Fuerte, 5 mil. Los cuatro principales ríos llevan al Golfo de México un total de 2.2 millones de litros de agua cada segundo. El Usumacinta contribuye con 900 mil, el Grijalva 700 mil, el Coatzacoalcos 400 mil y el Papaloapan con 200 mil. El río Bravo está considerado uno de los más largos del mundo, tiene 2 800 Kilómetros de longitud y un caudal máximo promedio cercano a los 120 mil litros por segundo. Fig. 1.4



Figura 1.4 Principales Ríos

Fuente: <http://mapserver.inegi.gob.mx/geografia/español/datosgeograf/basicos/hidrologia/rios/riosp.gif>

1.3.4 Regiones Hidrológicas

Una región hidrológica es la agrupación de varias cuencas hidrológicas con niveles de escurrimiento superficial muy similares. En México, las más húmedas son la número 30, llamada región del sistema Grijalva-Usumacinta; la número 29 o región del Coatzacoalcos; la número 28 o región del Papaloapan; y la número 23, llamada también región de la Costa de Chiapas. Las regiones hidrológicas más secas del país son la número 2, llamada región del Vizcaíno; la número 3 o región de la Magdalena; la número 4 o región de la Laguna Salada; la región 8 o región Sonora norte y la región 35, llamada comúnmente región del Mapimí. Las más densamente pobladas son la 29, llamada también región Tuxpan-Nautla y la región número 12, conocida como Lerma-Santiago. Uno de cada cuatro habitantes en localidades con más de 100 mil habitantes vive en estas regiones hidrológicas. Fig. 1.5



Figura 1.5 Regiones Hidrológicas

Fuente: <http://mapserver.inegi.gov.mx/geografia/español/datosgeograf/basicos/hidrologia/rios/regioneshidro.gif>

1.4 Desafíos en Materia de Agua

Al iniciarse el Siglo XXI México enfrenta diverso desafíos en materia de agua. Además de la desigual distribución regional y temporal de sus recursos hídricos, destacan los siguientes:

- La dotación de servicios de agua, alcantarillado y saneamiento para los sectores de la población que aún no cuentan con estos servicios en sus domicilios.
- Las aguas residuales que aún no reciben tratamiento previo antes de su regreso a los cauces naturales.
- El control de la erosión y la deforestación de las cuencas.
- El mantenimiento y operación de la infraestructura hidráulica existente y la construcción de la nueva. Actualmente se cuenta con más de 4,500 Presas de Almacenamiento y Derivación; 6.3 millones de hectáreas incorporadas al riego y 4,000 kilómetros de canales; 20 Grandes Presas de generación de energía eléctrica; más de 2,000 Plantas de tratamiento, y una infinidad de sistemas de abastecimiento, plantas de bombeo, etc.
- La prevención de los desastres asociados a los huracanes y las sequías
- La preservación de las aguas subterráneas
- La generación de inversiones suficientes para construir, rehabilitar y mantener la infraestructura hidráulica y sus servicios.

El agua es un asunto estratégico y de seguridad nacional.- Por su significado para el desarrollo de socioeconómico, administrar y manejar el agua racional y eficazmente se ha convertido en un asunto estratégico y de seguridad nacional. Sobre todo tomando en cuenta las enormes disparidades existentes entre la disponibilidad natural del recurso y la distribución de la población y la producción.

Por una parte la disponibilidad natural de agua per cápita en el sureste del país es siete veces mayor a la del resto del territorio nacional, por otra, en las regiones en donde se presenta el 32 por ciento del escurrimiento, habita el 77 por ciento de la población y se genera el 86 por ciento del Producto Interno Bruto.

Población y escasez de agua por cuenca (Año 2000).- Tomando como referencia las definiciones de escasez usualmente utilizadas por diversos organismos

internacionales, se calcula que las cuencas del Valle de México, Lerma Chapala y Altiplano disponen de menos de 1,000 m³ por persona y por año, y en ellas radica alrededor de 33 por ciento de la población nacional. Las cuencas de los Ríos Santiago y Bravo, así como la Península de Baja California también resienten ya la escasez de agua, pues en términos generales en esas regiones se dispone entre 1,000 y 1,700 m³ por persona y por año.

Población y escasez de agua por cuenca (Año 2020).- Para el futuro la situación de escasez de agua en México será mas compleja porque 57 millones de habitantes, equivalentes a 47 por ciento de la población estimada del año 2020, vivirá en cuencas de escasez extrema de agua; es decir, en cuencas cuya disponibilidad por persona y por año será menor a 1,000 m³. Otros 10 millones de personas vivirán en cuencas cuya disponibilidad por persona y por año será entre 1,000 y 1,700 m³.

1.5. La Gestión del Agua en México

El agua en México, junto con el petróleo y los minerales está considerada un “Bien público propiedad de la Nación” (Art. 27 de la Constitución) y no puede ser otorgada en propiedad, razón por la cual, el Gobierno Federal y específicamente el Presidente de la

República tiene la facultad de “permitir, asignar o concesionar el uso y aprovechamiento de las aguas superficiales y subterráneas”.

Para administrar las aguas nacionales el Gobierno Federal mexicano se auxilia de un organismo nacional denominado Comisión Nacional del Agua responsable de administrar, planificar y preservar las aguas nacionales.

Estructura del sistema actual de gestión del agua.- Actualmente, para llevar a cabo la gestión del agua, México cuenta con un sistema de carácter federal, estatal y municipal.

En el ámbito nacional federal, opera la Comisión Nacional de Agua con una organización central, regional y estatal basada.

En cada uno de los Gobiernos Estatales actualmente se promueve la constitución de una Comisión Estatal de Agua con el propósito de que a mediano plazo sea la receptora de las funciones y tareas que la federación dejará de realizar conforme avance el proceso de descentralización que se encuentra en marcha.

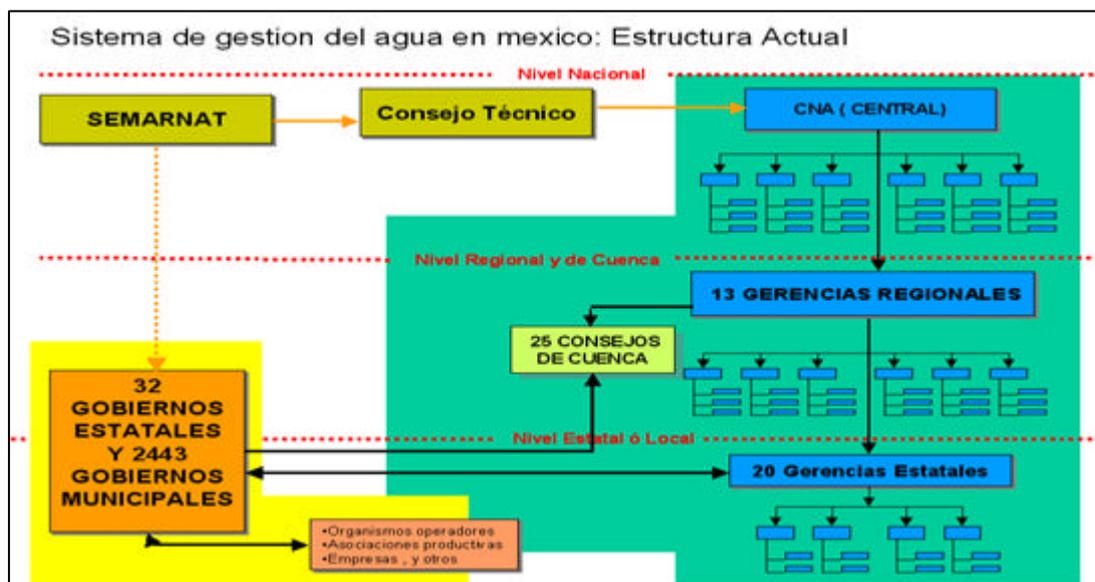


Figura 1.6. Sistema de Gestión del Agua en México: Estructura
Fuente: Director General de la Comisión Nacional del Agua,

En el ámbito municipal y de cada una de las localidades del país, operan organismos operadores de los servicios de agua y drenaje. Estos son empresas municipales y en algunos casos, empresas concesionarias de la prestación de estos servicios. Fig. 1.6

Sirven de vínculo entre el Gobierno Federal, los Gobiernos Estatales y Municipales y los usuarios, los Consejos de Cuenca creados como instancias de coordinación y de concertación.

Los Consejos de Cuenca, son órganos de coordinación entre los tres niveles de gobierno, federal, estatal y municipal, y entre la autoridad del agua, y los usuarios de las aguas nacionales, que se han constituido con el propósito de contribuir a mejorar la administración del agua por cuenca y acuífero, desarrollar la infraestructura Hidráulica y sus servicios y preservar las cuencas hidrológicas y los ecosistemas.

Los Consejos de Cuenca son una organización mixta de usuarios, entidades de gobierno y ciudadanos que, en el marco de la Ley de Aguas Nacionales busca la gestión sustentable del agua en cada cuenca. Son un Foro ordenado e institucional para ventilar los problemas de manejo del agua y de la cuenca por y con los protagonistas de la cuenca. Son una organización para vincular a los gobiernos con los ciudadanos; para transparentar los actos de gobierno y para facilitar la gestión integrada del agua y de la cuenca.



Figura 1.7. Los Consejos de Cuenca en México

Fuente: *Director General de la Comisión Nacional del Agua, México.*

Los Consejos de Cuenca no son un “organismo” porque su esencia es colectiva y su propósito es sumar voluntades y recursos y facilitar la toma de decisiones colegiadas, preservando la identidad y facultades de cada uno de sus integrantes. En cambio, sí son una “organización” que permite relacionar, de manera ordenada, a los gobiernos entre si y a estos con los usuarios directos del agua y con otros sectores de la sociedad, a fin de construir una nueva gobernabilidad de los recursos hídricos en la que el Estado cumple un papel normativo y regulatorio, y los diversos actores privados y sociales que actúan en las cuencas hidrográficas, asumen responsabilidades en el manejo del agua con pleno reconocimiento de que se trata de un recurso finito, variable en cantidad y vulnerable a la contaminación. Esta visión de la gobernabilidad de los recursos hídricos otorga prioridad a satisfacer las necesidades de la sociedad sin causar daños a la integridad del ciclo hidrológico, de tal manera que se otorga viabilidad al logro de objetivos de sustentabilidad y sostenibilidad. Fig. 1.7

Los Consejos de Cuenca en su condición actual, son principalmente un foro plural y abierto para analizar los problemas del agua en cada cuenca hidrográfica, proponer soluciones y llegar a la determinación de iniciar acciones conjuntas y sumar recursos para destinarlos a fines predefinidos y benéficos para el conjunto de la sociedad que radica en la cuenca, o para alguna de sus partes, en el entendimiento de que tanto la cuenca como los sistemas hídricos que la forman y dan unidad territorial, son indivisibles y por lo tanto, deben ser manejados integralmente. En este sentido, los Consejos han renovado y generalizado en

México la idea y el concepto de que las cuencas y los acuíferos deben ser las unidades de planeación y gestión del agua, con independencia de las divisiones político administrativas de estados y municipios. Los Consejos de Cuenca a medida que se consolidan como una organización estable y socialmente reconocida, se transforman en una instancia de negociación y consenso para dar sentido y viabilidad a la gestión ordenada e integrada del agua.

Cabe mencionar que siendo fundamental la voluntad política de abrir el manejo del agua a la participación de los usuarios directos del recurso, no es suficiente, toda vez que la gestión integrada y sustentable del agua, además de la voluntad y del involucramiento de los actores de mayor nivel político, requiere entre otras medidas, incluidas el desarrollo de capacidades institucionales para administrar y gestionar los recursos hídricos de un país; de sistemas de información confiables y accesibles; de la organización de los usuarios y de ciudadanos interesados en los asuntos del agua; de esquemas de planificación que permitan traducir las demandas y necesidades en actuaciones gubernamentales, privadas y sociales ordenadas conforme a prioridades, plazos y responsables, con mecanismos de seguimiento, evaluación y reformulación periódica; de sistemas financieros acordes con los requerimientos y demandas regionales y locales; de marcos regulatorios suficientes y claros que posibiliten la intervención de los agentes privados en un marco de certidumbre para sus inversiones, y de una gran variedad de instrumentos que deben ser puestos en práctica armónicamente para avanzar gradual y firmemente hacia una gestión ordenada e integrada de los recursos hídricos.

1.5.1 Objetivos Generales de los Consejos de Cuencas

Para cumplir su misión los Consejos tienen como objetivos generales, proponer y promover la ejecución de planes, programas y acciones para:

- a) Lograr el equilibrio entre oferta y demanda de agua en la Cuenca para sus diversos usos.
- b) El saneamiento de las cuencas, subcuencas, barrancas, acuíferos y cuerpos receptores de agua para prevenir, detener o corregir su contaminación.
- c) La conservación, preservación y mejoramiento de los ecosistemas de las cuencas con los que el agua forma sistemas naturales indivisibles.
- d) El uso eficiente y sustentable del agua en todas las fases del ciclo hidrológico.

- e) Impulsar una cultura del agua que considere a este elemento como un recurso vital y escaso, difundiendo su valor económico, social y ambiental y alentando la participación de la sociedad en su cuidado y uso sustentable.

1.5.2 Funciones Generales

Los Consejos de Cuenca se integran con funciones generales claramente definidas en la propia Ley. Las principales son:

- a) Formular y ejecutar programas que tiendan a mejorar la administración de las aguas nacionales.
- b) Desarrollar la infraestructura hidráulica necesaria y,
- c) Coadyuvar en la conservación y restauración de las cuencas hidrográficas.

En su sentido más amplio y general, los Consejos de Cuenca y sus organizaciones auxiliares son también:

- d) Instancias colegiadas para prevenir y dar cauce a los conflictos asociados a la distribución y usos del agua.
- e) Organizaciones plurales que se conforman para identificar, analizar, caracterizar, diagnosticar y pronosticar los problemas, situaciones, demandas y necesidades de agua en una cuenca hidrológica.
- f) Foros para conciliar propósitos, sumar voluntades y recursos y definir planes y programas que tienen la finalidad de aumentar la eficacia en la gestión del agua; mejorar su administración, procurar el saneamiento de sus corrientes, cauces y cuencas, y ordenar y eficientar sus usos, manejo y aprovechamiento.

1.5.3 Funciones Específicas

Son las siguientes:

- a) Conocer y difundir los lineamientos generales de la política hidráulica nacional y regional, y proponer aquellos que reflejen la realidad del desarrollo hidráulico a corto, mediano y largo plazos, en el ámbito territorial del Consejo de Cuenca;
- b) Promover la participación de las autoridades estatales y municipales, así como de los usuarios y grupos interesados de la sociedad, en la formulación, aprobación, seguimiento, actualización, y evaluación de la programación hidráulica de la cuenca o cuencas de que se trate en los términos de la ley;

- c) Promover la integración de comisiones de trabajo de diversa índole, que permitan analizar y en su caso, plantear soluciones y recomendaciones para la atención de asuntos específicos relacionados con la administración de las aguas, el desarrollo de la infraestructura Hidráulica y de los servicios respectivos, el fomento del uso racional del agua y la preservación de su calidad;
- d) Concertar con la Comisión Nacional del Agua las prioridades de uso y los demás instrumentos previstos en la programación hidráulica, conforme a lo dispuesto en la ley y su reglamento, así como los mecanismos y procedimientos para enfrentar situaciones extremas de emergencia, escasez, sobreexplotación, contaminación de las aguas o deterioro de los bienes a cargo de la Comisión;
- e) Apoyar las gestiones necesarias para la concurrencia de los recursos técnicos, financieros, materiales y tecnológicos que requiera la ejecución de las acciones previstas en la programación hidráulica;
- f) Participar en el desarrollo de los estudios financieros que lleve a cabo la Comisión, con objeto de determinar los montos de las contribuciones de los usuarios para apoyar la ejecución de los programas de la Comisión, que beneficien a los usuarios de la cuenca o cuencas comprendidas en el ámbito territorial del Consejo de Cuenca, y
- g) Participar o intervenir en los demás casos previstos en la Ley y su Reglamento para los Consejos de Cuenca.

El funcionamiento y alcances de los consejos, comisiones y comités de cuenca, así como de los "COTAS", por la naturaleza de sus trabajos y por la trascendencia y

complejidad social de los problemas que tratan, dependen de diversos factores. Destacan:

- La voluntad política de los gobiernos federal, estatal y municipal que participan en su creación y desarrollo.
- La decisión y voluntad expresada por diversos sectores representativos de la sociedad de participar en los procesos de gestión del agua.
- La dimensión y características de los problemas de escasez y contaminación de las aguas prevaletentes en las cuencas hidrológicas.

- Los recursos humanos y materiales susceptibles de ser movilizados para resolver los problemas del agua en cada cuenca. Fig. 1.8



Figura 1.8. Estructura de los Consejos de Cuenca.

Fuente: *Director General de la Comisión Nacional del Agua, México.*

1.6 Situación Actual

Actualmente se cuenta con una organización nacional para la gestión del agua compuesta de 25 Consejos de Cuenca, de los que dependen diversos órganos auxiliares constituidos por 6 Comisiones de Cuenca, 7 Comités de Cuenca, 54 Comités Técnicos de Aguas Subterráneas (COTAS), 25 Grupos de Seguimiento y Evaluación, 60 Grupos Especializados.

En los Consejos de Cuenca se han involucrado todos los gobernadores de las entidades federativas y el Jefe de Gobierno del Distrito Federal, los presidentes municipales de las más importantes ciudades del país y ciudadanos destacados de los ámbitos empresarial, académico, científico y cívico de las diferentes regiones, además de usuarios de todos los sectores de uso del agua.

El trabajo de los Consejos de Cuenca y de sus órganos auxiliares en la etapa inicial de consolidación se orientó para estudiar y analizar colectivamente entre otros, los siguientes temas:

- Reglas de Organización y Funcionamiento de los Consejos de Cuenca
- Balances y disponibilidades de agua por cuenca y acuífero
- Discusión de los sistemas de información necesarios para la gestión del agua
- Presentación y análisis de temas relevantes para la gestión integrada del agua destacando:
- Avances en la cuantificación de las disponibilidades de agua por cuenca y acuífero.
- El Registro Público de Derechos de Agua (REPGA) en la cuenca
- Elaboración y discusión de contenidos de diagnósticos y estrategias hidráulicas regionales
- Discusión de bases para la programación hidráulica nacional, regional y por cuenca
- Presentación de esquemas para facilitar la puesta en marcha del Movimiento Ciudadano por el Agua en el ámbito regional.
-

Los principales resultados alcanzados se resumen en:

- La constitución y desarrollo inicial de Consejos de Cuenca con bases organizativas y funcionamiento reglamentados, lo que deberá permitir y facilitar sus labores de planeación y gestión del agua.
- Una mayor y mejor participación de los usuarios, basada en el conocimiento de la información y documentación básica del agua.
- Involucramiento de los principales actores de cada cuenca (agricultores, industriales, prestadores de servicios, administradores públicos y representantes de gobiernos locales, representantes profesionales)
- La identificación de necesidades de capacitación acordes con las realidades regionales y orientadas al fortalecimiento de las capacidades locales para gestionar colectivamente el agua.
- Divulgación y asimilación del concepto de cuenca hidrográfica de la interdependencia existente entre los actores aguas abajo y aguas arriba.

- Elaboración de diagnósticos, planes y programas hidráulicos en consulta y colaboración con los usuarios.
- Mejor asimilación del concepto de Consejos de Cuenca y de su papel en la gestión integrada del agua por cuenca hidrológica, tanto en el ámbito de las instituciones del sector hidráulico como entre los gobiernos estatales y los usuarios de las aguas nacionales.
- Mayor receptividad y apoyo social y gubernamental para el financiamiento de los programas que se desprenden de la operación de los Consejos.
- Disminución de los conflictos entre usuarios y ente gobiernos por problemas de agua y apertura de canales de comunicación formalizados en comités y mesas de trabajo.
- Discusión y consenso de reglas para la distribución de las aguas superficiales.
- Desarrollo de nuevas actitudes y formas de relación entre administradores públicos y usuarios del agua.
- Identificación de disfuncionalidades y limitaciones para dar logro a la efectividad en los procesos de participación y toma de decisiones colectivas.

Al finalizar una primera etapa del proceso de creación y desarrollo de los Consejos de Cuenca, cada una de estas organizaciones cuenta con una Agenda Básica del Agua que señala los problemas y prioridades que deberán ser atendidos en los próximos años; igualmente se cuenta con Reglas de Organización y Funcionamiento expedidas por la Comisión Nacional del Agua como lo establece el Reglamento de la Ley de Aguas Nacionales, y se trabaja en un primer Proyecto de Plan o Programa de Gestión Integral del Agua para cada una de las principales cuencas hidrológicas del país, en el que habrán de señalarse objetivos, lineamientos de estrategia de mediano y largo plazo, e incluir la identificación de los principales proyectos y programas de inversión para los siguientes años. Todo ello ampliamente discutido entre los participantes.

Conforme los Consejos participen más amplia y directamente en el proceso de planeación de los usos y aprovechamientos del agua en cada cuenca, las nuevas formas de gestión colectiva de los recursos hídricos deberán ofrecer:

- Un nuevo orden para administrar y aprovechar mejor el agua,
- Programas mejor estructurados y más arraigados en la cultura regional para mejorar la eficiencia en el uso del agua; preservación y manejo de las cuencas; manejo y reglamentación de los acuíferos sobreexplotados; cuidado y cultura del agua, y otros de relevancia regional.

- Mayor conciencia de la responsabilidad compartida entre usuarios y gobiernos para resolver los problemas de disponibilidad del agua y de mejora de su calidad.
- Nuevas formas para financiar y alcanzar la autosuficiencia financiera en los servicios y sistemas de infraestructura hidráulica de la cuenca.

Los Consejos de Cuenca están llamados a ser organizaciones con facultades para actuar como Juntas de Gobierno de Organismos de Cuenca, y con capacidad para incidir en el destino de las inversiones que se canalizan al agua. En su actuación cotidiana deberán contar con mayor autonomía de gestión técnica y administrativa, articulando el esfuerzo de usuarios, organizaciones no gubernamentales y entidades de los tres órdenes de Gobierno para alcanzar la sustentabilidad del agua y contribuir al desarrollo económico de las regiones en que se ubican y operan.

2. Hidrología Básica

2.1 Introducción

Durante su vida sobre la tierra el hombre ha sido testigo, muchas veces sin entenderlo, del desarrollo del ciclo del agua en la naturaleza. La distribución de los climas, la formación de las nubes y su inestabilidad, la producción de las lluvias, la variación de los niveles de los ríos, y el almacenamiento de agua en depósitos superficiales o subterráneos son temas en cuyo estudio se ha venido profundizando a lo largo de los años, conformando una rama de la física que se conoce como Hidrología.

La Hidrología en su definición más simple es la ciencia que estudia la distribución, cuantificación y utilización de los recursos hídricos que están disponibles en el globo terrestre. Estos recursos se distribuyen en la atmósfera, la superficie terrestre y las capas del suelo.

Como ha ocurrido con otras ciencias, a medida que los estudios hidrológicos se fueron desarrollando fue necesario dividir el tema general en una serie de tópicos especializados e interdisciplinarios que se agruparon bajo el nombre de Planeamiento de los Recursos Hidráulicos. En el planeamiento se incluyen como temas principales la Meteorología, la Hidrología Superficial y la Hidrología del Agua Subterránea.

La Meteorología trata de los fenómenos que se desarrollan en la atmósfera y de la relación que existe entre los componentes del sistema solar. La Hidrología

Superficial estudia la distribución de las corrientes de agua que riegan la superficie de la tierra y los almacenamientos en depósitos naturales como lagos, lagunas o ciénagas. Por último, en la Hidrología del Agua Subterránea se incluyen los estudios de los almacenamientos subterráneos, o acuíferos, en lo referente a localización, volumen, capacidad de almacenamiento y posibilidad de recarga.

Los aspectos que tienen una relación muy estrecha con los anteriores en la planeación de proyectos de ingeniería son Geografía Física y Económica, Hidráulica Fluvial, Hidráulica Marítima, Hidrogeología, Geotecnia, Estadística, Teoría de Probabilidades, e Ingeniería de Sistemas.

La Hidrología Básica estudia los conceptos físicos del ciclo hidrológico, los métodos de recolección de información hidrológica y los procedimientos clásicos de procesamiento de datos estadísticos. Las técnicas que permiten la utilización de los recursos hidráulicos en proyectos de Ingeniería pertenecen al campo de la Hidrología aplicada

2.1.1 Objetivos de los Estudios Hidrológicos

Los proyectos que usan el agua como componente principal se clasifican de la siguiente manera:

1. Proyectos de suministro de agua.

Captan caudales de corrientes superficiales o de depósitos subterráneos para abastecer demandas de agua en áreas específicas.

Entre estos proyectos se cuentan los de acueductos y alcantarillados, de riego y drenaje de campos agrícolas.

2. Proyectos de suministro de energía hidráulica.

Captan caudales de corrientes superficiales y aprovechan diferencias de cota para entregar energía hidráulica a las turbinas de las centrales hidroeléctricas.

Las turbinas convierten la energía hidráulica en energía mecánica la cual se transmite a los generadores; éstos transforman la energía mecánica en energía eléctrica.

3. Diseño de obras viales, drenajes de aguas lluvias y estructuras de protección contra ataques de ríos.

Los estudios hidrológicos analizan los regímenes de caudales medios y extremos de las corrientes de agua en los tramos de influencia de las obras viales, en las zonas que requieren de alcantarillados de aguas lluvias, y en las zonas inundables adyacentes a los cauces .

Los caudales de creciente y las avalanchas que se generan por deslizamientos son las variables importantes en este tipo de proyectos. Estas variables se relacionan luego con los niveles de inundación, con las velocidades de flujo y con los procesos de socavación lateral y de fondo.

4. Proyectos de Navegación Marítima y Fluvial.

Los estudios de Hidrología en los proyectos de Navegación Marítima consisten en el análisis del estado del tiempo en mar profundo, en la plataforma continental y en los litorales. El estado del tiempo es una variable hidrológica que relaciona temperatura, humedad, presión atmosférica y vientos, y es responsable de la presencia de olas en la superficie del mar.

En los proyectos de Navegación Fluvial la Hidrología estudia los regímenes de caudales medios y extremos en los tramos navegables, las relaciones Caudal-Profundidad, y los volúmenes de sedimentos que se mueven como carga de fondo y en suspensión.

En el desarrollo de estos proyectos los estudios hidrológicos recolectan y procesan información histórica, programan y ejecutan programas de campo en topografía, batimetrías, aforos líquidos y sólidos, toma y análisis de muestras de sedimentos. Los resultados de los estudios producen información sobre los siguientes aspectos:

- Características climatológicas de las zonas que tienen influencia sobre el área del proyecto.
- Selección y capacidad de la fuente que suministrará el caudal que se entregará a los beneficiarios del proyecto. Se incluyen aquí los análisis sobre necesidad de almacenamiento.
- Magnitud de los eventos extremos, crecientes y sequías, que pueden poner en peligro la estabilidad de las obras civiles, o los procesos de navegación o el suministro confiable de agua a los usuarios.

Transporte de sedimentos hacia las obras de captación y almacenamiento.

¿Qué hacen los hidrólogos?

Los hidrólogos aplican el conocimiento científico y los principios matemáticos a la solución de problemas relacionados con el agua en la sociedad: problemas de cantidad, calidad y disponibilidad. Se encargan de encontrar los abastecimientos de agua para las ciudades o fincas con regadío, o de controlar las inundaciones por ríos o la erosión del suelo. También pueden trabajar en protección ambiental: prevención o limpieza de la contaminación o localización de lugares seguros para la eliminación de desechos peligrosos.

Las personas entrenadas en hidrología pueden tener una amplia variedad de ocupaciones. Algunas se especializan en el estudio del agua en solamente una parte del ciclo hidrológico: limnólogos (lagos); oceanógrafos (océanos); hidrometeorólogos (atmósfera); glaciólogos (glaciares); geomorfólogos (formas terrestres); geoquímicos (calidad del agua subterránea); e hidrogeólogos (aguas subterráneas). Los ingenieros que estudian hidrología pueden ser agrícolas, civiles, ambientales, hidráulicos, sanitarios, entre otros.

El trabajo de los hidrólogos es tan variado como los usos del agua y pueden variar desde proyectos multimillonarios hasta el aconsejar al propietario de una casa sobre sus problemas de drenaje. Algunos ejemplos son:

Aguas Superficiales.- La mayoría de las ciudades satisfacen sus necesidades de agua extrayéndola del río, lago o embalse más próximo. Los hidrólogos recogen y analizan los datos necesarios para predecir cuánta agua se dispone de las fuentes locales y si será suficiente para satisfacer las necesidades futuras proyectadas.

La gestión de los embalses puede ser muy compleja ya que, generalmente, tienen propósitos diversos. Los embalses aumentan la confiabilidad de los abastecimientos locales de agua. Los hidrólogos usan mapas topográficos y fotografías aéreas para determinar hasta dónde llegarán los niveles del embalse y así calcular las profundidades y la capacidad de almacenamiento. Este trabajo asegura que no ocurran inundaciones aún a su capacidad máxima.

La decisión de cuánta agua liberar y cuánto almacenar depende de la época del año, las predicciones de flujo para los próximos meses, y las necesidades de los regantes y las ciudades, al igual que las de los usuarios aguas abajo que dependen del embalse. Si también se usa el embalse para recreación o para la generación de energía hidroeléctrica, hay que tener en cuenta sus requerimientos. Los hidrólogos reúnen las informaciones necesarias y corren un modelo informático con ellas para tratar de predecir los resultados bajo varias estrategias de operación. En base a estos estudios, los administradores de los embalses pueden tomar las mejores decisiones.

Los posibles usos de las aguas superficiales (nadar, beber, industrial) a veces están restringidos debido a la contaminación; esta puede ser solamente un

inconveniente visual, o también puede ser una amenaza invisible, aunque letal, para la salud de las personas, plantas y animales.

Los hidrólogos ayudan en la vigilancia de los abastecimientos de agua para asegurarse que alcancen ciertos niveles de calidad. Cuando se descubre contaminación, los ingenieros ambientales trabajan con los hidrólogos para establecer el necesario programa de muestreo.

Aguas Subterráneas.- Con frecuencia, el agua subterránea es más barata, más conveniente y menos vulnerable a la contaminación que las aguas superficiales. Por lo tanto, estas aguas son comúnmente usadas para el abastecimiento de agua; en algunas áreas (regiones áridas), las aguas subterráneas pueden ser la única opción.

Los hidrólogos estiman el volumen de agua almacenada subterráneamente a través de mediciones de los niveles de agua en los pozos locales y estudiando la geología local. De esta manera, determinan la extensión, profundidad y espesor de los sedimentos y rocas con agua.

El agua subterránea es menos visible que las aguas de los ríos y lagos, pero es más insidiosa y difícil de limpiar. La contaminación de las aguas subterráneas resulta frecuentemente como resultado de una inadecuada eliminación de los desechos sobre el suelo. Entre las principales fuentes se encuentran los productos químicos industriales y del hogar, la basura en los rellenos sanitarios, las lagunas de desechos industriales, las colas y aguas usadas en las minas, los derrames de tanques de almacenamientos y tuberías, los lodos cloacales y sistemas sépticos.

Los hidrólogos dan lineamientos para la localización de pozos de vigilancia alrededor de lugares de eliminación y toman muestras de ellos a intervalos regulares para determinar si los lavados están contaminando las aguas subterráneas. En lugares contaminados, los hidrólogos pueden tomar muestras de suelo y agua para identificar el tipo y extensión de la contaminación.

Agricultura.- El principal usuario de agua es la agricultura. Es esencial la gestión eficiente del agua, especialmente en regiones áridas. Tradicionalmente, las familias campesinas aisladas tomaban el agua de arroyos, manantiales y pozos próximos. En la actualidad, se ejecutan vastos proyectos de irrigación de transportan el agua desde grandes distancias. En tales proyectos de tan gran escala, los hidrólogos intervienen determinando la mejor fuente de abastecimiento del agua.

Una finca que no se gestione adecuadamente puede ser una fuente de contaminación: sedimentos de los campos erosionados pueden obstruir los arroyos y represas; los fertilizantes, pesticidas y desechos animales pueden ser arrastrados hacia las aguas subterráneas o lavados hacia los arroyos, matando plantas, peces y otros animales. Los especialistas en conservación de suelos y

aguas trabajan con los agricultores en el desarrollo de planes de control de la erosión del suelo y la salinidad, y para la conservación del agua.

2.2 Ciclo Hidrológico

El fenómeno hidrológico tiene un comportamiento que responde a un complejo conjunto de interacciones entre diversos procesos físicos que dan origen a lo que se conoce como ciclo hidrológico. El ciclo hidrológico es un término descriptivo aplicable a la circulación general del agua en la tierra. Con mayor precisión se puede decir que es una sucesión de etapas que atraviesa el agua al pasar de la atmósfera a la tierra y volver a la atmósfera: evaporación desde el suelo, mar o aguas continentales, condensación de nubes, precipitación, acumulación en el suelo, deshielos, escurrimiento y nuevamente evaporación. Un esquema resumido de las distintas etapas y su interacción se muestra en la figura 2.1

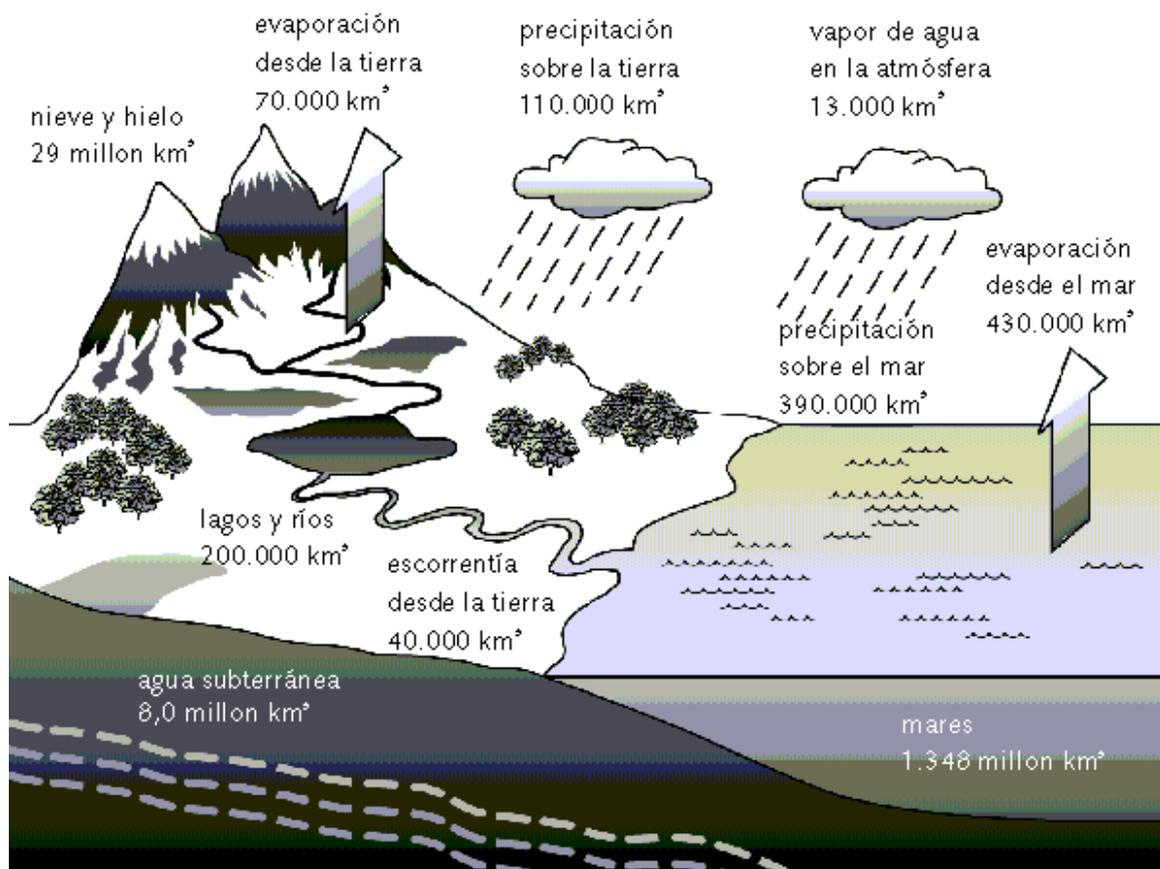


Fig.2.1 Esquema del ciclo hidrológico

Fuente: Clarke, R. *Water: The International Crisis*. Londres, Earthscan, 1991

Cada una de las etapas del ciclo hidrológico tiene implícita una fuerte componente estocástica.

Así, por ejemplo, las infiltraciones dependen de la porosidad del suelo, la que a su vez depende, en un rango amplio de valores, de la temperatura del suelo. Al mismo tiempo la evaporación y los deshielos dependen directamente de la temperatura ambiente, la que puede ser modelada como un proceso estocástico. Por otro lado, las precipitaciones están vinculadas a complejos procesos atmosféricos que poseen una fuerte componente estocástica. Luego, un modelo fenomenológico es difícil de implementar en la actualidad, debido a la gran cantidad de procesos aleatorios que están involucrados en el ciclo hidrológico, y a lo complejo de las interacciones que regulan el ciclo. Esto motiva a buscar métodos estadísticos para inferir el comportamiento futuro del fenómeno hidrológico, sin la necesidad de modelar el proceso físico.

El ciclo del agua, o ciclo hidrológico, explica el campo de aplicación de la Hidrología y su relación con otras disciplinas como son la Meteorología, la Oceanografía, la Hidráulica, la Geotecnia, las Ciencias naturales, etc.

El Ciclo comprende la circulación del agua desde los océanos hasta la atmósfera, luego a los continentes y nuevamente a los océanos.

En el desarrollo del Ciclo Hidrológico, el agua es transportada mediante procesos de Evaporación, Transpiración, Circulación Atmosférica, Condensación, Precipitación, Flujo Superficial y Subterráneo, y ocupa los almacenamientos que encuentra en su recorrido, tomando en cada caso el estado que corresponde a las condiciones imperantes de temperatura y presión. De esta forma, toma el estado líquido en océanos, lagos, embalses, ríos y acuíferos; el estado sólido en los glaciares y en los nevados, el estado gaseoso en la atmósfera y los estados sólido y líquido en las nubes. Fig. 2.2



Fig. 2.2 Proceso del Ciclo hidrológico
Fuente: ciencia.nasa.gov/.../y2002/01march_water.htm

El agua salada incluye los volúmenes almacenados en los océanos, en los acuíferos salados y en los lagos salados. Constituye el 97.47 % del total.

El agua dulce no utilizable es la que no está disponible en forma líquida para su aprovechamiento inmediato en los proyectos de ingeniería. Está comprendida por los glaciares, la nieve y la humedad atmosférica. Representa el 1.76 % del recurso hídrico.

En el agua dulce superficial se consideran los volúmenes que pertenecen a los ríos, lagos y pantanos: ocupa solamente el 0.0076 % del total de agua que hay en el globo terrestre.

Por último, el agua subterránea representa el 0.76 % del volumen total, lo cual indica que la cantidad de agua subterránea es 100 veces mayor que la de agua superficial. La utilización plena del agua subterránea, sin embargo, depende de factores económicos y técnicos que aproximadamente el 50 % del total de agua subterránea está confinada en acuíferos por debajo de 800 m de profundidad.

El agua no está distribuida uniformemente sobre la superficie del globo terrestre; existen factores de tipo meteorológico, astronómico, orográfico, geográfico y geológico que regulan el comportamiento de las variables del Ciclo Hidrológico.

En 1900 el doctor Vladimir Koppen presentó una clasificación climatológica que está relacionada con la Temperatura y la Precipitación; la clasificación permite

lograr una visión amplia acerca de la distribución cuantitativa del recurso hídrico sobre el globo terrestre.

La clasificación de Köppen reduce a 5 grandes grupos las diferentes variedades de climas que se presentan en el mundo. Estos grupos son los siguientes:

A. Climas húmedos tropicales. Característicos de zonas de alta precipitación, con temperaturas medias mensuales por encima de 18°C.

B. Climas secos. Zonas semiáridas y áridas, en las cuales la evaporación anual excede a la precipitación anual.

C. Climas húmedos mesotérmicos. En zonas lluviosas, con períodos cortos de invierno, y temperaturas medias mensuales que varían entre 0°C y 18°C en los meses más fríos.

D. Climas húmedos microtérmicos. En zonas lluviosas con períodos largos de invierno, y temperaturas medias menores de 0°C en los meses fríos y mayores de 10°C en los meses cálidos.

E. Climas polares. No tienen estaciones cálidas. Los meses más calientes tienen temperaturas inferiores a 10°C

2.3 Hidrología Estocástica

La Hidrología Estocástica aplica la teoría de los Procesos Estocásticos, que son parte de la Ingeniería de Sistemas, al estudio de los Eventos Hidrológicos. Los Eventos Hidrológicos, tales como aguaceros, caudales, niveles de embalse, etc., son Eventos Estocásticos. Se caracterizan porque, por un lado tienen un patrón medio de comportamiento a largo plazo, y por el otro el pronóstico de sus magnitudes en un momento dado tiene un mayor o menor grado de incertidumbre. El patrón medio corresponde a lo que se denomina la Tendencia General o Componente Determinística y la incertidumbre constituye la Componente Aleatoria del evento.

Los proyectos que se diseñan hoy serán construidos y operarán en los próximos años. Por esta razón, cuando los estudios hidrológicos definen caudales de diseño para abastecer un acueducto, o para proteger una zona contra las inundaciones, debe tenerse en cuenta que tanto la ocurrencia como las magnitudes de esos caudales son pronósticos hacia el futuro de eventos probables que están sujetos a un grado de incertidumbre.

Caudal real = Caudal calculado +/- Incertidumbre

La proyección hacia el futuro se basa en la combinación de la estadística y la teoría de la probabilidad. Con la estadística se realiza el análisis de frecuencias de los eventos históricos y se definen los parámetros que determinan el patrón general de comportamiento. Con estos parámetros y con la aplicación de la teoría de la probabilidad se hace el pronóstico de lo que puede esperarse en el futuro, dentro de cierto nivel de riesgo.

1. Generación estocástica de Series Hidrológicas.

Los sistemas que proveen de agua a los acueductos, los distritos de riego o las centrales hidroeléctricas utilizan generalmente los ríos como fuente de suministro. Estos sistemas constan de una captación, una conducción, un tanque de almacenamiento y una red de distribución, y se dimensionan con base en la capacidad de la fuente y en la demanda que requieren los futuros usuarios.

Mientras que los sistemas anotados operarán en el futuro, los datos hidrológicos que se tienen disponibles para realizar el diseño de las obras corresponden al pasado, es decir son historia.

Es necesario, entonces, hacer un pronóstico hacia el futuro sobre el comportamiento de la fuente durante la vida útil del proyecto de suministro de agua. Para lograr este propósito se pueden utilizar métodos que pertenecen a la Hidrología Estocástica, los cuales tienen su base en los conceptos de la Teoría de Probabilidades y de la Ingeniería de Sistemas.

Uno de estos métodos consiste en determinar la Serie Histórica de Caudales de la Fuente, y con base en ella generar N series estocásticas de Caudales Virtuales, cada una de longitud igual a la vida útil del proyecto. Las N series generadas tienen igual probabilidad de presentarse en el futuro, y una de ellas, no se sabe cuál, será probablemente parecida a la serie real futura. Las series estocásticas permiten, entonces, contar con un panorama de posibilidades de lo que puede ocurrir con la fuente en el futuro. Este panorama será más amplio entre mayor sea el número N. Fig. 2.3

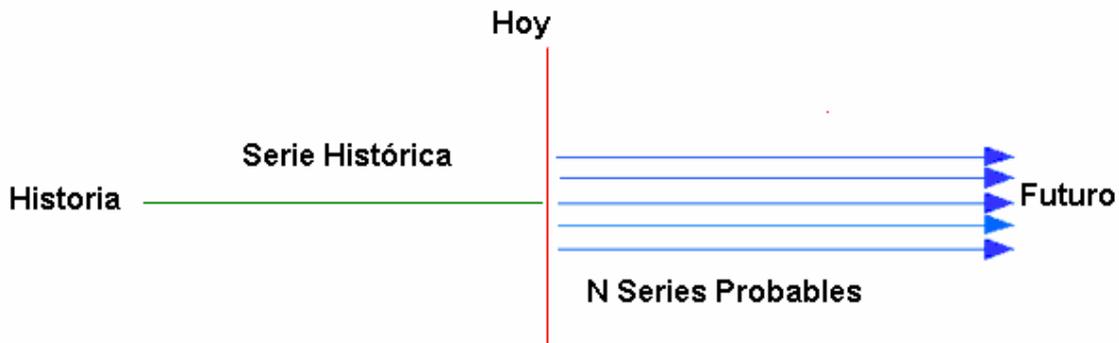


Fig.2.3 Series Hidrológicas
 Fuente: Elaboración propia basa en tesis Modelos Hidrológicos para Manejo de la Incertidumbre, Colombia, 2005

Como se mencionó anteriormente, para generar las series estocásticas es necesario realizar primero un análisis estadístico de frecuencias de la serie histórica. Si no existe serie histórica o si ésta es deficiente la Generación Estocástica no es aplicable.

2.3.1. Modelos Hidrológicos.

Existen diferentes tipos de Modelos Hidrológicos. Pueden ser Determinísticos o Estocásticos, de Simulación o de Optimización.

Son ejemplos de Modelos Determinísticos la Ecuación del Balance Hidrológico, la Fórmula Racional y el Hidrograma Unitario. Un modelo de este tipo corresponde a un algoritmo de cálculo que da un resultado único.

En los Modelos Estocásticos el algoritmo de cálculo incluye una o varias componentes probabilísticas. Con los Modelos Estocásticos se generan series futuras de lluvias, de caudales, de niveles de embalses, o de eventos extremos.

1. Componentes: Características físicas de la presa y del vaso del embalse.

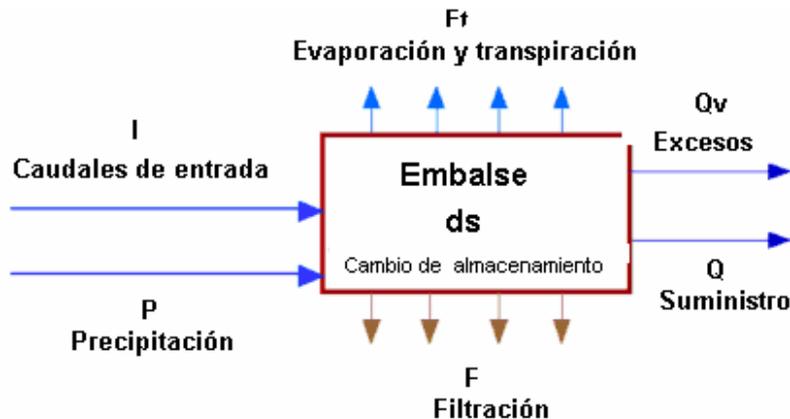
2. Variables:

Variable de Estado: Nivel del agua en el Embalse. Determina el volumen embalsado.

Variables de Entrada: Caudales que llegan al Embalse y Precipitación sobre el vaso del embalse.

Variables de Salida: Pérdidas por Evapotranspiración e Infiltración, Vertimiento de Excesos y Caudales de entrega al proyecto.

Las variables de Estado, Entrada y Salida se pueden tratar todas o algunas de ellas como Variables Estocásticas. Fig.4



Balance

$$I + P = Et + F + Qv + Q + ds$$

Fig. 2.4 Modelo Hidrológico

Fuente: Elaboración propia basa en tesis Modelos Hidrológicos para Manejo de la Incertidumbre, Colombia, 2005

El problema de diseño consiste en determinar el volumen de almacenamiento necesario para que el embalse sea capaz de suministrar la demanda con un nivel de probabilidad aceptable.

Por medio de la simulación se utilizan diferentes series hidrológicas generadas estocásticamente y se combinan con diferentes dimensiones del embalse. Esto permite obtener tantos volúmenes de almacenamiento probables como opciones que se analicen.

Con los resultados obtenidos se hace un análisis de probabilidades para determinar los niveles de riesgo de las diferentes soluciones. Estos análisis permiten al diseñador tener un buen criterio para tomar la decisión sobre el dimensionamiento de las obras y la operación futura del embalse.

4. Distribución de los Recursos Hidráulicos.

La distribución del recurso hidráulico disponible en una región puede obedecer a criterios políticos, sociales, técnicos o económicos. Es necesario tener una fundamentación muy clara para explicarle a la gente por qué se le suministró el agua disponible al proyecto de riego de una zona A y no al de una zona B. Como el volumen del recurso disponible es aleatorio la solución del problema técnico necesita de la Hidrología Estocástica para estimar volúmenes probables disponibles en los años futuros; en cambio los análisis económicos necesitan de otras herramientas de la Ingeniería de Sistemas. Entre estas herramientas están la Programación Matemática, la Programación Lineal y la Programación Dinámica. Fig. .2.5

Modelo determinístico

Dado un valor de entrada se obtiene un único valor de salida cada vez que se corre el modelo



Modelo estocástico

Dado un valor de entrada se obtiene una respuesta diferente cada vez que se corre el modelo



Fig. 5 Modelos Hidrológicos

Fuente: Elaboración propia basa en tesis Modelos Hidrológicos para Manejo de la Incertidumbre, Colombia, 2005

Los modelos de simulación permiten analizar diferentes situaciones para obtener un panorama amplio de posibles resultados de un problema. Aquí se incluyen los Modelos de Operación de Embalses, de Tránsito de Crecientes y de Avalanchas.

Por último los modelos de Optimización combinan la Simulación con los Costos y los Beneficios para determinar las soluciones más económicas.

Operación de Embalses. En los estudios de Operación de Embalses se utilizan Modelos de Simulación los cuales relacionan una serie de componentes y de variables.

2.4 Modelación de la incertidumbre Hidrológica

Los registros sintéticos de afluentes empezaron a ser usados con el comienzo del siglo XX por Hazen en 1914 en estudios de confiabilidad de sistemas de abastecimiento de agua, pero su construcción no estaba basada en la teoría de procesos estocásticos. Las primeras aplicaciones relacionadas con la hidrológica estocástica se hicieron alrededor del año 1940 en medio de la revolución de la física y las matemáticas. Entre otros avances, se establece el método de montecarlo, en manos de Stanislaw Ulam en 1946.

En el campo de los recursos hídricos, los aportes más importantes fueron los trabajos desarrollados por Barnes en 1954, sobre la generación de flujos anuales no correlacionados asumiendo una distribución normal, y los trabajos de Tomas y Fiering para la generación de flujos temporalmente correlacionados.

Alrededor del año 1970, el libro sobre series de tiempo escrito por Box y Jenkins se convirtió en uno de los principales aportes en la materia. Este libro, pese a ser escrito para aplicaciones en una amplia variedad de campos de la ciencia, adquirió gran popularidad para aplicaciones en hidrología estocástica.

Box y Jenkins desarrollaron un esquema de clasificación para una gran familia de modelos de series de tiempo. En esta clasificación se distinguen los modelos autoregresivos de orden p (AR(p)), los modelos de media móvil de orden q (MA(q)), combinaciones de los dos, llamados modelos "autoregresivos de media móvil" (ARMA(p,q)), y otras variantes como los modelos integrados. Sin embargo, la gran familia de modelos propuesta por Box y Jenkins, no cubre totalmente las necesidades impuestas por las series de tiempo hidrológicas, las cuales poseen algunas peculiaridades propias de los procesos geofísicos. Esto ha dado origen a una línea de investigación que ha resultado en un gran número de herramientas apropiadas para la aplicación en recursos hídricos.

Por otro lado, las primeras observaciones modernas de precipitaciones en el siglo XVIII llevaron a los investigadores a concluir que el proceso hidrológico muestra una fuerte aleatoriedad, como también una marcada ciclicidad dentro del año. Esta

evidencia de proceso estocástico y periodicidad dio origen a dos líneas de investigación que se mantienen en la actualidad:

1. El estudio de fluctuaciones aleatorias de los valores anuales de afluentes por métodos estadísticos, con el objetivo de encontrar una eventual periodicidad en la serie anual. Importantes esfuerzos se han invertido en el último siglo para dilucidar esta cuestión, encontrándose que el fenómeno hidrológico muestra la presencia de una periodicidad de largo plazo. Existen básicamente dos componentes de periodicidad de largo plazo; por un lado estudios hidrometeorológicos han estudiado la presencia de un ciclo de aproximadamente 11 años debido a la variación aproximadamente cíclica de la actividad solar.

2. Estudio de la estacionalidad (periodicidad) de la serie intra anual.

2.4.1 Propiedades de las Series Hidrológicas

Para la modelación de series de tiempo se deben tener en consideración las características propias de la serie que se está modelando, y así poder tomar una decisión lo más informada posible respecto de cuál es el modelo más apto para ser aplicado.

Las series hidrológicas poseen la característica fundamental de que, para un intervalo de tiempo de una duración correspondiente a una fracción del año, el valor medio y la desviación estándar de las series de caudales, son siempre periódicos. Esta periodicidad se fundamenta en el ciclo hidrológico, el cual está principalmente determinado por la cantidad de energía solar recibida por la tierra, la que es esencialmente aleatoria, y que posee un ciclo anual relacionado con el movimiento de la tierra en torno al sol. Adicionalmente, como es lógico, la serie sólo puede tomar valores positivos o nulos, lo que la convierte en una serie *no negativa*. Las características más distintivas de las series hidrológicas, se resumen a continuación:

1. *No Estacionaridad*: el proceso estocástico que da origen a la serie hidrológica es no estacionario, es decir, su media, varianza y autocovarianza varían con respecto al tiempo, mostrando periodicidad cuando se discretiza en intervalos de tiempo de duración inferior a un año. Para intervalos mayores que un año la serie de afluentes promedios tiene características estacionarias.

2. *Estacionalidad*: como se vio, cuando la escala de tiempo usada en la serie es una fracción del año, el proceso hidrológico deja de comportarse como estacionario debido a que las propiedades estadísticas de la serie tienen un comportamiento aproximadamente periódico. Este fenómeno se conoce como estacionalidad y se manifiesta en un comportamiento cíclico de la serie, en que los afluentes durante los meses calurosos (deshielos) están fuertemente ligados a las precipitaciones ocurridas en la cordillera durante las estaciones frías. Esto se

manifiesta estadísticamente en una fuerte correlación de los afluentes dentro del año.

3. *Dependencia Espacial*: las series hidrológicas muestran una componente de correlación espacial. Esto quiere decir que existe dependencia estadística entre el valor que toma un afluente en un lugar determinado, y el valor del afluente en otra ubicación geográfica.

4. *Intermitencia*: a escalas de tiempo pequeñas, algunas componentes de los caudales afluentes, como es el caso de las lluvias, pueden presentar intermitencia. Esto significa que el caudal medido puede alternar entre dos estados, un estado seco (caudal nulo) y uno húmedo (presencia de lluvias). Este aspecto, que adquiere gran relevancia en los afluentes de régimen pluvial, convierte a la incertidumbre hidrológica en una variable binaria que puede alternar entre dos estados posibles. En general, para etapas de duración superior a un mes, la intermitencia del fenómeno es en promedio despreciable.

2.4.2 Características de las Estadísticas Disponibles

Tradicionalmente en Chile los modelos para la incertidumbre hidrológica usados en el sector eléctrico, utilizan para su construcción la información contenida en una base de datos que almacena el registro histórico de los caudales afluentes semanales promedio, durante un período de 40 años. Se ha adoptado que los 40 escenarios hidrológicos históricos constituyen una estadística *representativa* del fenómeno real.

Las series de caudales contenidas en la estadística utilizada, se han obtenido principalmente de mediciones realizadas en las distintas cuencas del país. Sin embargo, para algunos afluentes no se cuenta con mediciones directas de los caudales, habiendo sido obtenidos por métodos indirectos, a partir de adición y substracción de otros afluentes de la misma cuenca.

En general, los afluentes asociados a los puntos de control de paso, son estimados a partir de la estadística de generación de éstas, utilizando la función de rendimiento de la central. Los datos obtenidos con esta metodología tienen dos serios inconvenientes. El primero, es que en la mayoría de los casos, las estadísticas de caudales son obtenidas asumiendo una función de rendimiento lineal para la central, lo que en general no se cumple. El otro inconveniente es que cuando las centrales operan a potencia máxima, un metro cúbico adicional de agua no se refleja en una potencia adicional generada, luego los caudales estimados se saturan en las puntas, y no reproducen adecuadamente la estadística real. Un ejemplo de esto se observa en la figura 2.6.

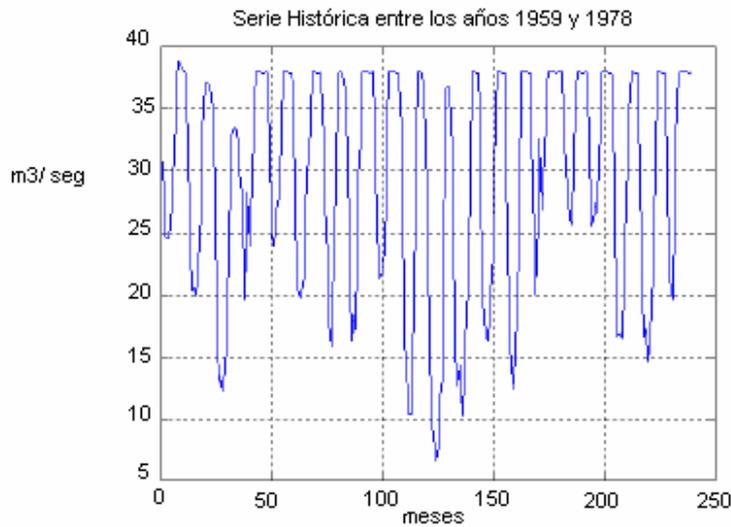


Figura 2.6: Afluente, obtenido a partir de matrices de energía generada que muestra recortes en las puntas en los períodos donde la central opera a potencia máxima.

Fuente: Tesis Modelo para manejo de la incertidumbre hidrológica en la planificación de la operación del SIC

Es importante señalar, que las estadísticas de caudales obtenidas de la forma descrita en el párrafo anterior, pueden ser útiles para representar escenarios de caudales cuando se simula la operación del sistema, pero resultan inadecuados si lo que se desea es ajustar un modelo estadístico a los datos.

Por otro lado, muchas series han sido completadas para disponer de un período de análisis común, desconociéndose los procedimientos empleados para ello. En general se dispone de unas pocas series largas, de más de 40 años, y el resto, sobre todo en los primeros años, se ha completado por algún procedimiento típico disponible en la literatura técnica habitual en hidrología. Entre los más comunes está el considerar la proporcionalidad entre las áreas de las cuencas respectivas o utilizar regresión lineal o múltiple entre caudales mensuales de cuencas vecinas. En estos casos, se debe ser cuidadoso con el empleo de estas series, ya que estos métodos de completéz reproducen adecuadamente los valores esperados, pero distorsionan las varianzas y covarianzas, de manera que aparecen dependencias espaciales más altas que las naturales. Al utilizar estos valores en la construcción de un modelo, éste heredará el mismo error.

2.5 Hidrología Aplicada

La Hidrología Aplicada utiliza la información básica y la procesa de acuerdo con las necesidades de los proyectos de aprovechamiento de los recursos hidráulicos, empleando las herramientas que ofrece la tecnología moderna.

Entre los temas que desarrolla la Hidrología Aplicada están los siguientes:

- Hidrología en cuencas pequeñas con información escasa
- Drenaje de aguas lluvias
- Hidrología en Proyectos de Riego y Drenaje
- Hidrología en Proyectos de Acueducto y Alcantarillado
- Hidrología en Proyectos de generación de Energía Hidráulica
- Diseño y Operación de embalses
- Hidrología para estudios de aprovechamiento de Aguas Subterráneas
- Control de inundaciones
- Estimativo de los volúmenes de sedimentos que pueden afectar el funcionamiento de las estructuras hidráulicas.

1. Hidrología en cuencas pequeñas con información escasa.

El problema de la información escasa es muy frecuente en las cuencas que están alejadas de los centros poblados y en las que pertenecen a zonas selváticas y montañosas. El tema principalmente se enfoca al manejo de la información hidrológica en estudios que utilizan las fuentes de agua para captaciones, y para diseño de obras en corrientes naturales.

2. Drenaje de aguas lluvias

El drenaje de aguas lluvias relaciona factores que tienen que ver con las lluvias intensas y con las características de las áreas de drenaje. En su estudio se combinan la Hidrología Aplicada y la Hidráulica general.

3. Hidrología en Proyectos de Riego y Drenaje.

En los proyectos de Riego y Drenaje los estudios de Hidrología tienen su desarrollo en los siguientes capítulos:

- Requerimientos de agua.

Los requerimientos de agua se refieren al volumen de agua que necesitan los cultivos para desarrollarse adecuadamente. Su valor depende de la relación que existe entre el clima, el suelo y el cultivo por una parte, y en el tamaño del área de proyecto, la eficiencia en la aplicación del riego y las pérdidas en las conducciones por la otra.

Los estudios hidrológicos que se ejecutan para determinar los requerimientos de agua comprenden análisis de Clima, Evapotranspiración y Lluvia en períodos cortos.

- Necesidades de riego.

Cuando los cultivos pueden desarrollarse adecuadamente dentro de las condiciones climáticas naturales de la zona del proyecto no hay necesidad de aplicar riego.

En caso contrario se estudia la necesidad de aplicar riego durante aquellos períodos que presentan deficiencias porque las condiciones naturales de la zona no tienen capacidad para suministrar los requerimientos de agua, y de instalar estructuras de drenaje de los campos agrícolas para evacuar los excesos de agua que se presentan durante los períodos de lluvias altas.

- Capacidad de la fuente seleccionada para suministrar la demanda.

Una vez que se ha determinado el valor de la Demanda de agua se analiza la fuente que va a suministrarla. Esa fuente puede ser una corriente natural o un depósito subterráneo.

El estudio hidrológico incluye análisis de Caudales Medios, Curvas de Duración de Caudales y Operación de Embalses.

- Sedimentación en captaciones.

Muchos sistemas de captación y conducción de aguas afrontan actualmente graves problemas de sedimentación en las estructuras de captación. Entre las causas de estos problemas pueden estar las siguientes:

1. No hubo suficiente información cuando se hicieron los cálculos de los volúmenes de transporte de sedimentos en las corrientes que alimentan las captaciones
 2. El uso de la tierra en las cuencas vertientes cambió luego de la construcción de las obras hidráulicas tanto por la instalación de asentamientos humanos como por la explotación no controlada de los recursos naturales.
- Magnitudes y Efectos de las crecientes sobre el funcionamiento de Captaciones, Embalses, Desarenadores y Conducciones.

Las obras que se construyen en los ríos o en sus riberas están expuestas a los ataques de las corrientes, tanto por la socavación del lecho y de las márgenes como por los desbordamientos en períodos de creciente.

- Drenaje de Aguas Superficiales y Drenaje de Suelos Agrícolas.

Tanto los excesos en la aplicación del riego como las lluvias intensas generan volúmenes indeseables de agua en los campos agrícolas. Se deben construir, entonces, canales para drenaje de aguas lluvias y subdrenes para drenaje de las aguas de infiltración.

4. Hidrología en Proyectos de Acueducto y Alcantarillado.

La Hidrología contempla los siguientes aspectos en los proyectos de Acueducto y Alcantarillado:

- Demanda.

El estudio de la demanda de un acueducto incluye el análisis del crecimiento de la población y la asignación de la dotación. Esta última se refiere al volumen medio de agua que necesita cada persona en su hogar o en hospitales, oficinas, establecimientos industriales o comerciales, etc.

- Capacidad de la fuente seleccionada para suministrar la demanda. Sedimentación en captaciones. Magnitudes y Efectos de las crecientes sobre el funcionamiento de Captaciones, Embalses, Desarenadores y Conducciones.

Los estudios hidrológicos que se necesitan para definir estos aspectos se realizan de acuerdo con las necesidades de los proyectos particulares.

- Caudales de aguas servidas.

Los caudales de aguas servidas representan un porcentaje de los caudales que se suministran al sistema de acueducto.

- Caudales de aguas lluvias.

El estudio hidrológico de los caudales de aguas lluvias se explica en el artículo sobre drenaje de aguas lluvias.

5. Hidrología en Proyectos de Generación de Energía Hidráulica.

El producto que entrega un proyecto de Generación de Energía Hidráulica es Energía en un tiempo dado. La Energía se expresa en Kilovatios-hora y tiene una fórmula matemática que responde a la siguiente expresión Fig. 2.7:

$$E = K * Q * H$$

donde K incluye el Tiempo, la Densidad del Agua y las Dimensiones. Q representa el Caudal y H la Cabeza Neta del Sistema Hidrológico de Generación.

El Sistema Hidrológico de Generación puede ser a Filo de Agua o con Embalse:

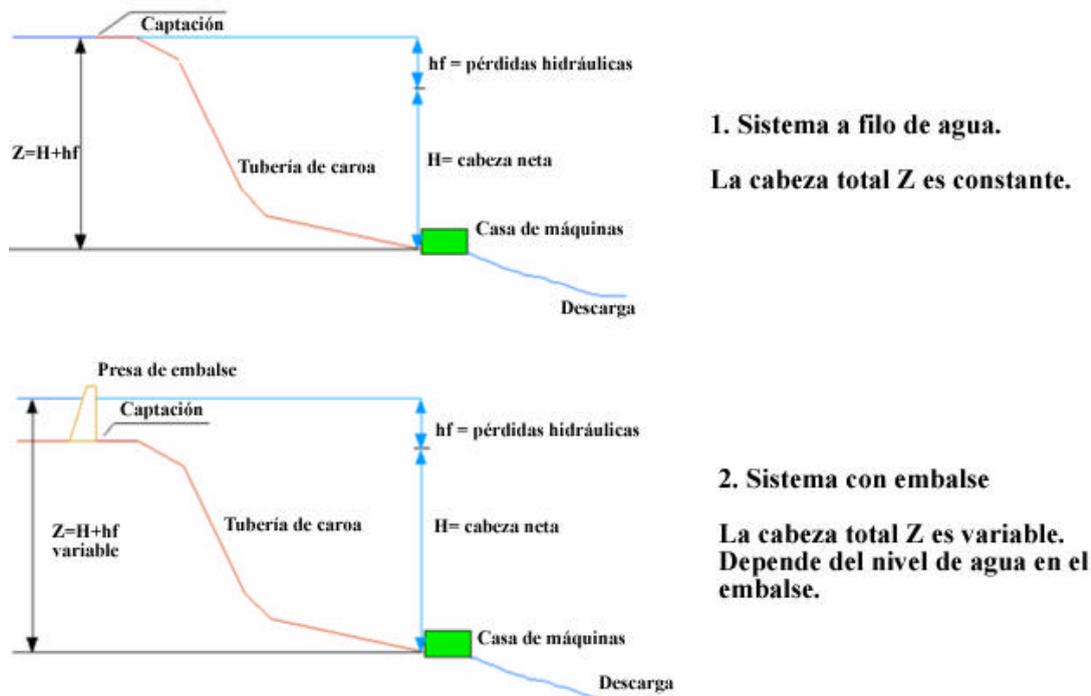


Fig. 2.7 Hidrología en Proyectos de Generación de Energía Hidráulica.
Fuente: ciencia.nasa.gov/.../y2002/01march_water.htm

Los estudios hidrológicos determinan la capacidad que tiene la fuente para suministrar la demanda de energía, analizan las magnitudes de las crecientes que pueden atacar las obras civiles, cuantifican los procesos de sedimentación y determinan las condiciones de la descarga.

Para cumplir con estos propósitos los estudios hidrológicos se realizan en coordinación con los estudios de Potencia y Energía.

6. Embalses

La necesidad de utilizar embalses en los proyectos de suministro de agua se analiza inicialmente con la Curva de duración de Caudales y posteriormente se

utiliza un Modelo de generación estocástica de caudales para afinar los resultados Fig. 2.8.

En la fuente de suministro, representada por la captación, se realizan los estudios hidrológicos de la corriente natural para determinar los caudales medios, máximos y mínimos que se esperan durante la vida útil del proyecto.

En el área de usuarios del proyecto se analizan las proyecciones de la demanda de agua.

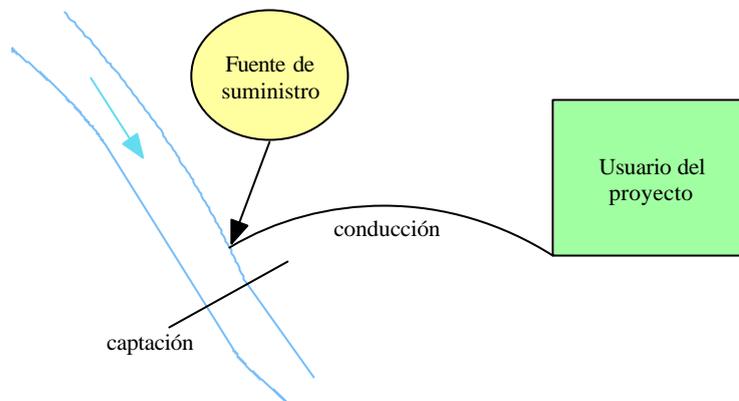


Fig. 2.8 Embalses

Fuente: ciencia.nasa.gov/.../y2002/01march_water.htm

Cuando la fuente tiene capacidad suficiente para suministrar la demanda durante el ciento por ciento del tiempo no es necesario utilizar embalses. Algunas veces, a pesar de que la fuente no tenga la capacidad suficiente, se asume el riesgo de que se presenten deficiencias en el suministro y se programan racionamientos para evitar los costos adicionales que representa la construcción de un embalse. Esta decisión no es recomendable cuando se trata de suministrar agua para acueductos pero puede ser factible en los suministros para riego o para generación de energía hidráulica.

La operación de un embalse o de una serie de embalses se simula mediante un modelo matemático que tiene como componentes las estructuras de descarga y las características geométricas de los embalses, y como variables las entradas de caudal, las entregas al proyecto, los niveles en el embalse y las pérdidas por evaporación e infiltración.

7. Hidrología para estudios de aprovechamiento de Aguas Subterráneas.

Dentro de los estudios de exploración que se realizan cuando se quiere evaluar la posibilidad de extraer aguas subterráneas de manera permanente y segura en una zona determinada, resulta indispensable analizar la relación que existe entre la Precipitación, el Caudal Superficial y la Infiltración.

Una parte del agua de Infiltración, conocida como Caudal de Recarga, llega hasta los depósitos de Aguas Subterráneas en la zona de estudio y los alimenta. Otra parte del agua de Infiltración es atrapada por el suelo y el resto drena en forma subsuperficial o se escapa como percolación profunda Fig. 9 .

ACUÍFERO NO CONFINADO:

Parte del agua de infiltración se queda en el suelo en forma de Humedad, otra parte se escapa como flujo subsuperficial y el resto drena hacia el depósito de agua subterránea.

Este depósito queda limitado en su parte superior por el Nivel Freático y en su parte inferior por una capa impermeable que impide el flujo vertical del agua a través de ella.

El Nivel Freático es variable y su posición depende del volumen de agua almacenada en el depósito.

ACUÍFERO CONFINADO:

El depósito de agua subterránea queda limitado en sus partes inferior y superior por dos capas impermeables.

La zona de recarga está restringida al área de terreno comprendida entre las capas impermeables.

El flujo del agua dentro del depósito subterráneo funciona a presión, como en el caso de una tubería.

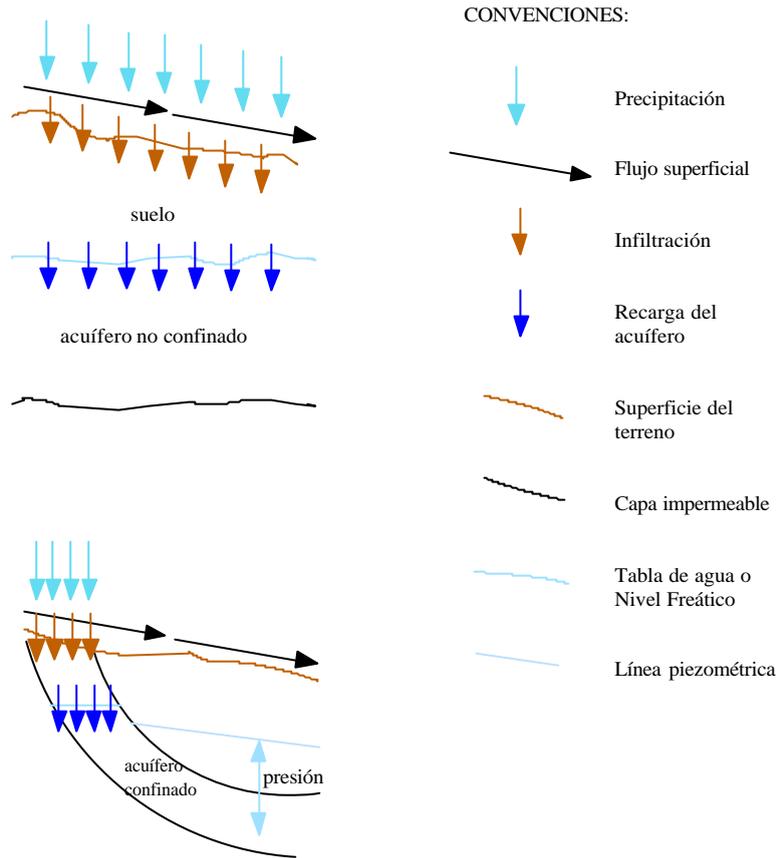


Fig. 2.9 Hidrología para estudios de aprovechamiento de Aguas Subterráneas.

Fuente: ciencia.nasa.gov/.../y2005/06march_water.htm

8. Control de Inundaciones

Las magnitudes y los efectos de las inundaciones dependen de las características de las crecientes que son generadas por lluvias intensas, y de otros eventos relacionados con ellas como son los deslizamientos de taludes, la formación y el rompimiento de presas naturales, y las obstrucciones al flujo por destrucción de obras civiles.

3. Modelos de Planificación de Sistemas Hidrológicos

3.1 Introducción

Los modelos para la planificación de la operación en corto plazo en sistemas hidrológicos, buscan definir una secuencia óptima de decisiones, que minimice el valor esperado de los costos futuros actualizados de operación y fallas del sistema.

La complejidad matemática y computacional de estos modelos aumenta a medida que se intenta ser más riguroso y apegado a la realidad en la representación matemática del sistema. El desarrollo que han alcanzado las computadoras en los últimos años ha permitido avanzar hacia modelos más precisos y fieles a la realidad.

Se han implementado varios modelos en las últimas décadas. Estos han servido de apoyo tanto en las funciones reguladoras como en las competitivas, permitiendo hacer estimaciones para el valor esperado del costo marginal de energía, valores del agua y estimación de costos marginales en el sector eléctrico, entre otros.

Modelos de Planificación de Sistemas Hidrológicos

Si se supone conocida la demanda en el horizonte de planificación, ésta corresponde a valores esperados sobre la hidrología. Si se modela la hidrología como una variable aleatoria, se puede observar al modelo de planificación como una función implícita de la variable aleatoria hidrológica, donde dado un escenario de caudales fijo, podemos calcular la política de operación óptima para ese escenario.

La forma en la que se modela la incertidumbre hidrológica, es determinante en la bondad de los resultados. La elección de un determinado modelo para la hidrología, dependerá también de cómo el modelo de planificación está calculando los valores esperados de las variables de interés. Es decir, la bondad de los resultados obtenidos no dependerá exclusivamente de las características del modelo para los caudales, si no que también dependerá de lo que se desea calcular y de cómo esto se está haciendo.

3.2 Esquema general de modelación

La construcción de un modelo estocástico que reproduzca las características de una serie cronológica, utiliza un esquema de modelación formado por varias etapas sucesivas a través de las cuales se va construyendo el modelo. Figura 3.1

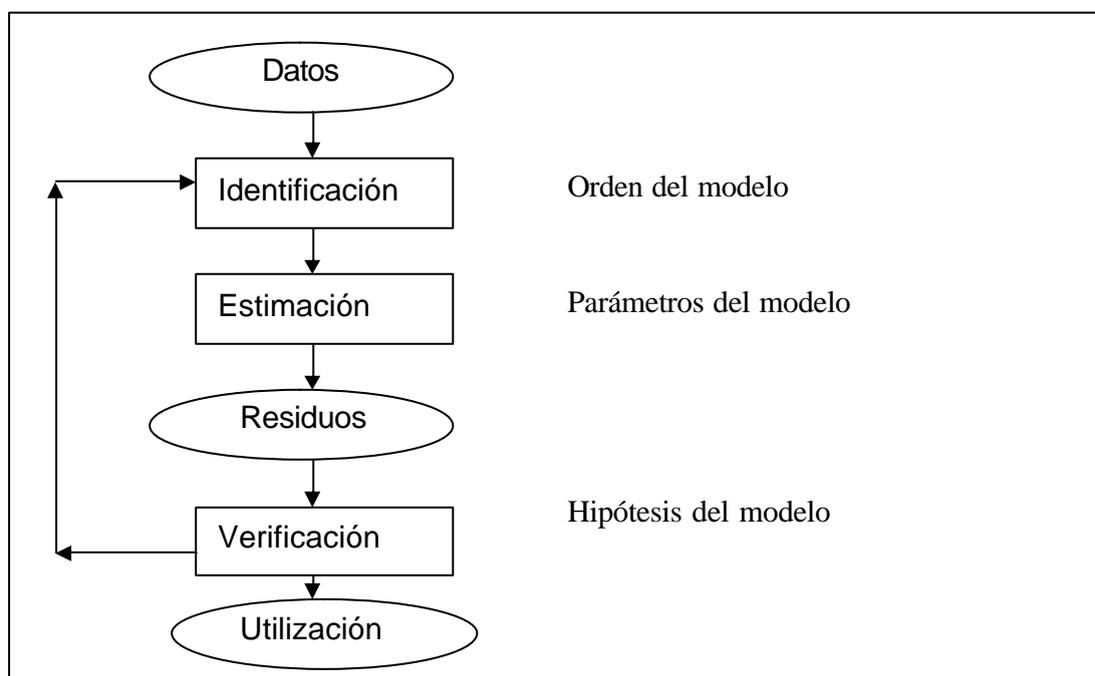


Figura.3.1 Esquema general de modelación

Fuente: Elaboración propia basa en Tesis Criterios para la Determinación de Caudales, Chile2005

Las etapas básicas comprenden la selección del tipo de modelo a emplear, la identificación del orden, la estimación de los parámetros y la validación del modelo. Para ello se deben tener en cuenta los tipos de modelos disponibles entre los cuales se pueden seleccionar los más adaptados según las propiedades que se desean modelar, así como las características de los datos o las variables a representar.

En la modelación de series cronológicas el modelo más utilizado es el de Box y Jenkins que consiste en varias etapas sucesivas y recurrentes. En realidad lo que se realiza no es una modelación sino más bien una selección del modelo más conveniente entre varias alternativas disponibles. En muchos casos ello no conduce a la obtención del mejor modelo sino más bien a la del menos malo. En general estas etapas son las siguientes:

- a) Selección del tipo. Para la simulación de un esquema multivariado se desea representar las relaciones entre las diferentes opciones. El esquema contemporáneo parece justificado dado que es el más sencillo y la dependencia espacial simultánea es la más importante. Sin embargo, el adoptar un esquema sólo autoregresivo para la dependencia temporal de cada serie no aparece igualmente justificado, pudiéndose pensar que los modelos tipo ARMA son más generales y flexibles y no es evidente un descarte preliminar de ellos.
- b) Identificación del orden del modelo. Dado un tipo de modelo se trata de identificar el orden más adecuado. Por ejemplo si se considera un modelo Periódico Autoregresivo (PAR) se determina para cada serie y en cada periodo el orden P_m de cada uno. La identificación del orden para cada período se realiza basándose en la comparación de las funciones de Autocorrelación Simple, FAS, y de Autocorrelación Parcial, FAP.
- c) Estimación de parámetros. Conocido el orden del modelo se estiman los parámetros respectivos, en base a las características de la muestra disponible. En el caso de que se conozcan las p coeficientes autoregresivo, existen varios procedimientos, siendo el más popular el método de los momentos, mediante el cual se elijen los coeficientes de manera que se reproduzcan los coeficientes de autocorrelación hasta el orden p . También es posible recurrir a minimizar la varianza de los residuos o métodos de máxima verosimilitud.
- d) Verificación de las condiciones supuestas. En esta etapa se verifica que el modelo propuesto satisfaga las condiciones de modelación. Entre ellas es importante la independencia temporal de los residuos y su normalidad. Si los residuos no son independientes se cambia el orden o se modifica el tipo de modelo propuesto. Si los residuos no son normales se pueden transformar para continuar con las etapas siguientes.

3.2.1 Modelos Típicos Disponibles

Para la modelación multivariada de series hidrológicas existen diversas alternativas. Los tipos básicos de dependencia entre varias series hidrológicas, para una modelación multivariada, se conoce como contemporánea, unidireccional y retroalimentada.

Dos o más series están contemporáneamente relacionadas si sus valores concurrentes, o contemporáneos, son independientes entre sí. En hidrología, la principal fuente de dependencia contemporánea se debe al efecto regional de la precipitación sobre varias cuencas, generadas por sistemas de extensión importante.

La dependencia unidireccional se define cuando los valores presentes y anteriores de una serie condicionan los valores presentes y futuros de otra. Normalmente esto significa una relación de causalidad entre ambas. Estos modelos también se denominan Función de Transferencia Multivariada.

Finalmente, existe dependencia retroalimentada si ambas series se condicionan mutuamente para valores pasados, presentes y futuros.

Existen tres esquemas de dependencia indicados en los cuales, mediante flechas, se señala la relación de dependencia entre las variables. Figura.3.2. Estos esquemas de modelación se pueden apreciar como relaciones lineales entre vectores que representan las series en cada instante con matrices de coeficientes responsables de la dependencia. Si la dependencia es contemporánea las matrices de coeficientes resultan del tipo diagonal, mientras en la unidireccional son triangulares y en la retroalimentada son completas. En cada caso de modelos multivariados adicionalmente a las dependencias, pueden existir dependencias temporales entre los valores de cada una de las series. La cantidad de parámetros de los modelos, así como la complejidad de la identificación, estimación, y operación de los modelos, crece enormemente entre el primer esquema y el tercero. La estructura de dependencia que puede apreciarse en series reales está condicionada por las propiedades físicas del sistema que las genera y también del intervalo de tiempo seleccionado para representar las series cronológicas involucradas.

Existen además otras proposiciones para considerar la dependencia multivariada, o espacial. Entre ellas se pueden mencionar las basadas en componentes principales, regresiones múltiples, esquemas de desagregación y de agregación.

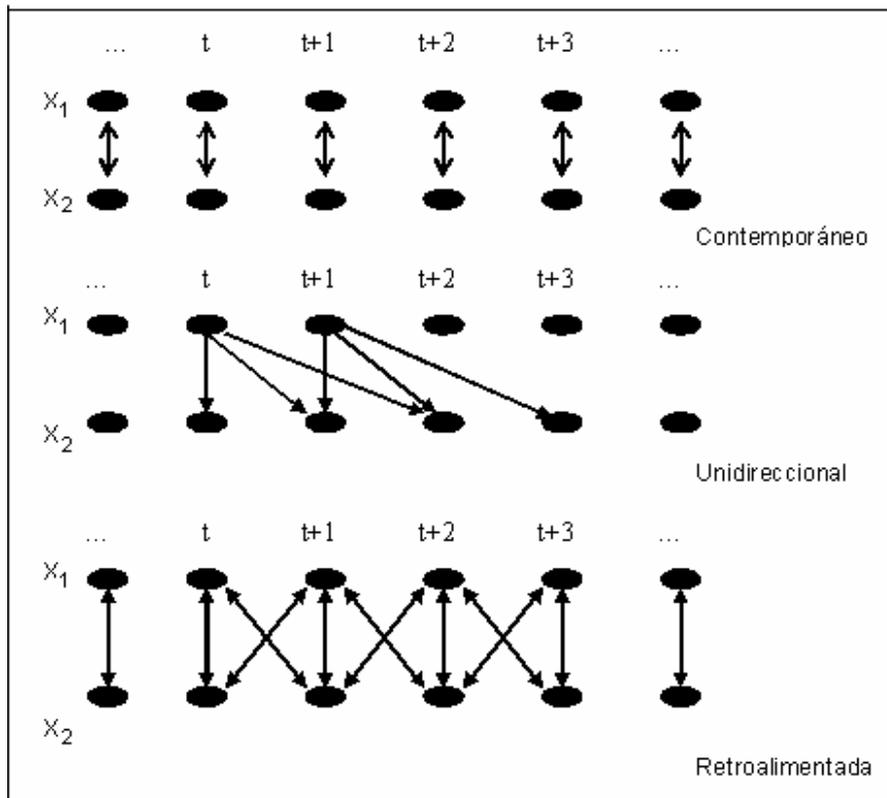


Figura. 3.2 Esquemas de dependencias multivariada.

Fuente: Tesis de Maestría en Estadística Aplicada, UNR. Directora: Msc. María Teresa Blaconá.

El esquema contemporáneo resulta muy adecuado para su empleo con modelos lineales del tipo ARMA, que capturan la dependencia temporal, ya que permite una representación adicional en etapas mediante la cual, en cada serie se remueve inicialmente la dependencia temporal y los residuos de ellas se modelan con un esquema contemporáneo para considerar el efecto multivariado. Se supone que toda la dependencia espacial no contemporánea se debe a la relación temporal impuesta en cada serie de manera independiente. Este esquema es el adoptado por Power Systems Research Inc. (PSRI) para la modelación de caudales en el sistema chileno.

3.2.2 Propiedades Estadísticas a Modelar

Los modelos de series temporales intentan reproducir algunas de las propiedades estadísticas de las series hidrológicas. En este caso las principales corresponden a la función de distribución marginal de cada una de las variables que intervienen,

Modelos de Planificación de Sistemas Hidrológicos

la dependencia temporal y la dependencia espacial. Los modelos lineales consideran la dependencia temporal como la correlación entre los valores de diferentes instantes, y la dependencia espacial como la correlación entre series en diferentes lugares.

El modelo de cada una de las series temporales es responsable de reproducir las siguientes propiedades: promedios de cada periodo (mes o semana), varianzas de cada período, asimetrías si la serie no es normal y autocorrelación temporal. Además la parte multivariada del modelo tiene en cuenta la correlación espacial entre series.

Se supone que los promedios, varianzas, asimetrías y dependencia temporal son periódicos, es decir, cambian de una semana a la otra o de un mes a otro, pero estacionarios, en el sentido que permanecen constantes para la misma semana de un año a otro.

Si $X_{m,n}$ representa el valor del caudal en un lugar cualquiera durante la semana m en el año n , con promedio \mathbf{m}_m y desviación típica \mathbf{s}_m , esta variable se puede transformar en una de promedio 0 y desviación típica 1 mediante:

$$Z_{m,n} = \frac{X_{m,n} - \mathbf{m}_m}{\mathbf{s}_m}$$

Con lo anterior todos los períodos presentan igual promedio y varianzas, lo que simplifica los cálculos posteriores.

Un modelo PAR (p_m) toma en cuenta la dependencia entre los períodos mediante una ponderación lineal p_m variables anteriores, con una relación como la siguiente:

$$Z_m = \mathbf{f}_{1,m} Z_{m-1} + \mathbf{f}_{2,m} Z_{m-2} + \dots + \mathbf{f}_{p,m} Z_{m-p} + a_m$$

Siguiendo una notación propuesta por Box y Jenkins se recurre al operador B definido de manera que $B^k Z_t = Z_{t-k}$, con lo cual la expresión anterior se convierte en:

$$\left(1 - \mathbf{f}_{1,m} B - \mathbf{f}_{2,m} B^2 - \dots - \mathbf{f}_{p,m} B^p\right) Z_m = a_m$$

Al polinomio de B de orden p se le designa como Φ^m para cada período m, con lo cual el modelo PAR (p_m) se representa de manera sintética como:

$$\Phi^m \left(\frac{X_m - \mathbf{m}_m}{\mathbf{s}_m} \right) = a_m$$

Modelos de Planificación de Sistemas Hidrológicos

Los valores a_m corresponden a los residuos del período m y se supone que son temporalmente independientes, pero pueden presentar una dependencia espacial contemporánea con los residuos similares de otras series, lo que modela posteriormente.

La consideración de la asimetría, o la no normalidad, se puede abordar de diferentes maneras. Una es suponer que en la modelación temporal se trabaja con valores normales y con la asimetría de los residuos, procediendo a su transformación si es necesario. La otra es efectuar la transformación de los valores Z antes del desarrollo del modelo temporal.

Con el esquema planteado, un modelo para w períodos tiene los parámetros que se representan en la tabla 3.1

Parámetros	Variable que lo representa	Cantidad
Promedios	m_1, m_2, \dots, m_w	w
Desviación típica	s_1, s_2, \dots, s_w	w
Asimetrías	g_1, g_2, \dots, g_w	w
Orden del modelo	p_1, p_2, \dots, p_w	w
Coef. Autoregresivo 1	$\Phi_{1,1}, \Phi_{1,2}, \dots, \Phi_{1,w}$	w
Coef. Autoregresivo 2	$\Phi_{2,1}, \Phi_{2,2}, \dots, \Phi_{2,w}$	w
...		...
Coef. Autoregresivo p_m	$\Phi_{p,1}, \Phi_{p,2}, \dots, \Phi_{p,w}$	w

Tabla 3.1. Parámetros implicados en un modelo para w períodos.
Tesis de Maestría en Estadística Aplicada, UNR. Directora: Msc. María Teresa Blaconá

3.2.3 Periodicidad y su Representación

El esquema planteado supone una gran cantidad de parámetros. Para el caso de modelos semanales de 48 semanas por año, ello significa 48 promedios, desviaciones, asimetrías y órdenes p_m del modelo. Además 48 parámetros autoregresivos. Para una serie cualquiera esto puede significar fácilmente unos 300 a 500 parámetros. Esto presenta dificultades de estimación y empleo del modelo. Para ello se recurre a una representación paramétrica mediante series de Fourier o promedios móviles.

3.3 Modelación de la Dependencia Temporal

3.3.1 Identificación de Modelos Temporales Periódicos

Si se ha seleccionado un modelo del tipo periódico lineal, como PAR(p_m) o PARMA(p_m, q_m), como el más adecuado para representar las características temporales de las series semanales, la etapa de identificación consiste en determinar el valor de p_m de cada semana, de manera que se logra que los residuos del modelo sean independientes, para ello se recurre a las funciones de Autocorrelación Simple, FAS y de Autocorrelación Parcial, FAP. Estas funciones son características de cada modelo y con ellas se trata de imitar el tipo de dependencia temporal detectado en la serie de datos disponibles que se desea modelar.

La función de Autocorrelación Simple, FAS, corresponde al coeficiente de correlación existentes entre los valores de un instante cualquiera y los de otro instante, separado por k períodos hacia atrás, lo que se designa como desfase o rezago k , esto es :

$$r^m(k) = E \left[\left(\frac{Z_m - m_m}{S_m} \right) \left(\frac{Z_{m-k} - m_{m-k}}{S_{m-k}} \right) \right]$$

Donde $E(.)$ corresponde al valor esperado. La función $r^m(k)$ en función de k se conoce con el nombre de correlograma o FAS de la semana m . Estas funciones presentan una forma característica para cada modelo y han sido propuestas por Box y Jenkins para identificar el orden del modelo por comparación. Así por ejemplo, un modelo AR(p) presenta valores diferentes de cero y determinados para $k=1,2,\dots,p$ y para valores de r decrece exponencialmente o en forma sinusoidal hasta desaparecer. Los valores $\bar{r}^m(k)$ pueden estimarse directamente de la muestra como el coeficiente de correlación entre los valores de una semana y los de otra que ocurrieron k períodos antes, como:

$$\bar{r}^m(k) = r^m(k) = \frac{1}{(N-k)S_m S_{m-k}} \sum_{n=k}^N (Z_n^m - m^m)(Z_n^{m-k} - m^{m-k})$$

Otra función útil para la determinación del orden de un modelo es la Función de Autocorrelación Parcial, FAP, que corresponde a los valores del coeficiente autoregresivo de orden k cuando se ajusta a la muestra un modelo AR(k). Estos valores para $k = 1, 2, 3, \dots$ en función de k , se denominan la función FAP. El comportamiento de esta función es tal, que un modelo AR(p) es diferente de cero sólo para k menor o igual que p y posteriormente es nula, de manera que es más simple de utilizar en la identificación de modelos AR que FAS. Si las variables son

Modelos de Planificación de Sistemas Hidrológicos

normales los valores de esta función se distribuyen normales con promedio cero y varianza $1/N$.

Para estimar los valores de la FAS es conveniente considerar que el modelo PAR (p_m) se expresa en forma reducida como:

$$\Phi^m(B) \left(\frac{Z_t - \mathbf{m}_m}{\mathbf{s}^m} \right) = a_t^m$$

Para simplificar la nomenclatura es conveniente designar a la variable centrada y reducida por Y_t donde:

$$Y_t = \left(\frac{Z_t - \mathbf{m}_m}{\mathbf{s}_m} \right)$$

De manera que el modelo PAR(p_m) corresponde a:

$$Y_t = \mathbf{f}_1^m Y_{t-1} + \mathbf{f}_2^m Y_{t-2} + \dots + \mathbf{f}_{p_m}^m Y_{t-p_m} + a_t$$

La estructura de dependencia temporal de este modelo puede encontrarse multiplicando por Y_{t-k} tomando valor esperado:

$$E(Y_t Y_{t-k}) = E(\mathbf{f}_1^m Y_{t-1} Y_{t-k}) + E(\mathbf{f}_2^m Y_{t-2} Y_{t-k}) + \dots + E(\mathbf{f}_{p_m}^m Y_{t-p_m} Y_{t-k}) + E(a_t Y_{t-k})$$

Como las variables están centradas y reducidas esta ecuación indica:

$$\mathbf{r}^m(k) = \mathbf{f}_1^m \mathbf{r}^{m-1}(k-1) + \mathbf{f}_2^m \mathbf{r}^{m-2}(k-2) + \dots + \mathbf{f}_{p_m}^m \mathbf{r}^{m-p_m}(k-p_m)$$

Estas ecuaciones definen la estructura de dependencia temporal de los modelos AR(p) y se conocen con el nombre de ecuaciones de Yule-Walker. Esta ecuación permite establecer un sistema de p ecuaciones que ligan los valores de los p parámetros del modelo con los valores de la función de autocorrelación.

La ecuación de Yule-Walker puede emplearse en diferentes etapas de la modelación, como se menciona a continuación:

a) Conocidos los parámetros de un modelo $\mathbf{f}_1^m, \mathbf{f}_2^m, \dots, \mathbf{f}_{p_m}^m$ y por supuesto su orden p_m , es posible determinar su estructura de dependencia encontrando los valores $\mathbf{r}^m(k)$ para $k = 1, 2, 3, \dots$. Estos valores se pueden comparar con los estimados de una muestra en la etapa de validación para verificar si un modelo seleccionado es adecuado. Este es el procedimiento basado en la FAS.

b) Si se dispone de estimaciones de la función de autocorrelación, $\hat{r}^m(k)$, obtenidas de una muestra que se desea modelar, se pueden calcular los valores de los parámetros $f_1^m, f_2^m, \dots, f_{p_m}^m$ de manera que se produzcan los primeros p valores de la función de autocorrelación. Esto es lo que se hace frecuentemente en la etapa de estimación de parámetros de modelos PAR cuando se emplea para ello el método de los momentos.

c) Finalmente si se dispone de valores estimados de la función de autocorrelación $\hat{r}^m(k)$ es posible obtener valores de la función de autocorrelación parcial, ajustando sucesivamente modelos de orden $p=1,2,3,\dots,k$, y calculando en cada caso el valor del último coeficiente autoregresivo de cada uno de ellos, esto permite emplear este sistema de ecuaciones en la etapa de identificación del orden del modelo.

Metodológicamente la secuencia lógica del empleo de esta ecuación es para la estimación de los coeficientes $f_{k,k}$ en la etapa de identificación, posteriormente en la etapa de estimación para calcular los parámetros $f_1^m, f_2^m, \dots, f_{p_m}^m$ dado el valor de p , y finalmente para establecer los valores de la función de autocorrelación del modelo en la etapa de verificación.

Es conveniente hacer notar que resulta difícil tomar decisiones totalmente automáticas respecto al orden del modelo con las herramientas mostradas. También se utiliza el denominado criterio de información Akaike, AIC, para seleccionar entre modelos similares, según el cual es conveniente elegir el modelo que minimice el AIC, calculado como:

$$AIC = 2p + N \ln \left(s_a^m \right)^2$$

3.3.2 Estimación de Parámetros

Una vez identificado el orden del modelo corresponde estimar sus parámetros. En cada semana m se conoce el orden p_m , y la etapa de estimación consiste en encontrar los valores que deben adoptar los coeficientes $f_1^m, f_2^m, \dots, f_{p_m}^m$ y las características de los residuos a_t , fundamentalmente su varianza, $s_m^2(a)$.

El sistema de ecuaciones para la estimación de parámetros de un modelo PAR(p_m) es el mismo que el necesario para calcular los valores de la FAP, con la única diferencia de que en este último caso se debe resolver varias veces para valores de $p=1,2,3,\dots,L$, siendo L un número adecuadamente elevado que asegure ser mayor que el más alto de los valores de p posibles. El sistema de ecuaciones

Modelos de Planificación de Sistemas Hidrológicos

se obtiene aplicando la ecuación de Yule- Walker para $k=1,2,\dots, p_m$, de lo cual resulta:

$$\begin{aligned} \mathbf{r}^m(1) &= \mathbf{f}_1^m + \mathbf{f}_2^m \mathbf{r}^{m-1}(1) + \mathbf{f}_3^m \mathbf{r}^{m-1}(2) + \mathbf{f}_{p_m}^m \mathbf{r}^{m-1}(1) \\ \mathbf{r}^m(2) &= \mathbf{f}_1^m \mathbf{r}^{m-1}(1) + \mathbf{f}_2^m + \mathbf{f}_3^m \mathbf{r}^{m-2}(2) + \mathbf{f}_{p_m}^m \mathbf{r}^{m-2}(2) \\ \mathbf{r}^m(3) &= \mathbf{f}_1^m \mathbf{r}^{m-1}(2) + \mathbf{f}_2^m \mathbf{r}^{m-2}(1) + \mathbf{f}_3^m + \mathbf{f}_{p_m}^m \mathbf{r}^{m-3}(3) \end{aligned}$$

$$\mathbf{r}^m(p_m) = \mathbf{f}_1^m \mathbf{r}^{m-1}(p_m-1) + \mathbf{f}_2^m \mathbf{r}^{m-2}(p_m-2) + \mathbf{f}_{p_m}^m$$

Se acostumbra en este sistema poner términos matriciales como:

$$\begin{bmatrix} 1 & \mathbf{r}^{m-1}(1) & \dots & \mathbf{r}^{m-1}(p_m-1) \\ \mathbf{r}^{m-1}(1) & 1 & \dots & \mathbf{r}^{m-2}(p_m-2) \\ \dots & \dots & \dots & \dots \\ \mathbf{r}^{m-1}(p_m-1) & \mathbf{r}^{m-2}(p_m-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} \mathbf{f}_1^m \\ \mathbf{f}_2^m \\ \dots \\ \mathbf{f}_{p_m}^m \end{bmatrix} = \begin{bmatrix} \mathbf{r}^m(1) \\ \mathbf{r}^m(2) \\ \dots \\ \mathbf{r}^m(p_m) \end{bmatrix}$$

Existen varios métodos para estimar los parámetros de modelos de series cronológicas. Los más empleados son: el método de momentos, el de mínimos cuadrados y el de máxima verosimilitud.

El método de los momentos consiste en igualar los momentos de la muestra con los del modelo, para lo cual una vez identificado el orden del modelo se intenta reproducir su estructura de dependencia temporal. Para un modelo PAR, esto se realiza estableciendo las ecuaciones de Yule-Walker para cada período de manera que se igualen los primeros p_m coeficientes de autocorrelación. Esto conduce, para cada, período a un sistema de p_m ecuaciones como las indicadas, pero remplazando los valores de \mathbf{r} por sus estimadores $\mathbf{r}^m(k)$, de manera que la solución se obtiene como:

$$\hat{\mathbf{f}} = \mathbf{R}^{-1} \mathbf{r}$$

Donde el vector $\hat{\mathbf{f}}$ es $(\mathbf{f}_1^m, \mathbf{f}_2^m, \dots, \mathbf{f}_{p_m}^m)^T$ para cada semana m . En el caso de modelos PAR cada semana tiene su sistema de ecuaciones independiente, de manera que la estimación por este procedimiento es sencilla. Si el modelo presenta términos de promedios móviles, el sistema de ecuaciones incluye una mezcla de semanas anteriores, lo que dificulta su utilización.

Una vez determinados los parámetros $\hat{\mathbf{f}}$ se puede encontrar el valor de la varianza de los residuos con la ecuación de Yule-Walker para $k=0$:

$$s_m^2(a) = 1 - \sum_{i=1}^{p_m} f_i^m r^m(i)$$

Otro método alternativo para la estimación de parámetros dado el orden del modelo consiste en minimizar la varianza de los residuos, sujeto a que los parámetros satisfagan las condiciones de estacionalidad e invertibilidad cuando corresponda. En el caso de variables normales ambos métodos conducen a resultados similares. Desde el punto de vista estadístico, la mejor opción de estimación es el empleo del método de máxima verosimilitud. Sin embargo, las funciones de verosimilitud no se conocen para todo tipo de modelos, lo que dificulta su utilización práctica.

3.4 Construcción del Modelo

Para encontrar el modelo más adecuado se sigue un esquema general propuesto por Box y Jenkins que se basa en la etapa de identificación, estimación y verificación, ordenadas como se ilustra en la figura 3.1

Adicionalmente, en la construcción del modelo se deben tomar decisiones en cuanto a la conveniencia de normalizar las variables involucradas y parametrizar o adoptar una representación parsimoniosa de los parámetros involucrados. Para períodos semanales, la parametrización es necesaria y se puede hacer representando mediante series de Fourier, el comportamiento periódico de los parámetros. Estos procedimientos se pueden realizar en diferentes etapas durante la construcción del modelo, pero se debe estar consciente de que los resultados finales estarán afectados por el esquema en el cual se realizan.

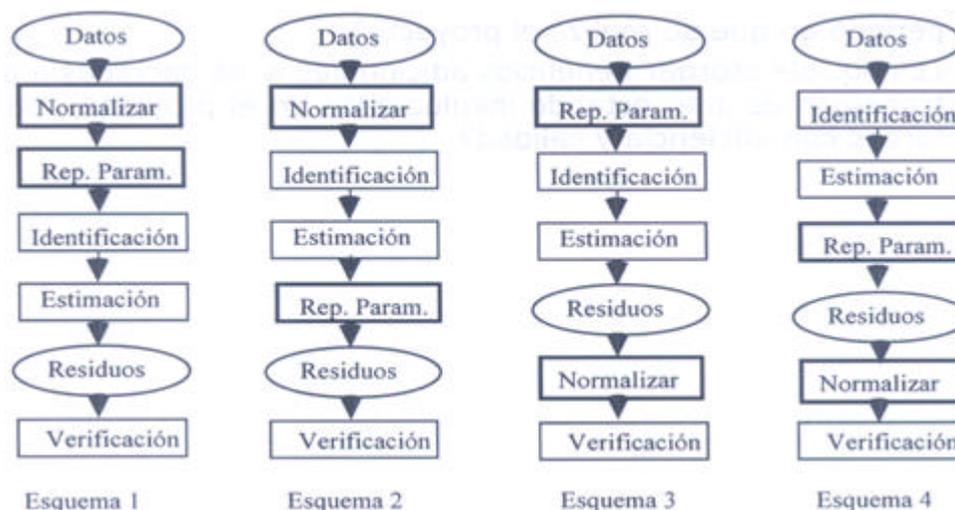


Figura.3.3 Esquemas alternativos del procedimiento para incorporar la normalización y la parametrización
 Fuente: Elaboración propia basa en el artículo Modelo de Utilización onjunta en la Cuenca Andreu y Marco,1983

Habitualmente las alternativas de normalización que se plantean son las siguientes: no normalizar los datos o normalizar los residuos. Similarmente las alternativas para la parametrización son no parametrizar, parametrizar las propiedades de los datos o parametrizar los parámetros del modelo. Las decisiones que se adopten en relación con estas alternativas, conducen a una gran cantidad de posibilidades de procedimiento de modelación. La más empleada, es la que se muestra en la figura 3.3, que considera sólo casos en que se normaliza y se parametriza, dejando de lado todas las situaciones en que se hace sólo uno de estos dos procesos o no se hace ninguno.

Dentro de los esquemas planteados en que se realizan los dos procedimientos, los más adecuados son los sindicados como esquemas 1 y 2, en la figura 3.3. Estos presentan la ventaja de que la identificación y estimación se realizan con variables normales, lo que asegura que los residuos son normales y satisface las condiciones de normalidad implícitas en los métodos empleados. Además al proceder así, se puede visualizar primero la calidad de los datos disponibles y decidir su utilización o modificación inmediata de acuerdo a ello.

En todo caso si se trabaja con caudales semanales es absolutamente necesario enfrentar el asunto de la normalización y la parametrización, siendo éste un aspecto esencial para lograr series sintéticas de calidad.

3.5 Diagnóstico o Verificación del Modelo.

Si el modelo es adecuado los residuos deben comportarse como ruido blanco. Para verificarlo se calculan los residuos como:

$$a_{v,m} = \mathbf{f}^m(B) X_{v,m}$$

La verificación o diagnóstico del modelo requiere comprobar que las hipótesis básicas realizadas respecto a los residuos son ciertas. Estos deben tener: media cero, varianza constante, falta de correlación para cualquier retardo, y distribución normal. Adicionalmente se pueden realizar otras comprobaciones como es la existencia de outliers, o la reproducción de propiedades de interés especial de acuerdo a los fines que se persiguen con el modelo.

Como los residuos provienen de modelos periódicos, la varianza de ellos también es periódica, de manera que es conveniente antes de proceder a las pruebas, efectuar una estandarización periódica mediante:

$$v_t = v_{w(n-1)+m} = \frac{a_{n,m-m}(a)}{s_m(a)}$$

Utilizando para ellos los promedios y las varianzas de los residuos de cada período.

3.5.1 Pruebas de Independencia

El primer contraste a realizar es si los residuos estimados están incorrelacionados. Para ello existen diferentes pruebas. Es conveniente realizar varias de ellas y utilizarlas en conjunto para establecer un criterio de rechazo o aceptación del modelo.

Estas pruebas se basan en la estructura de correlación existente en las series de residuos. Para ello se calcula la función FAS de cada período mediante:

$$r_a^m(k) = \text{Corr}(a_m, a_{m-k})$$

Donde a_m es el residuo del a semana m . Si éstos son independientes y normales los coeficientes $r_a^m(k)$ serán aproximadamente, para retardos altos, variables aleatorias con media cero, varianzas asintóticas $1/N$ y distribución normal. Por ejemplo, para un AR(1) la desviación típica asintótica de $r_a^m(1)$ es f/\sqrt{N} que puede ser mucho menor que $1/\sqrt{N}$.

El procedimiento habitual de verificar la independencia de los residuos es comparar sus funciones de autocorrelación simple con los límites $\pm 2/\sqrt{N}$ y comprobar si todos los coeficientes están dentro de estos límites de confianza. Como estos límites son, aproximadamente del 95%, en promedio uno de cada veinte coeficientes de autocorrelación estimados saldrán fuera, por lo que la aparición de un valor significativo en un retardo elevado es separable. Sin embargo, como según lo anterior estos límites sobrestiman la varianza en los pequeños, un valor próximo a los límites de confianza en los residuos finales debe considerarse un indicio claro de que el modelo es inadecuado.

Un contraste global de que los primeros L coeficientes son cero (L debe ser grande, del orden de $1/4$ a $1/3$ de la muestra) es el contraste de Ljung-Box. Si los residuos son realmente ruido blanco, el estadístico:

$$Q^m(L) = N(N+2) \sum_{j=1}^L \frac{(r_a^m(j))^2}{N-j}$$

Se distribuye asintóticamente como una ji- cuadrada con $(L-p_m)$ grados de libertad.

3.5.2 Prueba de Normalidad

En el caso de la modelación hidrológica es importante que los residuos de cada aprovechamiento sean normales, ya que ellos se emplearán para modelar la estructura espacial del sistema. Esto se puede verificar calculando el coeficiente de asimetría de los residuos de cada semana y contrastándolo con la hipótesis de que proviene de una muestra normal, para la cual el coeficiente de asimetría tiene promedio 0 y varianza $6/N$.

Para enfrentar esta situación se puede proceder a una transformación del tipo Box-Cox, según la cual:

$$e_i = \begin{cases} a_i^I & ;\text{si } I > 0 \\ I & \\ Ln(a_i) & ;\text{si } I = 0 \end{cases}$$

3.5.3 Promedio Nulo y otras Propiedades

Si bien en general los residuos de un modelo periódico no están en principio sujetos a la restricción de que sean nulos, es conveniente para facilitar la construcción del modelo. Para contrastar esta hipótesis en el caso general suponiendo N residuos y r parámetros, se puede hacer suponiendo que se trata de valores normales con promedio cero y varianza σ_a^2 . Entonces el promedio es también normal con valor esperado cero y varianza σ_a^2/N . este contraste debe aplicarse después de comprobar que los residuos están incorrelacionados.

La estabilidad de la varianza se comprueba estudiando el gráfico de los residuos a lo largo del tiempo. En caso de duda se puede dividir el intervalo muestral en tres o cuatro partes y aplicar un test de la razón de verosimilitud.

Finalmente, conviene siempre situar la gráfica de los residuos estimados a lo largo del tiempo. Su inspección se facilita situando en él los límites de 95%, 99% o 99.5% y estudiando los puntos que salen fuera de estos límites de control para detectar la existencia de outliers.

3.5.4 Reformulación

Un aspecto interesante en la verificación del modelo estimado es considerar a los residuos, a como una serie temporal y estudiar su estructura. Esto permite orientar las correcciones al modelo. Suponiendo que se ha identificado y estimado el modelo $\text{PAR}^m(p_m)$:

$$a_{n,m} = \mathbf{f}_X^m(B) X_{n,m}$$

Y al analizar la FAS y la FAP de los residuos estimados, a se aprecia que éstos no son independientes sino que siguen un modelo también autoregresivo:

$$e_{n,m} = \mathbf{f}_a^m(B) a_{n,m}$$

Donde los nuevos residuos e sí satisfacen todas las condiciones. Entonces, substituyendo en la ecuación anterior, resulta para la serie el modelo

$$e_{n,m} = \mathbf{f}_X^m(B) \mathbf{f}_a^m(B) X_{n,m}$$

Que correspondería al modelo adecuado.

3.5.5 Sobreajuste

Otra alternativa para comprobar si el modelo es adecuado, es utilizar la técnica del sobreajuste, que consiste en estimar un modelo de orden mayor que el analizado y comprobar si se obtienen coeficientes estimados significativos. Con esto es posible captar pequeñas estructuras remanentes que pueden mejorar las predicciones. En general, si se ha ajustado un $\text{PAR}(p)$ que parece adecuado, el sobreajuste se aplica estimando los modelos $\text{PAR}(p+1)$ o $\text{PAR}(p+2)$ y comprobando si los parámetros adicionales son significativos.

3.6 Modelado de la Dependencia Espacial.

Cuando existen varias series en varios lugares diferentes de una región de la misma variable, como es el caso de los aprovechamientos del SIC, o de diferentes variables en un mismo lugar, o combinaciones similares que presentan dependencia entre los valores de las series, se deben considerar modelos multivariados. En hidrología se han propuesto varios esquemas de modelación multivariada.

Entre las alternativas para sistemas complejos se pueden considerar los modelos Multivariados ARMA, de los cuales se desprenden los multivariados de Funciones

Modelos de Planificación de Sistemas Hidrológicos

de Transferencia y los contemporáneos ARMA. Los primeros son una generalización matricial de los ARMA. Como tales presentan una cantidad importante de parámetros que los hacen inmanejables cuando la cantidad de series supera el caso bivariado. El modelo más desarrollado corresponde al contemporáneo CARMA, en el cual se modelan las series temporales de cada aprovechamiento de manera separada y se mantiene la dependencia espacial mediante la relación entre los residuos de cada lugar. En este caso se propone una aplicación particular de modelos CARMA considerando para las series individuales, sólo modelos Autoregresivos periódicos, con lo cual se tiene realmente un modelo CAR.

Si cada serie se ha modelado de manera que se dispone de un vector de residuos para el instante t en los k lugares:

$$V_t = [V_t(1), V_t(2), \dots, V_t(k)]^T$$

Donde $V_t(j)$ corresponde al residuo normalizado del aprovechamiento j en el instante t . En este caso se supone que ya coexiste periodicidad entre los residuos de un lugar, de manera que $t=w(n-1)+m$, donde $n=1,2,\dots,N$ es el año y $m=1,2,\dots,w$ es el período. De esta forma existen wN valores de $V_t(j)$, con $j=1,2,\dots,K$. Además $V_t(k)$ es una serie de valores normales, independientes temporalmente, con promedio mVK y desviación sVK .

El comportamiento espacial de estos residuos se puede apreciar en matrices de correlación de desfases h entre las series D^h , cuyos elementos son:

$$\{D^h\} = \{d_{ij}^h\} \text{ con } d_{ij}^h = \frac{E[(V_t(i) - m_{Vi})(V_{t-h}(j) - m_{Vj})]}{s_{Vi}s_{Vj}}$$

Donde d_{ij}^h corresponde a la correlación entre los residuos del lugar i y los del lugar j rezagados en h intervalos de tiempo. En particular, para $h=0$ se tiene la matriz de correlaciones contemporáneas del sistema, que es una matriz simétrica con 1 en la diagonal:

$$\{D^0\} = \begin{bmatrix} 1 & r_{1,2}^0 & \dots & r_{1,k-1}^0 & r_{1,k}^0 \\ r_{2,1}^0 & 1 & \dots & \dots & r_{2,k}^0 \\ \dots & \dots & \dots & \dots & \dots \\ r_{k-1,1}^0 & & & 1 & \\ r_{k,1}^0 & r_{k,2}^0 & & & 1 \end{bmatrix}$$

Modelos de Planificación de Sistemas Hidrológicos

La modelación espacial se hace de manera de reproducir la dependencia contemporánea, simultánea, entre las series de $V_t(k)$. El modelo espacial contemporáneo para los residuos tiene la forma:

$$e_t = M w_t$$

Donde w_t representa un vector de K residuos independientes entre sí para el instante t , mientras e_t corresponde a los residuos de cada lugar centrados y reducidos, de manera que sus elementos son:

$$e_{t(i)} = \frac{V_t(i) - m_{Vi}}{s_{Vi}}$$

Entonces se puede demostrar que:

$$E[e_t e_t^T] = E[(M w_t)(M w_t)^T]$$

Donde $E(\cdot)$ corresponde al valor esperado, pero

$$E[e_t e_t^T] = \{D^0\}$$

Que es la matriz de covarianzas, o correlaciones si los valores están estandarizados, de los residuos temporales de cada serie. Mientras que como los valores de w son independientes, normales estándar, se cumple que:

$$E[w_t w_t^T] = I$$

Siendo I la matriz de identidad. Entonces, combinando las tres últimas ecuaciones, los coeficientes de la matriz M cumplen con la condición:

$$D^0 = MM^T$$

Para determinar los valores de los elementos de esta matriz, $\{m_{ij}\}$ es conveniente imponer las restricciones de que M sea triangular inferior, es decir que tenga valores nulos sobre la diagonal. Así se puede emplear la descomposición de Cholesky para resolver la ecuación anterior, según la cual:

$$m_{11} = \sqrt{r_{11}}$$

$$m_{1j} = r_{1j} / m_{11} \text{ para } j=2,3,\dots,k$$

$$m_{ii} = \sqrt{r_{ii} - \sum_{k=1}^{i-1} m_{ki}^2} \text{ para } i=2,3,\dots,k$$

$$m_{ij} = \frac{1}{m_{ii}} \left(r_{ij} - \sum_{k=1}^{i-1} m_{ki} m_{kj} \right) \text{ para } i=2,3,\dots,k; j=i+1,i+2,\dots,k$$

$$m_{ij} = 0 \text{ para } i > j$$

Entonces si se dispone de k valores $iid \sim N(0,1)$, $[w_t(1), w_t(2), \dots, w_t(k)]$, se pueden generar residuos dependientes temporalmente que reduzcan la dependencia espacial encontrada mediante:

$$e_t(i) = \sum_{j=1}^i m_{ij} w_t(j) \text{ para } i=1,2,\dots,k$$

Donde $e_t(1)$ corresponde al residuo $v_t(1)$, centrado y reducido. Esto permite expresar matricialmente el modelo espacial con la relación:

$$\begin{bmatrix} v_t(1) \\ v_t(2) \\ \dots \\ v_t(k) \end{bmatrix} = \begin{bmatrix} \mu_{v1} \\ \mu_{v2} \\ \dots \\ \mu_{vk} \end{bmatrix} + \begin{bmatrix} \sigma_{v1} & 0 & \dots & 0 \\ \dots & \sigma_{v1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{v1} \end{bmatrix} \begin{bmatrix} m_{11} & 0 & \dots & 0 \\ m_{21} & m_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ m_{k1} & m_{k2} & \dots & m_{kk} \end{bmatrix} \begin{bmatrix} w_t(1) \\ w_t(2) \\ \dots \\ w_t(k) \end{bmatrix}$$

3.7 Uso de los Modelos

Una vez construido el modelo, es decir, determinada su estructura y los parámetros, se puede utilizar para simulación o pronósticos. Uno de los principales usos del modelo es el análisis del sistema y de la información disponible, lo que se hace en la etapa de construcción.

3.7.1 Simulación

El conjunto de recursos hídricos del sistema interconectado que se modela mediante el modelo hidrológico, se emplea fundamentalmente en simulación. Esta simulación puede ser condicional a la situación actual o independiente.

La simulación independiente consiste en generar mediante el modelo, un conjunto de M nuevas alternativas de series hidrológicas en K lugares para N años en cada uno, que sean estadísticamente indiferentes a las series históricas disponibles como datos. Los nuevos valores son independientes de la condición actual y cada uno tiene valores iniciales diferentes. Normalmente se simula una cantidad de

Modelos de Planificación de Sistemas Hidrológicos

años similares a la de los registros históricos, es decir, del orden de 40 ó más años.

La simulación condicionada consiste en generar igualmente M nuevas alternativas de los K lugares y de N años, pero tomando como iniciales los valores registrados recientemente en cada lugar, de manera que se simulan posibles años futuros a partir de la condición actual. Normalmente en este caso, la cantidad de años que se simulan son menos que en la alternativa anterior, pudiendo ser del orden de 5 ó menos para representar la alternativa hidrológica del futuro mediato.

Para generar un nuevo conjunto de valores similares, se deben efectuar las siguientes etapas básicas: generar valores independientes espacial y temporalmente, introducir la dependencia espacial, transformar los residuos de acuerdo a su normalidad, incorporar la dependencia temporal en cada serie y desestandarizar los valores periódicamente. Cada una de estas etapas se explica detalladamente, utilizando como ejemplo un modelo espacial contemporáneo autoregresivo con una estructura temporal periódica también autoregresiva.

Se supone que se desea generar K series cada una de N años y w períodos. El primer paso consiste en generar para cada instante t, (con $t=1,2,\dots,w,w+1,\dots,Nw$) en cada lugar $K=1,2,\dots, K$, valores independientes idénticamente distribuidos normales de promedio cero y desviación típica unitaria, es decir valores Ind. (0,1).

La generación de un valor normal estándar se logra por varios procedimientos. Uno es la transformación de Box-Muller:

$$\eta_t^k = \left[2 \ln \left(\frac{1}{u_1} \right) \right]^{0.5} \cos(2\pi u_2)$$

Donde η_t^k es el valor para el instante t en el lugar k. Esto supone disponer para cada valor de dos números independientes Uniformes (0,1), u_1 y u_2 .

Así se puede obtener una tabla de valores normales estándar independientes temporalmente y espacialmente en los k lugares de interés.

Para incorporar la dependencia espacial entre los diferentes lugares para cada instante, se emplea el modelo contemporáneo desarrollado. Para ello se aplica dicho modelo, lo que supone realizar:

$$V_t = M h_t$$

O explícitamente para cada instante t:

Modelos de Planificación de Sistemas Hidrológicos

$$\begin{bmatrix} V_t^1 \\ V_t^2 \\ \dots \\ V_t^{k-1} \\ V_t^k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ m_{21} & m_{22} & 0 & \dots & 0 \\ m_{31} & m_{32} & m_{33} & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ m_{k1} & m_{k2} & \dots & \dots & m_{kk} \end{bmatrix} \begin{bmatrix} h_t^1 \\ h_t^2 \\ \dots \\ h_t^{k-1} \\ h_t^k \end{bmatrix}$$

Los elementos de esta matriz cumplen con la relación:

$\sum_{j=1}^k m_{ij}^2 = 1$; para $i = 1, 2, \dots, k$ los valores V deben ser independientes temporalmente en cada lugar, normales estándar y reflejar una dependencia espacial de acuerdo al modelo contemporáneo planteado. Es decir, debería verificarse en esta etapa, que se reproducen las matrices de dependencia espacial de desfase nulo, uno, dos, ..., verificando que sólo la primera es diferente de cero y que a su vez contiene valores estadísticamente iguales a los detectados en los datos.

Una vez que se dispone de los valores de residuos espacialmente dependientes en los k lugares, se debe reconstruir en cada lugar las series temporales.

Lo que se menciona a continuación corresponde al caso de una serie y debe por lo tanto, repetirse k veces para completar una simulación. Debe tenerse cuidado en mantener ordenados los valores de cada período para no distorsionar la dependencia espacial creada entre ellos.

Cada uno de los vectores de residuos $V^k = [V_1^k, V_2^k, \dots, V_{Nw-1}^k, \dots, V_{nw}^k]^2$ se pueden poner en forma de matriz de w períodos (48 semanas) y N años usando:

$$V_n^m(k) = V_{(n-1)w+m}^m$$

Con $m=1, 2, \dots, w$; $n=1, 2, \dots, N$. Normalmente $w=12$ meses o 48 semanas y $N=40$ años como es habitual.

Estos residuos corresponden a los que deben incorporarse en el modelo temporal de la serie k . por ahora son normales $(0,1)$ de manera que corresponde desestandarizarlos para incorporarles la varianza del residuo de cada período:

$$a_n^m(k) = \mathbf{m}_n(k) + \mathbf{s}_m(k) V_n^m(k)$$

Si fue necesario transformarlos en normales mediante alguna operación, son los pasos a seguir para su destransformación. Lo más típico es que en la etapa de

Modelos de Planificación de Sistemas Hidrológicos

modelación se procedió a tomar logaritmo de los valores de los residuos para normalizarlos, entonces ahora corresponde proceder de manera inversa.

Una vez que se dispone en cada lugar de las series de residuos temporales de cada semana para el conjunto de años deseado se debe incorporar en ellos la dependencia temporal, empleando el modelo PAR^m (p_m) correspondiente a cada período en cada serie:

$$Z_n^m = f_1^m Z_n^{m-1} + f_2^m Z_n^{m-2} + \dots + f_{p_m}^m Z_n^{p_m} + a_n^m$$

Si la simulación se hace para generar valores independientes de la condición histórica reciente, los valores se suponen nulos y se genera una cantidad razonable de años adicionales hasta que se independicen de las condiciones de origen. Por ejemplo No=5 a 10 años, los que posteriormente se descartan. Si la simulación es condicionada a los valores actuales, se emplean los últimos valores registrados como valores de partida para cada simulación.

Finalmente, se deben desestandarizar periódicamente los valores registrados de acuerdo al comportamiento semanal, o mensual, de la variable original:

$$x_n^m(k) = m_m(k) + s_m(k) Z_n^m(k)$$

Es conveniente revisar si estas series satisfacen las propiedades temporales y espaciales que se supone que el modelo reproduce. Además, es necesario revisar la existencia de valores negativos y extremos. La generación de valores negativos puede ser uno de los mayores inconvenientes de estos modelos para la generación de series sintéticas de caudales.

3.7.2 Pronóstico

Un pronóstico de interés consiste en estimar el valor del caudal medio mensual de las semanas siguientes a la fecha en que se dispone de información, contando con los datos de los días o de las semanas anteriores. Además de un valor estimado, es importante poder entregar un indicador de la calidad de este pronóstico, lo que normalmente se refleja en las características estadísticas de error que se espera cometer. Es conveniente que este error tenga promedio nulo, es decir, sea no sesgado, y mínima varianza. En estas condiciones se dice que el pronóstico es estadísticamente óptimo. Esta es una propiedad de los pronósticos realizados con los modelos PAR. La varianza del error es indicador de la calidad del pronóstico, y aumenta a medida que crece el intervalo de pronóstico, hasta hacerse igual a la varianza de la variable pronosticada para un tiempo muy largo, momento en el cual el disponer de información sobre lo ocurrido recientemente, no influye sobre los valores que se desea pronosticar.

Modelos de Planificación de Sistemas Hidrológicos

Las ecuaciones de pronóstico estiman el valor esperado de los caudales semanales condicionados a los últimos valores registrados en las semanas inmediatamente precedentes al día del pronóstico, y definen una banda de confianza en torno a esos valores.

Dado estos valores de Q_0, Q_1, Q_2, \dots se transforman y reducen aceptando que Q_0 corresponde a la semana m . Estos valores se obtienen I de Y_0 y Z_0 dados por

$$Y_0 = g_m(Q_0)$$

$$Z_0 = \frac{Y_0 - \bar{y}_m}{s_m}$$

Donde $g_m(.)$ corresponde a la transformación aplicada en la semana m para convertir los valores en normales. $\bar{y}_m s_m$ corresponden al promedio y la desviación típica de estos valores transformados. Esto mismo se aplica a todos los valores de las semanas anteriores registradas, necesarias para el pronóstico, que pueden ser dos o tres, dependiendo del orden del modelo, para obtener un conjunto de valores semanales (Z_0, Z_1, Z_2, \dots).

El pronóstico para la semana $k+m$, es decir, para la semana k siguiente está dado por:

$$\hat{Z}_k = f_{1,m+k-1} \hat{Z}_{k-1} + f_{2,m+k-1} \hat{Z}_{k-2} + \dots + f_{p,m+k-1} \hat{Z}_{k-p} \quad \text{para } k=1,2,3,\dots,L$$

Los valores de \hat{Z}_k corresponden a los medidos para $k < 1$, o a los estimados para $k=1,2,\dots,L$

La varianza del error del pronóstico para la semana $m+k$ está dada por $V(k)$ que se calcula como:

$$V(k) = \left(1 + \sum_{j=1}^{k-1} y_j^2 \right) s_a^2$$

Donde los coeficientes y_j se evalúan en función de los parámetros de los modelos involucrados en el pronóstico. En este caso como los modelos son periódicos deben combinarse diferentes modelos. Así por ejemplo, aceptando modelos AR(2) para las semanas involucradas a partir de la m , la varianza del error del pronóstico para la semana $m+k$ está dada por $V(k)$:

$$y_1 = 1$$

$$y_2 = f_{1,m+2}$$

Modelos de Planificación de Sistemas Hidrológicos

$$y_3 = f_{2,m+3} + f_{1,m+2} f_{1,m+3}$$

$$y_4 = f_{1,m+4} f_{2,m+3} + f_{2,m+4} f_{1,m+2} + f_{1,m+4} f_{1,m+3} f_{1,m+3}$$

A medida que aumenta el plazo de pronósticos se requieren coeficientes de orden superior, los que se obtiene siguiendo el mismo esquema. Cuando los modelos de las semanas involucradas son de menor orden, se deben eliminar o suponer nulos, los coeficientes autoregresivos correspondientes. El comportamiento de estos coeficientes es tal que la varianza del error para la semana $k=1$ es igual a la varianza del residuo del modelo, pero a medida que el plazo de pronóstico aumenta, $k=2,3,\dots$ la varianza aumenta hasta igualarse a la varianza de la variable pronosticada, es decir, el efecto del conocimiento del estado actual del sistema se pierde.

Como los errores son normales, el ancho de la banda de confianza para un nivel a está dado por:

$$\hat{Z}_k^+ = \hat{Z}_k + N_a \sqrt{V(k)}$$

$$\hat{Z}_k^- = \hat{Z}_k - N_a \sqrt{V(k)}$$

Donde N_a es la variable normal reducida con probabilidad de excedencia a de acuerdo a los valores que se indican en la tabla 3.2. El valor de a indica la proporción de casos en los que se espera que el ancho de la banda incluya en su interior a las situaciones reales que se presentarán en el futuro inmediato.

Probabilidad de No-excedencia, a	Valor N_a
0.95	1.960
0.90	1.645
0.80	1.288
0.70	1.036
0.60	0.842
0.50	0.674

Tabla 3.2. Variable normal reducida para bandas de diferente ancho.
Fuente: Elaboración propia basa en el artículo Modelo de Utilización onjunta en la Cuenca Andreu y Marco,1983

Este procedimiento se repite para obtener un segundo pronóstico con el modelo de la semana $(m-1)$ y la misma de caudales disponibles hasta el día del pronóstico.

3.8 Otros modelos

La modelación de una serie cronológica consiste en seleccionar un modelo que represente adecuadamente las características estadísticas de la serie.

En general, los modelos existentes para el análisis de las series de tiempo son bastante estándar, cuando se tiene que la distribución es normal y la estructura es lineal. Algunos ejemplos de estos tipos de modelos son AR (autoregresivos), FGN(modelos fraccionarios gaussianos), ARMA (autoregresivos de media móvil), BL (modelos de línea quebrada), modelos de ruido de dispar; modelos de procesos intermitentes; modelos de agregación y desagregación modelos markovianos mezclados, entre otros.

En cambio, si se trata de modelos no lineales y no normales, se acepta que es un área de suma dificultad en el modelado y que depende mucho de cada caso en particular. En especial, para las representaciones multivariadas no existen métodos de identificación ni estimación desarrollados.

En particular, en el caso de series no normales, se propone la utilización de modelos univariados gamma que incorporan dentro de la estructura del modelo la asimetría y correlación de las series. Estos modelos asumen una distribución marginal gamma y una estructura autoregresiva constante o periódica.

3.8.1 Modelo GOL y el Paradigma de la PDE

El modelo GOL (Gestión Óptima del Laja, 1974) se desarrolló como una aplicación computacional para estimar la demanda de carbón del sector eléctrico. En una primera etapa fue utilizado por Endesa (empresa energética española, electricidad, gas y energías renovables) para la planificación de la operación de sus centrales en la cuenca del Laja, para un horizonte de análisis de hasta 15 años con etapas trimestrales.

Este modelo considera la existencia de un único embalse, considerando en forma independiente para la modelación a las centrales de la misma cuenca (El Toro, Antuco, Abanico y Rucue), y el resto de las centrales hidroeléctricas existentes en el sistema se agrupan en una sola central. El GOL permite minimizar los costos térmicos de generación realizando una colocación óptima de los recursos hídricos del Laja a lo largo del horizonte de planificación, realizando un despacho uninodal, considerando la demanda como determinística.

Modelos de Planificación de Sistemas Hidrológicos

El modelo utiliza programación dinámica estocástica (PDE) para la resolución, donde para cada etapa del horizonte de tiempo y para cada escenario hidrológico, se determina la secuencia de decisiones óptimas de operación para cada estado posible del embalse. Estos estados corresponden a una discretización del rango continuo de estados admisibles.

El algoritmo comprende una fase de optimización y una de simulación. En la fase de optimización se procede en una recursión backward, partiendo desde la última etapa y avanzando en sentido inverso hacia la primera, donde el estado del embalse es conocido. Con esto se obtienen las funciones de costos futuros para los estados discretos en cada etapa, asociados a cada escenario hidrológico posible. En la fase de simulación, se establece una iteración forward en la que se obtienen las cotas óptimas del embalse para cada etapa, partiendo desde la etapa inicial con estado conocido, avanzando hacia la última.

En el GOL, la aleatoriedad hidrológica es tratada de la siguiente forma:

1. Se asume que existe independencia estadística entre los caudales afluentes de cada año, es decir, la ocurrencia de una hidrología durante un año no determina ni condiciona la hidrología del año siguiente. Al no ser posible prever lo que ocurrirá en los años siguientes, el valor estratégico asociado a una cota a fines de cada año hidrológico (abril a marzo) corresponde al valor esperado respecto a las hidrologías futuras posibles.
2. Se asume que para cada año del horizonte de planificación, la variable hidrológica toma un valor extraído directamente del registro estadístico de 40 años. Es decir, para un año determinado, la hidrología puede ser cualquiera de los 40 valores, los cuales se consideran equiprobables.
3. Para un año, la decisión de extracción del embalse supone conocida la hidrología del año n en un esquema azar-decisión, donde para cada hidrología anual posible, supuesta conocida, se calcula la decisión óptima. El valor estratégico a comienzos del año para cada cota, es calculado promediando los valores estratégicos obtenidos para todas las hidrologías posibles.
4. El año se ha dividido en subetapas trimestrales con el fin de presentar la estacionalidad de la demanda, los aportes hidrológicos y la disponibilidad de las centrales que varían a lo largo del año. Los aportes en cada trimestre se suponen perfectamente dependientes entre sí, es decir, la operación óptima a través de los cuatro trimestres se realiza como una programación dinámica determinística para cada hidrología.

En la fase de optimización, el modelo opera definiendo para cada año t un conjunto de n hidrologías $\{H_t^h\}_{h=1}^n$, donde n corresponde al número total de valores admisibles para la hidrología en el año t . Si se asume que ésta puede tomar

cualquier valor del registro histórico, tendríamos $n=40$. Para cada escenario H_t^h , se obtiene la decisión óptima de extracción para cada estado discreto S_t del embalse al comienzo del período t , resolviendo un problema determinístico dentro del año. Con esto, dada una hidrología, se calcula el valor estratégico del agua V_t de la etapa t , para cada estado discreto inicial S_t , es decir, se calcula

$V_t = V_t\left(\frac{S_t}{H_t^h}\right)$ Luego se calcula el valor estratégico esperado para cada estado inicial S_t como:

$$\bar{V}_t = \sum_{h=1}^{40} P(H_t^h) \cdot V_t(S_t/H_t^h) \quad (3.1)$$

Donde $P(H_t^h)$ es la probabilidad de ocurrencia del escenario H_t^h . Si se suponen escenarios equiprobables, extraídos directamente del histórico, se tiene que $P(H_t^h) = 1/40$. Toda la información respecto de la decisión óptima en cada etapa, para cada estado, y para cada hidrología es almacenada matricialmente para ser usada en la etapa de simulación.

La etapa de simulación está basada en el Método de Montecarlo realizando sorteo de escenarios Hidrológicos para todo el horizonte de planificación, suponiendo completa independencia estadística entre años. El número de escenarios sorteados suele ser del orden de 1000, el que es comparativamente muy bajo en relación con los 40^n escenarios posibles para un horizonte de n etapas. Luego, basándose en la información contenida en las matrices de decisión obtenidas en la fase de optimización, se determina la secuencia de decisiones óptima.

3.8.2 Modelo OMSIC

El modelo OMSIC (Operación Mensual del SIC (Sistema Interconectado Central)), se sustenta en las mismas bases conceptuales que el modelo GOL, basado en programación dinámica. Este modelo difiere del GOL básicamente en la definición de la etapa elemental (operando sobre etapas mensuales), en la extensión del horizonte de planificación (usualmente es de uno o dos años) y en el tratamiento que se le da a la incertidumbre hidrológica dentro del año. Durante los meses de invierno (abril-septiembre), al ser originadas fundamentalmente por lluvias, las hidrologías se consideran estadísticamente independientes entre meses, pudiendo presentarse de manera equiprobable, en un determinado mes, cualquiera de los escenarios observados históricamente para dicho mes en los 40 años del registro histórico. Por otro lado, para los meses de deshielo (octubre-marzo), se asume completa dependencia estadística entre etapas, resolviendo una optimización determinística para cada serie.

El algoritmo opera de la misma forma que en el GOL, con una fase de optimización en la que se determinan las matrices de decisiones óptimas, y una fase de simulación, en la que se realiza una simulación de montecarlo utilizando 1000 escenarios, los cuales son sorteados respetando los supuestos sobre la hidrología planteados en el párrafo anterior.

3.8.5 Cambio de Paradigma: De la PDE (Programación Dinámica Estocástica) a la PDDE(Programación Dinámica Dual Estocástica)

El esquema PDE, utilizado por GOL y OMSIC, ha sido utilizado por muchos años en la mayoría de los países con predominancia de generación hidráulica. Sin embargo, éste tiene una limitación muy importante, que se debe a la necesidad de enumerar todas las combinaciones posibles de las variables de estado (almacenamiento en los embalses y caudales en las etapas anteriores). Con esto, el esfuerzo computacional crece exponencialmente a medida que la discretización de los estados se hace más fina, y a medida que el número de embalses aumenta.

Esta es la llamada “Maldición de la Dimensionalidad”. Supongamos, por ejemplo, que los niveles de cada embalse y de cada caudal anterior han sido discretizados en 20 valores. El número de puntos a evaluar en el espacio de estados es por lo tanto $(20 \times 20)^I$, donde I es el número de embalses. Se tiene entonces. Figura 3.4:

Número de embalses	Número de combinaciones
1	$20^2 = 400$
2	$20^4 = 160 \text{ mil}$
3	$20^6 = 64 \text{ millones}$
4	$20^8 = 25 \text{ mil millones}$

Figura 3.4: Combinaciones a evaluar en un problema multiembalse.

Fuente: Elaboración propia basa en la tesis Modelo para Manejo de la Incertidumbre Hidrológica en la planificación de la operación del SIC Eugenio Sebastián Palacios Galindo, Chile 2004

El esquema de la PDDE (Programación Dinámica Dual Estocástica) se basa en la observación de que la Función de Costos Futuros (FCF) se puede representar como una función lineal por partes, es decir, para obtener la FCF completa, no es necesario interpolar los valores de la FCF obtenidos para un conjunto de estados discretos de los embalses, como ocurre con la PDE. Para entender el cambio de filosofía en la resolución, consideremos el problema bloque-etapa, el que en un esquema lineal simplificado puede escribirse como:

$$\begin{aligned}
 a_i(x_i) = \text{Min} \{ & C_i(x_{i+1}) + a_{i+1}(x_{i+1}) \} \\
 \text{s.a.} & \\
 A \cdot x_{i-1} \leq & h_i - E \cdot x_i \\
 x_{i+1} \geq & 0
 \end{aligned}
 \tag{3.2}$$

En la ecuación 3.2, $a_i(x_i)$ corresponde a la función de costos térmicos y de falla futuros, en función del estado de los embalses x_i al inicio de la etapa i . El término $a_{i+1}(x_{i+1})$ es la FCF (Función de Costos Futuros) vista desde la etapa $i+1$ (final de la etapa i) en adelante, en función del estado de los embalses en dicha etapa, y el término $C_i(x_{i+1})$ corresponde a los costos térmicos de la etapa.

Mediante un esquema de PDE clásica, el problema representado en la ecuación 2.2 se resuelve discretizando la variable de estado x_i en N valores. Con esto, se resuelven N problemas de minimización para obtener el valor de la FCF en los N puntos. Para obtener la FCF en otros puntos que no pertenecen a la discretización se utiliza interpolación. Esquemáticamente esto se muestra en la siguiente figura 3.5

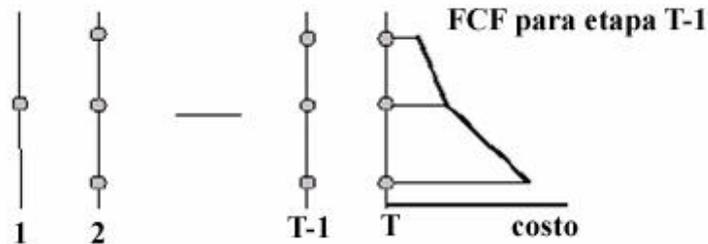


Figura 3.5 Cálculo de la FCF en la etapa T

Fuente: Elaboración propia basa en la tesis Modelación Hidrológica en el Sistema Interconectado Central, Eduardo A. Mihder Hetz, Chile, 1997

Como se mencionó antes, la PDDE utiliza una filosofía distinta para la resolución, aproximando la FCF como una función lineal por partes. Para ilustrarlo, se supone que la función $a_i(x_i)$, para el caso $x_i \in R$ (caso 1 embalse), tiene la forma mostrada en la figura 3.6.

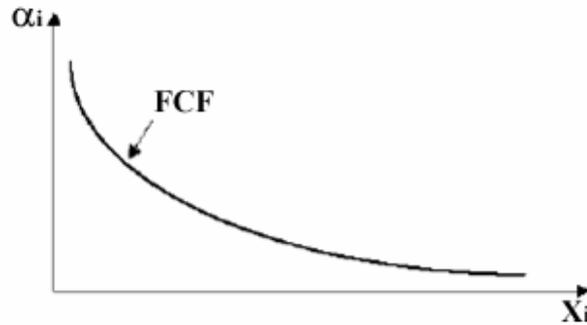


Figura. 3.6 Función de costos futuros en la etapa i

Fuente: Elaboración propia basa en la tesis Modelación Hidrológica en el Sistema Interconectado Central, Eduardo A. Mihder Hetz, Chile, 1997

Esta función es decreciente en X_i , en efecto, al aumentar el volumen X_i de agua almacenado en el embalse al comienzo de la etapa i, los costos totales de generación y falla vistos desde el inicio de la etapa i en adelante, disminuirán debido a una mayor disponibilidad de recursos hídricos para generación. La PDDE se basa en la observación de que la pendiente de la recta tangente en cada punto

X_i , $\frac{\partial a_i}{\partial X_i}(X_i)$ corresponde al multiplicador de Lagrange asociado a la restricción

de balance hidráulico de la ecuación 3.2. En efecto, la derivada en un punto de la curva está asociada a la sensibilidad de los costos futuros respecto de cambios infinitesimales en el estado X_i . Este concepto de "sensibilidad" es precisamente el significado que tienen los multiplicadores de Lagrange. Luego, si resolvemos la

ecuación 3.2 para un x_i^* fijo, obtendremos $a_i(x_i^*)$ y $I = \frac{\partial a_i}{\partial X_i}(x_i^*)$ Con esta

información podemos escribir una aproximación lineal de $a_i(x_i)$ en torno a x_i^* .
Figura 3.7

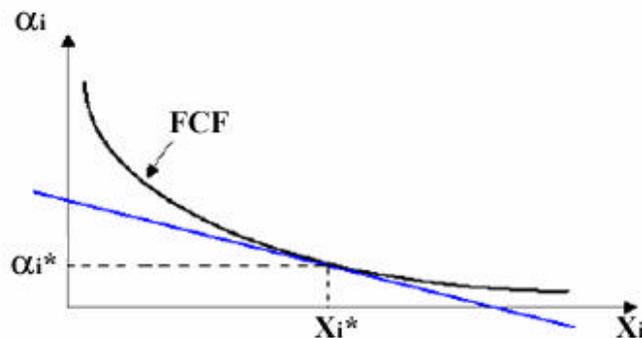


Figura. 3.7 Aproximación de primer orden de la FCF

Fuente: Elaboración propia basa en la tesis Modelación Hidrológica en el Sistema Interconectado Central, Eduardo A. Mihder Hetz, Chile, 1997

Si esto se realiza para N puntos, obtendremos una aproximación lineal por partes para la FCF (ver figura 3.8). El algoritmo mediante el cual se obtienen las aproximaciones tangentes se denomina Descomposición de Benders y las rectas reciben el nombre de planos minorantes o cortes de Benders. En el algoritmo de Benders los cortes se agregan de uno a uno por iteración, escogiendo convenientemente los X_i hasta alcanzar la convergencia. Este enfoque tiene la clara ventaja sobre la PDE clásica, de ofrecer en cada iteración una aproximación continua de la FCF (líneas rojas en la figura).

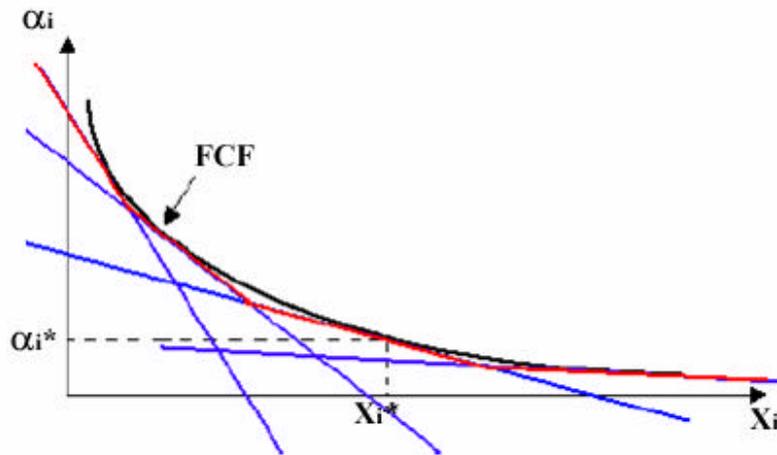


Figura. 3.8 Aproximación de la FCF como función lineal por partes.

Fuente: Elaboración propia basada en la tesis Modelación Hidrológica en el Sistema Interconectado Central, Eduardo A. Mihder Hetz, Chile, 1997

3.8.6 Modelo SDDP

El modelo SDDP (Stochastic Dual Dynamic Programming), está basado en programación dinámica dual estocástica, con búsqueda del óptimo mediante programación lineal.

Este modelo entrega una representación continua de la FCF en cada etapa, para el problema multinodal y multiembalse, haciendo una representación lineal de las pérdidas en el sistema de transmisión. Su horizonte de planificación es variable y suele ser de 10 años, más dos años de relleno. Las etapas de optimización pueden ser semanales, mensuales o trimestrales en todo el horizonte, no acepta combinaciones de ellas.

Para el tratamiento de la hidrología, el SDDP utiliza un modelo CPAR (Contemporaneous Periodic Auto Regressive), el que internaliza tanto la correlación temporal como espacial de los datos (se entiende como la dependencia

Modelos de Planificación de Sistemas Hidrológicos

estadística entre afluentes de distinta ubicación geográfica). Para el caso de la variable aleatoria hidrológica, un CPAR(p) (CPAR de orden p) representa el valor futuro de la hidrología como una combinación lineal de los p valores pasados de la serie y de los valores presentes (contemporáneos) de la serie ocurridos en otros lugares geográficos (correlación espacial). Al modelo, se le suma una variable aleatoria de distribución LogNormal e_t , que internaliza la componente estocástica del modelo.

El modelo CPAR(p) se clasifica dentro de los modelos de series de tiempo y posee la ventaja de capturar la tendencia estadística local de los datos. Éste se ajusta mediante métodos estadísticos, a partir de un registro histórico de valores medidos para cada uno de los afluentes del sistema durante un período determinado de años.

Debido a su estructura, el modelo CPAR(p) requiere conocer el valor de la serie en p instantes pasados para comenzar la generación de series sintéticas de caudales (una serie sintética corresponde a una secuencia de caudales ficticios, simulados por el modelo, para todo el horizonte de planificación). Dados estos valores, el SDDP sortea un conjunto de L secuencias de caudales iniciales de duración T (horizonte de estudio), $\{H_t^l\}_{t=1}^T$, $l=1, \dots, L$, que normalmente son 40, y que serán usadas en la recursión backward del SDDP. Una secuencia inicial de caudales debe entenderse como una trayectoria completa de la variable hidrológica a lo largo de todo el horizonte de planificación de largo T .

Estos L escenarios se obtienen mediante sorteo de la componente estocástica e_t del modelo CPAR, en cada etapa. Esquemáticamente esto se aprecia en la figura 2.9, donde las líneas azules corresponden a los L escenarios iniciales. Alternativamente, el SDDP permite obtener los escenarios iniciales directamente como sorteos del registro histórico, asumiendo independencia estadística entre años.

Una vez sorteados los L escenarios, basado en el método de Montecarlo, se determinan N aperturas condicionadas para cada nodo1 (figura 3.9), mediante sorteos de la componente estocástica del modelo CPAR. Para calcular la función de costos futuros en cada etapa y para cada escenario hidrológico, el SDDP resuelve el problema de optimización dado por la ecuación 3.2. para cada apertura. El valor esperado para la FCF en cada etapa, es calculado para cada serie inicial como un promedio de las funciones (en estricto rigor, de los multiplicadores de Lagrange que definen los cortes de Benders) de costo futuro entregadas por cada apertura.

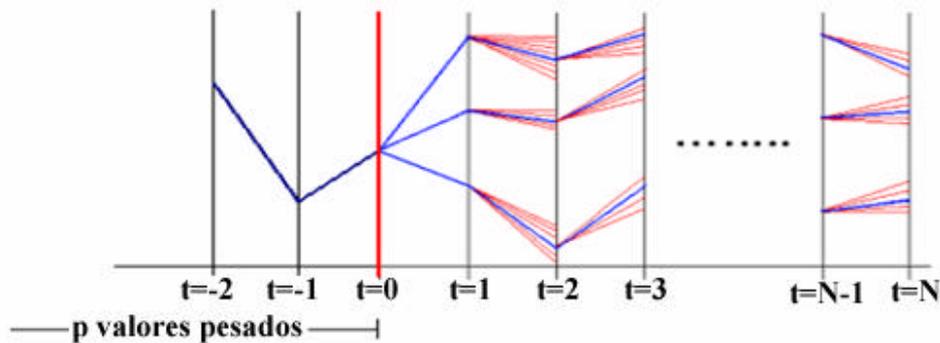


Figura 3.9 Árbol de escenarios para la variable hidrológica

Fuente: Elaboración propia basa en la tesis Modelación Hidrológica en el Sistema Interconectado Central, Eduardo A. Mihder Hetz, Chile, 1997

3.8.7 Modelo PLP (Programación de Largo Plazo)

Al igual que el SDDP, el modelo PLP utiliza programación dinámica dual estocástica en la resolución. En relación a la modelación de la incertidumbre hidrológica, éste ofrece gran flexibilidad para que la forma en la cual se definen las simulaciones de escenarios hidrológicos pueda ser manejada completamente por el usuario. Es importante notar que el modelo PLP no posee un modelo propio para manejo de hidrológicas, por medio de estos archivos se le entrega al modelo toda la información relacionada con los escenarios hidrológicos y la generación de aperturas. Estos archivos son generados externamente y son un input para el modelo PLP.

Una limitación de PLP en este aspecto es que tanto las simulaciones como las aperturas deben ser extraídas del registro estadístico histórico, con lo que no es tan directa la aplicación de modelos de series de tiempo para las series de caudales en el modelo PLP.

La forma en que PLP representa la incertidumbre hidrológica futura, es estructuralmente la misma que posee el modelo SDDP. Este define un conjunto de simulaciones iniciales (líneas azules en la figura 3.10) y aperturas en cada etapa (líneas rojas). PLP difiere de SDDP en la forma como son seleccionadas las simulaciones y las aperturas. Mientras SDDP se basa en un modelo de series de tiempo (CPAR) para determinar los escenarios, PLP sorteando tanto las simulaciones como las aperturas, directamente del registro histórico, asumiendo equiprobabilidad. Primero se sortean un conjunto de N (por defecto $N=40$) simulaciones iniciales (líneas azules), donde cada simulación representa una trayectoria completa de la variable hidrológica a lo largo de todo el horizonte de planificación.

Luego, el año hidrológico es dividido en dos períodos, un periodo de invierno, caracterizado por caudales que provienen principalmente de precipitaciones, y un período de deshielos, donde los afluentes provienen principalmente de dos derretimientos cordilleranos. Durante el período de invierno, se sortean para cada etapa y para cada serie inicial, un conjunto de M (por defecto $M=15$) aperturas (líneas rojas). Durante el período deshielo, PLP realiza una optimización unihidrológica (no se definen aperturas) para cada simulación inicial. Esto se basa en el hecho de que, durante los deshielos, los afluentes poseen una baja componente estocástica, comportándose en primera aproximación como determinísticos.

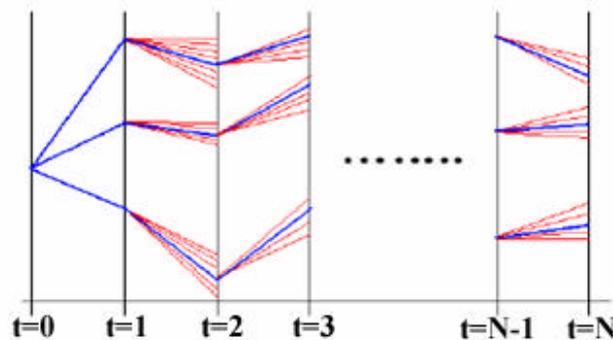


Figura 3.10 Representación de la hidrología futura implementada por PLP

Fuente: Elaboración propia basa en la tesis Modelación Hidrológica en el Sistema Interconectado Central, Eduardo A. Mihder Hetz, Chile, 1997

Las M aperturas para cada etapa durante el periodo de invierno, son sorteadas directamente desde el histórico, asumiendo a todos los escenarios como equiprobables. Por ejemplo, en el contexto de una planificación mensual, para determinar las aperturas del mes de julio, se sortean $M=15$ escenarios directamente de entre los afluentes observados para el mes de julio durante los 40 años del registro histórico, donde a cada escenario se le asocia la misma probabilidad de $1/40$.

Actualmente, las interfases que construyen los archivos de caudales de entrada a PLP, son modificados para tener un tratamiento distinto de la hidrología en las primeras semanas de la planificación de la operación. Este tratamiento, se conoce como incertidumbre reducida, define una hidrología fija para la primera semana del horizonte de planificación. Esta hidrología determinística se calcula a partir de los caudales afluentes de los 7 últimos días como:

$$Q_{1a.etapa} = [A \cdot Q_1 + (1 - A) \cdot Q_{2-7}]$$

Donde:

Q_1 : Para centrales de embalse, corresponde al caudal afluente del día anterior a la fecha en que se inicia el cálculo de la programación semanal.

Para centrales de pasada, corresponde al caudal afluente utilizable del día anterior a la fecha en que se inicia el cálculo de la programación semanal

Q_2 : Caudal afluente utilizable promedio de los días 2 al 7, tomados en sentido inverso al cronológico, considerando como día 1 al anterior al que se inicia el cálculo de la programación semanal.

A: Factor de ponderación igual a 0.5.

Para determinar los caudales de las semanas dos a cuatro, se calculan las correlaciones y las distribuciones de frecuencia, entre la primera etapa y cada una de las tres restantes, a partir de la estadística semanal de caudales históricos. Con esta información, y a partir del caudal Q_{1a} etapa calculado para la primera semana, se estiman las distribuciones condicionales de frecuencia acumulada (lognormal bivariada) para cada etapa. Con esta distribución se sortean n caudales afluentes por central para cada semana entre la 2 y la 4. Se reemplaza luego la estadística histórica en las semanas 2 a 4 por estas n muestras, a partir de las cuales PLP define tanto las simulaciones como las aperturas en dichas etapas.

3.9 Limitaciones de los Modelos

En un tratamiento tipo GOL, se sortean caudales anuales directamente desde el registro histórico, asumiendo independencia estadística entre años, pero completa dependencia dentro del año. Un tratamiento de este tipo asume implícitamente que, si en los primeros meses del año se presenta un determinado escenario hidrológico, entonces el año mantendrá la tendencia de esa serie en forma determinística.

Como contraparte, el modelo PLP sortea series iniciales, define aperturas en los meses de invierno y resuelve un problema determinístico, para cada serie, en el período de deshielo. Este tratamiento es muy similar al hecho por OMSIC.

En el invierno, las aperturas en PLP se sortean asumiendo completa independencia entre meses.

Por otro lado, la elección de una optimización uni-hidrológica durante el verano se basa en el comportamiento esencialmente determinístico que muestran los afluentes, que provienen básicamente de los deshielos.

Modelos de Planificación de Sistemas Hidrológicos

Se ve que, durante el invierno, ambos modelos realizan un tratamiento de la hidrología completamente contrapuesto, mientras el GOL asume correlación completa entre meses, PLP asume completa independencia (correlación nula). Lo cierto, es que ni lo uno ni lo otro es del todo correcto y va a depender, entre otras cosas, del régimen de la central. De este modo, las centrales de régimen exclusivamente pluvial mostrarán más independencia de sus afluentes, entre los meses de invierno, que aquellas de régimen pluvial, donde los efectos de la componente de deshielos presentan más correlación entre meses. En vista de esto, es necesario inspeccionar más a fondo el grado de validez de las hipótesis hechas por los modelos, respecto del tratamiento de la hidrología.

En la figura 3.11 se muestran los correlogramas (autocorrelaciones) para los 6 primeros meses del año hidrológico, para el afluente de lago Laja (afluente central El Toro). Estos gráficos muestran el grado de correlación que presentan los afluentes en cada mes, con respecto a los caudales de los 12 meses previos. Así, el primer gráfico muestra la correlación entre los afluentes observados en abril (mes 1) con los últimos 12 meses del año hidrológico anterior. El correlograma para el mes 3, muestra la correlación que presenta el mes tres, con los meses 2 y 1 del mismo año, y los 10 últimos meses del año anterior. El resto de los gráficos se interpretan de la misma forma.

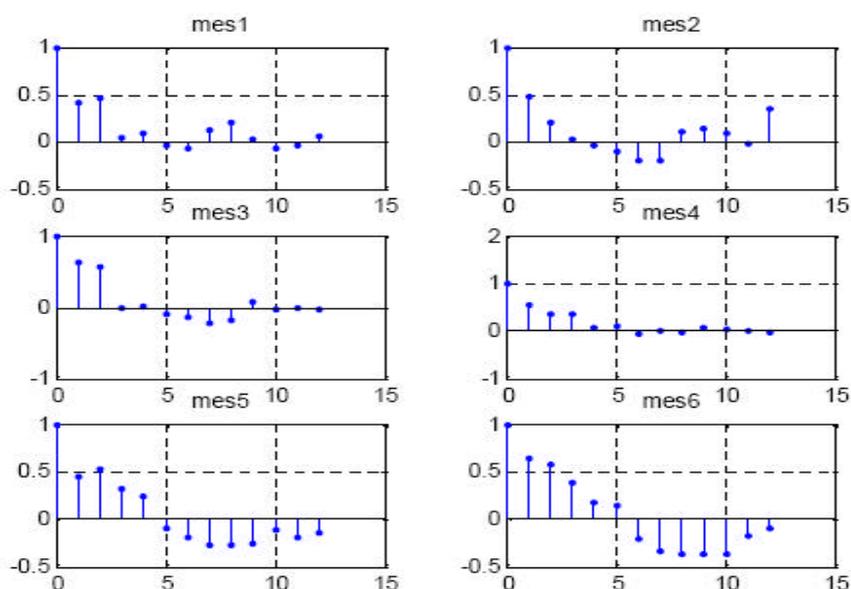


Fig. 3.11 Correlogramas para los meses de invierno, Afluente LAJA

Fuente: Elaboración propia basada en la tesis Modelo para Manejo de la Incertidumbre Hidrológica en la planificación de la operación del SIC Eugenio Sebastián Palacios Galindo, Chile 2004

Modelos de Planificación de Sistemas Hidrológicos

En la figura 3.12 se presentan los mismos gráficos pero para el afluente Capullo. Notemos que hay diferencias importantes en la información mostrada en ambos gráficos, mientras el afluente Laja muestra en casi todos los meses de invierno una fuerte correlación con los meses pasados, el afluente Capullo muestra, para el mismo período, una correlación más débil, observándose una correlación más marcada sólo en algunos meses. Estas diferencias podrían deberse a que los regímenes de las centrales son completamente distintos. Mientras Capullo es básicamente pluvial, Laja es de régimen nivo-pluvial, lo que implica que los efectos de los caudales se propagan a los meses posteriores, mostrando incluso una correlación importante al comienzo del año (mes 1) con los últimos meses del año hidrológico anterior. Con esto, en el contexto de una modelación mensual de los caudales, el suponer algún grado de dependencia entre años para los afluentes nivo-pluviales, no es completamente descartable en principio.

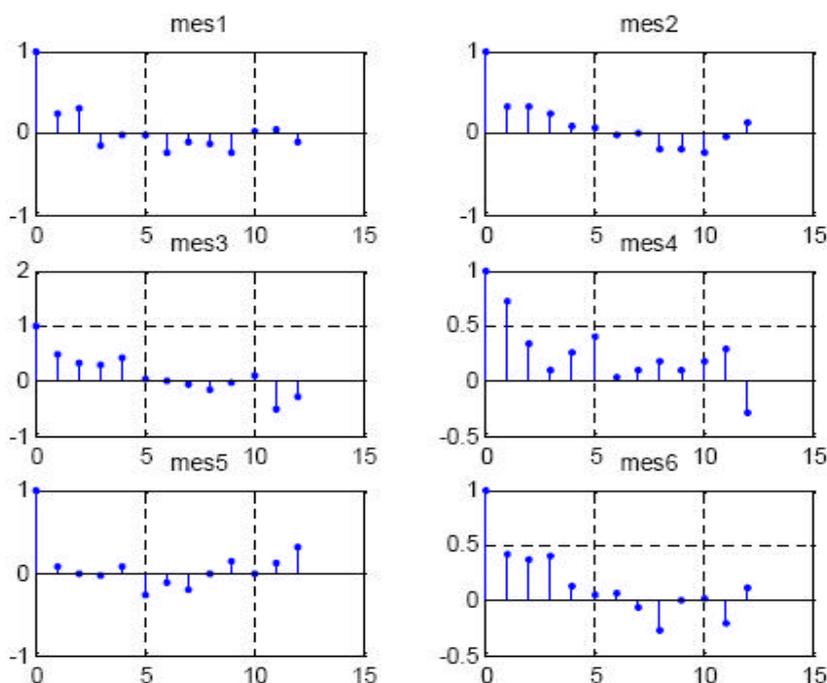


Figura. 3.12 Correlogramas para los meses de invierno. Afluente Capullo

Fuente: Elaboración propia basada en la tesis 'Modelo para Manejo de la Incertidumbre Hidrológica en la planificación de la operación del SIC' de Eugenio Sebastián Palacios Galindo, Chile 2004

Adicionalmente es necesario señalar, que existe poca consecuencia en la forma en la que PLP modela la hidrología en invierno y en verano. En el invierno sortea aperturas desde el histórico, en forma equiprobable. El sorteo de una apertura húmeda en el invierno, lleva implícito un almacenamiento mayor de nieve en la cordillera, lo que debería verse reflejado en un aumento de los caudales en la temporada de deshielos. Lo opuesto ocurre para una apertura seca. Dicho de otro

modo, la forma en la que se definen las aperturas en el invierno, es determinante en el valor esperado de los caudales durante el deshielo. El tratamiento uni-hidrológico, realizado por PLP en el período de deshielo, claramente no considera este efecto.

3.10 Generación de Escenarios para la Hidrología Futura

Para generar escenarios se adoptará una estructura de representación para la hidrología similar a la hecha por SDDP. Una representación de este tipo posee dos componentes:

1. Un conjunto de simulaciones iniciales, las que corresponden a una trayectoria completa de la variable hidrológica a lo largo de todo el horizonte de planificación
2. Aperturas en cada etapa (tanto en el invierno como en el deshielo), las que representan un conjunto de realizaciones alternativas posibles de la hidrología en cada etapa (líneas rojas en figura 3.10).

La diferencia fundamental entre el modelo propuesto, y el tratamiento realizado por SDDP, es básicamente la forma en que son seleccionadas tanto las simulaciones iniciales como las aperturas. Es necesario definir un criterio para seleccionar las simulaciones iniciales y las aperturas (tanto el criterio para seleccionarlas, como el número necesario de ellas).

Esto se realiza asignando probabilidades de ocurrencia a los escenarios históricos, analizando la tendencia estadística local de éstos con la ayuda de un modelo PARMA (es una extensión de los modelos ARMA (Proceso Autoregresivo de medias Móviles)). El esquema general de la propuesta se muestra en la figura 3.13.



Figura. 3.13 Esquema general del modelo propuesto

Fuente: Artículo, Desarrollo de Programas para el Tratamiento de Series Hidrológicas y Evaluación, España 2005

4. Modelado de Series de Tiempo Hidrológicas

4.1 Introducción

En las series de datos provenientes de procesos geofísicos, se establecen, en general, dos tipos de dependencia estadística, una dependencia temporal, que tiene relación con la dependencia estadística que muestran los valores que toma la serie en dos instantes de tiempo distintos, y la dependencia espacial, que guarda relación con la correlación que muestran las series de datos obtenidas en distintos lugares geográficos.

Los modelos de series de tiempo han cobrado fuerza en las últimas décadas en aplicaciones de series hidrológicas. La filosofía de estos modelos estadísticos se basa en el hecho de que la gran mayoría de las series temporales, que derivan de procesos del ámbito de la física y la economía, muestran una fuerte correlación con sus valores en instantes pasados. La metodología propuesta por Box y Jenkins en la década del setenta, para el análisis y modelación de series de tiempo, se convirtió en una de las herramientas más difundidas para el análisis de series

cronológicas, cuando se cuenta con un número grande de muestras. Básicamente, estos modelos permiten hacer estimaciones de la serie, expresando el valor futuro como una combinación lineal de los valores que tomó la serie en instantes precedentes.

El auge de estos modelos se ha traducido en un gran número de trabajos que utilizan esta estructura de modelación en el pronóstico de series hidrológicas. Se han desarrollado con éxito diversas aplicaciones para pronóstico de crecidas en ríos y en el pronóstico de lluvias.

Los principales problemas de estos modelos derivan del hecho de que los registros estadísticos de caudales afluentes suelen ser de mala calidad debido a la existencia de períodos de relleno¹, los que sólo reproducen correctamente las estadísticas de primer orden de la serie original. Otro aspecto a considerar es que estos modelos, al ser modelos estadísticos, serán tanto mejores mientras más grande sea el registro estadístico disponible para ajustar sus parámetros.

4.2 Modelos Autoregresivos

Existen modelos que derivan de la metodología propuesta por Box y Jenkins, conocidos como modelos autoregresivos y se evalúan las características de cada uno con el objeto de decidir cuál de todos ellos es el más apto para su aplicación a series hidrológicas de caudales.

En un modelo de series de tiempo autoregresivo de orden p , el valor real observado de una serie X_t en el instante t es expresado como una combinación lineal de los p valores previos del proceso, más un ruido blanco, conocido como innovación y que estructura al modelo como un modelo estocástico. Dicho de otro modo, un modelo autoregresivo consta de dos componentes, una determinística, construida como una combinación lineal de los valores pasados recientes de la serie, y una estocástica, determinada por la innovación. Denotemos a los valores que toma el proceso, a intervalos de tiempo equiespaciado, $t, t-1, \dots, t-p$, por $X_t, X_{t-1}, X_{t-2}, \dots, X_{t-p}$. Sean también $\bar{X}_t, \bar{X}_{t-1}, \bar{X}_{t-2}, \dots, \bar{X}_{t-p}$ las desviaciones de dichos valores con respecto a la media m , $\bar{X}_t = X_t - m$. Con esto escribimos:

$$\bar{X}_t = f_1 \cdot \bar{X}_{t-1} + f_2 \cdot \bar{X}_{t-2} + \dots + f_p \cdot \bar{X}_{t-p} + e_t \quad (4.1)$$

Si una serie temporal sigue una relación como (4.1), se dice que es un proceso Autoregresivo de orden p (AR(p)). El factor e_t es el ruido o innovación, y el modelo

¹ Es el período que se utiliza para interpolar los datos faltantes en el registro estadístico.

se construye de manera que sea un ruido blanco, es decir, se impone que sea no correlacionado en el tiempo y que se distribuya en forma normal $N(0, \sigma_e^2)$, en cada instante. Es importante recalcar que el modelo se ajusta de manera que estas condiciones sobre la innovación se satisfagan.

Otro aspecto importante, es que los modelos de series de tiempo se ajustan en general a series con valor esperado nulo. Si la serie a modelar no posee esta característica, existen dos caminos, el primero es agregar una constante aditiva a la ecuación 41, la que debe ser estimada junto con los otros parámetros del modelo, y el segundo es ajustar el modelo a la serie de datos centrados (restando el valor esperado). Como en general no se dispone a priori del valor esperado de la serie, este se puede estimar simplemente como el promedio histórico de ésta.

4.3 Modelos Autoregresivos de Medias Móviles

En un proceso Autoregresivo de Medias Móviles de orden (p,q) , ARMA (p,q) , el valor observado de la serie \bar{X}_t (serie centrada) en el instante t , no solo se expresa como una combinación lineal de los p valores pasados de la serie, si no que además de los q ruidos o innovaciones pasados de ésta, más la innovación e_t en el instante t . Así, la formulación matemática de este tipo de procesos es:

$$\bar{X}_t = \sum_{i=1}^p f_i^* \bar{X}_{t-i} + e_t - \sum_{i=1}^q q_i^* e_{t-i} \quad (4.2)$$

El signo menos delante de la componente de medias móviles es sólo cuestión de convención, pues los coeficientes q_i pueden tomar valores negativos.

4.3.1 Propiedades de los Procesos ARMA (p,q)

Recordemos que una de las principales características de las series de tiempo hidrológicas es que son no estacionarias y muestran periodicidad de sus estadísticas dentro del año. Esta característica invalida a los modelos ARMA para ser aplicados a series de caudales, debido a su estructura estática en el tiempo. En efecto, los parámetros del modelo tienen un valor fijo invariante en el tiempo, lo que se traduce en que sus propiedades estadísticas sean las mismas en todo instante. Debido a esto, una serie que sigue una relación como (4.2) es estacionaria.

Modelado de Series de Tiempo Hidrológicas

Por otro lado, resulta claro que para las series hidrológicas, el tipo de dependencia estadística que presentan los caudales con respecto a sus valores en los meses previos, dependerá fuertemente de la época del año en la que dicho mes se encuentre. Así por ejemplo, en el período de verano los afluentes son principalmente pluviales, con una baja correlación entre meses, en cambio, en primavera, los afluentes corresponden principalmente a los deshielos en la alta sierra, los que muestran una fuerte correlación entre meses y con respecto a las lluvias de verano. Esta dinámica invalida a los modelos ARMA para su aplicación en series hidrológicas debido a que establece una estructura de dependencia que no cambia a lo largo del año.

Existe un gran número de modelos dentro de la familia de los propuestos por Box y Jenkins que intentan superar las limitaciones de los modelos ARMA. Se distinguen los modelos ARIMA (Autoregressive Integrated Moving Average) diseñados para la modelación de series no estacionarias, los SARMA (Stational ARMA) para la aplicación en series que poseen una componente estacional, y los modelos mixtos SARIMA.

En los modelos SARIMA, la no estacionalidad de la serie es “borrada”, haciendo una diferenciación de la serie original, obteniéndose la serie diferenciada $W_t = \Delta X_t = X_t - X_{t-1}$ ². Luego, para la serie diferenciada, se ajusta un modelo SARMA. En los SARMA, la dependencia estacional se incluye escribiendo el valor X_t no sólo como una combinación lineal de lo ocurrido en las etapas previas, si no que además, se agrega una combinación lineal de los valores que tomó la serie en igual etapa del año anterior. De esta manera, un modelo SARMA tiene la siguiente forma funcional:

$$X_t = \sum_{i=1}^p f_i \cdot X_{t-i} + e_t - \sum_{i=1}^q q_i \cdot e_{t-i} + \sum_{i=1}^{\Pi} s_i \cdot X_{t-iT} \quad (4.3)$$

donde T corresponde al período de la serie ($T=12$ para etapas mensuales, $T=48$ en etapas semanales) y Π corresponde al número de años en el pasado con el cual la serie ésta correlacionada. Por ejemplo, mediante una expresión como (4.3) se puede pronosticar el caudal en el mes de julio, como una combinación lineal de lo que ocurrió en los meses previos (junio, mayo, abril, etc.), más una combinación lineal de los valores observados en la serie en el mes de julio en Π años anteriores.

² Si esta serie no resulta estacionaria debe realizarse una segunda diferenciación y así sucesivamente.

Los modelos SARIMA se sustentan en la filosofía de que el patrón de estacionalidad de la serie radica en una correlación entre los valores que la serie toma año tras año en cada período (mes, semana) del año.

Los modelos SARMA presentan básicamente las mismas limitaciones de los modelos ARMA clásicos, además de otros inconvenientes, para ser aplicados en series hidrológicas. Estos modelos, si bien permiten modelar estacionalidad, siguen estando limitados por su estructura estática en el tiempo y por lo tanto son incapaces de modelar la característica más distintiva de las series hidrológicas, a saber, la periodicidad de sus funciones estadísticas. Por otro lado, si se asume como hipótesis que para la serie existe completa independencia entre años, entonces los SARMA pierden toda validez en la modelación de series hidrológicas.

4.4 Modelo para Series de Tiempo Periódicas: El modelo PARMA

Como se ha señalado en las secciones precedentes, la característica más distintiva de las series hidrológicas, y en general de los procesos geofísicos, es que sus estadísticas tienen una periodicidad anual. Los procesos con estas características son conocidos como ciclo-estacionarios. Existe un tipo adicional de modelos de series de tiempo que permite abordar series de tiempo con dichas características, evitando las dificultades propias de los SARIMA. Estos modelos, denominados PARMA (Periodic ARMA) se basan en la idea de que las características periódicas de las series ciclo-estacionarias, pueden ser modeladas permitiendo que los coeficientes del modelo ARMA presentado en la ecuación (3.2) cambien también en forma periódica.

En un modelo ARMA clásico como el presentado en la ecuación (3.2) los coeficientes f_i, q_i permanecen fijos en el tiempo. En un modelo PARMA, se relaja esta condición permitiendo que estos valores varíen en el tiempo en forma periódica. Un modelo PARMA es básicamente una extensión de los modelos ARMA que permite a los parámetros depender de cada mes.

Sea $\{X_t\}$ una serie de tiempo ciclo-estacionaria que posee S estaciones³ por año (en el contexto de una modelación mensual, $S=12$). Supongamos que la serie posee un valor esperado nulo en todos los meses. Si esta condición no se satisface, basta con restar a la serie original su valor esperado en cada mes, obteniéndose la serie centrada, la que por definición posee media nula⁴.

³ Una estación corresponde a la fracción del año considerada como etapa de duración elemental. Por ejemplo un mes o una semana.

⁴ Notar que la media de la serie varía en forma periódica, mes a mes, con un período de 12 meses.

Dicho de otro modo, se define:

$$X_{nS+v} = X_{nS+v} - \mathbf{m}_v \quad v = 1, \dots, S \quad n = i, \dots$$

donde \mathbf{m}_v es el valor esperado de la serie en la estación (mes) v . El valor esperado en cada mes v , puede ser estimado como el promedio de los valores observados en el mes v durante los 40 años hidrológicos. La notación periódica $\{X_{nS+v}\}$ se refiere al valor que toma la serie $\{X_t\}$ durante el v -ésimo mes del año n . Se introduce esta notación de la variable temporal, para dar cuenta de la periodicidad anual de la serie.

Con esto, se dice que la serie de media nula, sigue un modelo PARMA, si para cada mes se puede representar con la siguiente ecuación de recurrencia:

$$X_{nS+v} = \sum_{i=1}^{p(v)} \mathbf{f}_i(v) \cdot X_{nS+v-i} - \sum_{i=1}^{q(v)} \mathbf{q}_i(v) \cdot \mathbf{e}_{nS+v-i} + \mathbf{e}_{nS+v} \quad \forall v = 1, 2, \dots, S \quad (4.4)$$

El modelo PARMA aquí presentado posee ventajas comparativas importantes para su aplicación en series de tiempo hidrológicas. Notemos que su estructura no es estática en el tiempo como ocurre con los modelos ARMA clásicos. Primero, el número de períodos pasados que se incluyen en la modelación (límite superior de las sumatorias de (4.4) depende de la estación (semana o mes) del año, lo que se denota explícitamente escribiendo $p(v), q(v)$. Segundo, el valor de los coeficientes de la combinación lineal cambian con la estación $(\mathbf{f}_i(v), \mathbf{q}_i(v))$. Dado que la estructura del modelo cambia etapa a etapa en forma periódica (periodo de duración igual a un año), los modelos PARMA son los más adecuados dentro de los modelos de series de tiempo, para reproducir las características periódicas de las series de tiempo hidrológicas.

En un modelo PARMA, la innovación aleatoria del modelo \mathbf{e}_{nS+v} también posee propiedades estadísticas que cambian con la estación. En particular su varianza $\mathbf{s}_e^2(v)$ es función explícita de la estación v

4.4.1 Propiedades de los Procesos PARMA

Como se señaló, los modelos del tipo PARMA son los más adecuados para la modelación de procesos ciclo-estacionarios. Este tipo de modelos permite interpretar el comportamiento futuro de una serie en función de su pasado reciente, con una estructura de dependencia que cambia en el tiempo. En particular nos interesa conocer el valor esperado y la desviación estándar de la distribución arrojada por el modelo, en cada instante de tiempo $nS + v$, en función

Modelado de Series de Tiempo Hidrológicas

de los valores que tomo la serie en los meses previos, es decir, queremos calcular los momentos de primer y segundo orden condicionados:

$$\mathbf{m}_{nS+v} = E(X_{nS+v} \mid X_{nS+v-1}, X_{nS+v-2}, \dots, X_{nS+v-p(v)}, \mathbf{e}_{nS+v-1}, \mathbf{e}_{nS+v-2}, \dots, \mathbf{e}_{nS+v-q(v)})$$

De la ecuación (3.4) es directo que:

$$\mathbf{m}_{nS+v} = \sum_{i=1}^{p(v)} \mathbf{f}_i(v) \cdot X_{nS+v-i} - \sum_{i=1}^{q(v)} \mathbf{q}_i(v) \cdot \mathbf{e}_{nS+v-i} + E(\mathbf{e}_{nS+v})$$

Pero por construcción del modelo, la innovación tiene valor esperado nulo a en cada instante, luego

$$\mu_{nS+v} = \sum_{i=1}^{p(v)} \phi_i(v) \cdot X_{nS+v-i} - \sum_{i=1}^{q(v)} \theta_i(v) \cdot \varepsilon_{nS+v-i} \quad (4.5)$$

Se nota que los términos de (4.5) son valores medidos en instantes anteriores, tanto de la serie como de la innovación, es decir, son componentes determinísticas del modelo. Ahora, para la varianza condicional, podemos escribir:

$$\mathbf{s}_{nS+v} = E((X_{nS+v} - \mathbf{m}_{nS+v})^2 \mid X_{nS+v-1}, X_{nS+v-2}, \dots, X_{nS+v-p(v)}, \mathbf{e}_{nS+v-1}, \mathbf{e}_{nS+v-2}, \dots, \mathbf{e}_{nS+v-q(v)})$$

De (4.4) podemos escribir:

$$\mathbf{e}_{nS+v} = X_{nS+v} - \sum_{i=1}^{p(v)} \mathbf{f}_i(v) \cdot X_{nS+v-i} - \sum_{i=1}^{q(v)} \mathbf{q}_i(v) \cdot \mathbf{e}_{nS+v-i} \Rightarrow X_{nS+v} - \mathbf{m}_{nS+v} = \mathbf{e}_{nS+v} \Rightarrow \mathbf{s}_{nS+v} = E(\mathbf{e}_{nS+v}^2) = \mathbf{s}_e^2(v)$$

Se ve entonces que si la serie sigue un modelo PARMA, entonces se distribuye como una normal de media \mathbf{m}_{nS+v} y una varianza igual a la de la innovación.

Para el análisis, en ocasiones será conveniente escribir el modelo PARMA con la notación simplificada:

$$X_{nS+v} = \mathbf{m}_{nS+v} + \mathbf{e}_{nS+v} \quad (4.6)$$

que corresponde a la suma de un valor medio que depende del pasado, más una innovación aleatoria.

Una vez ajustado un modelo como 4.4 a la serie de datos, es necesario obtener la serie de tiempo asociada a la innovación. Esta serie es, entre otras cosas, de gran utilidad para chequear las bondades del modelo ajustado. Esta serie se construye despejando la innovación de 4.4:

$$\mathbf{e}_{nS+v} = X_{nS+v} - \sum_{i=1}^{p(v)} \mathbf{f}_i(v) \cdot X_{nS+v-i} - \sum_{i=1}^{q(v)} \mathbf{q}_i(v) \cdot \mathbf{e}_{nS+v-i}$$

(4.7)

es decir, una muestra para la innovación en el instante $nS + v$, se obtiene restando el valor real medido X_{nS+v} con la estimación hecha por el modelo. Luego, se puede obtener la serie de tiempo de las innovaciones, aplicando esta expresión a la serie de datos contenida en el histórico.

4. 5. Modelado de la Dependencia Espacial

Las series hidrológicas, como la mayoría de las series extraídas de procesos geofísicos, muestran una fuerte dependencia espacial de los datos. Esto significa que el valor que adopta la serie en un determinado lugar geográfico esta correlacionado con el valor de la serie en otros sitios. Esta dependencia espacial de los valores de una misma serie, motiva la implementación de modelos multivariados.

En los modelos multivariados se consideran básicamente tres tipos de dependencia entre las distintas series:

1. Contemporánea: Dos o más series se dice que están contemporáneamente correlacionadas si solo sus valores presentes están correlacionados unos con otros. Esto implica en la modelación que cualquier interacción entre las variables es instantánea. Para series hidrológicas, la principal fuente de dependencia en este caso es la precipitación regional.
2. Unidireccional: Dos series presentan una dependencia causal unidireccional, si los valores pasados y presente de una de ellas condiciona la otra. Por ejemplo, existe una relación unidireccional entre los afluentes medidos en dos puntos a lo largo de un río, en el cual las mediciones hechas en el lugar que se encuentra más aguas arriba determinan causalmente las mediciones aguas abajo, pero no a la inversa.
3. Retroalimentada: Se dice que existe una relación de dependencia espacial retroalimentada si la causalidad definida en el punto anterior existe en ambas direcciones. En el campo de los recursos hídricos, relaciones de este tipo se pueden encontrar en el fenómeno de evapotranspiración.

Modelado de Series de Tiempo Hidrológicas

4. Mixta: Tres o más series poseen relaciones de dependencia mixta si se presentan entre pares ellas dos o más de las relaciones de dependencia mostradas antes.

Existen dos claras desventajas para el uso del modelo multivariado general, en su aplicación para la modelación de la dependencia espacial de las series hidrológicas:

1 En la práctica resulta complicado de implementar debido a que el número de parámetros involucrados crece exponencialmente con la dimensionalidad del modelo.

2 Se omite una importante característica de las series hidrológicas, y es que las propiedades físicas del sistema imponen restricciones en la estructura del modelo las que no son consideradas en la construcción del modelo multivariado.

Considérese por ejemplo el caso Chileno, en las cuencas no existe una dependencia retroalimentada de los caudales, debido a que los afluentes aguas abajo no generan efectos causales sobre los afluentes aguas arriba. Tampoco es clara la existencia de una dependencia unidireccional. En efecto, debido al tipo de características que se presentan en la cordillera chilena, la que es esencialmente superficial y de una gran pendiente, efectos de cambios en las afluentes aguas arriba se propagan rápidamente aguas abajo, de manera que en el contexto de una modelación semanal, la propagación de este efecto se puede considerar como instantánea.

Lo anterior permite argumentar que un modelo del tipo contemporáneo es suficiente para la modelación de la dependencia espacial en el contexto del presente trabajo, en el que se propone la implementación de un modelo CPARMA (Contemporaneous PARMA). En un modelo CPARMA se modelan las series temporales de cada afluente en forma separada, ajustando en cada caso un modelo PARMA como el descrito en la sección precedente. Luego de esto, se ajusta un modelo de dependencia contemporánea para la serie de innovaciones. La metodología para implementar un modelo contemporáneo indica que primero se debe ajustar un modelo PARMA univariado para cada uno de los afluentes a incluir en la modelación. Una vez que se tiene un modelo PARMA para cada afluente se debe generar la serie de tiempo de innovaciones en cada etapa, esto se realiza despejando e_t de 4.4:

$$e_{nS+v} = X_{nS+v} - \sum_{i=1}^{p(v)} f_i(v) \cdot X_{nS+v-i} - \sum_{i=1}^{q(v)} q_i(v) \cdot e_{nS+v-i}$$

Modelado de Series de Tiempo Hidrológicas

La serie de ruidos se construye, para cada afluente, a partir del modelo PARMA ajustado y a partir de los valores reales de la serie X_{nS+v} contenidos en el registro estadístico. La innovación en el instante $nS+v$ se calcula como la diferencia entre el valor real observado de la serie X_{nS+v} y el valor esperado que entrega el modelo PARMA. Para ilustrar esto, supongamos que se ajusto un modelo PARMA a la serie de datos mostrados en la tabla 4.1 ("Registro Histórico"), los que tienen una periodicidad anual, con cuatro estaciones dentro del año (con lo que el periodo de la serie equivaldría a cuatro trimestres). La tabla 4.2 muestra los parámetros de un modelo PARMA(1,0) ajustado a los datos.

t	X
1	0.7
2	0.25
3	0.14
4	0.15
5	0.5

Tabla 4.1

Estación	phi
1	1.96
2	0.45
3	0.5
4	0.98

Tabla 4.2

Partiendo desde el instante $t=2$, la innovación, para etapa se calcula como:

$$\begin{aligned}
 t=2 & \quad 0.25 - .45 \cdot 0.7 = -0.065 \\
 t=3 & \quad 0.14 - 0.5 \cdot 0.25 = 0.015 \\
 t=4 & \quad 0.15 - 0.98 \cdot 0.14 = 0.0128 \\
 t=5 & \quad 0.5 - 1.96 \cdot 0.15 = 0.4
 \end{aligned}$$

que corresponde a la serie de ruidos o innovaciones.

Luego de esto, se quiere encontrar un modelo que represente adecuadamente la correlación espacial que presentan las series de innovaciones calculadas para cada afluente del sistema. La intuición indica que el grado de correlación espacial que presentan las series de caudales, dependerá de la época del año. En efecto, las precipitaciones durante los meses de invierno, se distribuyen a nivel local (por ejemplo en una cuenca) de manera mas o menos uniforme. Por otro lado, en los meses de deshielo, los afluentes dependen principalmente de la temperatura ambiente, cuya distribución a nivel geográfico no es uniforme. Con lo anterior, se esperarí observar una correlación espacial alta en el la primera mitad del año y baja en la segunda. No puede descartarse en primera instancia, que esta propiedad sea heredada por la serie de ruidos.

Se desea proponer un modelo contemporáneo para las innovaciones, en el cual la variación dentro del año del grado de correlación a nivel geográfico de la serie de innovaciones no se considera, asumiendo que el tipo de dependencia espacial no

Modelado de Series de Tiempo Hidrológicas

cambia dentro del año, modelo que no es el adecuado para la representación de la correlación espacial.

Supongamos que se desea modelar la dependencia espacial de las series de innovaciones en K lugares geográficos distintos. Para esto, es conveniente definir el vector de ruidos para el instante $nS + v$ (mes v del año n), como el vector que contiene las innovaciones para dicho instante en los K lugares:

$$\bar{\mathbf{e}}_{nSv} = [\mathbf{e}_{nS+v}^1, \mathbf{e}_{nS+v}^2, \mathbf{e}_{nS+v}^3, \dots, \mathbf{e}_{nS+v}^k]$$

El comportamiento espacial de estos residuos se puede apreciar en la matriz de correlación espacial D, cuyos elementos son:

$$D_{ij}^{nS+v} = r_{ij} = \frac{E(\mathbf{E}_{nS+v}^i \cdot \mathbf{E}_{nS+v}^j)}{\mathbf{s}_{\mathbf{e}^i}^v \cdot \mathbf{s}_{\mathbf{e}^j}^v}$$

Donde $\mathbf{s}_{\mathbf{e}^i}^v$ corresponde a la varianza de la innovación en la i-ésima locación geográfica, en el mes v . Recordar que por construcción del modelo PARMA, las innovaciones tienen media nula. La matriz D es una matriz de KxK, cuyos elementos corresponden a las correlaciones cruzadas en el instante t de los K afluentes. Los elementos de D pueden aproximarse por sus valores muestrales:

$$D_{ij}^{nS+v} \approx r_{ij}^{nS+v} = \frac{1}{N} \sum_{n=1}^N \frac{\mathbf{e}_{nS+v}^i \cdot \mathbf{e}_{nS+v}^j}{\tilde{\mathbf{s}}_{\mathbf{e}^i}^v \cdot \tilde{\mathbf{s}}_{\mathbf{e}^j}^v}$$

Con:

$$\tilde{\mathbf{s}}_{\mathbf{e}^i}^v = \sqrt{\frac{1}{N} \sum_{n=1}^N (\mathbf{e}_{nS+v}^i)^2}$$

El modelo de dependencia espacial se construye imponiendo que las innovaciones para cada uno de los K afluentes no sean independientes entre ellas si no que interne de algún modo la información contenida en la matriz D. Esto se hace definiendo un nuevo vector auxiliar de K innovaciones $\bar{\mathbf{w}}_t \in R^K$, las cuales son no correlacionadas espacialmente. Las componentes del vector $\bar{\mathbf{w}}_t$ son independientes entre si y se distribuyen según una normal estándar N(0,1).

Por simplicidad, definamos el vector de ruidos reducidos como el vector en que cada componente se define por:

$$e_{nS+v}^i = \frac{\mathbf{e}_{nS+v}^i}{\mathbf{s}_v^i} \quad i = 1, \dots, K$$

Y establezcamos el modelo espacial contemporáneo para los ruidos como:

$$\bar{e}_{nS+v} = M \cdot \bar{w}_{nS+v}$$

dicho de otro modo, el vector de ruidos reducidos se escribe como el producto de un vector de ruidos no correlacionados espacialmente, y una cierta matriz M (por encontrar) que de algún modo guarda toda la información de la correlación espacial. De esta forma, se introduce en el vector de ruidos reducidos, el hecho de que sus componentes están correlacionadas espacialmente. Para encontrar M se escribe:

$$\begin{aligned} E\left(\bar{e}_{nS+v} \cdot \bar{e}_{nS+v}^T\right) &= E\left(\left(M \cdot \bar{w}_{nS+v}\right) \cdot \left(M \cdot \bar{w}_{nS+v}\right)^T\right) = E\left(\left(M \cdot \bar{w}_{nS+v}\right) \cdot \left(\bar{w}_{nS+v}^T \cdot M^T\right)\right) \\ &= M \cdot E\left(\bar{w}_{nS+v} \cdot \bar{w}_{nS+v}^T\right) \cdot M^T \end{aligned}$$

Notemos que por definición $E\left(\bar{e}_{nS+v} * \bar{e}_{nS+v}^T\right)$ Además tenemos que como los \bar{w}_t son independientes y normales estándar se cumple que $E\left(\bar{w}_{nS+v} * \bar{w}_{nS+v}^T\right) = I$, donde I es la matriz identidad. Con lo anterior se obtiene la relación:

$$D^{nS+v} = M \cdot M^T \tag{4.8}$$

De la ecuación (4.8) M puede ser obtenida mediante la descomposición de Cholesky de la matriz D^{nS+v} . Ésta descomposición permite separar una matriz definida positiva como el producto $R^T \cdot R$, donde R es una matriz triangular superior. Con esto, mediante descomposición de Cholesky podemos escribir:

$$D^{nS+v} = R^T \cdot R = M \cdot M^T \Rightarrow M = R^T \tag{4.9}$$

Se nota que dado que las propiedades estadísticas del vector de ruidos son periódicas (periodo igual a un año para la serie mensual), la matriz D^{nS+v} también será periódica. Con esto, podemos escribir $D^{nS+v} = D^v$. Del mismo modo, tendremos una matriz M para cada mes, con lo que en estricto rigor debemos escribir $M = M^v$, la que se obtiene con la ecuación (4.9).

4.6. Metodología de Ajuste de un Modelo PARMA

En el análisis de series de tiempo, se desea ajustar el modelo estocástico mas apropiado para representar un conjunto dado de datos. Sin importar cual sea el tipo de modelo que se desee ajustar, es recomendable seguir la metodología propuesta por Box y Jenkins, que comprende las etapas de identificación, estimación y diagnostico. La metodología general para la construcción de un modelo del tipo PARMA se resume en la figura 4.1.

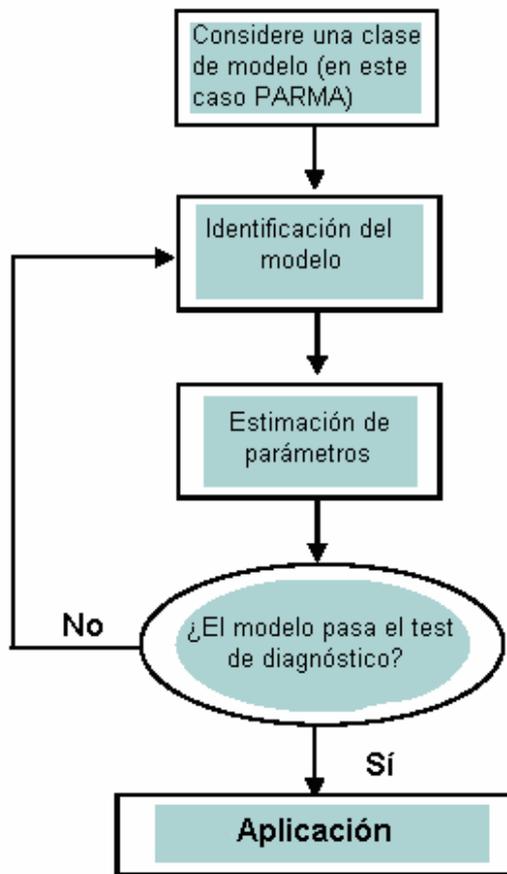


Figura 4.1: Metodología para construcción de modelo de series de tiempo
Fuente: Artículo, Desarrollo de Programas para el Tratamiento de Series Hidrológicas y Evaluación, España, 2005

Dado un tipo de modelo dentro de una familia de modelos, se identifica el orden de éste.

Determinado el orden del modelo se procede a la estimación de parámetros, si el ajuste del modelo es satisfactorio el modelo es usado para la aplicación, en caso contrario se debe volver nuevamente a la etapa de identificación, estableciéndose

un proceso iterativo hasta que un modelo satisfactorio es obtenido. Las distintas etapas se explican a continuación.

4.6.1 Identificación

Una vez seleccionado el modelo PARMA como el más adecuado para nuestra aplicación, en la etapa de identificación se debe determinar el orden del modelo, es decir los valores de $p(v)$, $q(v) \forall v=1, \dots, S$. Esto se realiza mediante la inspección visual de algunos tipos de gráficos, específicamente la función de autocorrelación simple (FACS) y la función de autocorrelación parcial (FACP).

4.6.1.1. Función de Autocorrelación Simple (FACS)

La función de autocorrelación simple corresponde a una representación gráfica de las autocorrelaciones para distintos valores del desplazamiento temporal. Esta función permite cuantificar el grado de dependencia lineal que existe entre los valores que toma un proceso estocástico en dos instantes de tiempo. Para procesos estacionarios, el valor de esta función sólo depende del intervalo de tiempo k y no del instante de tiempo en el que se está calculando. En un proceso cicloestacionario, como son las series hidrológicas, la función de autocorrelación depende del intervalo k y además depende del mes del año donde se calcula. Así, se define la autocorrelación de intervalo k , en la estación v , para un proceso cicloestacionario, como (4.10):

$$r_v(k) = E \left(\left(\frac{X_{nS+v} - m_v}{s_v} \right) \cdot \left(\frac{X_{nS+v-k} - m_{v-k}}{s_{v-k}} \right) \right) \quad (4.10)$$

Donde m_v , s_v corresponden a la media y varianza de la estación (mes, semana) v . Al hacer variar $k=0,1,2,\dots$ obtenemos la FACS.

La FACS posee tres propiedades importantes:

1. $r_m(0) = 1$
2. $-1 \leq r_m(k) \leq 1$
3. $r_{m-p}(k-p) = r_{m-k}(p-k)$

En la práctica se obtiene una estimación de la FACS mediante su valor muestral:

$$r_v(k) = \frac{1}{N-1} \sum_{i=1}^{N-1} \left(\frac{X_{iS+v} - \tilde{m}_v}{\tilde{s}_v} \right) \cdot \left(\frac{X_{iS+v-k} - \tilde{m}_{v-k}}{\tilde{s}_{v-k}} \right) \quad (4.11)$$

Donde \tilde{m}_v , \tilde{s}_v son las aproximaciones muestrales de la media y la varianza para el mes v cuyas expresiones son:

$$\tilde{\mu}_v = \frac{1}{N} \sum_{k=0}^{N-1} X_{kS+v}$$

$$\tilde{\sigma}_v = \sqrt{\frac{1}{N} \sum_{k=1}^N (X_{kS+v} - \tilde{\mu}_v)^2}$$

4.6.1.2 Función de Autocorrelación Parcial (FACP)

Para entender como se construye esta función, supongamos que se quiere ajustar un modelo periódico autoregresivo puro (PAR(p) ó PARMA(p,0)) a la serie centrada y ajustada por su desviación estándar, es decir, se desea obtener un modelo de la forma:

$$\left(\frac{X_{nS+v} - \mu_v}{\sigma_v} \right) = \sum_{i=1}^{p(v)} \phi_i(v) \cdot \left(\frac{X_{nS+v-i} - \mu_{v-i}}{\sigma_{v-i}} \right) + \varepsilon_{nS+v} \quad (4.12)$$

Observemos que al ser un modelo PAR puro, se omite la componente de medias móviles en la ecuación. Lo que se desea es encontrar una estimación preliminar de los coeficientes $f_i(v)$. Para esto, multiplicamos ambos lados de (4.12) por:

$$Z_{nS+v-k} = \left(\frac{X_{nS+v-k} - m_{v-k}}{s_v} \right)$$

Y luego obteniendo el valor esperado obtenemos:

$$E \left(\left(\frac{X_{nS+v} - m_v}{s_v} \right) \cdot \left(\frac{X_{nS+v-k} - m_{v-k}}{s_v} \right) \right) = \sum_{i=1}^{p(v)} f_i(v) \cdot E \left(\left(\frac{X_{nS+v-i} - m_{v-i}}{s_{v-i}} \right) \cdot \left(\frac{X_{nS+v-k} - m_{v-k}}{s_v} \right) \right) + E \left(\varepsilon_{nS+v} \cdot \left(\frac{X_{nS+v-k} - m_{v-k}}{s_v} \right) \right)$$

Notemos que el último término es nulo, debido a la independencia estadística de la innovación con lo observado en otros instantes. Identificando términos y sustituyendo en esta ecuación la expresión 4.10, obtenemos:

$$\mathbf{r}_v(k) = \sum_{i=1}^{p(v)} \mathbf{f}_i(v) \cdot \mathbf{r}_{v-i}(k-i) \quad (4.13)$$

Si escribimos la ecuación (4.13) para todos los valores de k entre 0 y $p(v)$, obtenemos un sistema de ecuaciones, el que escrito matricialmente tiene la forma:

$$\begin{bmatrix} 1 & \rho_{r-1}(1) & \dots & \rho_{r-1}(p(v)-1) \\ \rho_{r-1}(1) & 1 & \dots & \rho_{r-1}(p(v)-2) \\ \dots & \dots & \dots & \dots \\ \rho_{r-1}(p(v)-1) & \rho_{r-1}(p(v)-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1(v) \\ \phi_2(v) \\ \dots \\ \phi_{p(v)}(v) \end{bmatrix} = \begin{bmatrix} \rho_r(1) \\ \rho_r(2) \\ \dots \\ \rho_r(p(v)) \end{bmatrix} \quad (4.14)$$

Este conjunto de ecuaciones son conocidas como ecuaciones de Yule-Walker. Estas ecuaciones entregan una estimación preliminar de los parámetros autoregresivos cuando se ajusta un modelo PAR de orden $p(v)$. Notemos que existe un set de ecuaciones como (4.14) para cada mes.

Denotemos ahora f_{kj} al j -ésimo coeficiente (j -ésima componente del vector $[\mathbf{f}_1(v), \mathbf{f}_2(v), \dots, \mathbf{f}_{p(v)}(v)]^T$, en (4.14) obtenido cuando se ajusta un modelo PAR de orden k ($p(v) = k$) en la estación v . A medida que vamos moviendo el índice k , el valor que toma el último coeficiente f_{kk} como función de k corresponde a la FACP. Dicho de otro modo, la FACP entrega el valor del k -ésimo (último) coeficiente autoregresivo cuando se ajusta a la serie un modelo de orden k .

Es importante notar que existe una FACS y una FACP para cada mes del año, para cada afluente.

4.6.1.3 Procedimiento de Identificación

Las funciones FACS y FACP entregan información útil para poder identificar el orden del modelo a ajustar. Recordemos que tanto el orden del modelo, es decir, el número de periodos pasados que entregan información útil para la representación de la serie, como los coeficientes mismos depende de cada mes.

Modelado de Series de Tiempo Hidrológicas

Con esto, se debe identificar el orden del modelo para cada mes del año, lo que significa analizar un total doce funciones de autocorrelación y doce funciones de autocorrelación parcial.

Las funciones FACS y FACP poseen un comportamiento típico para modelos autoregresivos. Todo el análisis se basa en las dos observaciones siguientes

1. Si en el mes ν , una serie de tiempo sigue exactamente un proceso periódico autoregresivo puro $PAR(p)$ (o $PARMA(p,0)$), entonces los p primeros valores (retardos $k=1, 2, \dots, p$) de la FACP son significativamente distintos de cero, y el resto son todos nulos. Por otro lado, su FACS muestra un decaimiento exponencial o de una senoide amortiguada.

2. Si en el mes ν , la serie de tiempo sigue exactamente un proceso de medias móviles puro $MA(q)$ (o $PARMA(0,q)$), entonces los q primeros valores ($k=1, 2, \dots, q$) de su FACS poseen valores significativamente distintos de cero, el resto ($k > q$) son todos nulos. Por otro lado, su FACP mostrará un decaimiento exponencial o de una senoide amortiguada.

En base a esto podemos intuir que, si en el mes ν , una serie sigue un modelo $PARMA(p,q)$, entonces sus funciones FACS y FACP tendrán un comportamiento representado por una combinación de lo planteados en los puntos 1 y 2. En efecto, si la serie sigue un proceso $PARMA(p,q)$ su FACS mostrará un valor significativamente no nulo para los q primeros retardos y luego mostrará un decaimiento exponencial o de senoide amortiguada.

Por otro lado su FACP mostrará un valor significativamente distinto de cero para los p primeros valores, mostrando luego un decaimiento exponencial o una senoide amortiguada.

Utilizando esta información, se puede determinar el orden del modelo para cada mes, mediante una inspección visual de las funciones FAC y FACP respectivas.

Veamos por ejemplo las funciones FACS y FACP para el logaritmo de la serie correspondiente a algún afluente (figura 3.2), para el noveno mes del año hidrológico. En el gráfico superior se muestra la FACS y en el inferior la FACP. La FACS muestra un decaimiento exponencial a partir de $k=0$, además la FACP muestra un coeficiente significativamente no nulo para $k=1$, el resto de los coeficientes pueden ser estadísticamente interpretados como nulos. Con esto, un modelo tentativo para este afluente en el mes 9 podría ser un $PAR(1)$ ($PARMA(1,0)$). Fig.4.2

FAC y FACP para la estación una estación (mes 9)

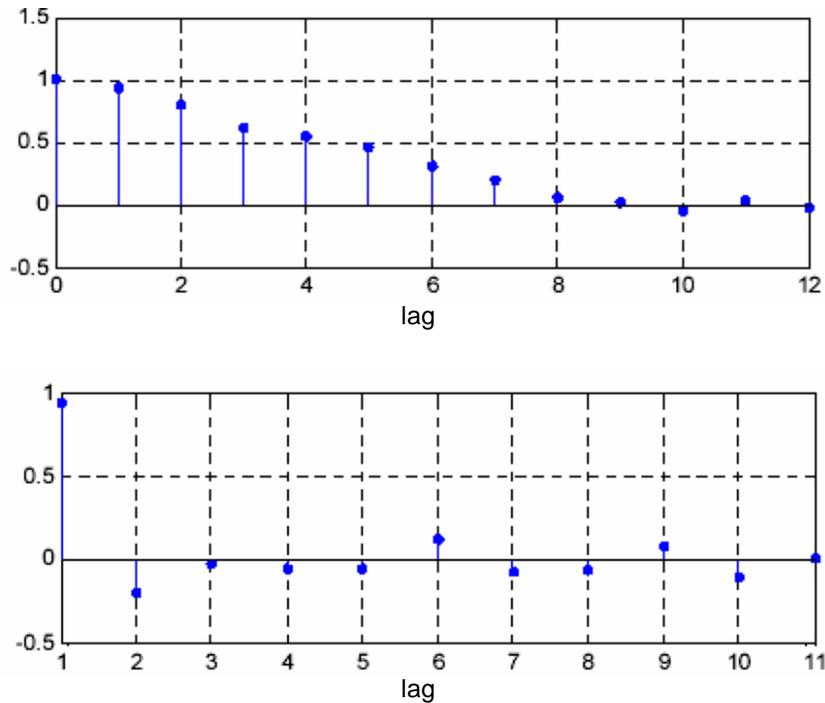


Figura 4.2: FACS y FACP para el afluente en el mes 9.

Fuente: Elaboración propia basando en la tesis Modelación Hidrológica en el Sistema Interconectado Central, Eduardo A. Mihder Het z, Chile, 1997

4.6.2 Estimación de Parámetros

Luego de la identificación, se debe ajustar a la serie de tiempo el modelo de orden señalado. Los parámetros serán ajustados por el método de mínimos cuadrados (Least Mean Square), el que consiste en minimizar la desviación cuadrática media del modelo respecto de la serie de datos. Formalmente, los coeficientes $f_i(v)$, $q_i(v)$ son estimados minimizando la suma de los cuadrados de los residuos:

$$F = \sum_{n=1}^N \sum_{v=1}^S e_{nS+v}^2 \quad (4.15)$$

donde los residuos, son obtenidos a partir de la ecuación (4.7).

Una vez que son obtenidos los parámetros del modelo, la varianza de la innovación puede ser estimada como:

$$s_e^2(v) = \frac{1}{N} \sum_{n=1}^N e_{nS+v}^2$$

4.6.3 Test de Diagnóstico

El objetivo de los test de diagnóstico es verificar si el modelo ajustado es el adecuado. Esto se realiza chequeando que las hipótesis asumidas en las etapas de identificación y estimación de parámetros son satisfechas.

Para un modelo PARMA estas hipótesis corresponden a la normalidad e independencia temporal de los residuos o innovaciones e_t , para cada mes del año. Si las innovaciones son independientes a lo largo del tiempo (no correlacionadas) la función de autocorrelación (FACS) para la serie de innovaciones e_t , construida a partir de la ecuación 47, debería mostrar una correlación casi nula para valores del retardo $k \geq 1$. En términos estadísticos, si las innovaciones son independientes y aproximadamente normales, para retardos altos, la autocorrelación sigue una distribución $N(0, 1/\sqrt{N})$. Esto quiere decir que si algún residuo cae fuera de la banda del intervalo de confianza de 95% $\left[\frac{-1.96}{\sqrt{N}}, \frac{1.96}{\sqrt{N}} \right]$, la hipótesis estadística de que dicho residuo sea nulo es rechazada con una probabilidad del 95%. La inspección visual de la FACS de las innovaciones, apoyada por el test de Porte Manteau

La hipótesis de normalidad de las innovaciones puede ser analizada dibujando la función de distribución acumulada de las innovaciones en un papel de distribución normal. Si los datos siguen una distribución normal, entonces estos deberían ajustarse a una línea recta en el dibujo.

4.6.4 Transformación de los Datos.

En el esquema de modelación propuesto por Box y Jenkins, los residuos o innovaciones son asumidos independientes a lo largo del tiempo y normalmente distribuidos. La suposición de independencia es la más importante de ellas, y su violación puede acarrear drásticas consecuencias. Sin embargo, si la hipótesis de normalidad no es satisfecha, esto puede corregirse mediante una transformación de los datos, conocida como transformación de Box-Cox.

En general, la hipótesis de normalidad de los residuos usualmente no es crítica para la obtención de una buena estimación de los parámetros del modelo. Si las innovaciones son independientes y poseen varianza finita, entonces se puede obtener una estimación razonable de los parámetros del modelo. Esto se puede demostrar experimentalmente, mediante simulaciones de Montecarlo a partir de un modelo de parámetros conocidos. Se genera sorteos de la innovación del modelo, suponiendo que estos se distribuían en forma normal, en forma uniforme, y como una normal contaminada, asumiendo el mismo valor de la varianza para

Modelado de Series de Tiempo Hidrológicas

los tres casos. Se ajusto un modelo a las series sintéticas generadas en los tres casos, obteniéndose que en todos ellos que los parámetros estimados tuvieran un valor muy cercano a los del modelo original.

Sin embargo, en la práctica resulta ventajoso satisfacer razonablemente bien la condición de normalidad de los ruidos, entre otras cosas, por que si las innovaciones no se distribuyen normalmente resulta imposible estimar intervalos de confianza que sean confiables. Como se señaló, se pueden transformar convenientemente los datos para satisfacer la normalidad, usando una transformación de Box-Cox. Sea x_t , la serie a modelar. La transformación de Box-Cox para la serie x_t se escribe:

$$x_t^\lambda = \begin{cases} \lambda^{-1} \cdot \left[(x_t + C)^\lambda - 1 \right] & \lambda \neq 0 \\ \ln(x_t + C) & \lambda = 0 \end{cases} \quad (4.16.)$$

Una vez transformados los datos se ajusta un modelo a la serie de datos transformados en lugar de la serie original. En la transformación, C es una constante seleccionada de manera que $x_t + C$ sea una cantidad estrictamente positiva para todo instante del tiempo.

Varias aproximaciones están disponibles para seleccionar adecuadamente los parámetros de una transformación de Box-Cox. A veces, es sabido de antemano que la serie en observación requiere cierto tipo de transformación. Por ejemplo, muchos autores han encontrado que en la práctica es a veces necesario aplicar logaritmo a los datos de las series hidrológicas, esto se traduce en hacer $\lambda = 0$, y a C se le asigna un valor pequeño, de manera que no haya valores nulos en la serie. En otros casos, los parámetros, C no son conocidos de antemano y deben ser estimados junto con los otros parámetros del modelo.

Para las series hidrológicas, en general basta con aplicar logaritmo natural, escogiendo C de manera que no haya observaciones nulas (para poder aplicar logaritmo). Existe además otra razón más de fondo que explica la necesidad de aplicar una transformación de este tipo a los datos de caudales, antes de ajustar un modelo del tipo PARMA, y es que si se ajusta directamente un PARMA a los datos sin transformar, las innovaciones, al distribirse en forma normal, pueden hacer que el modelo arroje valores negativos para la serie hidrológica durante la simulación, lo que no tiene sentido físico. Ajustar un modelo PARMA al logaritmo de los datos permite evitar esta dificultad

5. Modelo para Selección de Caudales Históricos

5.1 Introducción

Los modelos de planificación usados son incompatibles, en una primera aproximación, con una modelación en series de tiempo para la hidrología. Esto se debe a que los modelos usados en la planificación están obligados, a extraer los escenarios de caudales directamente del registro estadístico. Debido a esto, las interfases definidas para estos modelos sólo aceptan caudales de información histórica.

Como ejemplo, en el modelo PLP las simulaciones y aperturas son definidas previamente en archivos de texto, que guardan la información de los índices de los años hidrológicos que serán usados para la simulación.

Por lo anterior se modifican los modelos de planificación, implementando interfases que permitan simular escenarios de caudales sintéticos, que no estén estrictamente contenidos en el registro histórico. Esto es una tarea de más largo plazo, que impone barreras a la implementación inmediata del modelo CPARMA en los modelos de planificación. Buscando solucionar esta cuestión, se propone modelo alternativo, el que hace uso del modelo PARMA propuesto como un mecanismo para asignar probabilidades de ocurrencia a los escenarios históricos,

Modelo de Selección de Caudales Históricos

conociendo la información de la serie de caudales hasta el instante en que se inicia la planificación de la operación.

El modelo trabaja comparando la evolución pasada de la serie que se desea modelar, con la tendencia que mostraron los 40 años históricos en el mismo período de tiempo. Por ejemplo, si para el año en curso se desea implementar un modelo de pronóstico para los caudales de los meses de agosto en adelante, conociendo la información de caudales hasta el mes de julio, podemos inferir el comportamiento futuro ésta de la serie, comparándola con la evolución que mostraron cada uno de los 40 años hidrológicos en el período de marzo a julio. Con esta información, podemos asignar una mayor probabilidad de ocurrencia a aquellos escenarios históricos que mostraron una tendencia estadística similar a la serie actual.

Este modelo se construye sobre la hipótesis de que los 40 años estadísticos del registro histórico constituyen una muestra representativa del fenómeno hidrológico. Al extraer los escenarios de caudales directamente de la estadística, estamos asumiendo explícitamente que ésta contiene todas las realizaciones posibles del fenómeno.

La metodología propuesta esta sustentada en la idea de que las series presentan una tendencia pasada similar, entonces sus caudales futuros se distribuirán también de manera parecida. La característica distintiva del modelo propuesto, es que el modelo PARMA entrega la manera para comparar estadísticamente las series de datos.

Además, existe otra limitante que impide implementar en los modelos de planificación un modelo del tipo CPARMA de manera satisfactoria, y es que no es posible obtener modelos PARMA de calidad suficiente para todos los afluentes del sistema. El modelo que se propone en las secciones siguientes permite superar en parte esta limitación.

5.2 Modelo de Selección, Basado en un Modelo PARMA

Por ejemplo, si se desea modelar la evolución futura de los caudales de una serie, a partir del mes de agosto, del año en curso, en adelante. Y suponemos que para esto, se dispone de toda la información de afluentes hasta el mes de julio. Con esta información se puede construir una distribución normal para los caudales mensuales futuros con ayuda del modelo PARMA.

Modelo de Selección de Caudales Históricos

Las series sintéticas de caudales obtenidas mediante sorteos de ésta distribución normal no pueden ser aplicadas en el modelo de planificación, debido a limitaciones constructivas. En lugar de simular caudales mediante sorteos de la distribución normal asociada, se ocupará la distribución que se obtiene del modelo PARMA, para asignar probabilidades de ocurrencia a los caudales observados históricamente para el mes de agosto.

53							
54	AÑO	Abril	Mayo	Junio	Julio	Agosto	Se
55	63	22.0	25.6	43.6	56.2	60.5	
56	64	14.8	22.0	27.2	27.6	28.8	
57	65	49.5	80.1	139.8	90.0	89.6	
58	66	29.1	43.8	58.6	97.6	53.0	
59	67	27.7	44.5	34.7	32.7	51.8	
60	68	24.9	26.8	23.3	24.9	36.9	
61	69	30.0	61.7	148.4	80.2	81.7	
62	70	24.0	29.6	55.3	44.2	39.1	
63	71	21.9	78.2	50.7	97.9	87.4	
64	72	20.6	196.7	162.9	77.7	142.3	
65	73	20.5	60.6	65.9	61.3	43.8	
66	74	17.8	35.7	61.3	40.8	42.8	
67	75	35.5	75.3	86.8	79.5	45.4	
68	76	23.4	30.2	65.5	49.0		→ $N(\mu, \sigma)$
69	77	19.5	62.0	68.1	71.3		
70	78	22.3	50.7	62.6	148.0	67.3	
71	79	23.1	40.9	35.0	86.8	144.3	
72	80	92.0	168.2	165.4	109.1	73.7	
73	81	25.3	182.8	104.4	77.5	71.5	
74	82	15.4	53.3	104.4	117.6	69.4	

Figura 5.1 Distribución normal para los caudales de agosto en cada año histórico.

Fuente: Elaboración basada en Tesis Modelo para manejo de la Incertidumbre Hidrológica en la Planificación de la Operación del SIC

Esto se realiza estimando para cada uno de los 40 escenarios históricamente observados para el mes de agosto, la distribución normal para los caudales transformados¹, obtenida a partir del modelo PARMA (figura 5.1). Para esto se utiliza la información de caudales hasta el mes de julio.

Modelo de Selección de Caudales Históricos

Sea \bar{S} la distribución normal obtenida según el modelo PARMA, para los caudales de agosto de la serie que se quiere modelar (año actual). Sean además S_i , donde $i=1\dots 40$ las distribuciones asociadas, para el mismo mes, a cada uno de los 40 años históricos. Al comparar \bar{S} con la distribución del i -ésimo año S_k , resulta intuitivo que si ambas distribuciones son similares, entonces un conjunto de realizaciones de la distribución S_k será, en primera aproximación, una muestra representativa de la distribución \bar{S} .

Para cuantificar esta noción definamos para cada $i:1\dots 40$ los factores o pesos f_i según la expresión:

$$f_i = \frac{1}{|E(\bar{S}) - E(S_i)| + K} \quad (5.1)$$

Donde K es una constante positiva.

Para comprender el significado de los f_i , se debe notar primero que la cantidad $|E(\bar{S}) - E(S_i)|$ define, para el mes i , una noción de distancia o norma entre las variables aleatorias de distribución normal que fueron obtenidas a partir de un modelo PARMA. Con esto, a cada escenario histórico se le asigna un peso que depende directamente de la "distancia" que tiene dicho año con respecto a la serie que queremos representar. De esta forma, se le asigna un peso mayor, a aquellos escenarios estadísticamente más cercanos, y un peso menor a los más alejados.

Se asigna luego, una noción de probabilidad a cada escenario histórico, a partir de los pesos f_i .

Una forma de hacerlo es asignando como probabilidad el peso normalizado:

$$P_i = \frac{f_i}{\sum_{i=1}^{40} f_i}$$

Esta metodología muestra una forma de asignar probabilidades de ocurrencia a los caudales observados históricamente, en función de la tendencia local pasada que muestra la serie de datos que queremos modelar. Es decir, dada una serie de datos, es posible asignar para el mes $j:1,\dots,12$, una probabilidad a cada uno de los escenarios observados históricamente en el mes j .

Modelo de Selección de Caudales Históricos

En la figura 4.2 se muestran las probabilidades de ocurrencia para cada escenario histórico del mes de julio, para cierta serie de afluente. Supongamos que el caso que se modela, es una temporada de sequía muy severa. La estadística histórica utilizada en este caso corresponde a la ventana de tiempo comprendida entre los años hidrológicos 58-59 y 97-98. En el eje "Y" se encuentran las probabilidades calculadas, y en el eje "X" los años hidrológicos ordenados según hidrología creciente (de más seco a más húmedo, según los caudales observados para julio), se utilizó un valor para la constante K (ecuación 5.1) de 0.03.

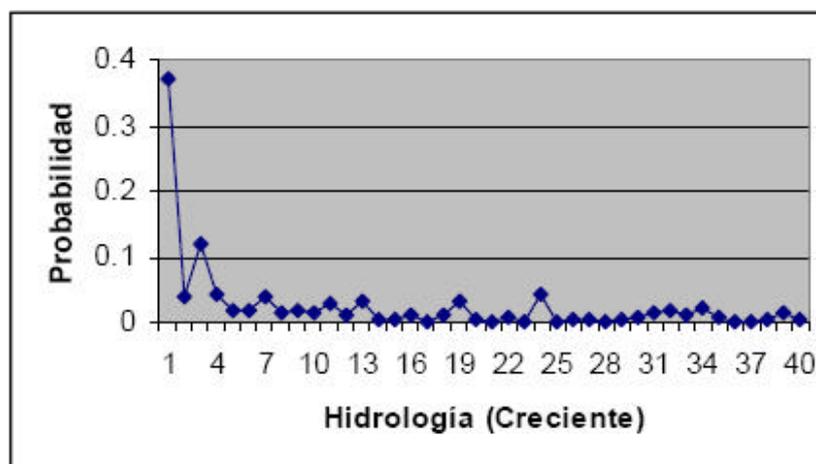


Figura 5.2 Probabilidades de caudales históricos, para mes de julio 98-99.

Fuente: Elaboración basada en Tesis Modelo para manejo de la Incertidumbre Hidrológica en la Planificación de la Operación del SIC

Se asume implícitamente un cierto grado de variabilidad en la hidrología futura, al asignar una probabilidad distinta de cero al resto de los escenarios históricos.

Si se repite el ejercicio anterior para la hidrología observada el año 2000-2001, las probabilidades obtenidas se muestran en la figura 5.2. La hidrología del año 00-01, puede ser considerada como una hidrología media-húmeda.

En la figura 5.3 se muestra la distribución de probabilidades para mes de julio del año hidrológico 59-60. Este año puede ser considerado en términos hidrológicos como un año medio.

Modelo de Selección de Caudales Históricos

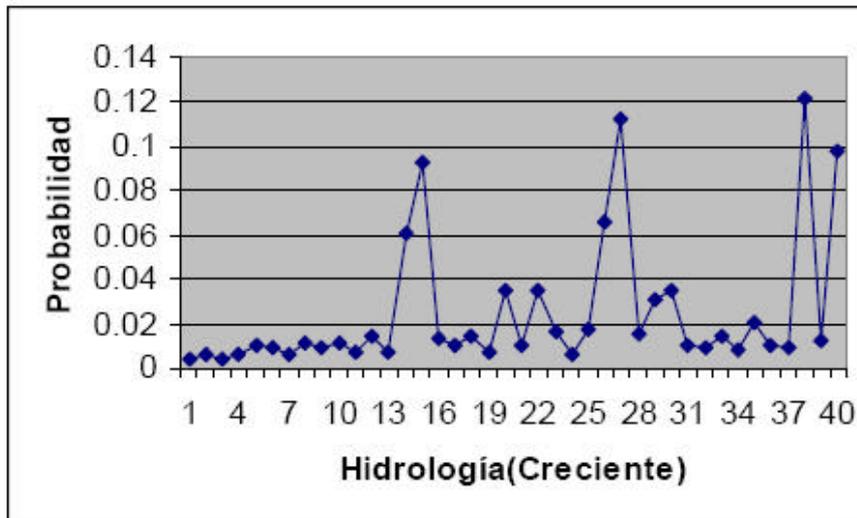


Figura 5.3 Probabilidades calculadas para el mes de julio, año 00-01.

Fuente: Elaboración basada en Tesis Modelo para manejo de la Incertidumbre Hidrológica en la Planificación de la Operación del SIC

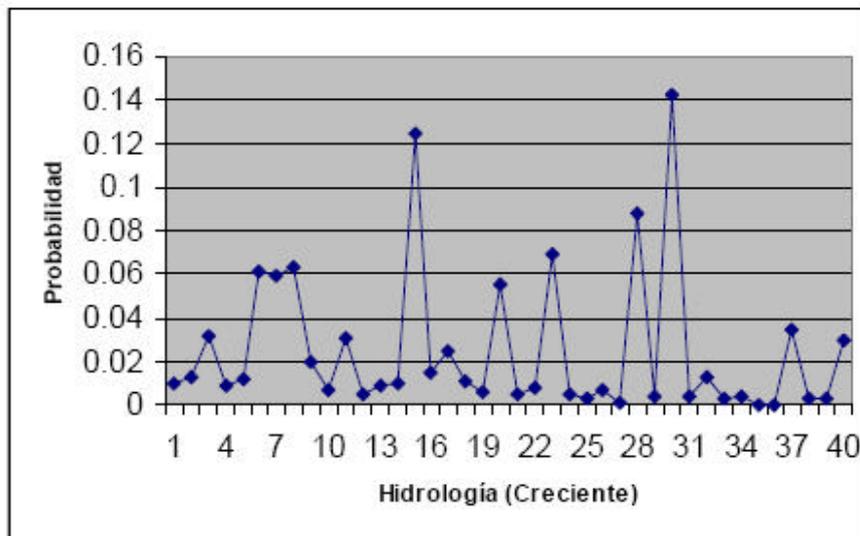


Figura 4.4 Probabilidades calculadas para el mes de julio, año 59-60.

Fuente: Elaboración basada en Tesis Modelo para manejo de la Incertidumbre Hidrológica en la Planificación de la Operación del SIC

Un aspecto importante de resaltar es que en las figuras 5.2, 5.3 y 5.4, en el eje x se ordenan en forma creciente las hidrologías observadas para el mes de julio. Es decir, el criterio para definir este orden es según el caudal promedio observado en julio, para los cuarenta años del registro.

Cuando decimos, a partir de la figura 5.2, que se le asigna una probabilidad alta al escenario más seco, nos referimos al caudal más seco observado estadísticamente en julio, y no al año más seco.

En las secciones siguientes, se plantea una metodología para generar un árbol de escenarios para la hidrología futura, y los resultados de la aplicación de esta representación en el modelo PLP.

5.3 Árbol de Escenarios

Si se inicia la planificación de la operación a comienzos del mes j , se puede obtener una distribución de probabilidad para los escenarios hidrológicos futuros, a partir del modelo PARMA.

Para generar un árbol de escenarios, el procedimiento utilizado es el siguiente:

1. Con la información de caudales de los meses previos al que se inicia la planificación de la operación, se asignan probabilidades de ocurrencia a cada escenario histórico para el primer mes del horizonte de planificación.
2. Para seleccionar las simulaciones iniciales se utilizan estas probabilidades calculadas (distribución de probabilidad discreta), y se sortean a partir de ella N escenarios de caudales históricos, para el primer mes de la planificación. Para cada uno de estos escenarios históricos sorteados, se toma la serie completa hasta el final del primer año del horizonte de planificación.
3. Para los años posteriores a cual se inicia la planificación de la operación, se asume completa independencia entre años hidrológicos y equiprobabilidad de los escenarios históricos anuales. Con estas hipótesis, para cada una de las N series obtenidas en el punto dos, se sortea una trayectoria completa para hidrología, desde el segundo año bajo estudio, hasta el final del horizonte de planificación. Con esto, se tienen N simulaciones iniciales.
4. Para cada simulación así definida, se generan M aperturas condicionadas en cada etapa las que se sortean según el mismo criterio, recorriendo cada simulación inicial y calculando las probabilidades en cada etapa, en función de los valores pasados observados, los que cambian a medida que se recorre cada secuencia.

5.4 Aplicación del Modelo

En las secciones previas se propone una metodología para asignar probabilidades de ocurrencia a los escenarios históricos, a partir de la información de caudales para los meses previos al que se inicia la planificación de la operación. Esta

Modelo de Selección de Caudales Históricos

modelación permite definir simulaciones iniciales y aperturas en cada etapa, entregando en cada sorteo, una mayor probabilidad de ocurrencia a aquellos escenarios cuya distribución futura de caudales son más cercanas a la de la serie cuyo comportamiento queremos representar.

Para la validación de esta metodología, se utilizó el modelo PLP. Para lo cual modificaron las interfases que generan los archivos de entrada al modelo, con el objeto de poder utilizar la metodología planteada.

El caso a simular corresponde a la planificación de la operación del SIC, con fecha de inicio del 1 de julio de 2004, pero con el estado inicial del sistema correspondiente al 1 de julio del año 1998, reconocido como el más seco de la estadística. Es decir, se simuló la operación óptima del sistema, con la topología, centrales, planes de obras y mantenciones proyectadas, para comienzos de julio de 2004, pero con las cotas iniciales para los embalses, observada a comienzos de julio de 1998.

El horizonte de estudio es de dos años, con un tratamiento semanal para el primer mes, y mensual para el resto del horizonte, con un sólo bloque de demanda por etapa.

Para este sistema, se simularon dos condiciones hidrológicas, la del año 98-99 y la del año 00-01, cuyas distribuciones de probabilidad para el mes de julio son las mostradas en las figuras 5.2 y 5.3. Para cada una de estas dos condiciones hidrológicas, se realizaron tres simulaciones:

1. Una simulación considerando un tratamiento hidrológico como el propuesto en las secciones 5.2 y 5.3. Con la información de caudales hasta el mes de junio, se calcula la distribución de probabilidad para los caudales en julio. El registro histórico utilizado, corresponde a la ventana de tiempo comprendida entre los años 58-59 y 97-98, y el afluente de referencia para calcular las probabilidades corresponde a la serie propuesta. Las distribuciones de probabilidad son calculadas según lo propuesto en la sección 5.2, a partir del modelo PARMA ajustados a los datos de caudales del afluente al lago Laja.
2. Una simulación, usando el tratamiento hidrológico actualmente implementado en PLP, el que considera incertidumbre reducida para las primeras cuatro semanas, e independencia hidrológica entre meses. Para calcular la incertidumbre reducida de las primeras cuatro semanas, es necesario conocer la información diaria de caudales de los siete días previos al inicio de la planificación de la operación.
3. Una simulación determinística para el primer año, la que asume conocida la hidrología entre julio y marzo del primer año bajo estudio. Luego asume independencia estadística y equiprobabilidad de los escenarios, para el segundo año de la planificación.

Modelo de Selección de Caudales Históricos

De esta forma se podrá comparar la metodología propuesta, con el tratamiento realizado actualmente por PLP, teniendo como referencia la situación que se hubiera alcanzado si se hubiese conocido de antemano la hidrología por el resto del año.

En la figura 5.5 se muestran los costos marginales obtenidos en cada etapa para los tres casos, bajo las condiciones hidrológicas del año 98-99, para la barra AJahuel220. Recordar que las primeras cuatro etapas son de duración semanal, y el resto del horizonte está discretizado en etapas mensuales.

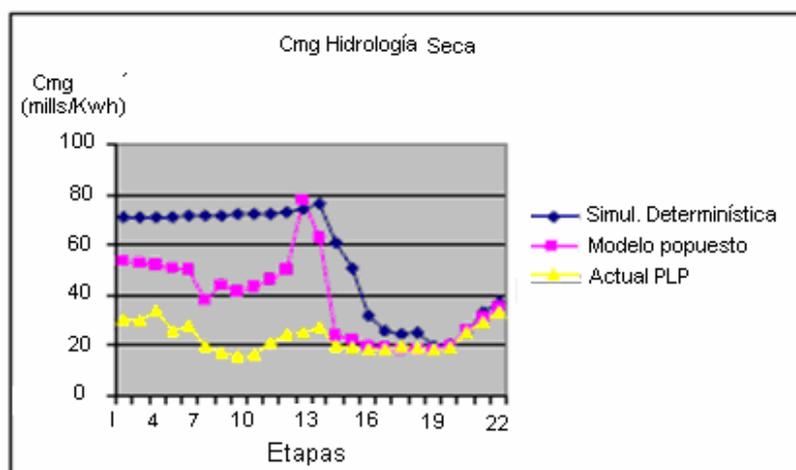


Figura 5.5 Costos marginales para los tres tratamientos hidrológicos.

Fuente: Elaboración basada en Tesis Modelo para manejo de la Incertidumbre Hidrológica en la Planificación de la Operación del SIC

En la figura 5.5 puede apreciarse que si la hidrología se conoce de antemano, el modelo PLP proyecta un costo marginal por sobre los 70 mills/Kwh para el primer año de operación del sistema. Se debe recordar que se está modelando el sistema al 1 de julio de 2004, pero con el estado inicial de los embalses observados a igual fecha del año 1998 y con los afluentes reales observados ese año.

Si se observa la trayectoria de los precios proyectada al usar la incertidumbre reducida e independencia hidrológica entre los meses de invierno (modelación que hace PLP en la actualidad), se ve que los costos marginales se sitúan muy por debajo del valor "real" obtenido en la modelación determinística, arrojando un precio que oscila en torno a los 30 mills/Kwh en las primeras etapas. De aquí podemos concluir dos cosas, la incertidumbre reducida usada por PLP en las primeras semanas no reconoce adecuadamente la tendencia hidrológica del año, debido a que se construye sólo con la ventana de tiempo de los 7 días previos al que se inicia la planificación. Por otro lado, la independencia estadística entre meses asumida por PLP, hace que la débil señal de escasez entregada por la incertidumbre reducida, se diluya inmediatamente al segundo mes del horizonte de estudio, haciendo que el precio caiga por debajo de los 20 mills/Kwh después de

Modelo de Selección de Caudales Históricos

la cuarta etapa¹. De lo anterior se concluye, que el tratamiento hidrológico implementado actualmente en el modelo PLP, es incapaz de anticiparse a un escenario de sequía severa.

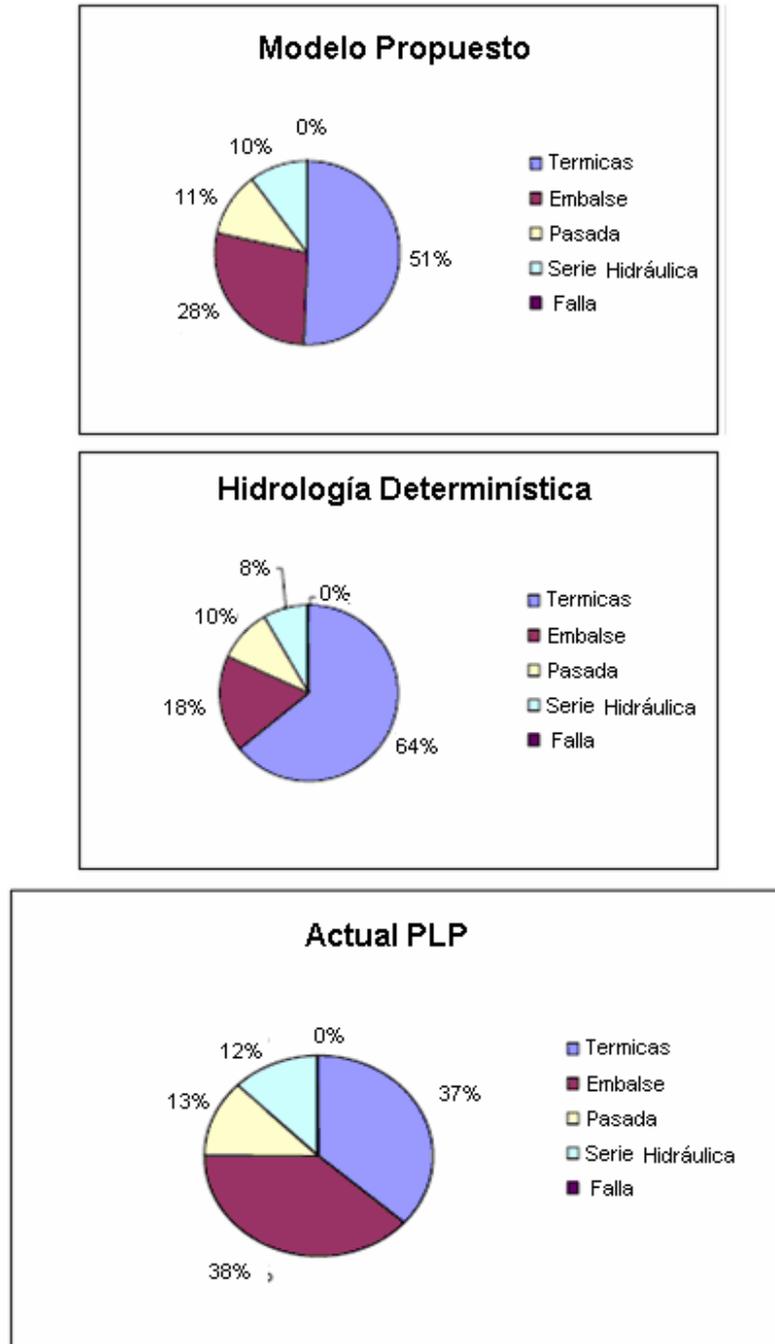


Figura 5.6 Aporte energético por tipo de central para los tres casos.

Fuente: Elaboración basada en Tesis Modelo para manejo de la Incertidumbre Hidrológica en la Planificación de la Operación del SIC

Lo anterior se puede ver con mayor claridad, si se observan los aportes energéticos por tipo de central, para el primer año del horizonte de planificación (figuras 5.6). El tratamiento actual de PLP, proyecta un 38% de generación térmica contra un 62% de generación hidráulica. Por su parte la simulación determinística (que muestra los valores que se habrían obtenido de conocerse de antemano la hidrología) proyecta un 64% de generación térmica, frente a un 36% de asignación hidráulica, colocando un 26% de generación térmica adicional, la que corresponde a centrales térmicas más caras, elevando los costos marginales.

Cuando se utiliza la modelación propuesta, el modelo identifica fuertemente (asigna una probabilidad alta) el año hidrológico simulado, con el escenario más seco contenido en la estadística, que en este caso corresponde al año 68-69. Los precios oscilan en torno a los 50 mills/kWh para el primer año, y los aportes hidráulicos son de un 49% frente a un 51% de generación térmica, mostrando un aumento de un 13% en la generación térmica, respecto del tratamiento actual de PLP. Con esto, se asigna un mayor valor al agua embalsada, ofreciendo las señales adecuadas frente a una sequía extrema.

En la figura 5.7, se muestran los costos marginales calculados para la barra Ajahuel220, obtenidos para los tres tratamientos hidrológicos, bajo las condiciones hidrológicas del año 00-01.

En la figura 5.8, se muestran los aportes energéticos según tipo de central para este caso, para el primer año del horizonte de estudio, para cada tratamiento de la hidrología.

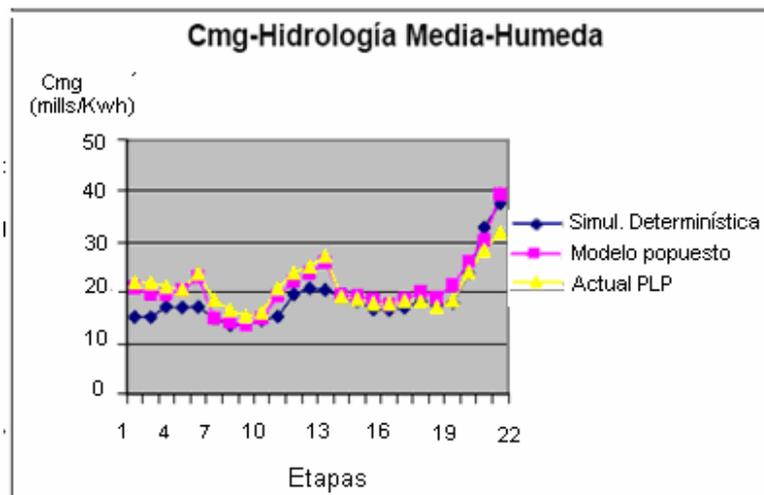


Figura 5.7 Costos marginales Ajahuel220 para los tres tratamientos hidrológicos. Hidrología Fuente: Elaboración basada en Tesis Modelo para manejo de la Incertidumbre Hidrológica en la Planificación de la Operación del SIC

Modelo de Selección de Caudales Históricos

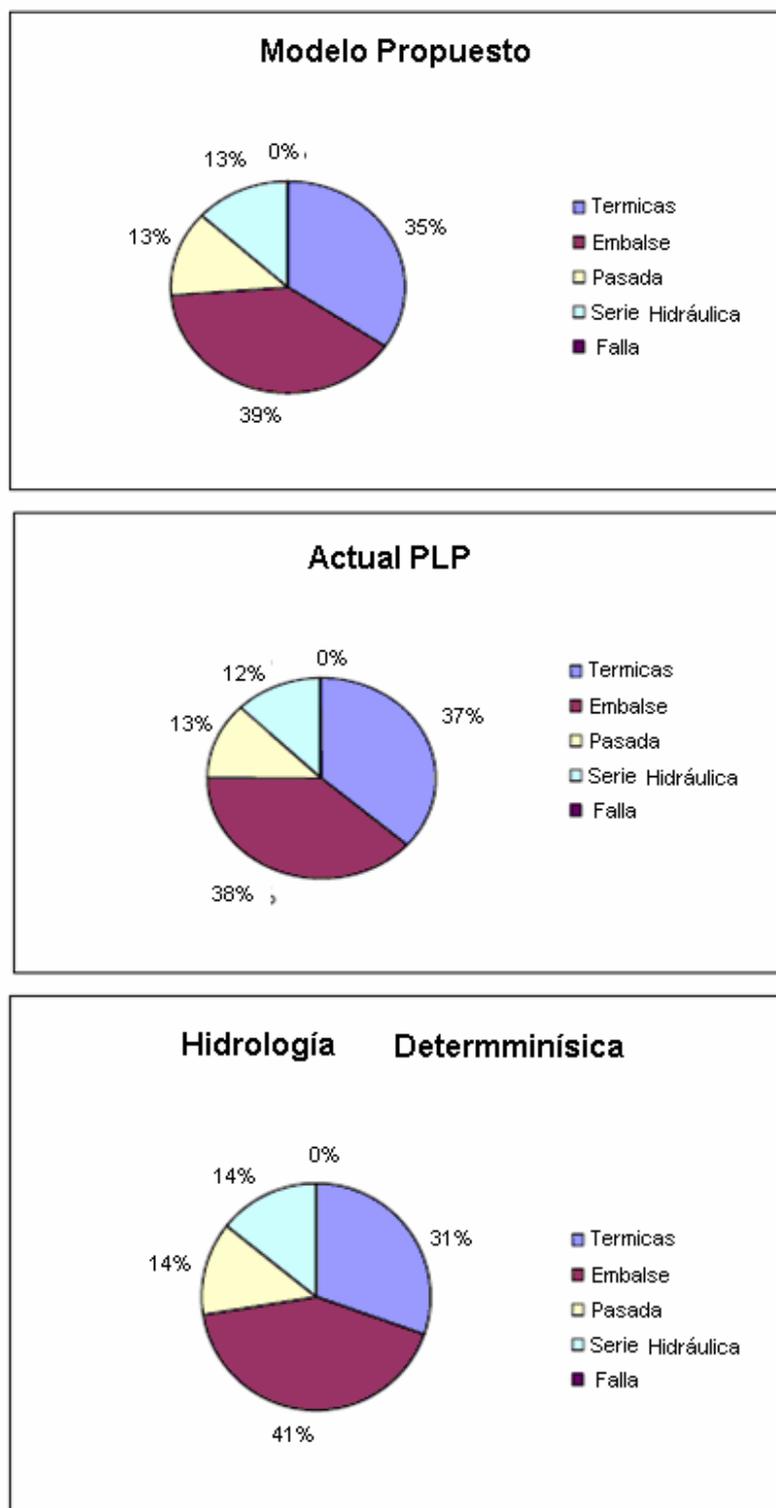


Figura 5.8 Aporte energético por tipo de central para los tres casos
Fuente: Elaboración basada en Tesis Modelo para manejo de la Incertidumbre Hidrológica en la Planificación de la Operación del SIC

Se observa que para el caso de una hidrología media-húmeda, los tres tratamientos entregan resultados similares, tanto en la proyección de precios como en la distribución de los aportes energéticos por tipo de central. En este caso, el modelo propuesto muestra resultados que difieren muy poco respecto de los valores que se obtienen al representar la hidrología con la metodología actualmente implementada en PLP.

Se podría anticipar entonces, que para el caso de hidrologías medias, el modelo propuesto entrega resultados similares a los que se obtienen con la modelación hidrológica actual de PLP, pero es capaz de anticiparse a un escenario de sequía extrema, entregando las señales económicas adecuadas, característica que no posee el tratamiento actual de PLP.

5.5 Validez y Limitaciones del Modelo

Se propone un método para asignar probabilidades a los escenarios históricos, en función de la tendencia que muestran los datos en los meses previos al que se inicia la planificación de la operación. Los resultados obtenidos en las simulaciones indican que el método propuesto entrega las señales adecuadas, obteniéndose valores esperados de las variables de interés (en este caso de los costos marginales) que son consecuentes con la tendencia de la serie hidrológica que se está modelando.

El modelo, asigna una mayor probabilidad a aquellas series históricas que muestran una tendencia futura similar a la de la serie hidrológica a modelar. Por tendencia futura similar, se entiende como aquellas series históricas que minimizan la *distancia* entre las variables aleatorias asociadas (a partir del modelo PARMA). Formalmente, la validez de la metodología propuesta, se sustenta en dos aspectos:

1. Es posible asignar una distribución de probabilidad normal, a las series de datos transformados de la serie del Laja, haciendo uso de un modelo PARMA. En este trabajo, el modelo PARMA fue ajustado utilizando el software SAMS, desarrollado por la universidad de colorado y que ha sido utilizado con éxito en diversas aplicaciones. En esta aplicación, las distribuciones describen el comportamiento de los caudales para el primer mes en el que se inicia la planificación de la operación.

2. Se puede definir una noción de distancia (norma) en el espacio de las variables aleatorias normales. En términos de esta noción de distancia, dos series de caudales son *cercanas*, si sus caudales futuros se distribuyen de manera similar.

No obstante lo anterior, los resultados obtenidos son exclusivamente de carácter preliminar, y no deben ser tomados como concluyentes. Una validación exhaustiva de este modelo requiere del desarrollo de una gran cantidad de simulaciones, estudiando distintas situaciones de operación del sistema.

Para la simulación del escenario hidrológico 98-99, parte de la información del estado inicial del sistema fue recuperada de los archivos de salida del modelo OMSIC, que era el modelo utilizado en ese entonces, y de la red interna del CDEC. Sin embargo no fue posible reconstruir fielmente el caso completo de ese año, debido principalmente a que la información disponible en las planillas OMSIC (que es un modelo uninodal y de un único embalse) no es directamente traducible a PLP (que es multinodal y multiembalse). Por esta razón, se optó por simular un escenario que consideraba la topología del sistema al primero de julio de 2004, pero con el estado inicial del sistema al primero de julio de 1998. Para esto fue necesario hacer algunas aproximaciones, relacionadas con el estado inicial de las centrales hidráulicas que no existían en esa fecha, como es el caso de la central Ralco, para la que se asumió una situación media, tanto para el estado inicial del embalse como para los afluentes iniciales.

El factor de tiempo no es menor, una simulación en PLP puede demorar varias horas dependiendo del número de bloques por etapa en los que se modela la curva de duración de la demanda, de la duración del horizonte de estudios y del número de etapas en que éste se discretiza. Por esta razón, las simulaciones se realizaron para un horizonte de estudio de sólo dos años, con las cuatro primeras etapas semanales y el resto mensual. La demanda se modeló con un solo bloque por etapa. Con estas simplificaciones, una simulación en PLP tarda aproximadamente una hora.

Debido a las simplificaciones introducidas en las simulaciones, los valores de costos marginales obtenidos en la sección anterior, no representan valores reales y sólo deben ser considerados como referenciales.

5.6 Perspectivas

El modelo propuesto, posee algunas limitaciones que podrían incidir directamente en la calidad de los resultados. Estas son:

1. La forma en la que se definen las probabilidades (pesos normalizados) para cada escenario, es completamente arbitraria.

2. La selección, tanto de las simulaciones como de las aperturas, se define a partir de la información entregada por el modelo PARMA ajustado para un sólo afluente, el que en este caso corresponde al afluente al lago Laja. Sólo con esta información se definen criterios para seleccionar los años hidrológicos a utilizar en las simulaciones y aperturas, para todos los afluentes del sistema.

Lo anterior motiva la búsqueda de criterios de selección, que incluyan a otros caudales afluentes del sistema, para los cuales también se dispone de un modelo PARMA de calidad satisfactoria.

Además, resultaría interesante inspeccionar otros métodos para definir las probabilidades o pesos.

No obstante lo anterior, lo valioso y rescatable del modelo propuesto es que utiliza la información estadística entregada por el modelo PARMA para modelar una serie de caudales, a partir de un conjunto de realizaciones históricas, y no mediante sorteos de una distribución.

5.6.1 Espacio Vectorial de las Variables Aleatorias Normales

Supongamos que se ha ajustado un modelo PARMA a las series de caudales de m afluentes del SIC. Estos modelos permitirán disponer de una distribución normal para los caudales de cada mes, y para cada afluente, en función de los datos de entrada al modelo (valores pasados de la serie que se desea modelar).

El tratamiento planteado se basa en el hecho de que el conjunto de las variables aleatorias que se distribuyen normalmente definen un espacio vectorial. Sea O dicho espacio. En el podemos definir un producto interno:

$$\langle , \rangle : O \times O \rightarrow \mathbb{R}$$
$$x, y \rightarrow E(x * y)$$

Para probar que esta aplicación efectivamente es un producto interno en O , hay que verificar lo siguiente:

- 1.- $\langle x, y \rangle = \langle y, x \rangle$
- 2.- $\langle \mathbf{a} \cdot x + \mathbf{b} \cdot y, z \rangle = \mathbf{a} \cdot \langle x, z \rangle + \mathbf{b} \cdot \langle y, z \rangle \quad \forall \mathbf{a}, \mathbf{b} \in \mathbb{R}, \forall z \in O$
- 3.- $\langle x, x \rangle > 0 \quad \wedge \quad \langle x, x \rangle = 0 \Leftrightarrow x = 0$

Las proposiciones 1 y 2 son triviales a partir de la definición y la linealidad del operador valor esperado.

Para probar 3 basta recordar que toda variable aleatoria normal $x \in N(\mathbf{m}, \mathbf{S}^2)$, se puede escribir de la forma:

Modelo de Selección de Caudales Históricos

$$x = \mathbf{s} \cdot e + \mathbf{m}$$

con $e \in N(0,1)$.

Luego, podemos escribir:

$$\begin{aligned}\langle x, x \rangle &= E\left[(\mathbf{s} \cdot e + \mathbf{m})^2\right] \\ &= \mathbf{s}^2 \cdot E(e^2) + 2 \cdot \mathbf{m} \cdot \mathbf{s} \cdot E(e) + \mathbf{m}^2\end{aligned}$$

Pero $e \in N(0,1)$, luego $E(e^2)=1$ y $E(e)=0$, con lo que finalmente tenemos:

$$\langle x, x \rangle = \mathbf{s}^2 + \mathbf{m}^2$$

Ahora, si $\langle x, x \rangle = 0$:

$$\Rightarrow \mathbf{s}^2 + \mathbf{m}^2 = 0 \Rightarrow \mathbf{m} = 0, \mathbf{s} = 0$$

$$\Rightarrow x \equiv 0$$

Luego, x corresponde a la variable aleatoria idénticamente nula.

A partir de este producto interno podemos definir una norma en O :

$$\|x\|_{\Omega} = \sqrt{\langle x, x \rangle} \tag{5.1}$$

Se incluye el subíndice O en la norma, para denotar explícitamente que ésta es una norma en O .

Con esta noción de distancia en O , podemos encontrar criterios para comparar la serie que queremos modelar, con aquellas contenidas en el histórico. Esto lo hacemos analizando distancia entre las distribuciones normales asociadas. Notemos que este criterio nos permite comparar la similitud del comportamiento futuro de dos series, en función de su comportamiento pasado

En efecto, siguiendo con el ejemplo de la sección previa, sea \bar{S} la distribución de los caudales de agosto, asociada a la serie cuyo comportamiento futuro queremos modelar, en función de los caudales en meses previos. Sea S_k la distribución normal para los caudales de agosto asociada a la k ésima serie histórica. Sean ahora x y x_k , variables aleatorias que tal que $x \in X$ y $x_k \in S$.

Se sigue que:

$$\|\bar{x} - x_k\|_{\Omega} = \sqrt{(\mathbf{s}_{\bar{x}} - \mathbf{s}_{x_k})^2 + (\mathbf{m}_{\bar{x}} - \mathbf{m}_{x_k})^2} \quad \sigma_{\bar{x}} = \sigma_{x_k}$$

Recordemos que por construcción del modelo PARMA, $\mathbf{s}_{\bar{x}} = \mathbf{s}_{x_k}$, con lo que obtenemos:

$$\|\bar{x} - x_k\|_{\Omega} = |\mathbf{m}_{\bar{x}} - \mathbf{m}_{x_k}|$$

es decir, el módulo de la diferencia de las medias entregadas por el modelo PARMA, es una noción de distancia para comparar la tendencia estadística de las distintas series históricas.

Para el desarrollo anterior se ha analizado el caso de un sólo afluente. En el caso general, los caudales definen una variable aleatoria vectorial, en que cada componente corresponde a un afluente del sistema. Con ésto, cada componente del vector aleatorio sigue una tendencia, la que puede ser modelada mediante un PARMA.

Supongamos que se tiene un modelo PARMA ajustado a m afluentes del sistema, con lo que tenemos que el vector aleatorio de afluentes $V \in \Omega^m$. Supongamos además que se desea modelar el comportamiento del vector de caudales para el j-ésimo mes del año hidrológico, para el año en curso. Suponga que para esto disponemos de toda la información de caudales hasta el mes j-1, para los m afluentes considerados. Con esta información, usando el modelo PARMA, podemos construir el vector $\bar{X}^j \in \Omega^m$ con $\bar{X}^j = (x_1^j, \dots, x_m^j)^T$, que distribuye al vector de caudales en el mes j, a partir de la información pasada de los m afluentes.

Se demuestra que si 4.1 es una norma en Ω , entonces:

$$\|\bar{X}^j\|_{\Omega^m} = \sqrt{\sum_{k=1}^m \|x_k^j\|_{\Omega}^2} \tag{5.2}$$

es una norma en Ω^m .

A partir de 5.2 se demuestra directamente que:

$$\|\bar{X}^j\|_{\Omega^m} = \sqrt{\sum_{k=1}^m (s_k^j)^2 + (m_k^j)^2}$$

donde s_k^j y m_k^j corresponden a la varianza y la media de la distribución asociada al k-ésimo afluente, en el j-ésimo mes.

Por construcción del modelo PARMA, para cada afluente, sólo la media es función de los valores pasados de la serie. La varianza, por su parte, no es función del pasado si no que es un valor predeterminado que cambia cada mes. Con esto, si se comparan en el mes j las variables aleatorias asociadas dos series vectoriales de caudales \bar{X} e \bar{Y} de años hidrológicos distintos, podemos escribir:

$$\|\bar{X}^j - \bar{Y}^j\|_{\Omega^m} = \sqrt{\sum_{k=1}^m (m_k^j - m_{y_k}^j)^2} \tag{5.3}$$

Donde m_k^j y $m_{y_k}^j$ corresponden a la media (valor esperado) para el k-ésimo afluente (k-ésima componente del vector de afluentes) en el j-ésimo mes, para \bar{X} y \bar{Y} respectivamente.

Con 5.3 tenemos una noción de distancia que permite comparar estadísticamente dos series de afluentes ocurridas en años hidrológicos distintos, incluyendo para la modelación, la información de todos los afluentes de interés. Una expresión de este tipo podría permitir, por ejemplo, definir una noción de probabilidad en función de la distancia entre la serie del año que se quiere modelar, y cada uno de los años históricos. Un tratamiento alternativo sería definir un método de clusters para clasificar mes a mes a los caudales históricos, en función de la tendencia del año presente para los distintos afluentes del sistema.

5.7 Número de Aperturas en Cada Etapa

A la luz del modelo PARMA, que asigna distribuciones normales a los caudales transformados, es razonable preguntar acerca del número de aperturas del que es necesario disponer en cada etapa, para obtener una representación adecuada de los valores esperados de las variables de interés. Se propone un modelo de sensibilidad de la función de costos futuros, respecto del vector de ruidos del modelo PARMA, para una futura validación e implementación. El desarrollo

propuesto trabaja sobre un sólo afluente. El caso para un vector aleatorio de afluentes es una extensión simple del modelo propuesto en esta sección.

5.7.1 Modelo de Sensibilidad de la FCF

Al ajustar un modelo PARMA a la serie de datos hidrológicos, la dependencia estocástica del problema de despacho queda limitada exclusivamente al vector de ruidos e_t . En efecto, el modelo PARMA internaliza toda la dependencia temporal de la hidrología en una combinación lineal de los eventos observados en instantes precedentes, los que son determinísticos. La componente estocástica del modelo está estrictamente contenida en e_t . Con esto podemos escribir la función de costos futuros en cada etapa, explícitamente como una función de e_t :

$$a_t = a_t(e_t) \quad \forall t$$

donde a_t corresponde a los costos futuros de operación observados desde el comienzo de la etapa t en adelante y e_t corresponde al ruido de la hidrología presentada en la etapa t.

Se observa que, si la varianza del ruido en la etapa s_{e_t} es pequeña, entonces el ruido e_t tomará un valor cercano a cero con una probabilidad muy alta. Con esto, podemos hacer una aproximación de primer orden de la FCF, con lo que se obtiene:

$$a_t = a_t(0) + \left. \frac{\partial a_t}{\partial e_t} \right|_{e_t=0} \cdot e_t + o(e_t)^2 \quad (5.4)$$

Por construcción del modelo PARMA, sabemos que el ruido se distribuye como una $N(0, s_{e_t})$ luego, podemos escribir:

$$e_t = s_{e_t} \cdot C$$

Donde x es una variable aleatoria que sigue una distribución normal estándar. Usando esto en la expresión (5.4):

Modelo de Selección de Caudales Históricos

$$x_t \approx \mathbf{a}_t(0) + \left(\frac{\partial \mathbf{a}_t}{\partial \mathbf{e}_t} \bigg|_{\mathbf{e}_t=0} \right) * \mathbf{S} \mathbf{e}_t \quad \forall t$$

Luego, en primera aproximación, si la varianza del ruido es pequeña, la función de costos futuros se distribuye:

$$\mathbf{a}_t \rightarrow N \left(\mathbf{a}_t(0), \frac{\partial \mathbf{a}_t}{\partial \mathbf{e}_t} \bigg|_{\mathbf{e}_t=0} \right) * \mathbf{S} \mathbf{e}_t \quad \forall t$$

Notemos que ésta expresión nos entrega una aproximación de primer orden de la varianza de la función de costos futuros en la etapa t:

$$\mathbf{s}_{\mathbf{a}_t} = \frac{\partial \mathbf{a}_t}{\partial \mathbf{e}_t} \bigg|_{\mathbf{e}_t=0} \quad \forall t$$

Para poder evaluar esta expresión, observemos que:

$$\frac{\partial \mathbf{a}_t}{\partial \mathbf{e}_t} = \frac{\partial \mathbf{a}_t}{\partial H_t} * \frac{\partial H_t}{\partial \mathbf{e}_t} \quad \forall t \tag{5.5}$$

donde H_t es el volumen de agua total entregado por el caudal afluente en la etapa t. Notar que el volumen afluente total captado en la etapa t, puede ser obtenido a partir del modelo PARMA.

Recordemos que el modelo PARMA es ajustado a la serie de datos transformados logarítmicamente, luego, para obtener datos de caudales a partir de la información entregada por el modelo es necesario aplicar la transformación exponencial inversa. Con esto, el volumen total de agua recibida desde el afluente H_t , durante la etapa de duración Δt , se determina con la expresión:

$$H_t = \Delta t * \exp(\bar{\mathbf{m}}_t + \mathbf{e}_t) \tag{5.6}$$

Donde

$$\bar{m} = \sum_{i=1}^{p(v)} f_i^{(v)} * X_{nS+v-i} - \sum_{i=1}^{q(v)} q_i * e_{nS+v-i}$$

corresponde a la media dada por la tendencia estadística de los datos, definida por el modelo

PARMA. Derivando H_t cr/a e_t :

$$\frac{\partial H_t}{\partial e_t} = \Delta t * \exp(\bar{m}_t + e_t)$$

Por otro lado, podemos escribir:

$$\frac{\partial a_t}{\partial H_t} = \Pi(v)$$

(5.7)

donde $p(v)$ es el multiplicador de Lagrange asociado a la restricción de balance hidráulico, o la pendiente del corte de Benders asociada al nivel de almacenamiento v al comienzo de la etapa $t+1$.

Con las ecuaciones (5.5), (5.6) y (5.7) podemos escribir:

$$S_a = p(v) * \Delta t * \exp(\bar{m}_t) \quad \forall t$$

(5.8)

Y la varianza para la función de costos futuros en la etapa t será:

$$S_a = p(v) * \Delta t * \exp(\bar{m}_t) * S_e \quad \forall t$$

Vemos que la varianza (y por lo tanto la sensibilidad) de la FCF en cada etapa depende de la duración de la etapa t , del valor esperado del caudal en la etapa $\exp(\bar{m}_t)$, de la sensibilidad local del modelo de optimización (que depende del estado del embalse) $p(v)$, y de la varianza del modelo de caudales.

La ecuación (5.4) permite escribir la FCF como una función explícita del ruido en cada etapa. Si tomamos valor esperado a la expresión (1), obtenemos:

$$E(a_t) = E(a_t(0) + 0(e_t^2)) + \left. \frac{\partial a_t}{\partial e_t} \right|_{e_t=0} * E(e_t)$$

Modelo de Selección de Caudales Históricos

Si se evalúa esta expresión con una simulación de Montecarlo mediante sorteos del ruido e_t se tiene:

$$E(a_t) \approx a_t(0) + \frac{1}{N} \sum_{k=1}^N \left. \frac{\partial a_t}{\partial e_t} \right|_{e_t=0} * \frac{1}{N} \sum_{k=1}^N e_{tk}$$

Donde e_{tk} corresponde a la k ésima muestra o sorteo del ruido, en el instante t . Notemos que para valores de e_t pequeños, el término más importante en orden de magnitud es el término de primer orden en e_t . Por otro lado, por hipótesis del modelo autoregresivo, $E(e_t)=0$. Con esto, es intuitivo que se habrá alcanzado una precisión aceptable en la estimación de la FCF cuando se cumpla:

$$\frac{1}{N} \sum_{k=1}^N e_{tk} \approx 0$$

En base a lo anterior, es posible definir el número N de muestras del proceso estocástico e_t para que este promedio sea tan cercano a cero como precisos queramos ser en la estimación del valor esperado de la FCF. Para esto, primero definamos la variable aleatoria:

$$h_t = \left. \frac{\partial a_t}{\partial e_t} \right|_{e_t=0} * e_t$$

Sabemos, por lo desarrollado en este capítulo, que h_t se distribuye con una normal de media nula y desviación estándar dada por (5.8). Con esto, la teoría estadística demuestra que la variable aleatoria definida por:

$$Z = \frac{\sqrt{N}}{S_a} * \bar{h}$$

Con $\bar{h} = \frac{1}{N} \sum_{i=1}^N h_i$, se distribuye como una normal estándar.

Modelo de Selección de Caudales Históricos

Con lo anterior, si queremos encontrar un intervalo de confianza para Z , podemos escribir:

$$P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1 - \alpha$$

De las tablas, para un nivel de confianza del 95%, obtenemos $Z_{\alpha/2} = 1.96$. Con esto se obtiene que:

$$\begin{aligned} P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) &= P(-1.96 \leq \frac{\sqrt{N}}{s_a} * \bar{h} \leq 1.96) \\ &= P\left(\frac{-1.96 * s_a}{\sqrt{N}} \leq \bar{h} \leq \frac{1.96 * s_a}{\sqrt{N}}\right) = 0.95 \end{aligned}$$

Esta expresión indica para cada valor de N , un intervalo que contendrá a \bar{h} con una probabilidad del 95%. Si se quiere un nivel de precisión d , entonces basta imponer:

$$\frac{1.96 * s_a}{\sqrt{N}} = d$$

Con lo que el número de muestras asociado será:

$$N = \left(\frac{1.96 * s_a}{d} \right)^2 \tag{5.9}$$

Esta expresión entrega el número mínimo de muestras de la hidrología que se deben generar en la etapa t , en función de la varianza de la FCF en la etapa t , que deben ser generadas para alcanzar una precisión d en el cálculo de la FCF en la etapa.

Es importante hacer notar que el número de aperturas determinado por la expresión 59 es válido sólo cuando la hidrología es representada mediante un modelo PARMA, y por lo tanto no es aplicable a la metodología propuesta en este trabajo. Esto se debe a que los caudales son sorteados desde el registro histórico, asignando probabilidades de ocurrencia a cada escenario con la ayuda de un modelo PARMA, luego los caudales no corresponden a series sintéticas sorteadas directamente a sorteos del modelo PARMA, si no que a series históricas.

Conclusiones y Recomendaciones

Las técnicas de programación estocástica constituyen, un atractivo campo de investigación, la gran importancia radica en que la función objetivo y las restricciones, suponen total o parcialmente información a priori, por ejemplo unimodalidad, convexidad, etc., no obstante, si tales condiciones no son fáciles de probar por la naturaleza estocástica de las funciones, la simulación es la única salida.

La Hidrología Estocástica aplica la teoría de los Procesos Estocásticos, que son parte de la Ingeniería de Sistemas, al estudio de los eventos hidrológicos.

Los Eventos Hidrológicos, tales como aguaceros, caudales, niveles de embalse, etc. son eventos estocásticos. Se caracterizan porque, de un lado tienen un patrón medio de comportamiento a largo plazo, y por el otro el pronóstico de sus magnitudes en un momento dado tiene un mayor o menor grado de incertidumbre. El patrón medio corresponde a lo que se denomina la tendencia general o componente determinística y la incertidumbre constituye la componente aleatoria del evento.

Los proyectos que se diseñan hoy serán construidos y operarán en los próximos años. Por esta razón, cuando los estudios hidrológicos definen unos caudales de diseño para abastecer un acueducto, o para proteger una zona contra las inundaciones debe tenerse en cuenta que tanto la ocurrencia como las

magnitudes de esos caudales son pronósticos hacia el futuro de eventos probables que están sujetos a un grado de incertidumbre.

La construcción de un modelo estocástico para representar datos hidrológicos se realiza de manera de imitar las propiedades estadísticas de las series de datos disponibles. En esta forma el modelo depende de los datos disponibles. Es por lo tanto, es importante que la información sea revisada cuidadosamente para asegurar que no existan valores que no aporten información o que la distorsionen. Este tipo de modelos tratan de reproducir los valores esperados, las varianzas, las covarianzas o dependencias temporales en una misma serie y las espaciales entre series cercanas.

En algunas ocasiones la estadística de caudales en algún lugar es relativamente corta ya que sólo se dispone de unas series completas para el período de análisis, de manera que muchas de ellas contienen datos creados mediante diferentes técnicas. Estas técnicas en general, mejoran la estimación de los valores esperados, pero introducen sesgos en las estimaciones de la varianza y las dependencias. Si no se revisa bien la información que se le proporcione al modelo reproducirá distorsiones. Es importante que en la construcción del modelo se verifiquen y se separen los datos reales de los datos creados y completados, y asegurar que las estimaciones de las varianzas y las dependencias son las correctas.

La formulación de un modelo estocástico para ser usado en la generación de series sintéticas es una tarea que debiera realizarse a partir de la información disponible una sola vez, de manera de definir las probabilidades del modelo y estimar sus parámetros que aseguren que se ha logrado el modelo más adecuado. En este caso, la tarea de modelación es diferente a la de simulación, debiendo el modelo ser único si los datos lo son.

Los procesos involucrados en la construcción de un modelo requieren adoptar decisiones en varias etapas intermedias del proceso que difícilmente pueden ser automáticas, menos sino siempre se tiene claro las consecuencias de seleccionar alguna de las opciones planteadas, esto se ve especialmente en la selección del modelo de orden, en la necesidad de normalizar las series, en las opciones entre modelos alternativos o la independencia de los residuos. Si bien todas estas decisiones pueden tomarse con criterios estadísticos de manera automática, es siempre útil mantener el control sobre los resultados y estar en condiciones de volver a etapas anteriores para rectificar las opciones adoptadas.

En algunas ocasiones las series de caudales son mensuales o semanales y estas representan características estadísticas relevantes de una alta dependencia temporal entre valores sucesivos de una misma serie, una dependencia espacial también importante entre series de una misma cuenca o de cuencas vecinas. Las propiedades de valores medios y varianza son periódicas con grandes diferencias entre los valores de una estación a otra en un año. Todo lo anterior hace necesario que el modelo sea del tipo multivariado, periódico y con dependencia temporal.

Cuando los valores de los caudales mensuales o semanales de los ríos presentan una asimetría, los estimadores de la asimetría son ineficientes. Es necesario poner mucha atención en la normalización de las series de tiempo, para evitar que el modelo genere valores irreales.

La periodicidad de las series mensuales y semanales genera una gran cantidad de parámetros, si estos se estiman directamente de los datos de cada período. Además estas estimaciones tienen una varianza de muestreo importante, de manera que es recomendable emplear una representación paramétrica de las características de las series o de los parámetros del modelo.

Los parámetros del modelo se estiman a partir de la muestra histórica que trata de reproducirse estadísticamente. Como estimadores estos valores tienen un valor esperado, que es el que se emplea en el modelo, pero también una varianza del estimador, que indica lo bueno o malo que puede ser el valor estimado y el margen en el cual podría variar.

En la etapa de simulación muchas de las propiedades que se transmiten en las series generadas provienen de los parámetros del modelo, mientras la aleatoriedad es principalmente consecuencia de la serie de ruidos inicial. Estos ruidos deben ser independientes normales desde el punto de vista estadístico, sin embargo, en algunas ocasiones pueden generarse series de ruidos que no satisfacen dichas propiedades. Para evitar este tipo de conflictos puede ser conveniente revisar al inicio del proceso de simulación, si las series de ruidos satisfacen estas propiedades básicas y descartar las que no lo cumplan antes de proseguir con el modelo.

Un aspecto que no es abordado en el modelo propuesto corresponde al sesgo en la estimación de algunos parámetros del modelo, lo que puede ser importante al usarlos para la simulación. Existen en la literatura procedimientos establecidos para corregir el sesgo de las principales propiedades de modelos autoregresivos

En México falta mucha información acerca del tema y es poca la población interesada a desarrollar este tipo de modelos, existe la posibilidad de obtenerlos gracias al interés y al apoyo que países como Argentina, Chile y Colombia, entre otros, han mostrado a la problemática de México y sus posibles soluciones, se han tenido acercamientos con estos países a través de foros y seminarios. México marcha de manera lenta en este tema, pero puede aplicar los modelos mencionados en este trabajo y aplicados con gran éxito en cada país ya que tiene características similares, por lo que se esperaría gran éxito.

Apéndice I - Series de Tiempo

1. Introducción

Una serie de tiempo esta dado por un conjunto de observaciones que están ordenadas en el tiempo, y que estas pueden representar el cambio de una variable ya sea de tipo económica, física, química, biológica, etc, a lo largo esa historia.

El objetivo del análisis de una serie de tiempo es el conocimiento de su patrón de comportamiento, para así poder prever su evolución en el futuro cercano, suponiendo por supuesto que las condiciones no variarán significativamente.

Los pronosticos que se puedan realizar en base al análisis de este tipo de datos sirvan para el desarrollo de nuevos planes para inversiones en agricultura por ejemplo, elaboración de nuevos productos por parte de las empresas, prevención de desastres por cambios en el clima, o captar turistas para la ciudad, etc.

En el ejemplo desarrollado se ha realizado el análisis de datos por series de tiempo, para un conjunto de datos de la temperatura del medio ambiente. Los datos fueron recolectados durante los años 1999, 2000, 2001, 2002, 2003, 2004. Para el análisis solo se han considerado la temperaturas de los cuatro primeros meses de cada año, además debo indicar que se ha tomado valores promedios mensuales.

2. Teoria

2.1 Series de Tiempo

Una serie temporal o cronologica es un conjunto e observaciones de una variable, ordenadas segu transcurre el tiempo.En una serie de tiempo las observaciones no se deben ordenar de mayor a menor debidoa que se perderia el grueso de la informacion debido a que nos intersea detectar como se mueve la variable en el tiempo es muy importante respetar la secuencia temporal de las observaciones. 2.2 Representacion de una Serie Temporal

Par realizar la reprsenyacion de una serie ytemporal se debe realizae mediante una gráfica de disprsió n x-y como se muestra en la fig.1

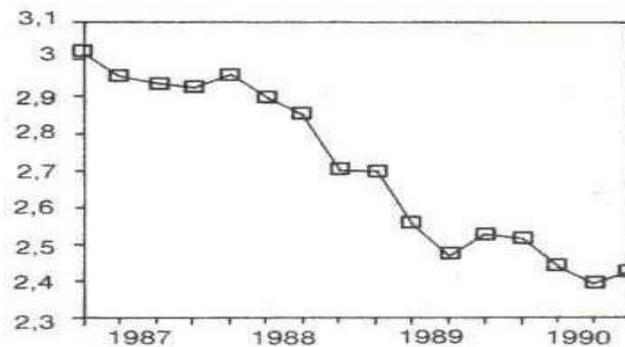


Fig.1. Representación de una serie temporal2.3 Componentes de una serie temporal

2.2 Tendencia

La tendencia es un movimiento de larga duración que muestra la evolucion general de la serie en el tiempo.

La tedencia es un movimiento que puede ser estacionario o ascendente, y su recorrido, una linea recta o una curva. Algunas de la posibles formas son las que se muestran en la fig.2

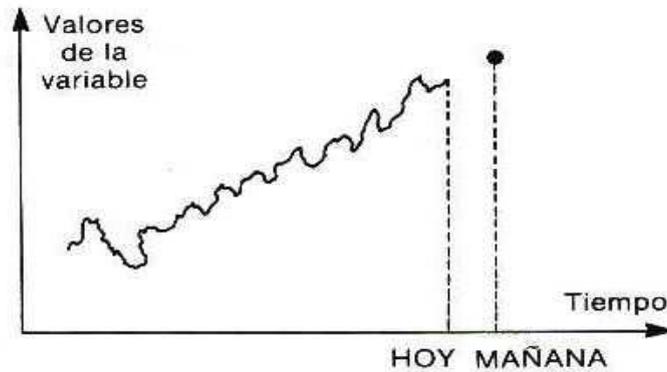


Fig.2. Representación de la tendencia

La tendencia es un movimiento que puede ser estacionario o ascendente o descendente como se indica en la fig.3

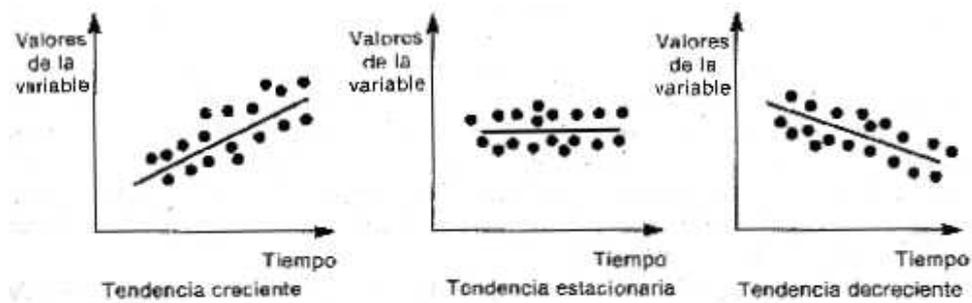


Fig. 3 Tendencias ascendente, estacionaria y descendente

También son posibles algunas formas para la tendencia, que no necesariamente tiene una distribución de puntos en forma aproximadamente lineal sino como las que se muestran en la fig. 4

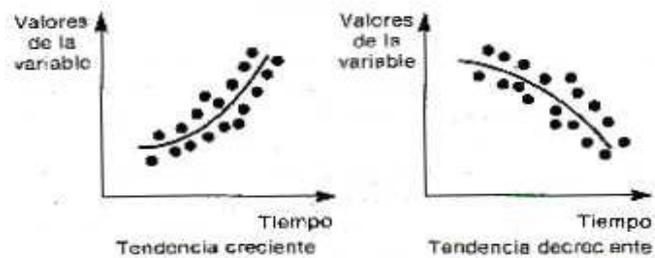


Fig.4 Líneas de tendencia de otras posibles formas.

2.3 Variaciones Estacionales.

Se habla de este tipo de variaciones usualmente cuando el comportamiento de la variable en el tiempo en un periodo esta relacionado con la época o un periodo particular, por lo general en el espacio cronologico presente.

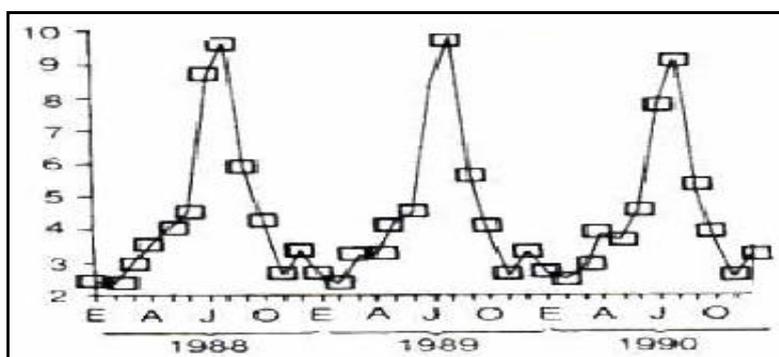


Fig. 5 Variaciones estacionales

2.3.3 Variaciones Cíclicas

Se llama así a las oscilaciones a lo largo de una tendencia con un periodo superior al año. El ciclo sugiere la idea de que este tipo de movimiento se repite cada cierto periodo con características parecidas. Los ejemplos más frecuentes se encuentran en el campo de las variables económicas, en estos casos se deben principalmente a la alternancia de las etapas de prosperidad y depresión en la actividad económica.

2.3.4 Variaciones Residuales

Cuando parecen hechos imprevistos, repentinos que afecten las variables en estudio acotando que no podemos prever nos hallamos frente a variaciones residuales provocadas por factores externos aleatorios.

Por ejemplo un día lluvioso y frío durante el verano es difícil de predecir y aunque perturbaría ciertas actividades diarias como la venta de helados no afectaría en este caso significativamente la serie.

3. Análisis de la Tendencia

En la práctica es difícil distinguir la tendencia del comportamiento cíclico. Por ejemplo la gráfica puede conducirnos a concluir que existe una tendencia ascendente en la parte de 1980 a 1982, pero esto es una parte de la serie de tiempo más grande.



Fig. 6 Tendencias cecrecientes, crecientes entre periodos de tiempo

3.1 Método Gráfico

Mediante este método muy elemental se determina la tendencia a partir de una representación gráfica de la serie. La aplicación de este método es como sigue

- Se representa gráficamente la serie cronológica
- Se unen los extremos superiores de la serie, se hace lo mismo con los inferiores
- Se obtiene dos líneas que encierran a la serie original
- Uniendo los puntos medios de las distancias entre las dos líneas o curvas se obtiene la tendencia. La línea o curva de tendencia obtenida tendrá un trazado mucho más suave que la serie original.

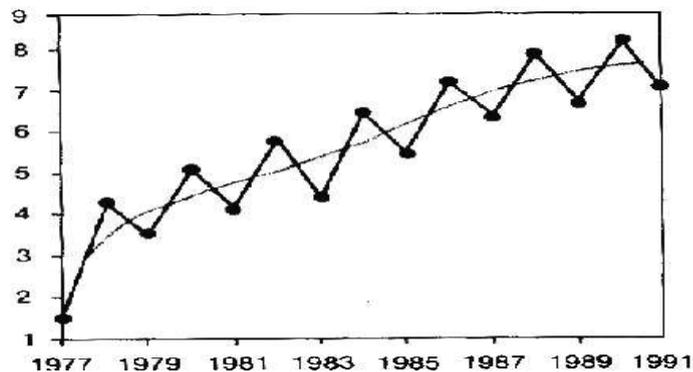


Fig. 7 Representación tendencia estacionaria

3.2 Método de las Medias Móviles

Para este método se deben de considerar los siguientes pasos que se detallan

- Observar con detenimiento la serie para determinar aproximadamente la fluctuación con periodo mas largo y llamamos q al numero de observaciones que *forman* una oscilación compleja.
- Se procede a calcular una serie de medias. La primera de ellas se calcula a partir de las q primeras observaciones de la serie pero eliminando la primera observación y añadiendo al instante posterior. Se prosigue así hasta calcular la media de las últimas q observaciones.
- Cada una de las medias obtenidas en el paso anterior se asigna al instante o momento central del periodo temporal que promedian.
- Uniendo las medias se obtiene la tendencia.

4. Aplicación

Caso 1: Producción de Motocicletas en una empresa japonesa, periodo 1974 - 1990

En la siguiente tabla se tiene la producción de motocicletas de una empresa (en millones de motos) en un periodo de 17 años que se muestra en la tabla N° 1

Tabla N°1
Venta de Motocicletas en un periodo de 17 años
(Producción en millones de motocicletas)

Años	Producción	Años	Producción	Años	Producción
	n		n		n
1974	2.1	1980	2.2	1986	2.1
1975	1.9	1981	2.0	1987	1.9
1976	1.7	1982	1.8	1988	1.5
1977	1.5	1983	1.7	1989	1.4
1978	1.6	1984	1.9	1990	2.5
1979	2.0	1985	2.4	----	-----

Se trasladan los datos a Microsoft Excel, ordenados en dos columnas, luego se realiza la gráfica de los datos.

Se obtiene la gráfica mostrada en la fig.8



Fig. 8 Representación de la serie de tiempo para las motocicletas por año

En la grafica se observa que los años donde se registra mayor producción son 1974, 1980, 1985,1990

Entonces podemos tomar cada cinco años como la cantidad de años para la cual la empresa realiza su mayor producción.

Sin embargo es conveniente encontrar una linea de tendencia tal que se pueda hallar una ecuación ajustada para los pronósticos de la producción en el tiempo.

Utilizando el método de la media móvil

Seconstruye una nueva tabla con las medias móviles

Esto es para suavizar la distribución de puntos

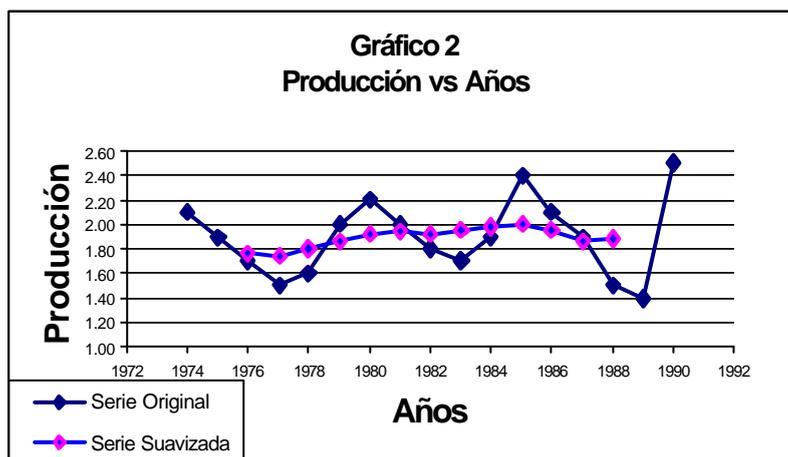


Fig. 9 Serie original y serie suavizada por los promedios móviles

Hallando la línea de tendencia

En Microsoft Excel, la línea de tendencia para la curva suavizada se obtiene fácilmente y se muestra en la Fig. 10

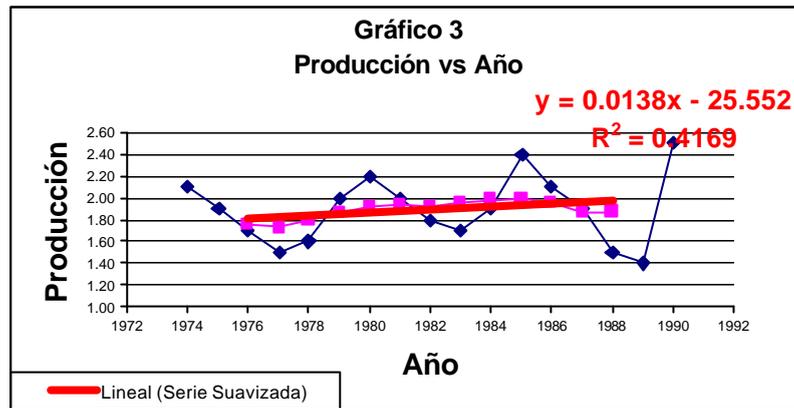


Fig. 10. Línea de tendencia con $R^2 = 0.4169$

El coeficiente de determinación es muy pequeño por lo que no se puede asegurar categóricamente que la ecuación lineal hallada es la que pronostica la producción en los años posteriores.

Será necesario realizar un segundo arreglo con medias móviles

El problema ahora es que el periodo donde alcanza la mayor producción es un número par de años, por lo que se hace difícil en la tabla hallar el año central, realizando el promedio de

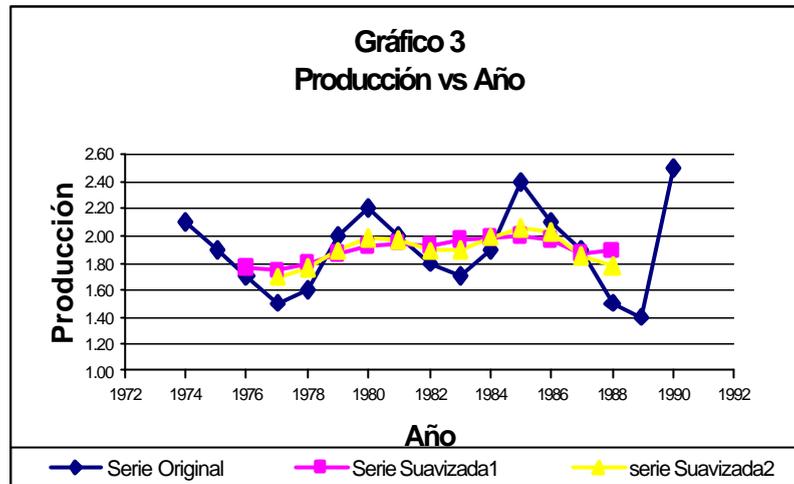


Fig.11 Suavizando la línea de tendencia por segunda vez

La fig. 11 muestra la segunda suavizada de la línea de tendencia, no ha variado mucho con respecto a la primera,.

Caso 2: Temperatura en Lima – Aeropuerto Internacional Jorge Chávez, periodo 2000- 2004

En la ciudad de Lima (Perú) el el Aeropuerto Internacional Jorge Chávez, las tempraturas registradas durante los años 2000, 2001, 2002, 2003, 2004 consideramos en este caso solo los primeros cuatro meses de cada año, las temperaturas registradas por cada mes promediados son las que se muestran en la tabla N°1

Tabla N° 2
Temperaturas de Lima – Aeropuerto Internacional Jorge Chávez (Lima –Perú)
2000 – 2004 ¹

	2000 T (°C)	2001 T (°C)	2002 T (°C)	2003 T (°C)	2004 T (°C)
Enero	21.835	21.694	21.132	22.257	20.443
Febrero	21.835	23.070	22.654	23.286	22.959
Marzo	21.113	22.181	22.654	22.053	21.887
Abril	21.113	20.440	21.270	19.340	20.443

Representación gráfica

Primero se organizan los datos de manera conveniente en la hoja de cálculo excel, Se obtiene la siguiente representación de los datos

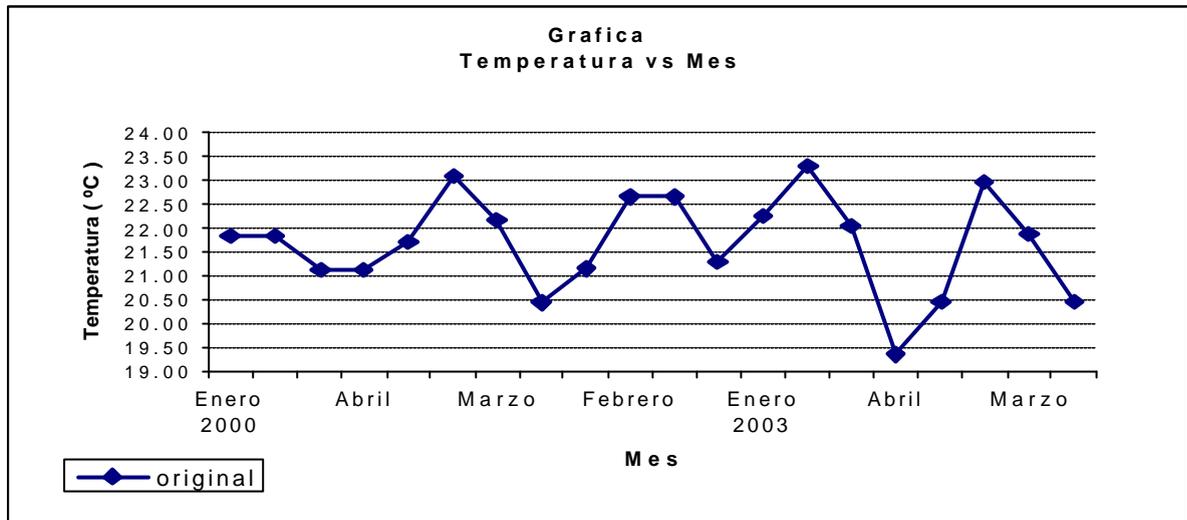


Fig. 12 Representación gráfica de los fatos de temperatura por cuatrimestre

CONCLUSIONES

- Las series temporales pueden servir para predecir acontecimientos futuros en base a ciertos comportamientos de determinadas variables
- Si tenemos mas observaciones que se puedan promediar, que ees el orden de la media móvil, se obtienen tendencias mas suaves. Este hecho no debe hacerno dvidadr que asunque hemos mejorado la tendencia con el suavizado, por el contrario perdemos información sobre los valores iniciales y finales de la tendencia estimada.
- Con el procedimientmo de medias móviles siempre es posible elegir el número de observaciones que se deben tomar para el promedio, esto no simepre es fácil, esto da el periodo de osilación
- Si se determina la función matemática de la tendencia lineal, esta no permitirá conocer los valores perdidos tanto al inicio como al final del proceso de busqueda de la línea detendencia,

Apéndice II.-Modelos ARIMA

1. Introducción

En 1970, Box y Jenkins desarrollaron un cuerpo metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega un papel fundamental. Una parte importante de esta metodología está pensada para liberar al investigador econométra de la tarea de especificación de los modelos dejando que los propios datos temporales de la variable a estudiar nos indiquen las características de la estructura probabilística subyacente. En parte, los procedimientos que vamos a analizar se contraponen a la "forma tradicional" de identificar y especificar un modelo apoyándonos en las teorías subyacentes al fenómeno analizado aunque, convenientemente utilizados, los conceptos y procedimientos que examinaremos constituyen una herramienta útil para ampliar y complementar los conocimientos econométricos básicos.

Se comenzará analizando los modelos en los que una variable es explicada utilizando exclusivamente una "exógena": su propio pasado. Podemos decir que la consideración exclusiva de los valores pasados de una determinada variable para explicar su evolución presente y futura supone, al mismo tiempo, una ventaja y un inconveniente:

- la ventaja radica en el hecho de no necesitar distintas series de datos (distintas variables) referidas al mismo período de tiempo (característica común a todos los modelos univariantes) y, al mismo tiempo, ahorrarnos la

- identificación y especificación del modelo en el sentido de la econometría tradicional,
- el inconveniente es que, al renunciar a la inclusión de un conjunto más amplio de variables explicativas, no atendemos a las relaciones que sin duda existen entre casi todas las variables económicas perdiendo capacidad de análisis al tiempo que renunciamos, implícitamente, al estudio teórico previo del fenómeno y a su indudable utilidad.

Dentro de estos modelos univariantes se desarrollarán suficientemente los conocidos con el nombre de ARIMA. Posteriormente se complementará esta perspectiva univariante añadiéndose a la especificación una o más variables exógenas al modelo "tradicional" aproximándonos al estudio de los conocidos como modelos de transferencia.

Como es habitual en economía, definiremos una estructura que nos permita, por sus características, cumplir el fin último de predicción: proceso estocástico estacionario. Diremos cuales son las condiciones que ha de cumplir esta función para que podamos calcularla y definiremos el proceso estocástico estacionario lineal y discreto. Posteriormente, analizaremos los modelos más simples (que emplean menos retardos) conforme a una serie de funciones características (covarianza, autocorrelación total y autocorrelación parcial), describiendo sus condiciones y planteando estructuras teóricas que luego puedan ser identificables con series temporales reales.

2 Definición y Conceptos Básicos de los Modelos ARIMA

2.1 Proceso Estocástico y Estacionalidad

Los modelos autoregresivos o de medias móviles que más tarde conceptualizaremos necesitan para su comprensión de la introducción del concepto de proceso estocástico.

Un proceso estocástico es una sucesión de variables aleatorias y_t , ordenadas, pudiendo tomar t cualquier valor entre $-\infty$ y ∞ . Por ejemplo, la siguiente sucesión de variables aleatorias puede ser considerada como proceso estocástico:

$$y_{-5}, y_{-4}, y_{-3}, y_{-2}, \dots, y_3, y_4$$

El subíndice t no tiene, en principio, ninguna interpretación a priori, aunque si hablamos de procesos estocásticos en el contexto del análisis de series temporales este subíndice representará el paso del tiempo.

Cada una de las variables Y_t que configuran un proceso estocástico tendrá su propia función de distribución con sus correspondientes momentos. Así mismo, cada par de esas variables tendrán su correspondiente función de distribución conjunta y sus funciones de distribución marginales. Esto mismo ocurrirá, ya no para cada par de variables, sino para conjuntos más amplios de las mismas. De esta forma, para caracterizar un proceso estocástico deberíamos especificar las funciones de distribución conjunta de cualquier conjunto de variables:

$$(Y_{t_1}, Y_{t_2}, Y_{t_3}, \dots, Y_{t_m})$$

cualesquiera que fueran los valores de (t_1, t_2, \dots, t_m) y cualquiera que fuera el valor de m ; por ejemplo:

$$y_1, y_2, y_3 \quad (t_1 = 1, y, m = 3)$$

$$Y_3, y_4, y_5, y_6 \quad (t_1 = 3, y, m = 4)$$

Habitualmente, conocer esas funciones de distribución resulta complejo de forma que, para caracterizar un proceso estocástico, basta con especificar la media y la varianza para cada y_t y la covarianza para variables referidas a distintos valores de t :

$$E[y_t] = m_t$$

$$s_t^2 = Var(y_t) = E[(y_t - m_t)^2]$$

$$g_t = Cov(Y_y, Y_s) = E[(y_t - m_t)(y_s - m_s)]$$

Las distribuciones de probabilidad podrían no estar completamente caracterizadas en algunas de las variables, los momentos podrían no coincidir incluso no existir para alguna de las variables aleatorias, lo mismo puede ocurrir con las distribuciones conjuntas o marginales. Sin embargo, de todos los tipos de procesos estocásticos posibles, nos interesan especialmente dos de ellos a los que la estadística ha dado nombres precisos:

- ruido blanco es una sucesión de variables aleatorias (proceso estocástico) con esperanza (media) cero, varianza constante e independientes para distintos valores de t (covarianza nula).
- proceso estocástico estacionario.

Decimos que un proceso estocástico es estacionario si las funciones de distribución conjuntas

son invariantes con respecto a un desplazamiento en el tiempo (variación de t). Es decir, considerando que $t, t+1, t+2, \dots, t+k$ reflejan períodos sucesivos:

$$F(Y_T, Y_{T+1}, \dots, Y_{T+k}) = F(Y_{t+m}, Y_{t+1+m}, \dots, Y_{t+k+m})$$

para cualquier t, k y m ; por ejemplo:

$$F(y_1, y_2, \dots, y_6) = F(y_{10}, Y_{11}, \dots, Y_{15})$$

Donde $t=1, k=5, m=9$

$$F(y_3, Y_4, Y_5) = F(y_7, Y_8, Y_9)$$

Donde $t=3, k=2, m=4$

Esta definición de estacionalidad se conoce como estacionalidad en sentido estricto o fuerte y puede relajarse sustancialmente utilizando la denominada estacionalidad en sentido amplio o débil. Decimos que un proceso estocástico es débilmente estacionario si:

- Las esperanzas matemáticas de las variables aleatorias no dependen del tiempo, son constantes:

$$E[Y_t] = E[Y_{t+m}] \quad \forall m$$

- Las varianzas tampoco dependen del tiempo y son finitas:

$$\text{Var}[Y_t] = \text{Var}[Y_{t+m}] \neq \infty \quad \forall m$$

- Las covarianzas entre dos variables aleatorias del proceso correspondientes a períodos distintos de tiempo (distintos valores de t) solamente dependen del lapso de tiempo transcurrido entre ellas:

$$\text{Cov}(Y_t, Y_s) = \text{Cov}(Y_{t+m}, Y_{s+m}) \quad \forall m$$

De esta última condición se desprende que, si un fenómeno es estacionario, sus variables pueden estar relacionadas linealmente entre si, pero de forma que la relación entre dos variables sólo depende de la distancia temporal k transcurrida entre ellas.

Lógicamente, la estacionalidad en sentido estricto garantiza la estacionalidad en sentido amplio pero no al revés.

Una vez introducido el concepto genérico de proceso estocástico puede decirse que una serie temporal cualquiera es, en realidad, una muestra, una realización concreta con unos valores concretos de un proceso estocástico teórico, real. El análisis de series que vamos a estudiar tratará, a partir de los datos de una serie temporal, inferir las características de la estructura probabilística subyacente, del verdadero proceso estocástico.

2.2 Modelos Autoregresivos

La palabra ARIMA significa Modelos Autoregresivos Integrados de Medias Móviles. Definimos un modelo como autoregresivo si la variable endógena de un período t es explicada por las observaciones de ella misma correspondientes a períodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda Y_t puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) más un término de error (innovación).

Los modelos autoregresivos se abrevian con la palabra AR tras la que se indica el orden del modelo: AR(1), AR(2),....etc. El orden del modelo expresa el número de observaciones retasadas de las series temporal analizada que intervienen en la ecuación. Así, por ejemplo, un modelo AR(1) tendría la siguiente expresión:

$$Y_t = f_0 + f_1 Y_{t-1} + a_t$$

El término de error de los modelos de este tipo se denomina generalmente ruido blanco cuando cumple las tres hipótesis básicas tradicionales mencionadas al principio del texto:

- media nula
- varianza constante
- covarianza nula entre errores correspondientes a observaciones diferentes
-

La expresión genérica de un modelo autoregresivo, no ya de un AR(1) sino de un AR(p) sería la siguiente:

$$Y_t = f_0 + f_1 Y_{t-1} + f_2 Y_{t-2} + \dots + f_p Y_{t-p} + a_t$$

pudiéndose escribir de forma abreviada como:

$$f_p(L) Y_t = f_0 a_t$$

donde $f_p(L)$ es lo que se conoce como operador polinomial de retardos:

$$f_p(L) = 1 - f_1 L - f_2 L^2 - \dots - f_p L^p$$

y donde, a su vez, el término L es lo que se conoce como operador retardo tal que, aplicado al valor de una variable en t , dé como resultado el valor de esa misma variable en $t-1$:

$$LY_t = Y_{t-1}$$

y aplicado sucesivamente p veces retarda el valor en p períodos

$$L^p Y_t = Y_{t-p}$$

Normalmente, se suele trabajar con modelos autoregresivos de órdenes bajos: AR(1) o AR(2), o bien con órdenes coincidentes con la periodicidad de los datos de la serie analizada (si es trimestral AR(4), si es mensual AR(12)....)

2.4 Modelo de Medias Móviles

Un modelo de los denominados de medias móviles es aquel que explica el valor de una determinada variable en un período t en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, ponderados convenientemente. Estos modelos se denotan normalmente con las siglas MA, seguidos, como en el caso de los modelos autoregresivos, del orden entre paréntesis. Así, un modelo con q términos de error MA(q) respondería a la siguiente expresión:

$$Y_t = m + a_t + q_1 a_{t-1} + q_2 a_{t-2} + \dots + q_q a_{t-q}$$

que de nuevo puede abreviarse utilizando el polinomio de retardos (como en caso de los modelos AR):

$$Y_t = q_q(L) a_t + m$$

Al igual que en el caso de los modelos autoregresivos, el orden de los modelos de medias móviles suele ser bajo MA(1), MA(2) o corresponderse con la periodicidad de los datos analizados MA(4), para series trimestrales, o MA(12) para series mensuales.

2.5 Interpretación de un Modelo de Medias Móviles

Así como un modelo autoregresivo es intuitivamente sencillo de comprender, la formulación de un modelo de medias móviles resulta sorprendente para el no iniciado. ¿Qué significa que una variable aleatoria se explique en función de los errores cometidos en períodos precedentes?, ¿De dónde proceden esos errores?, ¿Cuál es la justificación de un modelo de este tipo?.

En realidad, un modelo de medias móviles puede obtenerse a partir de un modelo autoregresivo sin más que realizar sucesivas sustituciones.

Efectivamente, un modelo AR(1), sin término independiente, tiene la expresión:

$$Y_t = fY_{t-1} + a_t$$

si consideramos $t-1$ en lugar de t el modelo sería en este caso: y sustituyendo queda:

$$Y_{t-1} = fY_{t-2} + a_{t-1}$$

si ahora sustituimos y_{t-2} por su expresión autoregresiva y así sucesivamente llegamos a un modelo del tipo:

$$Y_t = a_t + f a_{t-1} + f^2 a_{t-2} + f^3 a_{t-3} + \dots + f^j a_{t-j} + \dots$$

que es la expresión, sin término independiente, de un modelo de medias móviles como el planteado anteriormente. En realidad, de forma estricta, el paso de un modelo a otro debería realizarse al contrario (de un MA a un AR) utilizando el teorema general de descomposición de Wold.

2.6 Condiciones y Raíces Unitarias para los Modelos AR y MA

Hemos dicho anteriormente que, bajo condiciones generales, todo proceso estocástico estacionario se prestaba a una especificación tipo AR(p) y en consecuencia podía expresarse también como un MA(q). Es ahora el momento de especificar lo que antes hemos llamado "condiciones generales" y examinar en que casos es posible la realización de un proceso AR ó MA para representar un proceso estocástico estacionario.

Para que un proceso estocástico estacionario admita una formulación donde deben cumplirse dos condiciones accesorias:

- el proceso no debe ser anticipante (*hipótesis de recursividad temporal*); lo que quiere decir que los valores de una variable en un momento t no dependerán de los que esta misma tome en $t+j$, siendo jota cualquier valor superior a cero.
- el proceso ha de ser invertible; lo que supone que la correlación entre una variable y su pasado va reduciéndose a medida que nos alejamos más en el tiempo del momento para el que estamos considerando dicha correlación (*proceso ergódico*). La explicación intuitiva de esta situación derivaría de que si el especificáramos una variable en función de ciertos coeficientes que nos determinen su correlación con los valores pasados de ella misma,

los valores de dichos coeficientes deberían ser necesariamente inferiores a uno, porque sino el proceso de infinitos números sería "explosivo".

•

2.7 La Estacionalidad de las Series Temporales en la Realidad

Queda clara que la aproximación a los procesos estocásticos con modelos AR o MA está restringida, en términos generales, a aquellos procesos estocásticos que cumplan, al menos de forma débil, la restricción de estacionalidad. Cuando, en la realidad, queremos inferir a partir de una serie temporal (muestra) la estructura del proceso estocástico mediante modelos AR ó MA, debemos cubrir dos etapas:

- asegurarnos de que la serie temporal, como muestra del proceso estocástico, es estacionaria y, si no lo es,
- transformar la serie temporal original de forma que la nueva serie transformada si lo sea.

•

- a) ¿cómo verificamos si la serie a analizar es estacionaria en media? ¿cómo lograr que lo sea? Filtrado de la serie original

Para resolver la primera cuestión existen diversos métodos de aproximación y, de entre ellos, destacamos Podríamos subdividir la serie temporal en varios períodos de, aproximadamente, la misma longitud, y calcular su media. El proceso sería estacionario en el caso en que dichos estadísticos fueran prácticamente iguales para todos los subperíodos analizados. En la mayoría de los casos, un simple gráfico sirve para observar si existe o no una clara tendencia y, por tanto, si la serie es estacionaria o no.

Habitualmente, cuando una serie muestra tendencia, se subdivide dicha serie en dos componentes: una primera, la estimación de dicha tendencia, y, la segunda, el residuo o error que se comete cuando se utiliza dicha tendencia como valor estimado de la serie original.

$$y_t = T_t + r_t$$

Una vez estimada la tendencia, aproximada con una regresión lineal, parabólica, exponencial que sea más conveniente; trabajaremos con la serie del residuo, que entonces no mostrara tendencia y podremos decir que es estacionaria en media. Es sobre este residuo sobre el que llevaremos a cabo todo el proceso descrito como metodología de identificación ARIMA, sumando finalmente el valor de la tendencia estimada si queremos dar resultados de estimación de la serie original. Es decir:

1.- la identificación del proceso ARIMA se hará sobre esta serie del residuo $\hat{r}_t = y_t - \hat{T}_t$, estimada previamente la tendencia del modo más adecuado.

2.- Para obtener valores estimados de la serie original, sumaremos el componente tendencial al valor estimado del residuo mediante el modelo ARIMA $\hat{y}_t = \hat{T}_t + \hat{r}_t$

A este procedimiento se le conoce con el nombre de filtrado de la tendencia de la serie. Por supuesto, existen muy variadas formas de aplicar un filtro, siendo la que aquí hemos enunciado la más sencilla.

b) ¿Cómo se comprueba si una serie es estacionaria en varianza? Orden de integración

Sin duda alguna, el test más habitual a la hora de determinar la estacionalidad de una serie temporal, consiste en la aplicación del conocido como test de Dickey-Fuller (Test DF) o Dickey-Fuller Ampliado (Test ADF). Éste es un contraste de “No estacionalidad” ya que la hipótesis nula es precisamente la presencia de una raíz unitaria en el proceso generador de datos de la serie analizada.

Supongamos inicialmente, como modelo de partida para el análisis de una determinada serie y_t , el de un proceso estacionario autoregresivo de orden uno:

$$y_t = a_1 y_{t-1} + e_t \quad (\text{Ec. 1})$$

frente a este modelo se plantea, como hipótesis nula H_0 , el modelo alternativo de un paseo aleatorio no estacionario del tipo¹:

$$y_t = y_{t-1} + e_t$$

se trata por tanto de contrastar si el coeficiente a_1 es igual a la unidad o distinto de ella.

Sin embargo, para contrastar la nulidad del coeficiente a_1 , no podemos utilizar el contraste “t” habitual sobre una estimación por MCO del primer modelo. La razón es que la hipótesis nula que habitualmente se contrasta y, a partir de la cual se deriva la expresión y propiedades del test “t”, es la de nulidad del parámetro ($a_1=0$) de la (Ec.2), sin embargo, en nuestro caso, necesitaríamos contrastar $H_0: a_1 = 1$. Si la hipótesis nula fuera cierta, la varianza de y_t no sería estacionaria

¹ $y_t = y_{t-1} + e_t \rightarrow \Delta y_t = e_t$ en un proceso de este tipo, sustituyendo recursivamente se obtiene:

$$y_t = y_0 + \sum_{i=1}^t e_i, \text{ proceso con media constante } E[y_t] = E\left[y_0 + \sum_{i=1}^t e_i\right] = E[y_0] = y_0 \text{ y varianza estocástica}$$

$$V[y_t] + E\left[y_t - E[y_t]\right]^2 = E\left[y_0 + \sum_{i=1}^t e_i - y_0\right]^2 = E\left[\sum_{i=1}^t e_i\right]^2 = E[e_1^2 + e_2^2 + \dots + e_1 e_2 + e_1 e_3 + \dots] = E[e_1^2 + e_2^2 + \dots + e_t^2] = t = s_e^2$$

sino que crecería con los valores de “ t ” según la expresión de la varianza de un paseo aleatorio con deriva:

$$\text{Var}(y_t) = t \sigma_e^2$$

La estimación de a_1 en $y_t = a_1 y_{t-1} + e_t$ será siempre consistente sin embargo, su distribución variará según los valores que tome la estimación. Utilizando las palabras de Novales, la distribución de probabilidad asintótica del estimador de MCO del modelo AR(1) presenta una “discontinuidad” cuando $a_1 = 1$ y, como sustituto, deberán utilizarse las distribuciones derivadas de forma empírica mediante un procedimiento de Montecarlo realizado por Dickey. Más recientemente, MacKinnon realizó un número mayor de simulaciones que las tabuladas por Dickey y Fuller. Además, MacKinnon estimó la superficie de respuesta usando los resultados de la simulación, lo que permite calcular los valores críticos del test DF para cualquier tamaño muestral y cualquier número de variables en el lado derecho de la ecuación.

En la práctica, por cuestiones de sencillez operativa, el modelo utilizado para el contraste DF no es el expuesto al comienzo del epígrafe sino otro, equivalente al anterior, que se obtiene restando a uno y otro lado el término y_{t-1} .

$$\begin{aligned} y_t - y_{t-1} &= a_0 + a_1 y_{t-1} - y_{t-1} + e_t \\ \Delta y_t &= a_0 + (a_1 - 1) y_{t-1} + e_t \\ \Delta y_t &= a_0 + g \cdot y_{t-1} + e_t \end{aligned} \quad (\text{Ec. 2})$$

por tanto, la hipótesis nula inicial para la (Ec. 2), se transforma ahora en $H_0: g = 0$ frente a $H_1: g < 0$. Decir que g es nulo es lo mismo que decir que $a_1 = 1$, o sea, que existe una raíz unitaria, decir que es menor que cero equivale a decir que a_1 es menor que la unidad (proceso autoregresivo estacionario)².

El procedimiento básico para la aplicación simple del test DF es, a partir de aquí, aparentemente sencillo. Se estima el modelo propuesto y se calcula el valor estimado de la “ t ” del parámetro analizado. Una vez calculado se compara con el valor empírico de referencia obtenido con las tablas de Dickey y Fuller o de MacKinnon. Si el valor estimado para g es inferior en valor absoluto al tabulado dado un determinado nivel de confianza, admitiremos la hipótesis nula, o sea, la presencia de una raíz unitaria.

El modelo expuesto hasta el momento es el más simple posible, pero cabe que el modelo más adecuado a la realidad incluya otros términos, como una constante y/o una tendencia. Dolado *et al.* y Perron propusieron, entre otros autores, seguir

² No se considera el caso de procesos autoregresivos explosivos en que $a_1 > 1$

un proceso en etapas a fin de garantizar el éxito en la elección del modelo de referencia en el mayor número de ocasiones:

- En primer lugar se estimaría el modelo menos restringido (con término constante y tendencia determinista).
- Dado que el principal error de esta táctica inicial consistiría en la escasa potencia del contraste para el rechazo de la hipótesis nula por inclusión de variables irrelevantes, si los valores críticos indican rechazo (ausencia de raíz unitaria), terminaríamos el procedimiento.
- En el caso de no rechazarse la hipótesis nula de presencia de una raíz unitaria, es decir, en el caso en que admitamos la presencia de una raíz unitaria ($H_0: g=0$) pasaríamos ahora a examinar la significatividad del parámetro tendencial determinista a_2 . Dado que, en este punto, estaríamos bajo la hipótesis ya admitida de que $g=0$, utilizaríamos el valor de referencia de $t b_t$ incluso, para mayor seguridad, también el contraste conjunto $f_3(a_2=g=0)$.
- Si el término tendencial resulta significativo ($a_2 \neq 0$) contrastaremos de nuevo la presencia de una raíz unitaria ($H_0: g=0$) pero utilizando entonces las tablas de una normal estandarizada. Sea cual sea el resultado del test con las nuevas tablas finalizaríamos aquí el contraste admitiendo o rechazando la presencia de una raíz unitaria.
- Si el término tendencial es no significativo, deberá replantearse el modelo inicialmente estimado pasándose a examinar otro con término constante pero sin esta tendencia determinista. Con este modelo se vuelve a analizar la presencia de una raíz unitaria ($g=0$).
- En el caso en que, nuevamente, se sostenga la presencia de una raíz unitaria, se contrastará entonces la adecuación del término independiente a_0 bien con el contraste t_{am} , bien con f_1 . Si el término independiente resulta significativo usamos de nuevo las tablas de una normal para contrastar la presencia de la raíz unitaria, concluyendo de nuevo aquí el contraste.
- Sólo si entonces la constante a_0 es no significativa se utiliza el modelo más simple como modelo de referencia contrastándose, de nuevo, la presencia de raíz unitaria. En este caso, no tiene cabida el uso de la distribución normal estandarizada.
-

Está claro que lo expuesto hasta este momento permite contrastar la presencia de una o más raíces unitarias en una determinada serie temporal para la que se supone un proceso AR(1). Sin embargo, muchas serie temporales se ajustan más adecuadamente a procesos autorregresivos de orden superior AR(2) o AR(3). No parece, por tanto, muy correcto, contrastar la presencia de una o más raíces unitarias utilizando siempre la estructura de un modelo AR(1) ya que las raíces unitarias pueden aparecer también en estructuras más complejas. Este problema

da lugar a lo que se conoce como test de raíces unitarias de Dickey-Fuller Ampliado (DFA): si se quiere contrastar la presencia de una raíz unitaria en una serie que sigue un proceso AR(p), deberá aplicarse el procedimiento expuesto para el caso simple AR(1), pero suponiendo ahora del modelo:

$$\Delta y_t = a_0 + \mathbf{g} \cdot y_{t-1} + \sum_{i=2}^p \mathbf{b} \Delta y_{t-i} + \mathbf{e}_t$$

Donde:

$$\mathbf{g} = - \left(1 - \sum_{i=1}^p a_i \right)$$

Y:

$$\mathbf{b}_i = \sum_{j=1}^p a_j$$

3 Modelo ARIMA(p,d,q) SARIMA(P,D,Q)

En su forma más general el modelo ARIMA(p,d,q) ARIMA(P,D,Q,)S podría escribirse como:

$$Y_T = \mathbf{j}_1 Y_{T-1} + \mathbf{j}_2 Y_{T-2} + \dots + \mathbf{j}_{P+D_s+d} Y_{T-P-s-p-D_s-d} + \mathbf{d} + U_T + \mathbf{q}_1 U_{T-1} + \dots + \mathbf{q}_{Q_s+q} U_{T-s-Q-q}$$

Entendiendo que puede haber más de un proceso generador de la serie (en la parte regular y en la estacional) y escribiendo una combinación de los modelos MA(q) y AR(p) que han precisado de una serie de diferenciaciones "d" en la parte regular o "D" en la parte estacional para que fueran estacionarios.

3.1 Funciones de un Proceso Estocástico Estacionario

Definido un proceso estocástico como estacionario (ya sea de forma débil o fuerte), ya se ha comentado que si cumple las condiciones en sentido estricto, también cumple las condiciones en sentido débil. Siendo así, el proceso estaría perfectamente definido si conociéramos su media constante (\mathbf{m}), su varianza constante (\mathbf{s}) y la covarianza entre cada par de momentos diferentes en el tiempo.

Dicho ésto:

- La función de autocovarianza vendrá definida por los distintos valores que tomaría dicha covarianza cuando cambiamos el lapso temporal entre las observaciones de la serie que manejamos.

Análiticamente, se podría expresar como:

$$\mathbf{g}_j = Cov(y_t, y_{t-j}) = E[(y_t - \mathbf{m})(y_{t-j} - \mathbf{m})]$$

donde, evidentemente, cuando el valor de "j" es cero, tendríamos la varianza de la función:

$$\mathbf{g}_0 = Cov(y_t, y_{t-0}) = E(y_t - \mathbf{m})^2 = \mathbf{s}^2$$

- La función de autocorrelación se define igualmente como:

$$\mathbf{r}_j = \frac{cov(y_t, y_{t-j})}{\sqrt{var(y_t)}\sqrt{var(y_{t-j})}}$$

como nos encontramos ante un proceso definido como estacionario, la varianza es constante, por lo que podemos escribir:

$$\mathbf{r}_j = \frac{\mathbf{g}_j}{\mathbf{g}_0}$$

La función de autocorrelación de un proceso estocástico estacionario manifiesta las siguientes propiedades:

1.- $\mathbf{g}_0 = 1$

2.- $|\mathbf{g}_j| \leq 1$, ya que $|\mathbf{g}_j| \leq \mathbf{g}_0$

Esto asegura que el comportamiento de la función no sea explosivo.

3.- $\mathbf{g}_j = \mathbf{g}_{-j}$ (simetría)

dado que si un proceso es estacionario, la covarianza de dos variables aleatorias separadas por el mismo lapso de tiempo es la misma.

4.- $\lim_{j \rightarrow \infty} \mathbf{r}_j = 0.25$ (proceso ergódico)

Esta última propiedad, que define al proceso como "ergódico", es la que posibilita inferir valores de una serie en función de la información que sobre ella nos da su propio pasado, logrando estimadores consistentes. Sólo si se da esta propiedad, la pérdida de información al no considerar la influencia de los infinitos valores obtenidos en el pasado es cada vez más escasa e, incluirlos, añadiría poca información para la definición del proceso generador de datos que se intenta reproducir para aplicar al futuro.

Además de las dos funciones anteriores, se puede definir una tercera conocida como función de autocorrelación parcial, con el fin de tener en cuenta los valores de correlación entre dos variables aleatorias separadas entre sí "j" períodos y en función de los valores intermedios entre ellas. Es decir:

$$\Pi_j = \text{corr}\left(y_t, y_{t-j} \mid y_{t-1}, y_{t-2}, \dots, y_{t-j+1}\right)$$

Si planteamos las mejores predicciones de y_t e y_{t-j} como los resultantes de plantear Mínimos Cuadrados Ordinarios (MCO) siendo el primero de ellos del siguiente modo:

$$\hat{y}_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_{j-1} y_{t-j+1}$$

Se puede escribir la función de autocorrelación parcial, si la media es nula, como:

$$\Pi_j = \frac{\text{cov}\left[(y_t - \hat{y}_t)(y_{t-j} - \hat{y}_{t-j})\right]}{\sqrt{\text{var}(y_t - \hat{y}_t)}\sqrt{\text{var}(y_{t-j} - \hat{y}_{t-j})}}$$

Pudiendo demostrarse que:

$$\Pi_j = \frac{r_j - a_1 r_{j-1} - a_2 r_{j-2} - \dots - a_{j-1} r_1}{1 - a_1 r_1 - a_2 r_2 - \dots - a_{j-1} r_{j-1}}$$

3.2 Aplicación de estas Funciones a Muestras Concretas

En este apartado se pretender especificar estimaciones de los valores que caracterizan el proceso estacionario del tipo que se está describiendo ya no para el proceso estocástico general, sino para una manifestación concreta de éste traslucida en una serie temporal. Habrá que estimar la media (m), para lo que usaremos la media muestral; la varianza (g_0) y la función de autocovarianza (g_j), para lo que emplearemos la fórmula de la covarianza muestral) y la función de autocorrelación (r_j).

• Media muestral.

Como ya se ha dicho, el estimador $m = E(y_t)$ será la media muestral:

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

dicho estimador cumplirá dos propiedades:

a) Insesgadez.- la esperanza de la media de la serie será igual a m

Esto se demuestra ya que:

$$E(\bar{y}) = \frac{1}{T} \sum_{t=1}^T e(y_t) = \frac{1}{T} * T * m = m$$

b) Consistencia.- es decir, la varianza se anula cuando ampliamos la muestra a la población siendo el estimador insesgado, propiedad que se cumplirá siempre que se dé la siguiente condición, que no desarrollamos:

- Función de autocovarianza muestral.

El estimador de g_j se obtendrá según la siguiente expresión:

$$C_j = \frac{1}{T-j} \sum_{t=j+1}^T (y_t - \bar{y})(y_{t-j} - \bar{y})$$

que, a pesar de ser sesgado, dicho sesgo será determinable y cada vez más reducido según se aumente la muestra.

- Función de autocorrelación muestral.

Para su cálculo se recurrirá al cociente de funciones de autocovarianza del siguiente modo:

$$r_j = \frac{g_j}{g_0} = \frac{C_j}{C_0}$$

- El estimador para la función de autocorrelación parcial a emplear se calculará según el método recursivo de Durbin del siguiente modo:

$$\hat{f}_{11} = r_1$$

$$\hat{f}_{j+1,j+1} = \frac{r_{j+1} - \sum_{i=1}^j \hat{f}_{ji} r_{j+1-i}}{1 - \sum_{i=1}^j \hat{f}_{ji} r_i}$$

$$\hat{f}_{ji}, i = \hat{f}_{ji} - \hat{f}_{j+1,j+1} \hat{f}_{j,j+1-i} / i = 1 \dots j$$

Empleando estas fórmulas para los primeros casos, podremos escribir:

$$\hat{f}_{11} = r_1$$

$$\hat{f}_{22} = \frac{r_2 - \hat{f}_{11}r_1}{1 - \hat{f}_{11}r_1}$$

$$f_{21} = \hat{f}_{11} - \hat{f}_{22}r_1$$

$$\hat{f}_{33} = \frac{r_3 - \hat{f}_{21}r_2 - \hat{f}_{22}r_1}{1 - \hat{f}_{21}r_1 - \hat{f}_{22}r_2}$$

$$\hat{f}_{31} = \hat{f}_{21} - \hat{f}_{33}\hat{f}_{22}$$

$$\hat{f}_{32} = \hat{f}_{22} - \hat{f}_{33}\hat{f}_{21}$$

4. Proceso Estocástico Estacionario Lineal Discreto

Vamos a definir un nuevo caso especial de un proceso estocástico que nos permita luego intentar encontrar algo parecido en la realidad y que sea fácilmente identificable.

Se conoce por proceso estocástico estacionario lineal y discreto a aquel que puede expresarse como:

$$y_t = \mathbf{m} + a_t + \mathbf{q}_1 a_{t-1} + \mathbf{q}_2 a_{t-2} + \dots$$

donde:

- es lineal porque puede escribirse como combinación lineal de los errores.
- a_t es ruido blanco (esperanza y covarianza nulas y varianza constante).
- es discreto porque los lapsos temporales considerados son uniformes (no hay saltos temporales distintos entre las variables consideradas).

El siguiente paso será especificar las condiciones que nos aseguran que este proceso es estacionario, es decir que tiene media y varianza constantes y que su covarianza sólo varía cuando lo hace el espacio temporal que separa las observaciones empleadas para calcularla. Para ver estas condiciones, calcularemos los momentos de primer y segundo orden asegurando la estacionalidad en sentido débil.

- Media constante:

$$E(y_t) = E(\mathbf{m} + a_t + \mathbf{q}_1 a_{t-1} + \dots) = \mathbf{m} + \sum_{i=0}^{\infty} \mathbf{q}_i E(a_t)$$

$$\mathbf{q}_0 = 1 \text{ y } \sum_{i=1}^{\infty} \mathbf{q}_i = k < \infty$$

Es decir la media será constante en la medida en que exista la segunda parte del sumando que, al quedar multiplicada por la esperanza del "ruido blanco" será finalmente cero, y la media quedará igual a \mathbf{m} .

- Varianza constante:

$$\begin{aligned} \mathbf{g}_0 &= E(y_t - \mathbf{m})^2 = E(\mathbf{m} - \mathbf{m}^2 + a_t + \mathbf{q}_1 a_{t-1} + \mathbf{q}_2 a_{t-2} + \dots)^2 = \\ &= E(a_t^2 + \mathbf{q}_1^2 a_{t-1}^2 + \mathbf{q}_2^2 a_{t-2}^2 + \dots + 2\mathbf{q}_1 a_t a_{t-1} + 2\mathbf{q}_2 a_t a_{t-2} + \dots) = \\ &= \mathbf{s}_a^2 (1 + \mathbf{q}_1^2 + \mathbf{q}_2^2 + \dots + 0 + 0) = \\ &= \mathbf{s}_a^2 \sum_{i=0}^{\infty} \mathbf{q}_i^2 / \mathbf{q}_0 = 1 \end{aligned}$$

Siendo entonces condición necesaria para que la varianza exista que el último sumatorio sea calculable (converja).

- Covarianza constante:

$$\begin{aligned} \mathbf{g}_j &= E[(y_t - \mathbf{m})(y_{t-j} - \mathbf{m})] = \\ &= E[(a_t + \mathbf{q}_1 a_{t-1} + \mathbf{q}_2 a_{t-2} + \dots)(a_{t-j} + \mathbf{q}_1 a_{t-j-1} + \mathbf{q}_2 a_{t-j-2} + \dots)] = \\ &= \mathbf{q}_j E(a_{t-j}^2) + \mathbf{q}_1 \mathbf{q}_{j+1} E(a_{t-j-1}^2) + \mathbf{q}_2 \mathbf{q}_{j+2} E(a_{t-j-2}^2) + \dots = \\ &= \mathbf{s}_a^2 (\mathbf{q}_j + \mathbf{q}_1 \mathbf{q}_{j+1} + \mathbf{q}_2 \mathbf{q}_{j+2} + \dots) = \mathbf{s}_a^2 \sum_{i=0}^{\infty} \mathbf{q}_i \mathbf{q}_{i+j} \end{aligned}$$

Luego el proceso será estacionario en la medida en que se cumplan estas tres condiciones:

$$\sum_{i=0}^{\infty} \mathbf{q}_i E(a_t) / \mathbf{q}_0 = 1 \quad \text{y} \quad \sum_{i=1}^{\infty} \mathbf{q}_i = k < \infty$$

$$\sum_{i=0}^{\infty} \mathbf{q}_i^2 / \mathbf{q}_0 = 1$$

$$\sum_{i=0}^{\infty} q_i q_{i+j} < \infty$$

• Ventajas:

Una vez definido este proceso particular, vamos a ver resumidamente sus ventajas respecto a no contar con él.

En principio, si quisiéramos definir un proceso estocástico en general, tendríamos, al menos, que definir sus momentos de primer y segundo orden, para lo cual sería necesario estimar T varianzas,

T esperanzas y $(T^2 - T)/2$ covarianzas, lo que nos es imposible si sólo contamos con T datos.

Si el proceso fuera estacionario, ya sólo tendríamos que estimar una esperanza y una varianza (media y una varianza constantes) y (T-1) covarianzas ($\text{cov}(y_t, y_{t-1}), \text{cov}(y_t, y_{t-2}), \dots$); en total $1+1+(T-1)=T+1$ parámetros, lo que tampoco es posible. Si estamos ante un proceso estocástico estacionario lineal discreto, sólo necesitaremos contar con:

$$T > p + q + 2$$

siendo "p" y "q" los órdenes de los retardos de los modelos autorregresivos y de medias móviles que ya hemos definido anteriormente.

4.1 Modelos MA(1)

Una vez tenemos definidas las ventajas de contar con un proceso estocástico estacionario lineal y discreto, y que podemos calcular las funciones de autocovarianza y autocorrelación, puede resultar interesante ver que valores toman éstas en aquellos casos sencillos que luego nos permitan comprobar si series económicas generales pueden tener un comportamiento similar, simplemente acudiendo a la comparación de sus correlogramas (de la función de autocorrelación total y parcial).

El primer caso a analizar será el modelo de medias móviles de orden uno, que se define como:

$$y_t = m + a_t + q_1 a_{t-1}$$

Este modelo también se puede escribir en función del operador retardo, ya comentado, del siguiente modo:

$$y_t = \mathbf{m} + \mathbf{q}(L) a_t$$

$$\mathbf{q}(L) = 1 - \mathbf{q}_1 L$$

Se dice que un modelo MA(q) es invertible en la medida en que se pueda escribir como un proceso autoregresivo de orden infinito. Para que esta circunstancia pueda darse, será condición necesaria que las raíces de:

$$1 - \mathbf{q}(L) = 0$$

caigan fuera del círculo unitario, lo que se cumplirá siempre que $|\mathbf{q}| < 1$

Esta situación proviene de la conversión del modelo de medias móviles en modelo autoregresivo. Si escribimos el MA(1) como:

$$y_t = \mathbf{m} + a_t + \mathbf{q}_1 a_{t-1}$$

Podemos hacer sucesivas sustituciones hasta llegar al modelo autorregresivo:

$$a_{t-1} = y_{t-1} - \mathbf{m} - \mathbf{q}_1 a_{t-2}$$

$$y_t = \mathbf{m} + \mathbf{q}(y_{t-1} - \mathbf{m} - \mathbf{q}_1 a_{t-2}) + a_t$$

$$y_t = \mathbf{m} + a_t + \mathbf{q}_1 y_{t-1} - \mathbf{q}_1 \mathbf{m} - \mathbf{q}_1^2 y_{t-2} - \dots =$$

$$= \mathbf{q}_1 y_{t-1} + \mathbf{q}_1^2 y_{t-2} + \dots + \mathbf{m}(1 + \mathbf{q}_1 + \mathbf{q}_1^2 + \mathbf{q}_1^3 + \dots + \mathbf{q}_1^{n-1}) + a_t - \mathbf{q}_1^n a_{t-n}$$

si $n \rightarrow \infty$?

$$y_t = \mathbf{q}_1 y_{t-1} + \mathbf{q}_1^2 y_{t-2} + \dots + \mathbf{m} \frac{1}{1 - \mathbf{q}_1} + a_t$$

donde es necesario que $|\mathbf{q}| < 1$ para que la progresión geométrica que se produce sobre los parámetros \mathbf{q} sea calculable y no explosiva.

El siguiente paso, una vez definida la condición de invertibilidad, es definir las funciones que se han descrito para los procesos estocásticos en general para el caso del modelo MA(1).

• Esperanza:

$$E(y_t) = E(\mathbf{m} + a_t + \mathbf{q}_1 a_{t-1}) = E(\mathbf{m}) + E(a_t) + \mathbf{q}_1 E(a_{t-1}) = \mathbf{m}$$

• Varianza:

$$\text{var}(y_t) = E(y_t - \mathbf{m})^2 = E(\mathbf{m} + a_t + \mathbf{q}_1 a_{t-1} - \mathbf{m})^2$$

$$E = (a_t^2 + q_1^2 a_{t-1}^2 + 2q_1 a_t a_{t-1}) =$$

$$= s_a^2 + q_1^2 s_a^2 + 0 =$$

- Función de autocovarianza:

$$g_1 = \text{cov}(y_t, y_{t-1}) = E((y_t - m)(y_{t-1} - m)) =$$

$$E((a_t + q_1 a_{t-1})(a_{t-1} + q_1 a_{t-2})) =$$

$$E(a_t a_{t-1} + q_1 a_t a_{t-2} + q_1 a_{t-1} a_{t-1} + q_1^2 a_{t-1} a_{t-2}) =$$

$$q_1 E(a_{t-1})^2 = q_1 s_a^2$$

$$g_j = \text{cov}(y_t, y_{t-j}) = E((y_t - m)(y_{t-j} - m)) =$$

$$E((a_t + q_1 a_{t-1})(a_{t-j} + q_1 a_{t-j-1})) =$$

$$= E(a_t a_{t-j} + q_1 a_{t-1} a_{t-j} + q_1 a_t a_{t-j-1} + q_1^2 a_{t-1} a_{t-j-1})$$

$$E(a_t a_{t-j} + q_1 a_{t-1} a_{t-j} + q_1 a_t a_{t-j-1} + q_1^2 a_{t-1} a_{t-j-1}) = 0 = g_j$$

Luego la función de autocovarianza se anula para valores de "j" mayores que uno y es una fracción de la varianza del error para el valor de j=1.

- Función de autocorrelación:

Calculada como cociente entre la función de autocovarianza y la varianza, tal y como ya se vio antes, tendríamos:

$$r_j = \frac{g_j}{g_0}$$

$$r_1 = \frac{q_1 s_a^2}{(1 - q_1^2) s_a^2} = \frac{q_1}{(1 - q_1^2)}$$

$$r_j = 0 \quad j > 1$$

- La función de autocorrelación parcial se calcularía siguiendo la siguiente expresión, que no demostramos:

$$\hat{f}_{jj} = \frac{q_1^j (1 - q_1^2)}{1 - q_1^{2(j+1)}}$$

Definidas las funciones características de los procesos estocásticos para el caso concreto del MA(1), podemos enunciar las siguientes particularidades de este tipo de proceso:

- 1.- Siempre es estacionario.
- 2.- Para ser invertible, es necesario que $|q| < 1$.

3.- La r_j sólo tiene un punto significativo. El modelo "olvida" la correlación con períodos distintos al inmediatamente anterior y el correlograma sólo tendrá un punto significativo.

4.- La función de autocorrelación parcial no se anula, pero tendrá un comportamiento amortiguado hacia cero.

4.2 Modelos AR(1)

Definido el modelo AR(1) como:

$$y_t = f_0 + f_1 y_{t-1} + a_t$$

que también puede escribirse como:

$$\begin{aligned} f(L) y_t &= f_0 + a_t \\ f(L) &= 1 - f_1 L \end{aligned}$$

a diferencia de los modelos de medias móviles, los autoregresivos no son estacionarios por definición y, para que lo sean, ha de cumplirse que las raíces de la siguiente ecuación sean mayores que uno:

$$f(L) = 1 - f_1 L = 0$$

lo que nos permitiría escribirlo como un modelo de medias móviles y, en definitiva, esto supone que los coeficientes $|q|$ han de ser menores que 1.

Pasamos a describir las funciones definidas:

- Esperanza matemática:

$$\begin{aligned} E(y_t) &= E(f_1 y_{t-1} + f_0 + a_t) = \\ &= f_1 E(y_{t-1}) + f_0 + 0 \end{aligned}$$

donde, como el proceso es estacionario, las esperanzas $E(y_t) = E(y_{t-1}) = \dots = E(y_{t-j}) = m$ y puedo escribir:

$$\begin{aligned} E(y_t) &= m = f_1 m + f_0 \\ m &= \frac{f_0}{1 - f_1} \end{aligned}$$

• Varianza:

Para hacer los cálculos con mayor facilidad es conveniente poner el modelo autoregresivo en desviaciones a la media, sin que ello suponga ningún cambio en éste (se puede hacer la prueba escribiendo el modelo en desviaciones y llegando al modelo normal).

El modelo en desviaciones a la media lo definiremos como:

$$\tilde{y}_t = f_1 \tilde{y}_{t-1} + a_t$$

donde:

$$\tilde{y}_{t-j} = y_{t-j} - \mathbf{m} = y_{t-j} - \frac{f_0}{1-f_1}$$

Para calcular la varianza, escribimos el momento de segundo orden:

$$\begin{aligned} \text{var} = (y_t) = \mathbf{g}_0 &= E\left(f_1 \tilde{y}_{t-1} + a_t\right)^2 \\ &= E(f_1^2 \tilde{y}_{t-1}^2 + a_t^2 + 2f_1 \tilde{y}_{t-1} a_t) = \\ &= f_1^2 E(\tilde{y}_{t-1})^2 + E(a_t)^2 + 2f_1 E(\tilde{y}_{t-1} a_t) \end{aligned}$$

dado que el proceso es estacionario:

$$\begin{aligned} E(\tilde{y}_t)^2 &= E(\tilde{y}_{t-1})^2 = \mathbf{g}_0 \\ E(a_t)^2 &= \mathbf{s}_a^2 \end{aligned}$$

como el proceso en desviaciones a la media se puede escribir como un proceso de medias móviles y por lo que hemos visto anteriormente, podríamos escribir:

$$E(a_t \tilde{y}_{t-h}) = \mathbf{s}_a^2 \quad h=0 \quad \text{y} \quad 0 \quad \text{si} \quad h > 0$$

Por todo lo cual:

$$\mathbf{g}_0 = f_1^2 \mathbf{g}_0 + \mathbf{s}_a^2$$

$$\mathbf{g}_0 = \frac{\mathbf{s}_a^2}{1 - \mathbf{f}_1^2}$$

- Función de autocovarianza:

$$\begin{aligned} \mathbf{g}_1 &= \text{cov}(\tilde{y}_t, \tilde{y}_{t-1}) = E((\tilde{y}_{t-1}(\mathbf{f}_1 \tilde{y}_{t-1} + a_t))) = \\ &= \mathbf{f}_1 \mathbf{g}_0 = \mathbf{f}_1 \frac{\mathbf{s}_a^2}{1 - \mathbf{f}_1^2} \\ \mathbf{g}_2 &= \text{cov}(\tilde{y}_t, \tilde{y}_{t-2}) = E(\tilde{y}_{t-2}(\mathbf{f}_1 \tilde{y}_{t-1} + a_t)) = \\ &\mathbf{f} \mathbf{g}_1 = \mathbf{f}_1^2 \mathbf{g}_0 = \mathbf{f}_1^2 \frac{\mathbf{s}_a^2}{1 - \mathbf{f}_1^2} \end{aligned}$$

Lo que generalizando, se podría escribir como:

$$\mathbf{g}_j = \frac{\mathbf{s}_a^2}{1 - \mathbf{f}_1^2} \mathbf{f}_1^j$$

- Función de autocorrelación total

$$\mathbf{r}_j = \mathbf{f}_1^j \quad j \geq 1$$

- Función de autocorrelación parcial

$$\mathbf{f}_{jj} = \mathbf{r}_1 \quad j = 1 \quad \text{y } 0 \text{ si } j > 1$$

Definidas todas estas funciones, podemos caracterizar el proceso autoregresivo del siguiente modo:

- 1.- Siempre es invertible (está directamente invertido).
- 2.- Para ser estacionario, ha de cumplirse que $|\mathbf{q}| < 1$.
- 3.- La función de autocorrelación total no se anula, pero se va amortiguando hacia cero.
- 4.- La función de autocorrelación parcial se anula para retardos superiores a uno.

	FAC	FAP
MA(q)	Se anula para retardos superiores a q	Decrecimiento rápido sin llegar a anularse
AR(P)	Decrecimiento rápido sin llegar a anularse	Se anula para retardos superiores a p
ARMA(p,q)	Decrecimiento rápido sin llegar a anularse	Decrecimiento rápido sin llegar a anularse

5. Identificación del Modelo

Aunque podríamos seguir definiendo las características de otros procesos ARIMA de órdenes mayores, no tiene mayor interés una vez se ha entendido el procedimiento y sí interesa precisar la forma que adoptarían los correlogramas de estas funciones porque, fruto de su comparación con los que obtendremos con nuestras series de interés, podremos asociar a nuestra serie temporal de estudio un proceso ARIMA que identifique su proceso generador de datos, tanto a pasado como a futuro.

Los correlogramas o funciones de autocorrelación total y parcial están disponibles en el libro de Pulido, A: "Predicción Económica y Empresarial" Editorial Pirámide.

De forma muy poco académica, el proceso de identificación consistirá en calcular las funciones de autocorrelación total y parcial de nuestra serie (una vez estamos seguros de que ésta cumple las condiciones que definen un proceso estocástico estacionario) y comparar sus correlogramas con los correspondientes a los modelos teóricos AR(p), MA(q) o ARMA(p,q).

En principio, si el proceso está bien identificado, procederemos a su estimación y, si analizamos los correlogramas de los residuos obtenidos en la estimación, serán "ruido blanco". Si esto no es así, habrá que realizar una nueva estimación incorporando la estructura más parecida al modelo teórico que podamos intuir con la comparación con los modelos teóricos.

Para saber cuando estamos ante un "ruido blanco", se pueden hacer las siguientes comprobaciones:

- Media nula

Puede observarse en el gráfico de residuos si el error se mueve en torno al valor cero o bien calcularse el cociente entre la media y la varianza muestral de los residuos. Si ese ratio es inferior a 2, podemos concluir (con un $\alpha=0,05$) que la media no es significativamente distinta de cero.

- Varianza constante

Observando el gráfico de los residuos puede analizarse la constancia de la varianza del error. En caso de heterocedasticidad y es recomendable una transformación logarítmica en la serie original.

- Incorrección

Deben observarse los coeficientes de autocorrelación muestral de los residuos y comprobar que ninguno de ellos supera el valor de las bandas de significativa al 5% ($\pm 1,96(1/T^{1/2})$). El valor $T^{1/2}$ es una aproximación de la varianza asintótica pero resulta sólo adecuada para valores grandes de "j". Se aconseja, por tanto, utilizar distinta amplitud de bandas como por ejemplo $\pm(1/T^{1/2})$ para los términos más cercanos a cero.

El estudio de las funciones de autocorrelación muestral y autocorrelación parcial muestral de los residuos, pueden servirnos fácilmente para el replanteamiento del modelo inicial.

El contraste de la "Q" de Box-Pierce analiza la hipótesis nula de que:

$$H_0 : r_1(a) = r_2(a) = r_3(a) = \dots = r_M(a) = 0$$

suponiendo que la expresión:

$$Q = T \sum_{j=1}^M r_j^2(\hat{a})_t$$

o la alternativa propuesta por Ljung-Box:

$$Q^* = T(T+2) \sum_{j=1}^M (T-j)^{-1} r_j^2(\hat{a})_t$$

se distribuye como una chi-cuadrado con Mk grados de libertad. Otros contrastes alternativos son los de Vandaele, que analiza las autocorrelaciones muestrales de los residuos transformados mediante una diferencia regular y el de Peña .

6 Breve Reseña sobre la Estimación de los Modelos ARIMA

Se analizará a continuación el proceso de estimación de los modelos ARMA(p,q) x ARMA(P,Q), centrando los desarrollos en el caso específico de un modelo sin componente estacional, es decir, un ARMA(p,q) ó ARIMA (p,d,q):

$$(1 - f_1 L - f_2 L^2 - \dots - f_p L^p) w_t = d + (1 - q_1 L - q_2 L^2 - \dots - q_p L^q) a_t$$

donde se entiende por w_t la serie ya en diferencias y donde, como siempre, a_t se supone un ruido blanco con media cero y varianza constante (σ_m^2).

Definida esta función genérica, el objetivo principal es estimar el vector formado por los parámetros correspondientes a la parte autoregresiva \mathbf{q}_j y de medias móviles \mathbf{q}_j (incluido, si fuera necesario, el término independiente) así como la varianza residual.

6.1 Problemas Iniciales: los Valores Iniciales y la No Linealidad.

La naturaleza del modelo implica que la variable a explicar se hace depender de valores pasados de la misma y errores cometidos en la estimación de dichos valores pasados. De esta forma el planteamiento de minimización de los errores como procedimiento de estimación lleva necesariamente aparejada la necesidad de conocer valores pasados de la variable endógena y de los errores ya que la expresión del error, por ejemplo para un período "t" sería:

$$a_t = w_t - \mathbf{f}_1 w_{t-1} - \dots - \mathbf{f}_p w_{t-p} - \mathbf{d} + \mathbf{f}_1 u_{t-1} - \dots - \mathbf{f}_q a_{t-q}$$

Al conjunto de valores iniciales requeridos de la variable endógena desde "t-1" a "t-p" y de los errores desde "t-1" a "t-q" los notaremos por los vectores:

$$W^0 = (W_{t-1}, W_{t-2}, \dots, W_{t-p})$$

$$a^0 = (a_{t-1}, a_{t-2}, \dots, a_{t-q})$$

El procedimiento de estimación que lleva implícita la especificación a priori de unos valores iniciales se denomina "enfoque condicional", mientras que aquel en el que se estiman simultáneamente los valores iniciales y los parámetros se denomina "enfoque exacto".

Además de este primer problema, se señala el de la no linealidad del modelo cuando este incluye medias móviles, lo que puede comprobarse fácilmente a partir de una transformación de la especificación de un modelo sencillo [por ejemplo MA(1)]:

$$Y_t = a_t - \mathbf{q}_1 a_{t-1} \rightarrow a_t = Y_t + \mathbf{q}_1 a_{t-1}$$

↓

$$\begin{aligned} \text{para } t=1 \quad a_1 &= Y_1 + \mathbf{q}_1 a_0 \\ \text{para } t=2 \quad a_2 &= Y_2 + \mathbf{q}_1 a_1 \\ &= Y_2 + \mathbf{q}_1 (Y_1 + \mathbf{q}_1 a_0) = \\ &= Y_2 + \mathbf{q}_1 Y_1 + \mathbf{q}_1^2 a_0 \end{aligned}$$

Partiendo de esta expresión, se observa como al minimizar:

$$\sum_{t=1}^T \hat{a}_t^2$$

en el proceso de estimación, esta expresión no será lineal. Por ello, una primera conclusión es que sea cual sea el método de estimación utilizado (Mínimos cuadrados o Máxima verosimilitud) deberán aplicarse algoritmos de resolución no lineales.

6.2 Predictor Óptimo.

"La predicción es el fin último y primordial del análisis univariante de series temporales"⁴ Una vez identificado y estimado el modelo ARIMA, se plantea su utilización para conseguir la mejor predicción de los valores a futuro de una serie a partir de su propia historia . El primer interrogante que surge se referirá a la determinación del predictor óptimo para este fin.

Intuitivamente, el mejor predictor posible será "el que menos se equivoca" o, en términos estadísticos, aquel que minimiza el error cuadrático medio respecto a otro potencial predictor alternativo. Esto se puede expresar:

$$E[(Y_{T+1} - \hat{Y}_T(l))^2 / I_T] \leq E[(Y_{T+1} - \hat{Y}_T^*(l))^2 / I_T]$$

donde $\hat{Y}_T(l)$ sería el valor de predicción de la serie para el período $(T+1)$, condicionado a los valores históricos de Y_t ($Y_t = Y_{T-1}, Y_{T-2}, \dots$)

Se demuestra que el predictor elegido es óptimo cuando su valor esperado es igual al valor real de predicción condicionado a la información existente en el período T respecto a la serie que nos ocupa; es decir:

$$\hat{Y}_T(l) = E[Y_{T+1} / I_T]$$

El error cuadrático medio de un predictor arbitrario siempre es mayor que aquel cuyo valor coincide con la esperanza del valor real en el período que estamos considerando. Esta propiedad será fundamental para el posterior desarrollo de la predicción puntual.

6.3 Predicción Puntual

Partiendo de un modelo ARIMA sobre el que se han realizado una serie de diferenciaciones para lograr una serie estacionaria, el planteamiento de la predicción se hace sobre los valores reales de la serie, por entender que es de éstos de los que se quiere obtener valores a futuro. Esta circunstancia deberá ser tomada en cuenta a la hora de interpretar los subíndices que acompañan a las fórmulas de este apartado.

La aparición de órdenes autoregresivos superiores a "p" será debida a dos circunstancias: traslación del modelo a períodos fuera del espacio temporal conocido y aplicación de las expresiones obtenidas para $W_T = (1-L)^d Y_T$ directamente a los valores de la serie sin transformar (sustituir WT por su valor en cada caso).

En su forma más general el modelo ARIMA(p,d,q) ARIMA(P,D,Q,)S podría escribirse como:

$$Y_T = j_1 Y_{T-1} + j_2 Y_{T-2} + \dots + j_{p_s+p+DS+d} Y_{T-p_s-p-sD-d} + d + a_T + q_1 a_{T-1} + \dots + q_{Q_s+q} a_{T-sQ-q}$$

El desarrollo teórico se hace para un modelo sin parte estacional, sin que ello suponga pérdida de generalidad en los resultados. El modelo podría escribirse entonces como:

$$f_p^*(L)W_T = d + q_q^*(L)a_t$$

como sabemos que $W_T = (1-L)^d Y_T$, la expresión anterior se puede reescribir como:

$$j_p^*(L) = Y_t = d + q_q^*(L)a_t$$

donde:

$$j(L)Y_T = 1 - j_1 L - j_2 L^2 + \dots + j_{p+d} L^{p+d}$$

$$q(L) = 1 - q_1 L + \dots + q_q L^q$$

El modelo ARIMA correspondiente sería:

$$Y_T = j_1 Y_{T-1} + j_2 Y_{T-2} + \dots + j_{p+d} Y_{T-p-d} + d + a_T + q_1 a_{T-1} + \dots + q_q a_{T-q}$$

En la medida en que aparece la perturbación aleatoria en la definición de cada valor de predicción de la serie Y_T , el modelo es susceptible de ser escrito como función infinita de los valores de la perturbación aleatoria del período considerado y de los anteriores (no habría más que despejar el valor de la perturbación aleatoria de la expresión general de un modelo ARIMA).

Para hacer predicción hay que tener en cuenta dos supuestos iniciales:

- 1.- Los parámetros de las funciones presentadas son conocidos.
- 2.- Las perturbaciones aleatorias se conocen en el período muestral y tiene carácter de ruido blanco para los períodos de predicción.

Teniendo en cuenta ambos supuestos, se puede especificar un modelo ARIMA para definir el valor de predicción en función de la serie en "p" períodos anteriores (por la parte autoregresiva) y de los "q" errores cometidos al estimar la serie en los "q" períodos previos (por la parte de las medias móviles). Dado que son los valores reales aquellos sobre los que es verdaderamente interesante hacer predicción, se plantea la ecuación de predicción de un período ($T + l$) como:

$$\hat{Y}(l) = j_1 \hat{Y}_T(l-1) + j_2 \hat{Y}_T + j_{l-1} \hat{Y}_T(l) + \dots + j_{l+1} \hat{Y}_{T-1} + \dots + j_{p+d} \hat{Y}_{T+l-p-d} + d - q_{l+1} a_{T-1} - \dots - q_q a_{T+l-q}$$

Es conveniente remarcar tres aspectos de la predicción siguiendo esta formulación:

- Los valores de predicción se calculan de forma secuencial. Para hacer posible la aplicación de la parte autoregresiva del modelo en períodos distintos al primero de predicción, se toma el valor predicho en el inmediatamente anterior ($Y_T(l-1)$).
- Por la condición necesaria para estar ante el predictor óptimo (el valor esperado de la serie en el período de predicción es igual al predicho si es óptimo), se demuestra que las perturbaciones aleatorias empleadas en la predicción son sólo las de períodos anteriores, es decir:

$$a_{T+j} \quad \forall j \leq 0$$

ya que se ha supuesto que en el período de predicción, las perturbaciones aleatorias tienen esperanza nula.

Para realizar las sucesivas predicciones de $Y_t(1)$ tenemos que contar con unos valores para $a_T, a_{T-1}, \dots, a_{T+1-q}$. Como tales se tomarán:

$$a_{T+j} = Y_{T+j} - \hat{Y}_{T+j-1}(l) \quad \forall j \leq 0$$

- Las predicciones ARIMA son adaptativas y los resultados obtenidos para $(T + l)$, con la información disponible hasta el período T , son los mismos que las que obtendríamos para el mismo período tomando como base informativa hasta $T - 1$, y añadiendo un término de error.

7 Características de las Predicciones Realizadas con Modelos ARIMA

- Modelos AR(p): la predicción tiende a m (media del proceso) a medida que aumenta el horizonte temporal de la predicción.
- Modelos MA(q): dada la memoria limitada que caracteriza a estos procesos, la predicción es igual a m (media del proceso) cuando el horizonte temporal de la predicción es mayor que el orden del proceso (q).
- Modelos ARMA(p,q): a partir de "q" períodos futuros la predicción tiende a m (media del proceso) a medida que aumenta el horizonte temporal de la predicción.
- Modelos ARI(p,d) e IMA(d,q): la predicción ya no tiende a m sino que será una línea recta que parte de $\hat{Y}(1)$ con pendiente igual a la media del proceso W_T (serie resultante de las transformaciones necesarias para hacerla estacionaria).

7.1 Selección de Modelos

Si entendemos que una predicción es mejor que otra cuando comete menor error, los criterios de selección de modelos serían el error cuadrático medio (ECM), error absoluto medio (EAM) y error absoluto porcentual medio (EAPM). Estos indicadores se calcularían a período histórico, es decir, se calcularían los valores que el modelo ofrece para las H últimas observaciones y se compararían con el valor real, del siguiente modo:

$$ECM(H) = \frac{1}{H} \sum_{l=1}^H e_{T-H}^2(l) = \frac{1}{H} \sum_{l=1}^H [y_{T-H+l} - \hat{y}_{T-H}(l)]^2$$

$$EAM(H) = \frac{1}{H} \sum_{l=1}^H |e_{T-H}(l)| = \frac{1}{H} \sum_{l=1}^H |y_{T-H+1} - \hat{y}_{T-H}(l)|$$
$$EAPM(H) = \frac{1}{H} \sum_{l=1}^H \frac{|e_{T-H}(l)|}{y_{T-H+1}} * 100 = \frac{1}{H} \sum_{l=1}^H \frac{|y_{T-H+1} - \hat{y}_{T-H}(l)|}{y_{T-H+1}} * 100$$

El problema es que estos indicadores no tienen en cuenta la estructura estocástica del modelo, no informan sobre alguna característica estocástica supuesta sobre el período extramuestral.

Referencias

Hidrología Básica

1. Ayllon, T, Gutiérrez, G. 1983. **INTRODUCCION A LA OBSERVACION METEOROLOGICA**. Limusa.
2. Chorley, R. 1979. **INTRODUCTION TO GEOGRAPHICAL HYDROLOGY**. Methuen.
3. Crowe, 1971 **P. CONCEPTS IN CLIMATOLOGY**. Longman Group Ltd..
4. Eagleson, 1970. **P. DYNAMIC HYDROLOGY**. McGraw-Hill.
5. Hardy, R. 1983. **EL LIBRO DEL CLIMA**. Herman Blume Ediciones.
6. Henao, J. 1988. **INTRODUCCION AL MANEJO DE CUENCAS HIDROGRAFICAS**. U. Santo Tomás. Bogotá.
7. Linsley, Kohler, Paulhus 1985. **HYDROLOGY FOR ENGINEERS**. 3rd. Ed. McGraw-Hill..
8. Nemeck, J. 1960 **ENGINEERING HYDROLOGY**. McGraw-Hill..
9. Shaw, E. 1983 **HYDROLOGY IN PRACTICE**. Van Nostrand..
10. Silva M Gustavo A. 1998 **HIDROLOGIA BASICA**. Facultad de Ingeniería. Universidad Nacional de Colombia. Bogotá..
11. U.S. Soil Conservation Service. **HYDROLOGY**. Section 4. National Engineering Handbook.
12. Viessman, W; Knapp; Lewis. 1989. **INTRODUCTION TO HYDROLOGY**. 3rd. Ed. Harper & Row.
13. Ward, R. 1975. **PRINCIPLES OF HYDROLOGY**. McGraw-Hill. 2nd Ed.
14. Wiesner, C, J. 1970. **HYDROMETEOROLOGY**. Chapman & Hall.

Hidrología Aplicada.

1. Chow, Ven Te et al. 1988. **APPLIED HYDROLOGY**. McGraw-Hill.
2. G.G. Suanidze, **MATHEMATICAL MODELING OF HYDROLOGIC SERIES**, Ed. Water Resources Publications
3. Hoggan, D. 1989. **COMPUTER ASSISTED FLOOD PLAIN HYDROLOGY AND HYDRAULICS**. McGraw-Hill.
4. J.D. Salas ,J.W. Delleur V Yevjevich and WL Lane, **APPLIED MODELING OF HYDROLOGIC TIME SERIES**, Ed. Water Resources Publications
5. Linsley, Franzini. 1992. **WATER RESOURCES ENGINEERING**. 4th. Ed. McGraw-Hill.

Hidrología Estocástica.

1. Bellman and Dreyfus 1962. **APPLIED DYNAMIC PROGRAMMING**. Princeton University Press.
2. Bras, R; Rodríguez-Iturbe, 1993. I. **RANDOM FUNCTIONS AND HYDROLOGY**. Dover.
3. Ciriani, T; Maione, W. 1977. **MATHEMATICAL MODELS FOR SURFACE WATER ENGINEERING**. Wiley.
4. Haan, Ch. 1977. **STATISTICAL METHODS IN HYDROLOGY**. Iowa State University Press.
5. Hall and Dracup 1970.. **WATER RESOURCES SYSTEMS ENGINEERING**. McGraw-Hill Series.
6. Hjermfelt, A; Cassidy, J. 1975. **HYDROLOGY FOR ENGINEERS AND PLANNERS**. Iowa State University Press.
7. Maas et al. 1962. **DESIGN OF WATER RESOURCE SYSTEMS**. Mc Millan.
8. Shahin, M. 1993. **STATISTICAL ANALYSIS IN WATER RESOURCES ENGINEERING**. Applied Hydrology Monographs. A. Balkema.
9. Shumway, R. 1988. **APPLIED STATISTICAL TIME SERIES ANALYSIS**. Prentice Hall.
10. Taha, H. A. 1982. **OPERATIONS RESEARCH**. Collier McMillan. 3rd Ed.
11. Vujica Yevjevich, **STOCHASTIC PROCESSES IN HYDROLOGY**, Ed. Water Resources Publications
12. Yevchevich, V. 1972. **PROBABILITY AND STATISTICS IN HYDROLOGY**. Water Resource Publications. Fort Collins Co.
13. Yevchevich, V. 1972. **STOCHASTIC PROCESSES IN HYDROLOGY**. Water Resource Publications. Fort Collins Co.

Modelos ARIMA

- Aznar, a. Y trivez, f.j.1993 **MÉTODOS DE PREDICCIÓN EN ECONOMÍA II. ANÁLISIS DE SERIES TEMPORALES** Editorial Ariel Economía, Barcelona
- Espasa, a. Y cancelo, j.r. 1993: **MÉTODOS CUANTITATIVOS PARA EL ANÁLISIS DE LA COYUNTURA ECONÓMICA** Alianza Editorial

PULIDO, A. Y PÉREZ GARCÍA, J.2001: **MODELOS ECONÓMICOS**
Editorial Pirámide, Madrid
Enders, w. 1995. **APPLIED ECONOMETRIC TIMES SERIES**. John Wiley &
Sons, Inc. United States.

Consultas de internet

www.uam.es/docencia/predysim/prediccion_unidad3/unidad3.htm - 43k -

www.uam.es/personal_pdi/economicas/rarce/pdf/Box-Jenkins.PDF

www.einsteinnet.com/econometria/ARIMA/arima

www.geocities.com/gsilvam/hidrobas.htm

http://es.wikipedia.org/wiki/Geograf%C3%ADa_de_M%C3%A9xico

http://cambio_climatico.ine.gob.mx/sabycono/sabyconoporque.html#hidrologia

http://cecadesu.semarnat.gob.mx/biblioteca_digital/biodiversidad/bio07.shtml

http://www.geofisica.unam.mx/atlas/mexico/elaqua_dr.htm

http://www.desarrollolatino.org/web2/box_derecha.htm

<http://www.ideam.gov.co/>

<http://www.senamhi.gov.bo/>

<http://www.astromia.com/glosario/hidrologia.htm>

<http://www.inegi.gob.mx/geo/default.asp?c=394>

http://agronomia.uchile.cl/web/rodrigo_fuster/sitiohidrologia/

<http://www.canaltiempo21.com/hidrologia.htm>

http://db.doyma.es/cgi-bin/wdbcgi.exe/doyma/hemeroteca_web.resultado_cat?p_selector=5&p_categoria=1320

<http://www.miliarium.com/Paginas/Soft/pid.asp>

<http://www.hobeco.net/nova/hidrologia/hidrologia.html>

http://www.uv.es/~sala/pagina_n5.htm

- <http://www.uv.es/asepuma/jornadas/malaga/8.pdf>
- <http://www.fing.edu.uy/inco/eventos/elavio/resumenes.htm>
- www.fmat.ull.es/php/fichaasignatura.php
- www.personales.upv.es/arodrigu/IDI/Escenarios.pdf
- www.doi.icaei.upco.es/simio/apuntes/a_sp.pdf
- www.iqcelaya.itc.mx/~vicente/CV-VRR.pdf
- www.uandes.cl/ver_articulo.asp?id=1009&seccion_padre=288
- www.ing.puc.cl/gescopp/Sergio_Maturana/DISCUSSIONBIBLIOGRAFICA.html
- www.uned.es/guias12c/08/085140.pdf
- www.celogis.com/maestria_ejecutiva_logistica.htm
- www.insma.udg.es/ambientalitzacio/linies/curricular/webciutats/TEXTMEXI.HT
- www.acabtu.com.mx/guerrero/hidrologia.html
- www.mapserver.inegi.gob.mx/geografia/espanol/datosgeogra/basicos/hidrologia/rios/regiones_hidrologicas
- www.apuntes.rincondelvago.com/series-de-tiempo_1.html
- www.bccr.fi.cr/ndie/Documentos/DIE-02-2002-NT-ASPECTOS%20CONCEPTUALES%20SOBRE%20SEATS.pdf
- www.cesma.usb.ve/~lillas/PS8/SERIES/index.html
- www.monografias.com/trabajos19/series-de-tiempo/series-de-tiempo.shtml
- www.monografias.com/trabajos30/software-ilegales/software-ilegales.shtml
- www.siap.sagarpa.gob.mx/TS.html
- www.stockssite.com/pf/algebra_matricial.html
- www.mdk.ideam.gov.co/Manual_Uusuario_CAM/node10.html
- www.cema.edu.ar/postgrado/dae/download/2005/landro_ast.pdf
- www.stockssite.com/pf/st/seriesdetiempo1.html
- www.economia.uanl.mx/licenciatura/materias/series_tiempo.html
- www.desarrollolatino.org/web0/INDEXIZQ.htm

www.mat.puc.cl/~wilfredo/espanol/cursos/pro_series.htm

www.ci.ulsal.mx/~jgarcia/docencia/time.pdf

www.uam.es/personal_pdi/economicas/rarce/pdf/Box-Jenkins.PDF

www.icesi.edu.co/~jcalonso/ER/modulos/ARIMA/ARIMA.HTM

www.einsteinnet.com/econometria/ARIMA/arima1.htm

www.quipu.uni.edu.pe/OtrosWWW/webproof/public/revistas/tecnica/vol9n1/11art/index.html

www.monografias.com/trabajos19/series-de-tiempo/series-de-tiempo.shtml

www.seh-lelha.org/tseries.htm

www.gl.wikipedia.org/wiki/Modelos_Arima

www.rincondelvago.com/modelo-econometrico.html

www.psico.uniovi.es/REMA/v4n2/a4/p1.html