



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

# **MODELACIÓN HIDROLÓGICA EN EL SISTEMA INTERCONECTADO CENTRAL**

**EDUARDO ALEJANDRO MINDER HETZ**

Tesis para optar al grado de  
Magister en Ciencias de la Ingeniería

Profesor Supervisor:  
**BONIFACIO FERNÁNDEZ L.**  
**HUGH RUDNICK V.**

Santiago de Chile, 1997



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA  
Departamento de Ingeniería Eléctrica

# **MODELACIÓN HIDROLÓGICA EN EL SISTEMA INTERCONECTADO CENTRAL**

**EDUARDO ALEJANDRO MINDER HETZ**

Tesis presentada a la Comisión integrada por los profesores:

**BONIFACIO FERNÁNDEZ L.**

**HUGH RUDNICK V.**

**SAMUEL JERARDINO E.**

**JOSÉ MUÑOZ P.**

Para completar las exigencias del grado  
de Magister en Ciencias de la Ingeniería

Santiago de Chile, 1997

A mis padres, quienes con cariño y esfuerzo siempre me dieron lo mejor de ellos. Gracias.

## **AGRADECIMIENTOS**

A mi esposa Paulina quien en todo instante me brindó su apoyo incondicional y me comprendió en los momentos más difíciles del desarrollo de esta tesis.

A mis padres, Martita y Eduardo, y hermanas por su permanente interés y preocupación. A mi fallecida abuelita Violeta quien siempre fue muy cariñosa y estuvo orgullosa de mis logros.

Al profesor Bonifacio Fernández por el apoyo siempre brindado, el permanente interés en revisar los borradores y la constante preocupación en los momentos en que se tornó más difícil el desarrollo de esta tesis. También quisiera extender el agradecimiento al Departamento de Ingeniería Hidráulica y Ambiental de la Pontificia Universidad Católica de Chile que me acogió como si fuera un alumno de su especialidad.

Al profesor Hugh Rudnick por su excelente disposición y continuo interés en el avance logrado en la tesis.

A mis amigos quienes siempre se preocuparon por el estado del desarrollo de esta tesis, en especial a Fernando, Luis, Nelson y Vanessa.

A la Empresa Eléctrica Colbún Machicura S.A. y a la Empresa Nacional de Electricidad S.A. por su apoyo a la investigación; y a la Unidad de Investigación y Desarrollo en la Pontificia Universidad Católica de Chile.

## ÍNDICE GENERAL

	Pág.
DEDICATORIA .....	ii
AGRADECIMIENTOS.....	iii
ÍNDICE DE TABLAS.....	vii
ÍNDICE DE FIGURAS .....	ix
RESUMEN .....	xi
ABSTRACT .....	xii
I. INTRODUCCIÓN .....	1
1.1 Objetivos.....	2
1.2 Características de los sistemas eléctricos de potencia.....	3
1.3 Requerimientos de despacho .....	7
II. EL SISTEMA ELÉCTRICO CHILENO.....	9
2.1 Características generales de los sistemas eléctricos chilenos.....	9
2.1.1 Sistema Interconectado del Norte Grande (SING).....	9
2.1.2 Sistemas de la Zona Austral (XI y XII Regiones).....	10
2.1.3 Sistema Interconectado Central (SIC).....	10
2.2 Comparación con otros sistemas latinoamericanos: sus marcos regulatorios	11
2.2.1 El sector generación .....	12
2.2.2 El sector transmisión.....	14
2.2.3 El sector distribución .....	16
2.3 Organización del SIC.....	16
2.3.1 Empresas participantes en el SIC.....	16
2.3.2 Centro de Despacho Económico de Carga (CDEC).....	19
2.4 Necesidades de modelación para la operación del SIC .....	23
2.4.1 Modelos empleados por las empresas generadoras y el CDEC .....	23
2.4.2 Modelos empleados por la CNE .....	25

2.4.3	Validez de los datos que se utilizan .....	26
2.4.4	Optimización a largo plazo .....	27
2.4.5	Pronóstico a corto plazo .....	30
III.	MODELACIÓN DE SISTEMAS HIDROLÓGICOS.....	32
3.1	Esquema general de modelación .....	32
3.1.1	Modelos típicos disponibles.....	34
3.1.2	Propiedades estadísticas a modelar .....	36
3.1.3	Periodicidad y su representación.....	39
3.2	Modelación de la dependencia temporal .....	39
3.2.1	Identificación de modelos temporales periódicos .....	39
3.2.2	Estimación de parámetros .....	42
3.3	Construcción del modelo.....	44
3.4	Diagnóstico o verificación del modelo.....	46
3.4.1	Pruebas de independencia .....	47
3.4.2	Prueba de normalidad.....	48
3.4.3	Promedio nulo y otras propiedades .....	48
3.4.4	Reformulación.....	49
3.4.5	Sobreajuste.....	49
3.5	Modelación de la dependencia espacial.....	50
3.6	Usos de los modelos .....	53
3.6.1	Simulación .....	54
3.6.2	Pronóstico.....	58
3.7	Otros modelos.....	61
IV.	ANÁLISIS DE LAS SERIES DISPONIBLES PARA LA MODELACIÓN.....	62
4.1	Características de los aprovechamientos del SIC .....	62
4.1.1	Selección de las series cronológicas de interés. ....	63
4.1.2	Período común recomendado para todas las cuencas del SIC.....	79
4.2	Propiedades hidrológicas de interés .....	80
4.2.1	Características de las series hidrológicas .....	80
4.2.2	Análisis de series de caudales del SIC .....	86
4.3	Observaciones acerca de las series disponibles en el SIC .....	93

V.	MODELACIÓN HIDROLÓGICA DE LOS RECURSOS DISPONIBLES.....	96
5.1	Construcción de un modelo .....	96
5.2	Principales modelos utilizados por el CDEC .....	97
5.3	Ejemplo de modelación hidrológica propuesto .....	100
5.3.1	Uso del modelo en simulación de la operación.....	100
5.3.2	Uso del modelo en pronóstico.....	106
5.4	Aplicación del ejemplo de modelación hidrológico propuesto al modelo OMSIC .....	119
VI.	CONCLUSIONES Y RECOMENDACIONES.....	122
6.1	Consideraciones acerca de los datos.....	122
6.2	Comentarios del modelo PRSI-SDDP.....	123
6.3	Acerca del uso del modelo en pronóstico.....	126
6.4	Acerca del uso del modelo en simulación de la operación.....	128
	BIBLIOGRAFIA .....	129
	A N E X O S.....	134
	ANEXO A: CENTRALES QUE COMPONENTEN EL SING Y EL SIC Y MAPA DE LA ZONA ESTUDIADA.....	135
	ANEXO B: REGLAMENTO DEL FUNCIONAMIENTO DEL CDEC.....	138
	ANEXO C: DATOS DE LAS SERIES UTILIZADAS EN LOS EJEMPLOS Y ESTADÍGRAFOS CALCULADOS.....	141
	ANEXO D: DETALLE MATRICES SINTÉTICAS Y COSTOS MARGINALES OBTENIDOS DE CORRIDAS DEL MODELO OMSIC .....	167

## ÍNDICE DE TABLAS

	Pág.
Tabla 1.1: Etapas de un SEP y sus principales características. ....	4
Tabla 1.2: Tipos de centrales utilizadas en la generación de energía en Chile y el mundo. [Vara95] .....	4
Tabla 1.3: Comparación entre centrales hidroeléctricas y termoeléctricas. ....	7
Tabla 2.1: Participación de empresas en el SIC. [Cdec95] .....	17
Tabla 3.1: Parámetros implicados en un modelo para w períodos. ....	38
Tabla 3.2. Variable normal reducida para bandas de diferente ancho. ....	60
Tabla 4.1: Series hidrológicas registradas en la cuenca del Laja. ....	64
Tabla 4.2: Series hidrológicas registradas en la cuenca del Maule. ....	68
Tabla 4.3: Series hidrológicas registradas en la cuenca del Rapel. ....	71
Tabla 4.4: Series hidrológicas registradas en la cuenca del Chamiza. ....	73
Tabla 4.5: Series hidrológicas registradas en las cuencas del Limarí, Maipo, Pilmaiquén y Huanehue. ....	75
Tabla 4.6: Registros de información en las estaciones Colorado antes de la junta con Olivares y Olivares antes de la junta con Colorado. ....	78
Tabla 4.7: Series consideradas para el análisis de las propiedades hidrológicas. ....	87
Tabla 4.8: Propiedades estadísticas de las series anuales. ....	88
Tabla 4.9: Valores máximos de las asimetrías semanales de las series en estudio. ....	88
Tabla 5.1: Años empleados para la comparación en cada serie. ....	111
Tabla 5.2: Comparación de errores de pronósticos para invierno. ....	112

Tabla 5.3: Comparación de errores de pronósticos para verano.....	113
Tabla 5.4: Comparación de modelos PAR, ARMA y CDEC. ECM de los errores. ....	117
Tabla 5.5: Comparación de modelos PAR, ARMA y CDEC. Desviación típica de los errores.....	117
Tabla 5.6: Resultados de las corridas realizadas al modelo OMSIC. ....	120
Tabla A.1: Potencia bruta instalada en el SING. [Elec96] .....	136
Tabla A.2: Centrales termoeléctricas del SIC. [Cdec95].....	136
Tabla A.3: Centrales hidroeléctricas del SIC. [Cdec95] .....	137

## ÍNDICE DE FIGURAS

	Pág.
Figura 3.1: Esquema general de modelación.....	33
Figura 3.2: Esquemas de dependencia multivariada. ....	35
Figura 3.3: Esquemas alternativos del procedimiento para incorporar la normalización y la parametrización.....	45
Figura 4.1: Diagrama unilineal resumido de la cuenca del Laja. ....	66
Figura 4.2: Diagrama unilineal resumido de la cuenca del Maule. ....	69
Figura 4.3: Diagrama unilineal resumido de la cuenca del Rapel. ....	72
Figura 4.4: Diagrama unilineal resumido de la cuenca del Chamiza. ....	73
Figura 4.5: Diagrama unilineal resumido de la cuenca del Limarí.....	76
Figura 4.6: Diagrama unilineal resumido de la cuenca del Maipo.....	76
Figura 4.7: Diagrama unilineal resumido de la cuenca del Pilmaiquén. ....	77
Figura 4.8: Diagrama unilineal resumido de la cuenca del Huanehue. ....	77
Figura 4.9: Serie de valores anuales de Colbún.....	89
Figura 4.10: Promedios semanales de la serie Colbún. Se agrega la representación de los valores periódicos mediante series de Fourier.....	89
Figura 4.11: Desviaciones típicas semanales de la serie Colbún. Se agrega la representación de los valores periódicos mediante series de Fourier.....	90
Figura 4.12: Asimetrías semanales de la serie Colbún. Se agrega la representación de los valores periódicos mediante series de Fourier.....	91
Figura 4.13: Valores promedio de gasto en Laja en Tucapel. ....	95

Figura 4.14: Valores promedio de gasto en Estero El Toro en captación. ....	95
Figura 5.1: Pronóstico para las semanas siguientes al día 200, utilizando el modelo de la semana 26. El pronóstico se realiza con los datos de la semana 4.....	109
Figura 5.2: Pronóstico para las semanas siguientes al día 200, utilizando el modelo de la semana 27. El pronóstico se realiza con los datos de la semana 4.....	110
Figura 5.3: Pronóstico para las semanas siguientes al día 200, utilizando los modelos de las semanas 26 y 27 ponderados. El pronóstico se realiza la semana 4. Se incluye pronóstico del CDEC para la semana 5. ....	111
Figura 5.4: Errores de los pronósticos semanales con modelos PAR y CDEC. Serie Colbún en invierno para el año 1988-89. ....	113
Figura 5.5: Errores de los pronósticos semanales con modelos PAR y CDEC. Serie Colbún en verano para el año 1988-89.....	114
Figura 5.6: Errores de los pronósticos semanales con modelos PAR y CDEC. Serie Colbún en invierno para el año 1988-89. ....	118
Figura 5.7: Errores de los pronósticos semanales con modelos PAR y CDEC. Serie Colbún en verano para el año 1988-89.....	118
Figura A.1: Sistemas eléctricos chilenos [Cdec95].....	135

## RESUMEN

En esta tesis se desarrolla un estudio detallado de las características hidrológicas que definen la disponibilidad de recursos hídricos empleados por las centrales que conforman el Sistema Interconectado Central (SIC). El objetivo es validar la metodología de simulación y pronóstico actualmente en uso, así como visualizar alternativas para mejorar su modelación.

Inicialmente se realiza una revisión de los principales aspectos involucrados con el sistema eléctrico chileno, las empresas participantes y el rol del Centro de Despacho de Carga (CDEC) y de la Comisión Nacional de Energía (CNE), entre otros.

Resulta fundamental revisar de manera detallada las series cronológicas de interés en busca de períodos de “relleno” en ellas. El problema no lo presentan los rellenos en sí, sino que la forma en que éstos fueron realizados, ya que lo buscado era reproducir momentos de primer orden para poder representar la disponibilidad promedio de recursos. No obstante, con esto se introduce una mayor dependencia en los momentos de mayor orden, lo que lleva a que series de una misma cuenca o de cuencas vecinas resulten más dependientes entre sí de lo que realmente son.

Luego se aborda, de manera detallada, la forma de construir los modelos Periódicos Autoregresivos (PAR) utilizados como las alternativas de mejora de la modelación de los aspectos hidrológicos estudiados. Esto implica desde el análisis de las propiedades estadísticas a modelar hasta sus usos en la simulación de la operación y en el pronóstico. Como resultado es posible generar varias matrices hidrológicas sintéticas a partir de una serie única de valores históricos registrados. Esto permite obtener un perfil de costos marginales al utilizar modelos de operación, en particular, se realizaron corridas del modelo OMSIC, empleando como hidrologías de entrada las series generadas. Por otra parte, los modelos PAR son ocupados para realizar pronósticos de caudales entregando varios períodos con límites de confianza que permiten tener una idea precisa de la bondad del pronóstico. Adicionalmente, se hace una comparación de los ejemplos propuestos con los utilizados en el módulo hidrológico del modelo PRSI-SDDP.

## ABSTRACT

The appropriate modelation of hydrological time series for their use in power system operational studies and forecasting is an important issue. The main objective of the research reported herein was to validate methodologies for simulation and forecasting that are employed in the main Chilean electric interconnected system, as well as to visualize alternative methodologies.

First, a revision of the main characteristics of the Chilean electric power system is done, in particular of the main entities that are involved in the system operation : the electricity companies, the National Load Dispatch Center (CDEC), the National Energy Commission (CNE), etc.

An important aspect in the operation of the main Chilean system is related to the data that is used to feed the hydrothermal dispatch models. Since many of the hydrological series, that are related to the hydroelectric power plants, have been measured and registered since the early fifties or even later, missing data needs to be filled up. It is done through regression methods that are able to reproduce the relevant aspects to estimate the resource's availability. However, they underestimate higher order momenta, such as standard deviation, skewness, curtosis, etc. As a result, the filled up series are more dependant among each other than the real ones.

The building of Periodic Autoregressive models (PAR) is reviewed and an alternative methodology is proposed to consider the hydrological aspects involved in operational simulation and forecasting electric models. They are particularly relevant in the OMSIC model, which determines monthly operation of the central interconnected system. As a result, it is possible to obtain many streamflow sequences for a unique hydrological series. This allows the obtention of a set of marginal costs, as the output of the OMSIC model, instead of a unique signal. Several runs of the OMSIC model were made with the generated series.

The PAR models are also employed to forecast streamflows sequences. Finally, the result of a comparison with the hydrologic module of PRSI-SDDP model is presented.

## I. INTRODUCCIÓN

El crecimiento del país en los últimos años ha llevado a un aumento sostenido en la demanda de energía. Las principales fuentes de energía en Chile son el petróleo, la hidroelectricidad, la leña, el carbón y el gas. A partir de 1997, se espera un fuerte repunte del gas natural debido a la entrada en funcionamiento del gaseoducto de la empresa Gas Andes que trae gas desde la República Argentina.

El rol que ha jugado el sector eléctrico ha sido muy importante en el desarrollo nacional. Las altas tasas a las que ha crecido el PIB en los últimos 15 años han impactado de manera significativa el crecimiento de la demanda de energía eléctrica, a una razón promedio del 7% anual, llegando a prácticamente duplicar la energía demandada cada 10 años. Lo anterior ha influido significativamente en el desarrollo de los sistemas eléctrico de potencia (SEP) del país, en lo que respecta a su tasa de crecimiento y características propias. Un SEP es un conjunto de instalaciones que permite generar, transmitir y distribuir la energía eléctrica, en condiciones determinadas de tensión, frecuencia y disponibilidad. Se distinguen tres etapas principales por las cuales debe pasar la energía antes de llegar al consumidor final: generación, transmisión y distribución.

La primera etapa, que corresponde a la generación, puede realizarse mediante la utilización de centrales hidroeléctricas, térmicas, nucleares u otro tipo menos convencional. En Chile, la principal fuente de generación eléctrica son las centrales hidroeléctricas, las que dependen para su operación de la disponibilidad de recursos hídricos. Es por ello que en esta tesis se enfoca la atención sobre la forma y los métodos empleados en el Sistema Interconectado Central (SIC) para representar el comportamiento de la oferta de recursos con que cuenta el sistema.

Como anteriormente se indicó, la principal fuente de aleatoriedad respecto del funcionamiento de las centrales hidroeléctricas proviene de la disponibilidad del recurso hídrico, por esta razón resulta fundamental disponer de modelos que representen adecuadamente las series hidrológicas asociadas a las centrales de generación.

Los modelos estocásticos para representar series hidrológicas tienen por objeto reproducir las características estadísticas básicas de las series cronológicas de las variables de interés. Esta información está contenida en series históricas que se emplean para identificar los modelos más adecuados y estimar sus parámetros, de manera que existe una estrecha relación entre los modelos propuestos y las propiedades estadísticas que muestran la información disponible. Esto se traduce en que la validación de los modelos y su comportamiento necesariamente significan un análisis de los datos de las diferentes etapas de la modelación mediante la observación de resultados numéricos.

## **1.1       Objetivos**

El objetivo básico de esta tesis consiste en analizar las características hidrológicas que definen la oferta y disponibilidad de recursos hídricos empleados por las centrales que conforman el SIC, de manera de validar la metodología de simulación y pronóstico actualmente en uso, así como visualizar alternativas para mejorar su modelación.

Para cumplir este objetivo general se plantean los siguientes objetivos específicos: estudiar y analizar los aspectos hidrológicos en el modelamiento que se realiza en el SIC y proponer metodologías alternativas; realizar un estudio de las series hidrológicas que caracterizan la oferta de agua al SIC para poder observar la forma en que se comportan sus propiedades hidrológicas de interés, distinguir el efecto del relleno y la extensión de las estadísticas; analizar el modelo estocástico de caudales del modelo OMSIC (Operación Mensual del Sistema Interconectado Central), con el objeto de hacer observaciones tendientes a mejorar la solución utilizada en la actualidad o proponer un modelo alternativo; validar el desempeño de las metodologías en uso en la actualidad con las metodologías propuestas en base a la obtención de un perfil de costos marginales, utilizando el modelo OMSIC con matrices de hidrología actuales y matrices sintéticas; analizar el nuevo modelo que ha desarrollado el Centro de Despacho de Carga del Sistema Interconectado Central (CDEC-SIC), mediante consultoría externa; en especial, el módulo de hidrología, en el cual se usa un modelo estocástico de caudales afluentes a los embalses y aprovechamientos del SIC; y analizar los modelos de pronóstico de corto plazo de

caudales afluentes al sistema y comparar modelaciones alternativas en base a modelos estocásticos clásicos.

## **1.2 Características de los sistemas eléctricos de potencia**

En la generación de energía eléctrica es necesario destacar que la potencia eléctrica que es producida por una central se genera a una tensión que varía entre los 6.300 y 13.000 Volts. Este nivel de tensión es relativamente bajo y requiere ser elevado, ya que mientras menor es el voltaje, mayor es la intensidad de corriente, por lo que se necesitarían conductores de gran sección para llevar esta alta intensidad, lo que haría aumentar los costos. De esta manera se adecúa la tensión a la que se genera la energía eléctrica, en concordancia con la potencia de las centrales y con las distancias hacia donde se transporta, mediante transformadores que llevan el voltaje a: 500.000, 220.000, 154.000, 110.000 o 60.000 Volts.

Una vez que se ha obtenido el voltaje adecuado y la intensidad se ha reducido, la energía se reparte a los grandes centros de consumo por medio de líneas de transmisión de alta tensión. Antes que la electricidad llegue a los usuarios debe producirse una nueva transformación porque no es económico distribuir la energía a tan altas tensiones, para ello se cuenta con transformadores en subestaciones, que son los encargados de efectuar esta segunda modificación, al reducir la alta tensión de transmisión a media tensión, es decir, 13.200 ó 23.000 Volts.

La distribución en los centros urbanos se hace desde las subestaciones en las que se originan las líneas de media tensión. Estas líneas pueden ser aéreas, subterráneas o una combinación de ambas. Sin embargo, antes de llegar a los usuarios, sufre una última transformación donde la media tensión se reduce a baja tensión: 380 y 220 Volts. Esta última fase se efectúa por medio de transformadores de distribución que se instalan sobre postes, en cámaras subterráneas o en recintos cerrados para quedar en condiciones de alimentar las redes de distribución a nivel domiciliario.

Las principales características de cada una de las etapas reseñadas son las presentadas en la tabla 1.1.

Tabla 1.1: Etapas de un SEP y sus principales características.

Etapa	Principales Características
Generación	<ul style="list-style-type: none"> <li>•Diversidad de fuentes de generación (libertad de producción)</li> <li>•Aleatoriedad del recurso hidrológico</li> <li>•Restricción tecnológica de almacenamiento de energía</li> <li>•Largo período de gestación de proyectos</li> <li>•Economías de escala</li> <li>•Intensidad en el uso de capital</li> <li>•Alta tasa de crecimiento de la demanda</li> </ul>
Transmisión	<ul style="list-style-type: none"> <li>•Ventajas de los sistemas interconectados</li> <li>•Economías de escala</li> <li>•Intensidad en el uso de capital</li> </ul>
Distribución	<ul style="list-style-type: none"> <li>•Independencia entre áreas geográficas</li> <li>•Tendencia al monopolio natural en cada área</li> </ul>

La primera etapa que compone un SEP, la generación de energía eléctrica, puede ser realizada a través de distintos tipos de centrales, a saber, hidroeléctricas, termoeléctricas, nucleares y otras. De acuerdo a datos actualizados a 1988 [Vara95] la composición de la energía generada según el tipo de central, tanto en Chile como el promedio mundial, es la que se presenta en la tabla 1.2.

Tabla 1.2: Tipos de centrales utilizadas en la generación de energía en Chile y el mundo. [Vara95]

Tipo de central	Chile	Promedio Mundial
Hidroeléctrica	68%	19%
Termoeléctrica	30%	64%
Nuclear	0%	17%
Otros	2%	0%

Como se puede observar en la tabla anterior, Chile es un país con generación preponderantemente hidroeléctrica, siguiéndola en importancia la energía termoeléctrica y con nula presencia de centrales nucleares. Las principales características de las centrales hidroeléctricas y termoeléctricas se presentan a continuación:

#### a) Centrales hidroeléctricas

Las centrales hidroeléctricas o hidráulicas son las que utilizan como elemento motriz la fuerza mecánica del agua, la que actuando directamente sobre la turbina acoplada a un generador, pone a éste en movimiento, generando la energía eléctrica. Las centrales hidroeléctricas presentan la desventaja de tener limitaciones en la energía generable si hay limitaciones en el caudal disponible. Es precisamente ésto lo que le da la aleatoriedad al sistema de generación y hace que sea fundamental modelar el recurso hídrico de la mejor forma posible.

Las centrales hidroeléctricas se pueden clasificar de acuerdo a su capacidad de regulación, que es el volumen útil del embalse (volumen comprendido entre los niveles mínimo y máximo de operación bajo condiciones normales), distinguiendo de menor a mayor capacidad de regulación las siguientes situaciones típicas:

- i) Centrales de pasada: no poseen capacidad de almacenamiento, los aportes son generados a medida que los afluentes aportan sus caudales.
- ii) Centrales con regulación diaria o mensual: tienen un pequeño estanque de regulación en comparación con su capacidad de producción que les permite almacenar agua en las horas de menor demanda y generarla en las horas de máxima demanda. En algunos casos, pueden hacer transferencias desde los fines de semana a los días de trabajo.
- iii) Central con regulación mensual o estacional: el volumen del embalse es suficiente para almacenar agua por períodos largos: de un mes a otro, de una estación a otra. En general son capaces de almacenar las crecidas.
- iv) Central con regulación interanual: presentan grandes embalses, que permiten transferir agua entre años húmedos a secos.

Otra característica importante de las centrales hidroeléctricas son sus costos variables, que en general son muy bajos. En efecto, el agua tiene un valor económico, sin embargo, ninguna central debe realizar desembolsos asociados a su

generación. Por otra parte, su costo fijo es bastante alto debido, generalmente, a las obras civiles que es necesario realizar.

#### b) Centrales termoeléctricas

Las centrales termoeléctricas se diferencian por el tipo de combustible que usan, así se distinguen:

- i) Centrales a vapor: son plantas en que el elemento motriz es el vapor generado en una caldera, por el calor producido por la combustión de petróleo, gas o carbón. El vapor a presión actúa sobre una turbina a vapor acoplada a un generador, poniéndolo en movimiento y produciendo energía eléctrica.
- ii) Centrales Diesel: son el tipo más simple de instalaciones generadores, consta de un generador movido por un motor de combustión interna a petróleo diesel.
- iii) Turbinas a gas: son un generador accionado por una turbina que es movido por la presión de gases de combustión de petróleo y sus derivados.
- iv) Ciclo combinado: el funcionamiento de la central consiste en dos ciclos. En el primero se quema el gas y mediante la combustión produce gases de alta temperatura y energía que mueven una turbina (a gas), la que a su vez mueve un generador eléctrico que entrega a la red. Los gases de escape que salen de esa turbina tienen una temperatura muy alta. Esa energía se aprovecha en una caldera recuperadora de calor para producir vapor, el que alimenta otra turbina, la que mueve un generador a vapor.

Los costos variables de éstas son más altos que los de las centrales hidroeléctricas y están estrechamente relacionados con el tipo de combustible que ocupa la central. Su costo fijo es significativamente menor que el de las centrales hidroeléctricas.

Una comparación entre las centrales hidroeléctricas y termoeléctricas respecto de sus características de flexibilidad y disponibilidad se presentan en la tabla 1.3.

Tabla 1.3: Comparación entre centrales hidroeléctricas y termoeléctricas.

Concepto	Central hidroeléctrica	Central termoeléctrica
Flexibilidad	<ul style="list-style-type: none"> <li>•Parten (se detienen) y alcanzan plena carga en minutos</li> <li>•Prácticamente funciona a cualquier potencia</li> <li>•Debe estar ubicada en las inmediaciones del recurso hídrico</li> </ul>	<ul style="list-style-type: none"> <li>•Estando fría, demora horas en partir</li> <li>•No son estables bajo la potencia mínima de la unidad</li> <li>•Existe grado de flexibilidad en la ubicación física</li> </ul>
Disponibilidad	<ul style="list-style-type: none"> <li>•Gran limitante es el aporte hidrológico</li> <li>•Con capacidad de regulación, pueden dar su potencia por algunas horas</li> <li>•En sistemas predominantemente hidroeléctricos, los problemas de abastecimiento se producen por falta de energía, y no de potencia</li> <li>•Disponibilidad mecánica (falla o mant. programada) no afecta prácticamente a la generación (80 a 82%)</li> </ul>	<ul style="list-style-type: none"> <li>•Requerimientos de mantención periódica (cada 1 a 2 meses)</li> <li>•Salidas intespectivas de servicio</li> <li>•Disponibilidad disminuye a medida que aumenta el tamaño de la central</li> <li>•Problemas en los primeros años de operación de la central (problemas de “juventud”)</li> </ul>

### 1.3 Requerimientos de despacho

El despacho consiste en abastecer la demanda (cargas eléctricas) de acuerdo a las disponibilidades de centrales generadoras. Dado que no se puede acumular energía eléctrica, ésta debe ser generada para su consumo de manera instantánea. Las cargas son consumos que varían en el tiempo en forma aleatoria (a nivel individual), pero que a nivel de red de transmisión/distribución son de carácter

predecible, debido a que las variaciones son relativamente lentas (régimen permanente); son dependientes de la tensión y de la frecuencia; trifásicas equilibradas, en general; y las cargas típicas consumen siempre potencia reactiva.

## **II. EL SISTEMA ELÉCTRICO CHILENO**

### **2.1 Características generales de los sistemas eléctricos chilenos**

Los primeros sistemas de alumbrado público eléctrico [Sala94] se establecieron en Londres en 1882 y ya en 1883 se estrenaban en Chile, iluminando la Plaza de Armas y sus cercanías; sin embargo, fue en 1897 cuando se instala la primera central hidroeléctrica, Chivilingo, a 10 kms del mineral de Lota.

Por otra parte, en 1943, por un acuerdo del Consejo de CORFO para llevar a cabo la gran misión de ejecutar un Plan de Electrificación de Nacional, se crea así la Empresa Nacional de Electricidad S.A., ENDESA. Entre 1944 y 1948 se instalan las primeras tres centrales hidroeléctricas a cargo de ENDESA: Pilmaiquén. El Abanico y Sauzal. En 1952 se construyeron Los Molles y Cipreses. La puesta en marcha de esta última central marca el inicio del primer sistema eléctrico de potencia: el SIC al unir desde La Ligua hasta Victoria.

A continuación se presentan los principales sistemas eléctricos de Chile, estudiando con más detalle el Sistema Interconectado Central.

#### **2.1.1 Sistema Interconectado del Norte Grande (SING)**

Sólo con 40 MW de potencia contaba la zona norte del país a comienzos de siglo, cuando en 1915 entró en operaciones Tocopilla, la primera central de energía eléctrica del extremo norte. [Elec96]

En la actualidad, con cuatro empresas generadoras de servicio público (ver tabla A.1 ANEXO A), el Sistema Interconectado del Norte Grande posee una potencia bruta instalada del orden de 1119 MW y con las estimaciones de crecimiento de la demanda se prevee el surgimiento de nuevos proyectos de generación que en la primera década del próximo siglo duplicarían la actual potencia.

El fenómeno observado en el SING durante los últimos años se ha debido, fundamentalmente, a la demanda generada por las compañías mineras. En la zona, el 88% del consumo energético corresponde a clientes libres, siendo casi la totalidad de ellos empresas mineras.

En el ANEXO A se presentan las centrales que componen el SING.

### **2.1.2 Sistemas de la Zona Austral (XI y XII Regiones)**

En la zona sur del país hay otros sistemas de generación de energía eléctrica de servicio público. Uno abastece la XI Región, con minicentrales y el otro abastece la XII Región. Éstos también quedan fuera del sistema en estudio, por no constituir parte del SIC.

### **2.1.3 Sistema Interconectado Central (SIC)**

El Sistema Interconectado Central es el sistema eléctrico de servicio público más importante del país, cubre aproximadamente 2.100 kms del territorio nacional, desde Taltal por el norte, hasta la Isla Grande de Chiloé por el sur, como se muestra en la figura A.1 del ANEXO A. El SIC está constituido por la interconexión de los sistemas eléctricos de transmisión y las centrales generadoras de las empresas eléctricas CHILGENER S.A., COLBÚN S.A., ENDESA, PEHUENCHE S.A. y GUACOLDA S.A. A este sistema se interconectan, además, otras empresas generadoras, tales como Hidroeléctrica Guardia Vieja, Empresa Hidroeléctrica Pullinque S.A., Empresa Eléctrica Pilmaiquén S.A., y otras de menor capacidad de generación.

La potencia instalada del SIC era de 4.107 MW en 1995, de los cuales 77% corresponde a instalaciones hidroeléctricas y el resto a termoeléctricas, lo que representa cerca del 75% del total instalado para el servicio público de la nación. El saldo lo suministran otros sistemas eléctricos.

El SIC puede visualizarse como una red compuesta por más de 35 centrales generadoras que se unen por líneas de transmisión, que se emplean, a su vez, para llevar la energía a todos los puntos de consumo. En las tablas A.2 y A.3 del ANEXO A se presentan las centrales que componen el SIC y un mapa de la zona referenciada.

## **2.2 Comparación con otros sistemas latinoamericanos: sus marcos regulatorios**

En los últimos años, en Chile y en otros países latinoamericanos, se han llevado a cabo cambios en los sectores eléctricos de manera de modificar su estructura. Se ha pasado de una propiedad mayoritariamente estatal y de decisiones centralizadas a una participación principalmente privada en la propiedad de las empresas y a una gran descentralización de las decisiones generadas por las condiciones del mercado.

No obstante, los objetivos que han perseguido los distintos marcos regulatorios han sido comunes, las soluciones adoptadas en cada uno de los países considerados en el estudio no son idénticas [Gati96]. Esta diferencia en las soluciones adoptadas resulta razonable si se considera que los sistemas eléctricos considerados difieren en tamaño, importancia relativa de la generación hidroeléctrica, grado de electrificación y topología del sistema.

En [Gati96] se efectúa un análisis del marco conceptual que permite identificar los principales elementos a considerar y una presentación detallada de las soluciones adoptadas en Argentina, Bolivia, Chile, Colombia y Perú.

Dada las características técnicas y económicas propias del sector, en la nueva estructura que se ha generado no se ha logrado una completa desregulación, por lo que se han establecido normas que tratan de garantizar y promover la existencia de mercados competitivos cuando se ha estimado que ello es posible. En tanto, para los segmentos del negocio para los que no es posible lograr esto por la existencia de monopolios naturales, se han establecido normas destinadas a preservar la eficiencia económica, manteniendo la propiedad en agentes privados.

En todos los casos analizados se reconoce la existencia de tres negocios separados: generación, transmisión y distribución. También se acepta que el mercado de la generación es competitivo y, la transmisión y distribución son monopolios naturales. Tomando en cuenta lo anterior, en todos los países se han establecido condiciones para que el propietario se obligue a otorgar libre acceso a terceros, mediante el pago de peajes, al uso de sus instalaciones de transmisión y

distribución. En este sentido cabe destacar que en todos los países, excepto Chile, se han impuesto limitaciones a la propiedad, en cuanto a porcentajes máximos de propiedad de un determinado segmento (límites a la integración horizontal) y a la propiedad en más de un segmento en forma simultánea (límites a la integración vertical), con la intención de resguardar la competencia en el sector.

La normativa que regula al sector se plantea en dos niveles: i) leyes: requieren aprobación del Poder Legislativo, situación que dificulta posteriores modificaciones; ii) reglamentos: son dictados por el Poder Ejecutivo. Contienen normas y procedimientos detallados. Relacionado con el contenido y alcance de leyes y reglamentos se identifican distintos estilos: desde leyes detalladas que especifican procedimientos de cálculo (por ejemplo, en Chile y Perú), hasta leyes de tipo general que delegan el detalle de lo operativo en los reglamentos (Argentina y Colombia). En este último caso, las reglas pueden modificarse más fácilmente, otorgando así mayor flexibilidad y un continuo perfeccionamiento. Como contraparte a lo anterior, una excesiva tendencia a modificar las reglas del juego puede ser percibida como una señal de uso discrecional del poder que en definitiva se traduce en un mayor riesgo para los inversionistas.

### **2.2.1 El sector generación**

En todos los casos se establece que el desarrollo de la generación se hará a mínimo costo por iniciativa de agentes privados. En el caso de Colombia, la expansión debe corresponder a la de mínimo costo social. En los casos de Bolivia, Chile y Perú, los organismos encargados de definir la política sectorial publican planes indicativos de generación-transmisión.

Respecto de la operación y despacho, se acepta la necesidad de una operación coordinada de las centrales generadoras, por razones económicas y de seguridad de abastecimiento. Las decisiones de operación son determinadas por un organismo encargado del despacho (OED) y obligan a todos los generadores conectados que participen en dicho organismo. Existen distintas concepciones de los OED, las cuales difieren en la naturaleza de los integrantes (desde OED constituidos por representantes del Gobierno, generadores, transportistas, distribuidores y consumidores en Argentina y Bolivia, hasta OED conformado sólo por parte de los

generadores de servicio público como en Chile), grado de autonomía operacional (personal y activos propios o aportados por las empresas integrantes), mecanismos de decisiones (unanimitad, mayoría simple) y de soluciones de divergencias.

Dadas las proyecciones de demandas a abastecer, la disponibilidad del parque generador, el estado de los embalses y las proyecciones hidrológicas, el despacho económico es consecuencia de los costos variables de operación de las unidades térmicas y del valor marginal del agua almacenada en los embalses. Para los costos variables de los generadores térmicos existe el caso en que los precios de los combustibles están asociados a costos internacionales relevantes (Bolivia), casos en que los generadores declaran costos reales de producción auditados (Chile y Perú) y casos con declaraciones de costo de producción no auditados, ya sea con restricciones en torno a valores de referencia (Argentina) o sin limitaciones (Colombia). Por su parte, el valor marginal del agua de las centrales con embalses puede ser determinado por el OED mediante una optimización centralizada de la operación, utilizando modelos matemáticos ad hoc (caso de Bolivia, Chile y Perú) o bien ser declarado por el generador (caso de Argentina y Colombia).

A nivel de generación es posible distinguir dos tipos de mercados:

- i) mercado *spot*: en el cual se transan excedentes de generación con respecto a los compromisos de cada generador. Distintas concepciones quedan en evidencia al comparar las normas de Chile y Colombia. En el primer caso sólo los generadores participan en el mercado *spot* y las transacciones sólo se hacen entre ellos. En Colombia, los generadores venden su producción al precio *spot* en la bolsa de energía y se ha definido formalmente la función de comercialización, la que puede ser realizada tanto por agentes especializados (*broker*) como distribuidores o generadores.
- ii) mercado de contratos: también denominado mercado a término, en el cual un generador o un *broker*, establece contratos de abastecimiento.

Para efectos de comercialización, sólo en Colombia se reconoce la figura de comercializadores puros. En Argentina se ha dictado recientemente una resolución que permitiría esta figura, no obstante no está reglamentada su puesta en

práctica. A fin de establecer contratos se exige en todos los casos, salvo en Colombia, contar con respaldo de generación firme propia o contratada con otros generadores.

En el mercado spot, el precio de la energía corresponde al costo marginal de corto plazo, varía horariamente y depende de la ubicación geográfica y del nivel de tensión. Estos precios son determinados por el OED. El costo marginal de potencia corresponde al costo de desarrollo de capacidad de punta (turbina a gas), salvo en Argentina en que el precio ha sido fijado por la autoridad.

En el mercado de los contratos, los precios pueden ser libremente negociados entre las partes o regulados por la autoridad.

### **2.2.2 El sector transmisión**

Como se indicó anteriormente, la obligatoriedad de otorgar libre acceso a terceros al uso de las instalaciones de transmisión es un factor clave para crear condiciones de competencia en el sector.

La aplicación de la tarificación a costo marginal consiste en pagar a los generadores por la potencia inyectada al sistema y cobrar a los comercializadores por lo retirado al costo marginal instantáneo en la barra correspondiente. El monto total pagado por los comercializadores supera los pagos a los generadores, debido al efecto derivado de pérdidas marginales superiores a las pérdidas medias y a desacoplamientos de costos marginales en los extremos de líneas copadas. Esta diferencia permite remunerar parcialmente al sistema de transmisión. Se llama peaje al pago complementario requerido para cubrir los costos de los sistemas de transmisión.

En el caso argentino, las líneas privatizadas hundieron parte de sus costos, y perciben el ingreso correspondiente a los costos marginales de energía, ingresos por calidad de la vinculación entre extremos de línea, además de cargos fijos destinados a rentar instalaciones de conexión y costos fijos de operación. Los nuevos sistemas logran financiarse mediante el pago de un canon por parte de los beneficiarios, generadores y consumidores, el cual resulta de una licitación pública.

En Bolivia se determinan áreas de influencia de responsabilidad de generadores y/o consumidores. El valor máximo es calculado por la autoridad, considerando un sistema adaptado. En el precio regulado se incluye el cargo que es responsabilidad del consumo. Los generadores pagan su parte correspondiente en proporción a su potencia firme.

En Chile, se define un área de influencia para cada generador. El pago de peaje en las líneas que son área de influencia de varios generadores se reparte en proporción a la potencia máxima transitada. Las líneas que no pertenecen al área de influencia de ningún generador deben ser pagadas por los generadores que abastecen a los consumidores conectados a ellas. La metodología para el cálculo de los peajes está determinada en la ley, pero tanto la proporción a pagar, como el valor de las instalaciones está sujeto a acuerdo entre las partes o arbitraje. Los precios regulados no incluyen peajes.

En Colombia se calculan, para cada nudo, cargos por conexión y por uso, sin considerar la señal dada por las pérdidas marginales. Los generadores pagan por inyectar potencia a la red y los comercializadores por retirar la energía. Los cargos deben ser tales que un 50% del costo anual de la red sea pagado por generadores y un 50% por comercializadores.

En el caso peruano se distingue el sistema de transmisión principal y el sistema de transmisión secundario. El sistema principal está constituido por instalaciones que tienen un rol de interconexión, en la que se producen intercambios entre generadores y libre comercialización de la energía eléctrica. El pago al sistema principal se basa en los costos de inversión y operación de un sistema económicamente adaptado. El peaje anual está regulado y es pagado por todos los generadores en proporción al reconocimiento de potencia firme. El sistema secundario está asociado a roles de conexión de generadores al sistema principal y de suministro a clientes a partir de éste. Su pago es negociado entre las partes. El peaje se agrega al cargo por potencia cobrado a los clientes regulados.

### **2.2.3 El sector distribución**

Se reconoce que la distribución es una actividad monopólica que debe ser regulada. Se otorgan conexiones, en la mayoría de los casos exclusivas, para la instalación de redes de distribución destinadas a proveer este servicio en un área geográfica determinada. Como contrapartida existe obligación de dar servicio en ella. En este sentido se diferencia de las actividades de generación y transmisión en las cuales, dada la mayor competencia, no existe obligación de servicio.

En general se impone la obligación de dar acceso a los generadores o comercializadores para abastecer clientes dentro del área de concesión, sujeto al pago de peaje por las instalaciones utilizadas.

Se han planteado dos modalidades para determinar la remuneración de la actividad de distribución: a) Comparación con una empresa tipo o modelo. La empresa de distribución compite contra una empresa eficiente, con lo cual la obliga a maximizar su eficiencia; b) Reconocimiento de costos efectivos.

Se establecen precios libremente negociados para consumidores mayores que un determinado tamaño (fijado en 100 KW en Argentina y en 2 MW en Bolivia, Chile y Colombia) y precios regulados para los consumidores menores.

Los precios regulados se forman a partir de un precio de compra en alta tensión más los costos correspondientes al servicio de distribución. El costo de las compras en alta tensión puede estar valorizado al precio regulado (Argentina, Chile y Perú), al precio de los contratos de suministro si éstos resultan de licitaciones públicas supervisadas (Bolivia) o a un precio de referencia no necesariamente relacionado con los costos reales de abastecimiento. Sólo en Argentina y Bolivia las empresas distribuidoras tienen acceso al mercado spot.

## **2.3 Organización del SIC**

### **2.3.1 Empresas participantes en el SIC**

Las empresas que participan en el SIC son CHILGENER, COLBÚN, ENDESA, PEHUENCHE y GUACOLDA y otras de menor generación. En la tabla

2.1, se presenta la participación de cada una de ellas, en términos de capacidad de generación de electricidad.

Tabla 2.1: Participación de empresas en el SIC. [Cdec95]

Empresa	Termoeléctricas (MW)	Hidroeléctricas (MW)	TOTALES (MW)
Chilgener	511,5	244,9	756,4
Colbún (1)	0,0	538,6	538,6
Endesa	252,7	1602,7	1855,4
Pehuenche	0,0	585,0	585,0
Guacolda	150,0	0,0	150,0
Otras (2)	17,4	198,0	215,4
SUBTOTALES	931,6	3169,2	4100,8
Autoprodutores (3)	0,0	6,5	6,5
TOTALES	931,6	3175,7	4107,3

(1) Incluye Colbún, Machicura y Pullinque.

(2) Incluye Pilmaiquén, Guardia Vieja, Florida, Carbomet, Sauce Andes, Aconcagua, E Verde y Capullo.

(3) La potencia aportada por los autoprodutores se estimó suponiendo una entrega de energía promedio de ocho horas.

a) Empresa Nacional de Electricidad S.A., ENDESA.

ENDESA nació en 1943 y en la actualidad participa en el país con más de 16 centrales, con una potencia total que representa un 60% del total del SIC. En 1980 se modificó su estatuto social, dejando la responsabilidad del Plan de Electrificación a cargo de la Comisión Nacional de Energía, CNE, y pasó de ser una empresa del estado a una privada que hoy en día cuenta con más de 50 mil accionistas, 2300 trabajadores y sigue un proceso de crecimiento a nivel nacional e internacional. Es así como, a comienzos de esta década adquiere el 70% de las acciones de PEHUENCHE S.A. a CORFO, pasando a ser la propietaria de un 95%, y en 1992, ENDESA como parte de un consorcio, se adjudica un 60% de la propiedad de la Central Costanera en Buenos Aires. Luego se incorpora en EDESUR, empresa distribuidora en Argentina, para en 1993 adjudicarse una participación en el Complejo Hidroeléctrico El Chocón - Arroyito. Además, ha adquirido centrales en

Colombia. En el ámbito nacional se han puesto en funcionamiento las centrales Curillín y Mejillones, y la central Pangue con 450 MW, en el Alto Bío-Bío.

b) CHILGENER S.A.

CHILGENER o Chilectra Generación nace de la Compañía de Electricidad Ltda., empresa que surgió como una empresa privada de generación y distribución de energía eléctrica el 1º de Septiembre de 1921, de la fusión de la Chilean Electric Trainway and Light Co., fundada en 1889 y de la Compañía Nacional de Fuerza Eléctrica, que operaba desde 1919.

El 14 de Agosto de 1970, Chilectra es estatizada mediante la promulgación de la Ley N°17.323 que autorizó a la Corporación de Fomento de la Producción para adquirir todas o parte de las acciones y bienes de la compañía, mediante un contrato de compraventa, eliminándose la participación privada en la empresa.

En el año 1981 se produce una reestructuración que contempló un esquema funcional tipo holding mediante la creación de una matriz Chilectra S.A., y tres empresas filiales: Compañía Chilena Metropolitana de Distribución Eléctrica S.A. (CHILECTRA METROPOLITANA) para distribuir energía eléctrica en la Región Metropolitana, Compañía Chilena Metropolitana de Distribución Eléctrica Quinta Región S.A. (CHILECTRA QUINTA REGION), para servir a Valparaíso y el Valle del Aconcagua, y Compañía Chilena de Generación Eléctrica S.A. (CHILECTRA GENERACIÓN), que mantuvo las funciones de generación y transmisión. Posteriormente, estas tres empresas se separan y privatizan independientemente.

La empresa, integrando un consorcio internacional se adjudicó el 60% de las acciones de la Central Costanera S.A. de Buenos Aires, Argentina, además de otras inversiones, en Colombia por ejemplo, que manifiestan el dinamismo de este mercado. Durante el transcurso del presente año se incorporará la central Nueva Renca con 370 MW.

c) Empresa Eléctrica Colbún-Machicura S.A.

La creación de la Empresa Eléctrica Colbún Machicura S.A. tiene su origen en la división de la Empresa Nacional de Electricidad S.A., ENDESA, que en su Vigésima Séptima Junta Extraordinaria de Accionistas celebrada el 29 de Abril de 1986, acordó dividir la Empresa, con efecto retroactivo al 1° de enero del mismo año, en una sociedad continuadora de la misma y en una nueva sociedad anónima, que se constituyó bajo la razón social de Empresa Eléctrica Colbún Machicura S.A. En la actualidad posee tres centrales: Colbún, Machicura y San Ignacio.

Recientemente se ha iniciado su privatización definitiva con la entrada de capitales privados con paquetes significativos dentro de la propiedad de la empresa.

d) Empresa Eléctrica Pehuenche S.A.

La Empresa Eléctrica Pehuenche S.A. se constituyó por escritura pública otorgada con fecha 1° de abril de 1986. Su principal accionista es la Empresa Nacional de Electricidad S.A., ENDESA, cuya participación alcanza a un 92,55% de la propiedad. Posee dos centrales: Pehuenche y Curillinque. Durante el presente año entrará en funcionamiento la central Loma Alta de esta empresa.

e) Empresa Eléctrica Guacolda S.A.

Esta central entró en funcionamiento en el año 1995 con la puesta en marcha de la unidad n°1 de 150 MW. Durante Agosto de 1996 entra en funcionamiento la segunda unidad de idénticas características.

Su principal accionista es CHILGENER que posee el 50% de la propiedad de la empresa.

### **2.3.2 Centro de Despacho Económico de Carga (CDEC)**

La creación del CDEC-SIC obedece a la necesidad de compatibilizar, por una parte la operación del sistema de tarificación a costo marginal establecido en la Ley General de Servicios Eléctricos y por otra, el cumplimiento de suministros

suscritos por las empresas generadoras integrantes con la operación a mínimo costo del conjunto del parque generador del sistema.

Los antecedentes legales que llevaron a la constitución del CDEC-SIC se originan en el Decreto con Fuerza de Ley N° 1 de 1982 del Ministerio de Minería que corresponde a la Ley General de Servicios Eléctricos, que establece en su Artículo N° 81, entre otras disposiciones, que los concesionarios de cualquier naturaleza que operen sistemas eléctricos interconectados entre sí, deberán coordinar dicha operación con el fin de: i) preservar la seguridad del servicio en el sistema eléctrico; ii) garantizar la operación más económica para el conjunto de las instalaciones del sistema eléctrico; y iii) garantizar el derecho de servidumbre sobre los sistemas de transmisión establecidos mediante concesión.

El Artículo 91° del Decreto mencionado, establece que las transferencias de energía entre empresas eléctricas generadoras que resulten de la aplicación de la coordinación de la operación a que se refiere el Artículo N° 81, serán valorizadas de acuerdo a los costos marginales instantáneos del sistema eléctrico.

Posteriormente, el Decreto N° 6 de febrero de 1985 del Ministerio de Minería aprueba el Reglamento de Coordinación de la Operación Interconectada de Centrales Generadoras y Líneas de Transporte, disponiendo las condiciones que deben cumplir las empresas que están obligadas a coordinar su operación, las funciones básicas y la organización del Comité de Operación que es denominado Centro de Despacho Económico de Carga CDEC. Dicho Reglamento establece además, que para cumplir con los propósitos señalados en los Artículos 81° y 91° del D.F.L. 1/82, el CDEC realizará, entre otras, las siguientes funciones:

- i) Planificar la operación de corto plazo del sistema eléctrico, considerando la operación actual y esperada para el mediano y largo plazo, y comunicarla a los integrantes del CDEC para que éstos operen sus instalaciones de acuerdo a los programas resultantes.
- ii) Calcular los costos marginales instantáneos de energía eléctrica que se derivan de la planificación de la operación.

- iii) Coordinar la mantención preventiva mayor de unidades generadoras.
- iv) Verificar el cumplimiento de los programas de operación y de mantención preventiva mayor, adoptando las medidas correctivas que se requieran.
- v) Determinar y valorizar las transferencias de electricidad entre los integrantes del CDEC.

En materia de organización, el Reglamento establece que el CDEC estará compuesto por un Directorio y una Dirección de Operación, ambos integrados por representantes de las respectivas empresas participantes.

El organismo operativo, es decir, la Dirección de Operación, tiene como misión principal adoptar los acuerdos de operación que permitan cumplir las funciones básicas establecidas para el CDEC señaladas anteriormente, acordar programas, metodologías, modelos matemáticos, esquemas de medida, valores de parámetros, etc., utilizados para definir la operación del SIC.

En conformidad con las diversas disposiciones legales antes expuestas, en mayo de 1985 se formalizó la constitución del CDEC-SIC, quedando integrado por ENDESA y CHILGENER. Como resultado de la constitución de la Empresa Eléctrica Colbún-Machicura S.A. como filial de CORFO el 30 de abril de 1986, esta empresa fue incorporada al CDEC-SIC en mayo de 1986. En abril de 1991 la Empresa Eléctrica PEHUENCHE S.A. pone en servicio la central homónima y más recientemente se incorpora al CDEC-SIC la Empresa Eléctrica GUACOLDA S.A., quedando entonces el organismo integrado por cinco empresas.

Cada empresa productora integrante del CDEC-SIC, debe estar en condiciones de satisfacer en cada año su demanda de energía aún bajo condiciones de hidrología seca, en sus centrales hidroeléctricas, y considerando disponibilidad promedio en sus unidades térmicas, ésto para asegurar la calidad de servicio a que está comprometido el CDEC. Con este objeto, anualmente cada empresa integrante debe presentar al CDEC un balance de Potencia y Energía Firme, que demuestre que está en condiciones de asegurar el abastecimiento de las demandas de sus clientes, con los criterios de seguridad exigidos.

Como la capacidad instalada de las empresas no coincide con la demanda que tienen que satisfacer, hay transferencias de energía entre las empresas generadoras, de modo de aprovechar la capacidad instalada o no tener demanda insatisfecha. Por este efecto, para una empresa, además de considerar la capacidad de generación propia, se toman en cuenta los suministros de otras entidades generadoras, sean o no miembros del CDEC, que operen en sincronismo con el sistema y cuya producción parcial o total le haya sido asegurada mediante un contrato suscrito a precio libremente convenido.

La demanda de energía a considerar corresponde entonces a la suma de todos los suministros de energía previstos, ya sea para sus propios clientes o bien para otros integrantes del CDEC, además de considerar las pérdidas de transmisión. Por lo tanto, la suma de los aportes de energía de centrales hidráulicas para condiciones de hidrología seca, de los aportes de centrales termoeléctricas bajo condiciones de disponibilidad promedio, de los contratos de compra de energía firme a otras entidades generadoras conectadas al sistema, constituyen la energía firme disponible de cada empresa integrante y debe ser suficiente para abastecer su demanda anual de energía.

Una vez alcanzado el equilibrio entre las disponibilidades firmes y contratos de suministro a clientes de acuerdo a los balances hechos, la producción de las centrales generadoras del sistema se decide independientemente de los compromisos de la empresa propietaria de dichos medios de generación.

El criterio de optimización que aplica el CDEC-SIC para operar el sistema, como se ha mencionado, consiste en minimizar el costo global actualizado de operación y falla, preservando la seguridad del sistema eléctrico. Este criterio implica independencia de las preferencias particulares de cada empresa generadora, tomándose las decisiones en función de parámetros que son aceptados por todos los participantes, tales como: cotas de los embalses, precios de combustibles, demandas, etc. En consecuencia, se programa la generación de las unidades que se justifiquen económicamente, en virtud de que su costo marginal sea inferior o igual al costo marginal del sistema, costo calculado por el CDEC.

La optimización de la operación del conjunto del parque generador y la existencia de contratos de suministro de cada empresa generadora con clientes directos, origina transferencias de energía entre las empresas productoras. Estas transferencias de energía se valorizan al costo marginal del sistema, según se indica en el Decreto N°6 en el Título IV, que trata sobre las transferencias de energía eléctrica entre integrantes del CDEC. Mayores detalles del funcionamiento del CDEC y la normativa legal que lo rige se presentan en el ANEXO B.

## **2.4 Necesidades de modelación para la operación del SIC**

Existen necesidades de modelación tanto para la planificación del sistema a largo plazo, como la operación de las centrales a corto y mediano plazo.

En términos generales, el objetivo que se persigue con el desarrollo de los modelos de planificación es simple y consiste en determinar un programa de instalación de obras de generación y transmisión que conduzca a que el costo de abastecimiento (la suma de los costos de inversión, operación y distribución) del sistema eléctrico sea mínimo. La operación de un sistema eléctrico, en general, tiene por objetivo abastecer la demanda a un mínimo costo, respetando condiciones de seguridad, de calidad de servicio y de entorno.

Las necesidades de modelación para la operación del SIC son satisfechas por las empresas generadoras y por el CDEC mediante la utilización de los modelos GOL (Gestión Óptima del Laja) y OMSIC (Operación Mensual del Sistema Interconectado Central). Más recientemente se ha buscado incorporar un modelo de operación económica multiembalse usando Programación Dinámica Dual Estocástica (SDDP).

### **2.4.1 Modelos empleados por las empresas generadoras y el CDEC**

Los principales modelos utilizados por las empresas generadoras y el CDEC son el GOL y el OMSIC. Adicionalmente, se ha desarrollado el modelo multiembalse SDDP-PRSI.

Dado que en la actualidad, la generación en el SIC es principalmente de origen hidroeléctrico, complementada con generación proveniente de centrales a

carbón y petróleo. Dentro de las centrales hidroeléctricas, las hay de pasada, de embalse con pequeña capacidad de regulación y finalmente aquellas asociadas al embalse del Lago Laja, cuya capacidad de regulación es interanual. Resulta entonces evidente que la gestión de este embalse condiciona la generación termoeléctrica y en consecuencia los costos de operación del sistema. En líneas gruesas, la operación del sistema a costo mínimo significa determinar qué volumen del Lago Laja debe ser empleado en la generación en un período determinado de tiempo, dada su incidencia en los costos de operación presente y futura. Naturalmente, lo anterior supone una gestión adecuada del resto de las centrales: máxima utilización posible de las centrales de pasada, traspasos óptimos en las centrales con menor capacidad de regulación y utilización de las centrales térmicas según sus costos marginales de operación crecientes. En el caso de Chile, la hidrología juega un papel de mucha importancia y debe tomarse en consideración su carácter aleatorio.

En relación con lo anterior el Modelo Gestión Óptima del Laja (GOL) [Cne93] fue concebido como un modelo de programación dinámica estocástica, donde el consumo está centrado en una barra única, optimiza extracciones del Lago Laja, minimiza el costo actualizado de generación termoeléctrica y falla en todo horizonte de planificación. Utiliza criterio azar-decisión y considera etapas trimestrales. Se deja de usar por su prolongado tiempo de reacción ante cambios en las reservas del Lago Laja, ya que se actualizaba cada tres meses; no se representaba la estacionalidad de los consumos dentro de los tres meses y no consideraba la capacidad de regulación de los embalses menores.

Por otra parte, el Modelo Operación Mensual del Sistema Interconectado Central (OMSIC) [Cdec91] es un modelo de operación del SIC con etapas mensuales que utiliza programación dinámica y que optimiza las extracciones del Lago Laja y la generación de las centrales térmicas, considerando las centrales de pasada. En este modelo la hidrología se considera independiente entre meses en el período de invierno (Abril-Septiembre) y dependiente en el período de verano (Octubre-Marzo, meses de deshielo), y las decisiones se toman en la modalidad azar-decisión. Este modelo considera dos etapas, una de optimización y otra de simulación. Adicionalmente, existe una serie de modelos para las centrales que cuentan con embalses (modelo Rapel, de la Laguna Invernada, del embalse Melado, de los

embalses Colbún-Machicura y Canutillar. Estos modelos reciben una señal de costo marginal de producción del modelo OMSIC y a partir de ella entregan la energía que económicamente resulta óptima generar a ese precio. Con esa energía se verifica si se satisface la demanda. De ser posible el modelo ha convergido y ese es el costo marginal a utilizar. De no ser así, se inicia una nueva iteración del modelo. A la interacción entre los modelos de embalses y el modelo OMSIC se le denomina proceso de convergencia.

Adicionalmente, el Modelo de Despacho Hidrotérmico con Restricciones de Transmisión Programación Dinámica Dual Estocástica desarrollado por Power Systems Research Inc. (SDDP-PSRI) [Powe94] se compone de dos módulos principales, el primero efectúa un cálculo de la política de operación hidrotérmica, para lo cual determina la política operativa más económica para los embalses, teniendo en cuenta las incertidumbres en las afluencias futuras y las restricciones en la red de transmisión. El segundo realiza una simulación de la operación del sistema a lo largo del período de planificación, para distintos escenarios de secuencias hidrológicas. Calcula índices de desempeño tales como el promedio de los costos operativos, los costos marginales por barra y por bloque de carga, y los intercambios óptimos entre empresas.

Mayor detalle de los modelos anteriormente descritos se presenta en el capítulo V en la sección 5.2 de esta tesis.

#### **2.4.2 Modelos empleados por la CNE**

La Comisión Nacional de Energía (CNE) es la encargada de dar información indicativa de la situación del sector. Debe entregar las estimaciones de la autoridad en cuanto a crecimiento; ésto resulta fundamental en el sector eléctrico donde los proyectos tienen largos períodos de maduración. Es importante destacar que, dado que el sector opera en un marco competitivo y abierto, no le corresponde a la CNE determinar cuándo y qué tipo de inversiones realizar y los planes indicativos son sólo referenciales.

Para realizar la planificación indicativa, la CNE utiliza principalmente los modelos [Cne93]: MGI, WASP y GOL. El Modelo Global de Selección de

Inversiones (MGI) es un modelo matemático que permite determinar, en líneas gruesas, un programa de instalación de centrales y líneas de transmisión de largo plazo (próximos 15 a 20 años). Su principal ventaja reside en que posibilita el análisis de un gran número de centrales alternativas, cada una con tamaño variable; no obstante, la ubicación cronológica de las obras del programa óptimo resultante no es precisa. Lo anterior lleva a utilizar modelos más finos, como el WASP y el GOL, para el objeto de adoptar las decisiones de ejecución de las primeras obras del programa. El *Wien automatic system planning package* (WASP) es un modelo computacional de planificación eléctrica diseñado por el Organismo Internacional de Energía Atómica para determinar, desde un punto de vista económico, la secuencia óptima de instalación de centrales generadoras; y el modelo GOL que corresponde al descrito anteriormente dentro de los modelos utilizados por las empresas generadoras y por el CDEC.

#### **2.4.3 Validez de los datos que se utilizan**

La operación de los modelos actualmente en uso para colaborar en la gestión del sistema interconectado central requiere de información hidrológica que represente la oferta de agua al sistema. Esta información representa caudales medios en diferentes secciones de las cuencas que participan con su oferta al sistema de centrales y embalses que forman el SIC. La selección de estas secciones ha sido hecha a base de la información histórica disponible, así como a las propiedades de cada una de las series hidrográficas, de manera que sea posible representar, o calcular, con precisión razonable los aportes a las principales obras que deben operarse.

En la actualidad los modelos en uso utilizan como información básica series cronológicas de valores medios semanales de 40 años de extensión, representativas del período 1941-1980 organizadas por años hidrológicos (de Abril a Marzo). Estas series han sido formadas por 48 valores pseudosemanales, con meses de cuatro semanas, de manera que la cantidad de días asignados a cada semana fluctúa entre 7 y 8 de acuerdo a una convención adoptada por el CDEC. Para las series mensuales se han considerado 12 meses que coinciden con los meses del año, lo que implica que tengan una duración variable. Si bien varias de estas series

corresponden a datos medidos, algunas de ellas completan la información a partir de relaciones simples de continuidad, o consideran los aportes medidos en varias estaciones. Un aspecto que debe destacarse por sus posibles efectos sobre el empleo de esta información es que muchas de dichas series han sido rellenadas y extendidas de manera de completar todos los años del período de datos común. Al analizar los procedimientos utilizados para completar la información disponible se han detectado problemas desde una perspectiva de la modelación estocástica multivariada. Entre ellos, que el proceso de relleno y estimación se ha basado en la mayoría de los casos en correlacionar estas estaciones con otras conocidas de manera de proyectar el valor esperado de un período, lo que provoca una subestimación de las varianzas reales en las series rellenadas y una sobrevaloración de la dependencia espacial de las series por el relleno de datos correlacionados.

#### **2.4.4 Optimización a largo plazo**

La operación de un sistema eléctrico tiene por objetivo abastecer la demanda a un mínimo costo, respetando condiciones de seguridad, de calidad de servicio y de entorno. La seguridad de operación corresponde a la capacidad que posee un sistema para resistir fallas en sus instalaciones sin que se produzcan caídas parciales o totales de servicio. La calidad de servicio, en tanto, es la capacidad del sistema para proporcionar a sus usuarios en todo momento un servicio continuo, sin limitaciones, y en niveles predeterminados de voltaje y frecuencia. Las restricciones de entorno son las derivadas de las exigencias de riego y turísticas, control de crecidas, emisión de gases y lo más importante en lo que es la cantidad de energía generada en varias de las cuencas utilizadas. Esta última restricción determina un nivel mínimo de energía que se debe generar, o equivalentemente un nivel mínimo de caudal que se debe entregar, por lo que no puede ser embalsado o derivado fuera del cauce.

El período de optimización da un marco de referencia para las decisiones de más corto plazo. Los cursos de acción posibles de tomar a fin de definir la operación óptima diaria también están condicionados por decisiones tomadas anteriormente acerca del manejo de los embalses, programas de mantenimiento de las centrales, stock de combustible, compromisos de compra-venta de energía, etc.

a) Características de la optimización

Las características de los modelos de optimización de sistemas eléctricos, ocupados por las empresas y el CDEC, son las siguientes:

- i) Horizonte multianual.
- ii) Período elemental de análisis mensual, trimestral o anual.
- iii) Se recurre a análisis probabilísticos para tomar en cuenta los efectos derivados de la hidrología.
- iv) Se hace una representación simplificada de oferta y demanda.

En la oferta se toman en consideración las indisponibilidades de las centrales por mantenimiento y la incorporación de futuras obras que amplían la capacidad de generación.

La operación del sistema en un horizonte multianual queda definido principalmente por la operación de los embalses más grandes, de regulación interanual. Actualmente, los modelos de optimización en Chile consideran sólo el embalse de regulación interanual del Laja.

La operación de un embalse está condicionada por la operación en el período anterior, ya que la cantidad de agua con la que se cuenta en el período en curso,  $V(t+1)$ , es igual a la cantidad de agua que había en el período anterior menos el volumen ocupado en la generación y las pérdidas por evaporación y otros, más el caudal proveniente de deshielos y lluvias (condición hidrológica):

$$V(t+1) = V(t) + Q(t) - G(t) - P(t)$$

Con  $V(t)$ : volumen de agua disponible al comienzo del período  $t$ .

$Q(t)$ : caudal afluente al embalse en el período  $t$ .

$G(t)$ : volumen de agua extraído en el período  $t$ .

$P(t)$ : pérdidas por evaporación y otros en  $t$ .

Como existe esta dependencia de un período con respecto al anterior, al optimizar una etapa se debe considerar que el volumen extraído condiciona el beneficio que se obtendrá a futuro. Esta condición se resuelve considerando que el agua tiene un valor estratégico dada por los beneficios que con ella se pudiesen obtener a futuro. El criterio de optimización es entonces minimizar los costos de generación, sujeto a las restricciones de demanda, de capacidad instalada, de riego, etc y considerando el valor estratégico del agua.

Como en la optimización debe "conocerse" el valor estratégico del agua que resulta de la operación de los períodos futuros, el método con el que se lleva a cabo la optimización es de programación dinámica, mediante el cual se resuelve el problema de la optimización, comenzando con el último período del horizonte hasta llegar al comienzo.

#### b) Manejo de las condiciones hidrológicas

La optimización además tiene que manejar la aleatoriedad de las condiciones hidrológicas. El análisis del comportamiento aleatorio de los caudales afluentes a las centrales, requiere disponer de información histórica. Los registros hidrológicos utilizados en estudios de operación pueden corresponder a: series cronológicas de caudales, o series cronológicas de energías generables.

El uso de series de caudales es obligado en el caso de centrales hidroeléctricas capaces de realizar traspasos entre períodos, los cuales deben ser optimizados. En el caso de centrales que no cuentan con regulación suficiente para efectuar traspasos de agua, es usual utilizar series de energía generable.

Las series de energías generables se obtienen a partir de las series de caudales. En el caso de centrales de pasada, se aplica una constante propia de la central al caudal afluente que depende de las características de la central.

El número de registros estadísticos para cada central con los que se cuenta es variable. En algunos casos se cuenta hasta con 50 años hidrológicos considerados desde Abril de un año calendario a Marzo del año siguiente.

No hay que olvidar que las estadísticas de energías generables no necesariamente coinciden con la generación real de una central, debido a eventuales rebases por falta de consumo, indisponibilidades por unidades fuera de servicio, unidades aún no incorporadas, o cambios físicos posteriores como aumento de capacidad, rendimiento, etc.

Desde el punto de vista estadístico, las series de caudales y de energías generables constituyen una de las posibles ocurrencias de un fenómeno que puede en el futuro adoptar valores diferentes a los históricos. La hipótesis básica que hasta el momento se ha ocupado, es que los valores históricos como los que se darán en el futuro están regidos por las mismas leyes de probabilidad y que las hidrologías futuras son equiprobables a las acontecidas históricamente. Además, el modelo de optimización considera estas estadísticas y supone los años independientes entre sí. Este punto está sujeto a discusión, porque es un supuesto de mucha trascendencia en lo que es la modelación de los sistemas eléctricos y no es muy próximo a la realidad.

En consecuencia, las series cronológicas de caudales y energías generables constituyen una muestra y su análisis estadístico resulta fundamental a efectos de establecer escenarios de simulación coherentes.

En resumen, los modelos de optimización resuelven la incógnita de cuánto deben generar los embalses dada cierta cota inicial del embalse de regulación interanual del SIC (el embalse del Laja), y dada una condición hidrológica. Ésta es la entrada principal al modelo de simulación que calculará valores esperados de variables que no están incluidas en el modelo de optimización, por considerar al SIC como una unidad y analizar sólo sus costos.

#### **2.4.5 Pronóstico a corto plazo**

El pronóstico consiste en estimar el valor esperado del caudal en cada estación para el período de previsión, así como la varianza del estimador, lo que permite presentar los resultados con asignación de probabilidades, definiendo en cada caso bandas alrededor del valor esperado que encierran los posibles eventos futuros dada la historia hasta el instante en que se efectúa el pronóstico.

El sistema de pronóstico empleado por el CDEC consiste en la ponderación de los datos de caudales en los 7 días anteriores al pronóstico [Fern96]. El esquema de ponderación utilizado para el caudal de la semana siguiente considera el del último día con información ponderado por un factor y el promedio de los seis días anteriores con el factor complementario, es decir:

$$Q_1 = A(q_{j-1}) + (1-A) (q_{j-2} + q_{j-3} \dots + q_{j-7})/6$$

donde A toma valor 0,3 ó 0,5 de acuerdo a si se trata del período de verano o invierno, respectivamente. Este valor es igual para todas las cuencas involucradas en el pronóstico a corto plazo.

Como puede apreciarse se trata de un modelo del tipo autorregresivo de los caudales diarios, con coeficientes de ponderación seleccionados arbitrariamente.

### **III. MODELACIÓN DE SISTEMAS HIDROLÓGICOS**

En este capítulo se presentan los aspectos teóricos relacionados con la metodología de modelación propuesta en esta tesis.

#### **3.1 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, como se muestra en la figura 3.1. Las etapas básicas comprenden la selección del tipo de modelo a emplear, la identificación del orden, la estimación de parámetros y la validación del modelo. Para ello se debe tener en cuenta los tipos de modelos disponibles entre los cuales se pueden seleccionar los más adecuados, las propiedades que se desean modelar así como las características de los datos o las variables a representar.

En la modelación de series cronológicas se ha popularizado un esquema de trabajo propuesto por Box y Jenkins [Box76] 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 situación del SIC la selección de un esquema multivariado es obvia si se desea representar las relaciones entre las diferentes cuencas de la región. 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 autorregresivo para la dependencia temporal de cada serie no aparece igualmente justificado, pudiéndose pensar que modelos del 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 período 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.

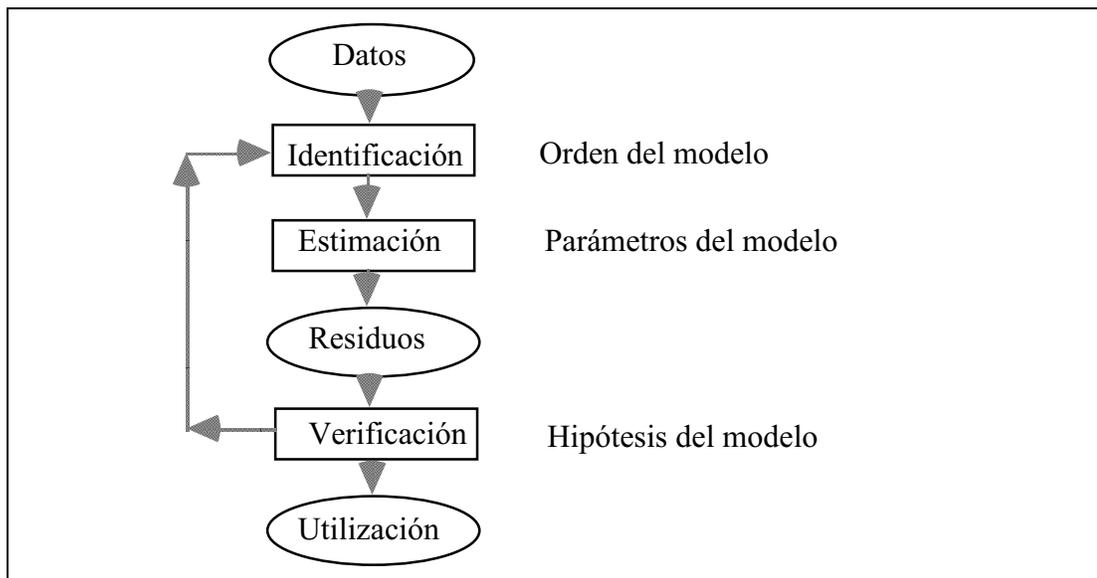


Figura 3.1: Esquema general de modelación.

- 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 del ejemplo se trata de los  $p$  coeficientes autorregresivos de cada período, dado que se conoce el valor de  $p$ . Existen varios procedimientos, siendo el más popular el método de los momentos, mediante el cual se eligen los coeficientes de manera de reproducir los coeficientes de autocorrelación hasta el de 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.

### **3.1.1 Modelos típicos disponibles**

Para la modelación multivariada de series hidrológicas existen en la literatura diversas alternativas. Un buen resumen de ellas se presentan y discuten en [Sala85]. Los tipos básicos de dependencia entre varias series hidrológicas, para una modelación multivariada, se conocen como contemporánea, unidireccional y retroalimentada.

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

En la figura 3.2 se muestran los tres esquemas de dependencia indicados en la cual, mediante flechas se señala la relación de dependencia entre las variables. 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. Ésto conduce a que los modelos

multivariados respectivos se conozcan también como diagonales, triangulares o completos, respectivamente. En cada caso adicionalmente a las dependencias indicadas en la figura 3.1 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.

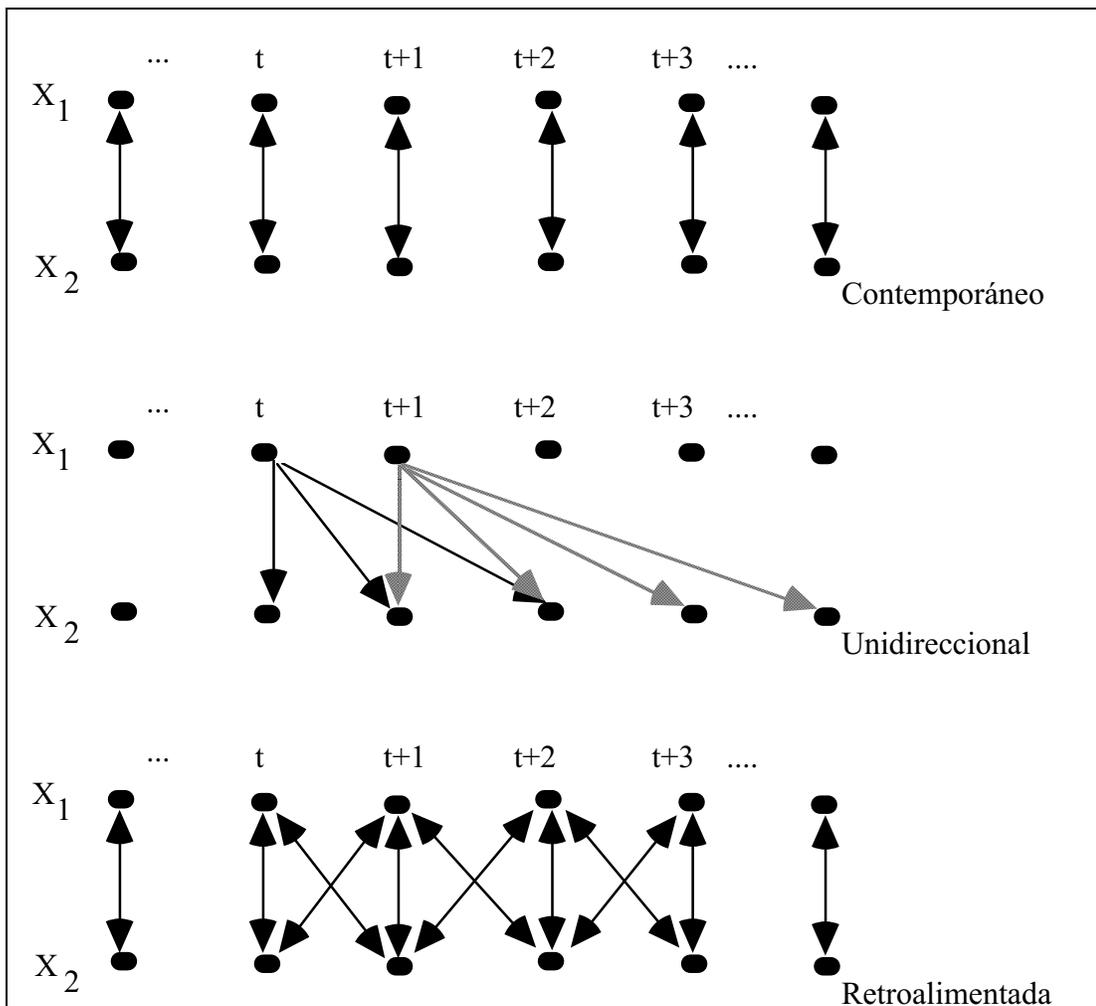


Figura 3.2: Esquemas de dependencia multivariada.

Existen además otras proposiciones para considerar la dependencia multivariada, o espacial. Entre ellos se pueden mencionar las basadas en componentes principales, regresiones múltiples, esquemas de desagregación y de agregación. Una visión general de ellos se presenta en [Sala85], y más recientemente en [Hipe94].

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. [Fern95/1]

### **3.1.2 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, 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 período (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 las 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. En los ejemplos que se adjuntan a este informe se pueden apreciar estas propiedades y su estimación, así como las variaciones típicas que presentan entre los distintos períodos del año.

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

$$Z_{m,n} = \frac{X_{m,n} - \mu_m}{\sigma_m}$$

con lo anterior todos los períodos presentan igual promedio y varianza, lo que simplifica los cálculos posteriores.

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

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

Siguiendo una notación propuesta por Box y Jenkins (1970) se recurre al operador de retardo  $B$  definido de manera que  $B^k Z_t = Z_{t-k}$ , con lo cual la expresión anterior se convierte en:

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

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

$$\Phi^m \left( \frac{X_m - \mu_m}{\sigma_m} \right) = a_m$$

Los valores  $a_m$  corresponden a los residuos del período  $m$  y se suponen que son temporalmente independientes, pero pueden presentar una dependencia espacial contemporánea con los residuos similares de otras series lo que se 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 hacerse cargo de la asimetría en 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 presentan en la tabla 3.1.

En total se trata aproximadamente de  $(4+p_m)*w$  parámetros, que ponderado por 48 semanas del año hidrológico (modelos con intervalo de tiempo semanal) y 3 órdenes de regresión implican la no despreciable cantidad de 336 parámetros.

Tabla 3.1: Parámetros implicados en un modelo para  $w$  períodos.

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

### 3.1.3 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 por  $p$  parámetros autorregresivos. 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. La representación mediante este esquema se explica en [Fern85].

## 3.2 Modelación de la dependencia temporal

### 3.2.1 Identificación de modelos temporales periódicos

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 de lograr 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 disponible que se desea modelar.

La Función de Autocorrelación Simple, FAS, corresponde al coeficiente de correlación existente 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:

$$\rho^m(k) = E \left[ \left( \frac{Z_m - \mu_m}{\sigma_m} \right) \left( \frac{Z_{m-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right]$$

Donde  $E(\cdot)$  corresponde al valor esperado. La función  $\rho^m(k)$  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 (1976) para identificar el orden del modelo por comparación. Así por ejemplo, en un modelo AR(p) presentan valores diferentes de cero y determinados para  $k = 1, 2, \dots, p$  y para valores mayores de  $p$  decrece exponencialmente o en forma sinusoidal hasta desaparecer. Los valores de  $\rho^m(k)$  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:

$$\hat{\rho}^m(k) = r^m(k) = \frac{1}{(N-k)s_m s_{m-k}} \sum_{n=k}^N (Z_n^m - \mu^m)(Z_n^{m-k} - \mu^{m-k})$$

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 autorregresivo 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 en 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 la FAS. Si las variables son 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:

$$\Phi^m(B) \left( \frac{Z_t - \mu_m}{\sigma_m} \right) = a_t^m$$

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

$$Y_t = \left( \frac{Z_t - \mu_m}{\sigma_m} \right)$$

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

$$Y_t = \phi_1^m Y_{t-1} + \phi_2^m Y_{t-2} + \dots + \phi_{p_m}^m Y_{t-p_m} + a_t$$

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

$$E(Y_t Y_{t-k}) = E(\phi_1^m Y_{t-1} Y_{t-k}) + E(\phi_2^m Y_{t-2} Y_{t-k}) + \dots + E(\phi_{p_m}^m Y_{t-p_m} Y_{t-k}) + E(a_t Y_{t-k})$$

Como las variables están centradas y reducidas esta ecuación indica:

$$\rho^m(k) = \phi_1^m \rho^{m-1}(k-1) + \phi_2^m \rho^{m-2}(k-2) + \dots + \phi_{p_m}^m \rho^{m-p_m}(k-p_m)$$

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 mencionan a continuación:

- a) Conocidos los parámetros de un modelo  $\phi_1^m, \phi_2^m, \dots, \phi_{p_m}^m$  y por supuesto su orden  $p_m$ , es posible determinar su estructura de dependencia encontrando los valores  $\rho^m(k)$  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,  $\hat{\rho}^m(k)$ , obtenidas de una muestra que se desea modelar, se pueden calcular los valores de los parámetros  $\phi_1^m, \phi_2^m, \dots, \phi_{p_m}^m$  de manera de reproducir 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,  $\hat{\rho}^m(k)$ , es posible obtener valores de la función de autocorrelación parcial,  $f_{k,k}$ , ajustando sucesivamente modelos de orden  $p=1,2,3,\dots,k,\dots$  y calculando en cada caso el valor del último coeficiente autorregresivo de cada uno de ellos,  $f_{k,k}$ . 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  $\phi_1^m, \phi_2^m, \dots, \phi_{p_m}^m$  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 de Akaike, AIC, para seleccionar entre modelos similares, según el cual es conveniente elegir el modelo que minimice el AIC, calculado como:

$$AIC = 2p + N \ln(\sigma_a^m)^2$$

### 3.2.2 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  $\phi_1^m, \phi_2^m, \phi_3^m, \dots, \phi_{p_m}^m$  y las características de los residuos  $a_t$ , fundamentalmente su varianza,  $\sigma_m^2(a)$ .

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

$$\begin{aligned} \rho^m(1) &= \phi_1^m + \phi_2^m \rho^{m-1}(1) + \phi_3^m \rho^{m-1}(2) \cdots + \phi_{p_m}^m \rho^{m-1}(p_m-1) \\ \rho^m(2) &= \phi_1^m \rho^{m-1}(1) + \phi_2^m + \phi_3^m \rho^{m-2}(1) \cdots + \phi_{p_m}^m \rho^{m-2}(p_m-2) \\ \rho^m(3) &= \phi_1^m \rho^{m-1}(2) + \phi_2^m \rho^{m-2}(1) + \phi_3^m \cdots + \phi_{p_m}^m \rho^{m-3}(p_m-3) \\ \rho^m(p_m) &= \phi_1^m \rho^{m-1}(p_m-1) + \phi_2^m \rho^{m-2}(p_m-2) \cdots + \phi_{p_m}^m \end{aligned}$$

Este sistema se acostumbra poner en términos matriciales como:

$$\begin{bmatrix} 1 & \rho^{m-1}(1) & \cdots & \rho^{m-1}(p_m-1) \\ \rho^{m-1}(1) & 1 & & \rho^{m-2}(p_m-2) \\ \cdots & & \cdots & \\ \rho^{m-1}(p_m-1) & \rho^{m-2}(p_m-2) & & 1 \end{bmatrix} \begin{bmatrix} \phi_1^m \\ \phi_2^m \\ \cdots \\ \phi_{p_m}^m \end{bmatrix} = \begin{bmatrix} \rho^m(1) \\ \rho^m(2) \\ \cdots \\ \rho^m(p_m) \end{bmatrix}$$

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, ésto se realiza estableciendo las ecuaciones de Yule Waker para cada período de manera de igualar los primeros  $p_m$  coeficientes de autocorrelación observados en la muestra con los que entregaría el modelo. Ésto conduce, para cada, período a un sistema de  $p_m$  ecuaciones como las indicadas, pero reemplazando los valores de  $\rho$  por sus estimadores  $\rho^{m(k)}$ , de manera que la solución se obtiene como:

$$\hat{\phi} = R^{-1}r$$

donde el vector  $\hat{\phi}$  es  $(\phi_1^m, \phi_2^m, \phi_3^m, \dots, \phi_{p_m}^m)^T$  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  $\hat{\phi}$  se puede encontrar el valor de la varianza de los residuos con la ecuación de Yule-Walker para  $k=0$ :

$$\sigma_m^2(a) = 1 - \sum_{i=1}^{p_m} \phi_i^m \rho^m(i)$$

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.

### 3.3 Construcción del modelo

Para encontrar el modelo más adecuado se sigue un esquema general propuesto por Box y Jenkins [Box76] que se basa en la etapas 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 que los resultados finales estarán afectados por la oportunidad en la cual se realizan.

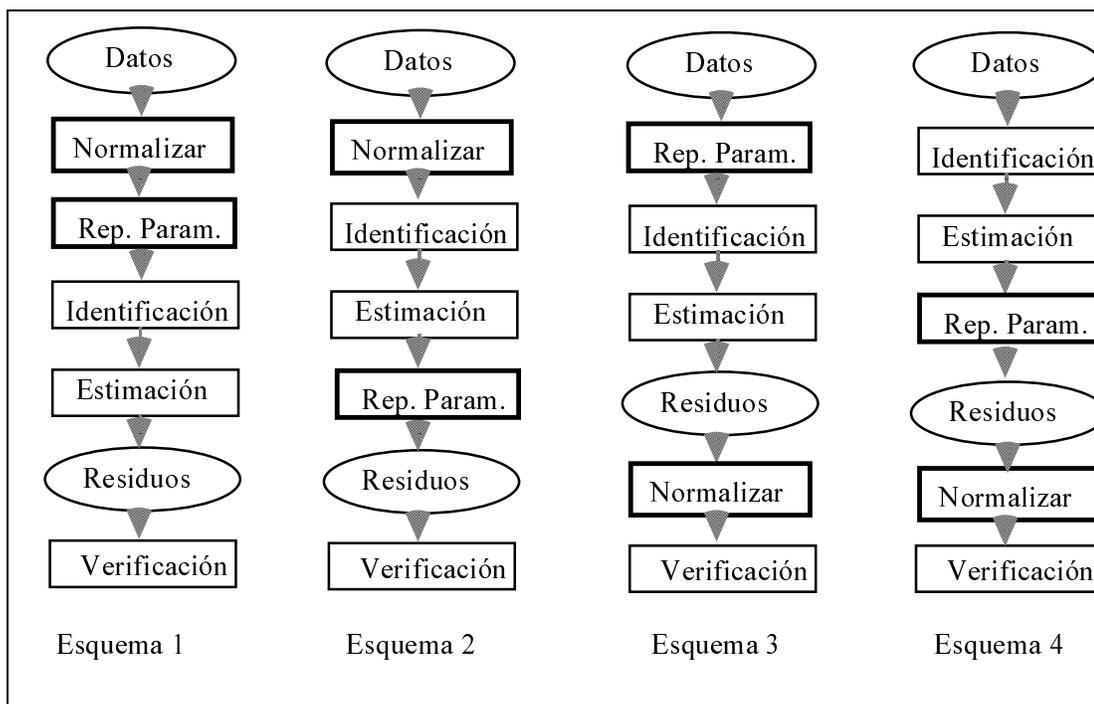


Figura 3.3: Esquemas alternativos del procedimiento para incorporar la normalización y la parametrización.

Habitualmente las alternativas de normalización que se plantean son las siguientes: no normalizar, 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, que se muestran en la figura 3.3, es la que considera sólo casos en que se normaliza y se parametriza, dejando de lado todas las situaciones en que se hace sólo una 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 indicados 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 en la primera etapa 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.

### 3.4 Diagnóstico o verificación del modelo

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

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

La verificación o diagnóstico del modelo requiere comprobar que las hipótesis básicas realizadas respecto a los residuos son ciertas. Éstos 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:

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

utilizando para ello los promedios y las varianzas de los residuos de cada período.

### 3.4.1 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 de la 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. La varianza asintótica es válida para  $k$  grande, pero no para los primeros retardos. Por ejemplo, para un AR(1) la desviación típica asintótica de  $r_a^m(1)$  es  $\phi_1/\sqrt{N}$  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 saldrá fuera, por lo que la aparición de un valor significativo en un retardo elevado es esperable. Sin embargo, como según lo anterior estos límites sobrestiman la varianza en los retardos pequeños, un valor próximo a los límites de confianza en los retardos iniciales 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 [Sala85]. 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 Chi-cuadrado con  $(L-p_m)$  grados de libertad. Se concluye que un modelo es inadecuado si  $Q^m(L)$  es mayor que  $\psi^2(L-p_m)$ . Esta prueba se conoce también como del estadígrafo de Portmanteau.

### 3.4.2 Prueba de normalidad

En el caso de la modelación hidrológica del Sistema Interconectado Central es importante que los residuos de cada aprovechamiento sean normales, ya que ellos se emplearán para modelar la estructura espacial del sistema. Ésto se puede verificar calculando el coeficiente de asimetría de los residuos de cada semana y contrastarlo 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:

$$\varepsilon_t = \begin{cases} \frac{a_t^\lambda}{\lambda} & ; \text{si } \lambda > 0 \\ \text{Ln}(a_t) & : \text{si } \lambda = 0 \end{cases}$$

### 3.4.3 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 su promedio sea nulo, es conveniente para facilitar la construcción del modelo. Para contrastar esta hipótesis en el caso general suponiendo  $N$  residuos y  $p$  parámetros, se puede hacer suponiendo que se trata de valores normales con promedio cero y varianza  $\sigma_a^2$ . Entonces el promedio es también normal con valor esperado cero y varianza  $\sigma_a^2/N$ . Este contraste debe aplicarse después de comprobar que los residuos están incorrelados.

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 verosimilitudes.

Finalmente, conviene siempre estudiar el gráfico de los residuos estimados a lo largo del tiempo. Su inspección se facilita situando en él los límites del 95%, 99% o 99,5% y estudiando los puntos que salen fuera de estos límites de control para detectar la existencia de *outliers*.

#### 3.4.4 Reformulación

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

$$a_{n,m} = \phi_X^m(B)X_{n,m}$$

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 autorregresivo:

$$e_{n,m} = \phi_a^m(B)a_{n,m}$$

donde los nuevos residuos  $e$  si satisfacen todas las condiciones. Entonces, sustituyendo en la ecuación anterior resulta para la serie el modelo

$$e_{n,m} = \phi_X^m(B)\phi_a^m(B)X_{n,m}$$

que correspondería al modelo adecuado.

#### 3.4.5 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 ésto es posible captar pequeñas estructuras remanentes que pueden mejorar las predicciones. En general, se ha ajustado un PAR(p) que parece adecuado, el sobreajuste se aplica estimando los modelos PAR(p+1) o PAR(p+2) y comprobando si los parámetros adicionales son significativos. Para seleccionar entre estos modelos alternativos se puede recurrir al ya mencionado criterio de Akaike.

### **3.5 Modelación 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 Sistema Interconectado, 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. Los principales de ellos se exponen en [Sala85]. Recientemente en [Hipe94] se presenta un esquema detallado de modelación que incluye el caso de diferentes alternativas para la modelación multivariada.

Entre las alternativas para sistemas complejos se pueden considerar los modelos Multivariados ARMA, de los cuales se desprenden los multivariados de Funciones de Transferencia y los contemporáneos ARMA. Los primeros son una generalización matricial de modelos 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 ha propuesto una aplicación particular de modelos CARMA considerando para las series individuales sólo modelos Autorregresivos 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 no existe 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  $\mu_{VK}$  y desviación  $\sigma_{VK}$

El comportamiento espacial de estos residuos se puede apreciar en matrices de correlación de desfase  $h$  entre las series  $\mathbf{D}^h$ , cuyos elementos son:

$$\{\mathbf{D}^h\} = \{d_{ij}^h\} \text{ con } d_{ij}^h = \frac{E[(V_t(i) - \mu_{vi})(V_{t-h}(j) - \mu_{vj})]}{\sigma_{vi}\sigma_{vj}}$$

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:

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

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 ruidos tiene la forma:

$$\mathbf{e}_t = \mathbf{M}\mathbf{w}_t$$

donde  $\mathbf{w}_t$  representa un vector de  $K$  residuos independientes entre sí para el instante  $t$ , mientras  $\mathbf{e}_t$  corresponde a los residuos de cada lugar centrados y reducidos, de manera que sus elementos son:

$$e_t(i) = \frac{V_t(i) - \mu_{V_t}}{\sigma_{V_t}}$$

Entonces se puede demostrar que:

$$E[\mathbf{e}_t \mathbf{e}_t^T] = E[(\mathbf{M}\mathbf{w}_t)(\mathbf{M}\mathbf{w}_t)^T] = E[(\mathbf{M}\mathbf{w}_t)(\mathbf{w}_t^T \mathbf{M}^T)] = \mathbf{M}E(\mathbf{w}_t \mathbf{w}_t^T) \mathbf{M}^T$$

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

$$E[\mathbf{e}_t \mathbf{e}_t^T] = \{\mathbf{D}^0\}$$

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  $\mathbf{w}$  son independientes, normales estándar se cumple que:

$$E[\mathbf{w}_t \mathbf{w}_t^T] = \mathbf{I}$$

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

$$\mathbf{D}^0 = \mathbf{M}\mathbf{M}^T$$

Para determinar los valores de los elementos de esta matriz,  $\{m_{ij}\}$  es conveniente imponer la restricción de que  $\mathbf{M}$  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 reproduzcan 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. Ésto permite expresar matricialmente el modelo espacial con la relación:

$$\begin{bmatrix} V_t(1) \\ V_t(2) \\ \dots \\ V_t(K) \end{bmatrix} = \begin{bmatrix} \mu_{V1} \\ \mu_{V2} \\ \dots \\ \mu_{VK} \end{bmatrix} + \begin{bmatrix} \sigma_{V1} & 0 & \dots & 0 \\ 0 & \sigma_{V1} & & 0 \\ \dots & & \dots & \dots \\ 0 & 0 & \dots & \sigma_{V1} \end{bmatrix} \begin{bmatrix} m_{11} & 0 & \dots & 0 \\ m_{21} & m_{22} & \dots & 0 \\ \dots & & \dots & \dots \\ m_{K1} & m_{K2} & \dots & m_{KK} \end{bmatrix} \begin{bmatrix} w_t(1) \\ w_t(2) \\ \dots \\ w_t(K) \end{bmatrix}$$

### 3.6 Usos de los modelos

Una vez construido el modelo, es decir, determinada su estructura y los parámetros, se puede utilizar para simulación o pronóstico. A continuación se explica brevemente la forma en que se abordan estos usos. No está de más recalcar que 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.

### 3.6.1 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 condicionada 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 simulan una cantidad de años similar a la de los registros históricos, es decir, del orden de 40 o más años.

La simulación condicionada consiste en generar igualmente  $M$  nuevas alternativas en los  $K$  lugares y de  $N$  años, pero tomando como instantes 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.

A continuación se describe cómo generar un nuevo conjunto de valores similares a la muestra histórica, es decir, para ser empleados en la primera alternativa. Los procedimientos son similares en la otra alternativa. Para ello 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 a continuación, 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 iidN(0,1).

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

$$\eta_t^k = \left[ 2Ln\left(\frac{1}{u_1}\right) \right]^{0,5} \cos(2\pi u_2)$$

donde  $\eta_t^k$  es el valor para el instante t en el lugar k. Ésto 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 temporal 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:

$$\mathbf{V}_t = \mathbf{M}\boldsymbol{\eta}_t$$

o explícitamente para cada instante t:

$$\begin{bmatrix} V_t^1 \\ V_t^2 \\ \dots \\ V_t^{K-1} \\ V_t^K \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ m_{21} & m_{22} & 0 & \dots & 0 \\ m_{31} & m_{32} & m_{33} & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ m_{K1} & m_{K2} & \dots & \dots & m_{KK} \end{bmatrix} \begin{bmatrix} \eta_t^1 \\ \eta_t^2 \\ \dots \\ \eta_t^{K-1} \\ \eta_t^K \end{bmatrix}$$

Los elementos de esta matriz cumplen con la relación:

$$\sum_{j=1}^K m_{ij}^2 = 1 \quad ; \text{ para } i = 1, 2, \dots, K$$

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, debiera verificarse en esta etapa que se reproducen las matrices de dependencia espacial de desfase nulo, uno, dos, ..., verificando que sólo la primera es diferentes 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 sola serie  $k$  y debe por lo tanto, repetirse  $K$  veces para completar una simulación. Debe tenerse cuidado en mantener ordenados los valores en cada período para no distorsionar la dependencia espacial creada entre ellos.

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

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

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 en el caso del SIC.

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:

$$a_n^m(k) = \mu_m(k) + \sigma_m(k)V_n^m(k)$$

Si fue necesario transformarlos en normales mediante alguna operación es el momento de proceder a su destransformación. Lo más típico es que en la etapa de 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:

$$z_n^m = \phi_1^m z_n^{m-1} + \phi_2^m z_n^{m-2} + \dots + \phi_{p_m}^m z_n^{m-p_m} + a_n^m$$

Si la simulación se hace para generar valores independientes de la condición histórica reciente, los primeros 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:

$$x_n^m(k) = \mu_m(k) + \sigma_m(k)z_n^m(k)$$

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.

### 3.6.2 Pronóstico

El pronóstico de interés para la operación del SIC 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 del 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 un 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.

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  $m$ . Éstos se obtienen los valores de  $Y_0$  y  $Z_0$  dados por:

$$Y_0 = g_m(Q_0)$$

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

donde  $g_m(\cdot)$  corresponde a la transformación aplicada en la semana  $m$  para convertir los valores en normales.  $\bar{y}_m, \sigma_m$  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:

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

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:

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

donde los coeficientes  $\Psi_j$  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  $V(k)$ :

$$\psi_1 = 1$$

$$\psi_2 = \phi_{1,m+2}$$

$$\psi_3 = (\phi_{2,m+3} + \phi_{1,m+2} \phi_{1,m+3})$$

$$\psi_4 = (\phi_{1,m+4} \phi_{2,m+3} + \phi_{2,m+4} \phi_{1,m+2} + \phi_{1,m+4} \phi_{1,m+3} \phi_{1,m+3})$$

A medida que aumenta el plazo de pronóstico 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 autorregresivos 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  $\alpha$  está dado por:

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

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

donde  $N_\alpha$  es la variable normal reducida con probabilidad de excedencia  $\alpha$  de acuerdo a los valores que se indican en la tabla 3.2. El valor de  $\alpha$  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.

Tabla 3.2. Variable normal reducida para bandas de diferente ancho.

Probabilidad de no-excedencia, $\alpha$	Valor, $N_\alpha$
0,95	1,960
0,90	1,645
0,80	1,288
0,70	1,036
0,60	0,842
0,50	0,674

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

### 3.7 Otros modelos

La modelación de una serie cronológica consiste en seleccionar un modelo que represente adecuadamente las características estadísticas de la serie. [Mind95]

En general, los modelos existentes para el análisis de las series de tiempo son bastante estándar [Sala80] cuando se tiene que la distribución es normal y la estructura es lineal. Algunos ejemplos de tipos de modelos son: AR (autoregresivos); FGN (modelos fraccionarios gaussianos); ARMA (autoregresivos de media móvil); BL (modelos de línea quebrada); modelos de ruido de disparo; 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 modelamiento 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 [Fern86] 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.

## **IV. ANÁLISIS DE LAS SERIES DISPONIBLES PARA LA MODELACIÓN**

### **4.1 Características de los aprovechamientos del SIC**

Las series de información de caudales disponibles provienen de mediciones que se han realizado en las diversas cuencas del país. Los valores considerados son obtenidos fundamentalmente de tres formas:

- a) medición directa de la serie en estaciones fluviométricas ubicadas en el lugar considerado.
- b) obtención de los valores a través de la adición y substracción de flujos medidos en otros lugares de la cuenca (dado que en el lugar de interés no se cuenta con mediciones directas), de manera de completar la información necesaria, en base a relaciones de continuidad.
- c) estimaciones a partir de otras mediciones (Ej.: filtraciones, superficies de hoyas, etc.), empleando para ello regresiones, o transformaciones de distintos tipos.

A modo de ejemplo, es posible considerar la serie “Laja en Tucapel”, que es medida en la estación fluviométrica del mismo nombre en el lugar del mismo nombre y se ha registrado directamente desde Abril de 1916. Por otra parte, la “Cuenca intermedia entre Captación Rama Laja, Central Antuco y Lago Laja” ha sido medida sólo desde Abril de 1981. No obstante éstas dos series se han medido de manera directa, la disponibilidad de registros es absolutamente distinta. Luego, para poder utilizar la segunda de las series nombradas en los modelos construidos, es necesario efectuar un relleno y extensión a un período común.

Debido a la existencia de estos períodos de relleno en las series de caudales de las cuencas chilenas, en general, los datos a utilizar deben ser cuidadosamente filtrados de modo de trabajar con los datos más fidedignos posibles. Cabe destacar que el relleno de datos en sí no es el problema, sino que el problema es que éste fue realizado sin la preocupación de reproducir momentos de mayor orden que el primero. En efecto, con los rellenos hechos en el pasado se reprodujeron

promedios, pero no desviación estándar, asimetría y curtosis; por lo que se tiene como consecuencia que las series rellenadas son menos fluctuantes (menor desviación estándar) que las reales. Además, las series rellenadas presentan una mayor dependencia temporal y espacial debido a que las técnicas de regresión y uso de relaciones empíricas no toman en cuenta la aleatoriedad.

#### **4.1.1 Selección de las series cronológicas de interés.**

Para la selección de las series de tiempo de interés se ha usado la metodología que considera las siguientes etapas: estudio detallado de las series consideradas como información base y comparación con las utilizadas en la modelación; revisión exhaustiva de las series disponibles para la detección de los períodos de “relleno” y extendidos; y selección de un período de información realista común.

Las series consideradas como información base son aquellas a partir de las cuales se construyen los datos que son alimentados al modelo OMSIC (en particular a través del archivo OMSHID). La forma de obtener dichos datos es una de las tres explicadas en el inicio de la sección 4.1 de este capítulo.

Esta metodología se ha aplicado al estudio de las series de tiempo asociadas a las principales cuencas del SIC con sus respectivas centrales de generación:

- i) Cuenca del Laja (Centrales El Toro, Abanico y Antuco)
- ii) Cuenca del Maule (Centrales Cipreses, Isla, Curillinque, Pehuenche, Colbún y Machicura)
- iii) Cuenca del Rapel (Centrales Rapel, Sauzal y Sauzalito)
- iv) Cuenca del Chamiza (Central Canutillar)
- v) Otras cuencas menores (Cuenca del Limarí: central Los Molles; cuenca del Maipo: centrales Alfalfal, Maitenes, Queltehues y Volcán; cuenca del Pilmaiquén: central Pilmaiquén; y cuenca del Huanehue: central Pullinque)

A continuación se presenta el análisis bajo la metodología antes descrita para cada una de las cuencas consideradas.

i) Cuenca del Laja

Interesa realizar un estudio detallado de las series consideradas como información base y compararlas con las utilizadas en la modelación.

Tabla 4.1: Series hidrológicas registradas en la cuenca del Laja.

N	Nombre	Lugar de medición	Período
1	pf-30005	Filtraciones Lago Laja	Estimadas según Ley de Filtraciones
2	q2-30007	Estero Malalcura en Captación	Feb. 54 - May. 82
3	q2-30018	Estero El Toro en Captación	May. 41 - Jul. 82
4	q2-30019	Cipreses en Captación	Nov. 63 - Abr. 83
5	q2-30020	Pichipolcura en Captación	Dic. 61 - Abr. 82
6	q2-30021	Trubunleo en Captación	Abr. 63 - Abr. 90
7	q2-30055	Cuenca Int. entre Captación Rama Laja, C. Antuco y Lago Laja	Abr. 81 a la fecha
8	q2-30212	Canales del Lago Laja	Abr. 47 a la fecha
9	q2-30305	Cuenca Int. Laja en Tucapel - Lago Laja - Cap. Alto Polcura.	Se obtiene por diferencia
10	q4-30002	Polcura en Bocatoma Central Antuco	Nov. 40 a la fecha.
11	q4-30005	Afluentes del Lago Laja	Estimación según superficie de la hoya
12	q4-30041	Laja en Tucapel	Abr. 16 a la fecha
13	q4-30047	Captación Alto Polcura	Feb. 56 a la fecha
14	q4-30257	Laja en Bocatoma Central Abanico	Abr. 41 a la fecha
15	q4-30309	Cuenca Int. Bocatoma Central Antuco - Captación Alto Polcura	Diferencia entre la serie número 10 y 13
16	q4-30317	Cuenca Int. Entre Bocatoma Central Abanico y Lago Laja.	Se obtiene por diferencia

En esta cuenca se han considerado 16 series de datos de caudales mensuales de años hidrológicos (Abril a Marzo de cada año) de diferentes extensiones. En la tabla 4.1 se presentan los nombres de las series indicadas y los períodos de medición.

En la figura 4.1 se presenta un diagrama unilineal de la cuenca que muestra la relación entre las series y los aprovechamientos.

A partir de los datos presentados en la tabla 4.1 se construyen las series hidrológicas que ocupa el modelo OMSIC y que resumen la información considerada de interés para la operación en la optimización que son las 5 que se presentan a continuación:

- Pasada Abanico: corresponden a la estadística de caudales de igual nombre (R.N.), su medición es directa.
- Alto Polcura: corresponden a la estadística de caudales de igual nombre (R.N.), su medición es directa.
- Pasada Antuco: se obtiene de la manera que se indica a continuación de acuerdo a la referencia [Epop83].

$$\text{- Pasada Antuco (Lim. } 190 \text{ m}^3/\text{s)} = \text{Aducción Laja} + \text{Aducción Polcura} + \text{Captación Pichipolcura (Lim. } 11 \text{ m}^3/\text{s)}.$$

Ahora bien, a su vez los valores de aducción Laja y aducción Polcura se calculan de acuerdo a:

$$\text{- Aducción Laja (Lim. } 30 \text{ m}^3/\text{s)} = \text{Cuenca Int. Bocatoma Antuco y Lago Laja} + \text{Estero El Toro (Lim } 6.6 \text{ m}^3/\text{s)} + \text{Estero Malalcura (Lim. } 5.4 \text{ m}^3/\text{s)}.$$

$$\text{- Aducción Polcura (Lim. } 190 \text{ m}^3/\text{s)} = \text{Río Polcura en Boc. (R.N.)} - \text{Captación alto Polcura}.$$

Sin embargo, al sumar dichas matrices los valores obtenidos difieren de los usados en el OMSIC aproximadamente en  $2 \text{ m}^3/\text{s}$  en todas las mediciones. Lo que aparentemente sucede es que se le ha sumado a la aducción Laja, los Esteros

Cipreses en Captación y Trubunleo en Captación. Luego de sumar estas matrices a las anteriores, los valores obtenidos difieren de los usados en el OMSIC en promedio en  $0,3 \text{ m}^3/\text{s}$ .

- Afluente Laja + Alto Polcura: corresponde a la suma de las mediciones directas de ambas series (R.N.)
- Laja en Tucapel: corresponde a la estadística de caudales de igual nombre (R.N.), su medición es directa.

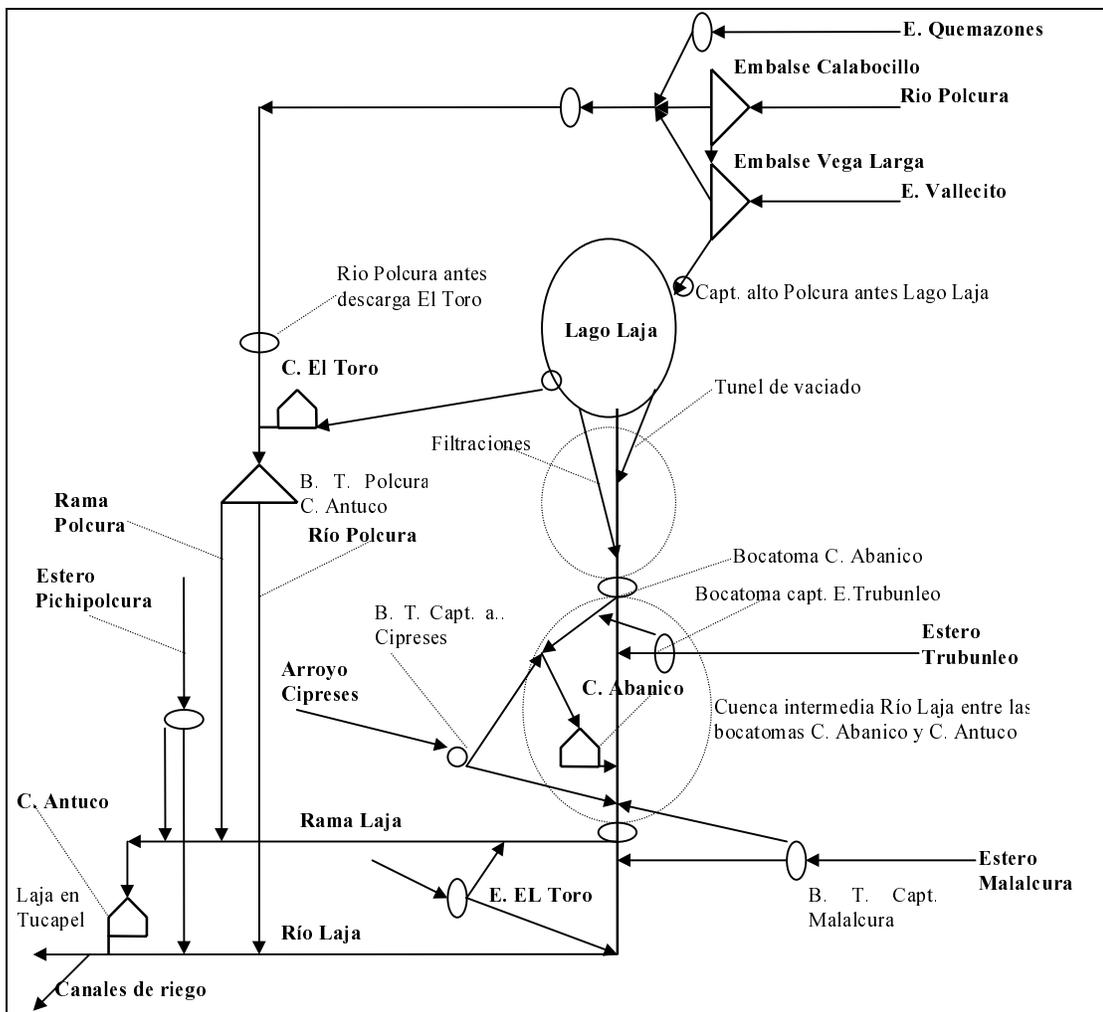


Figura 4.1: Diagrama unilineal resumido de la cuenca del Laja.

En todos los casos, salvo en el explicado de Antuco, los valores usados en el OMSIC coinciden con los datos de la información base.

Una revisión exhaustiva de cada una de las series disponibles permite la detección de los períodos de “relleno” de estadísticas, así como la forma en que se han realizado.

Como el modelo OMSIC utiliza matrices de cuarenta años hidrológicos correspondientes al período 41/42 a 80/81, en los lugares donde existe sólo parte de éstos se ha rellenado a través de determinar su correlación con los datos disponibles en otras series relacionadas de la misma cuenca. En este caso en particular la estadística de Laja en Tucapel, que es la más antigua de la cuenca, ha servido para reconstruir las restantes.

#### ii) Cuenca del Maule

Como se señala anteriormente se realiza un estudio detallado de las series consideradas como información base para luego compararlas con las utilizadas en la modelación. De acuerdo a la información recopilada [Divi75] en la hoya hidrográfica del río Maule se han instalado alrededor de 82 estaciones fluviométricas existiendo en funcionamiento sólo 52 a esa fecha. En esta cuenca se han considerado 6 series de datos de caudales mensuales de años hidrológicos (Abril a Marzo de cada año) de diferentes extensiones. En la tabla 4.2 se presentan los nombres de las series indicadas y los períodos de medición.

En la figura 4.2 se presenta un diagrama unilineal de la cuenca que muestra la relación entre las series y los aprovechamientos.

A partir de los datos presentados en la tabla 4.2 se construyen las series hidrológicas que ocupa el modelo OMSIC y que resumen la información considerada de interés para la operación en la optimización que son las 7 que se presentan a continuación (ver [Divi82] y [Divi75]):

- Afluentes a la Laguna Invernada: corresponden a la estadística de caudales de igual nombre (R.N.), su medición es directa.

Tabla 4.2: Series hidrológicas registradas en la cuenca del Maule.

N	Nombre	Lugar de medición	Período
1	q1-28052	Quebrada el Ciego	1956-1974
2	q2-28301	Cuenca Int. entre Laguna Invernada y Bocatoma Central Isla.	Se obtiene como ponderado del caudal afluente a l. Invernada
3	q4-28009	Afluentes a la Laguna Invernada (R.N.)	1941- a la fecha
4	q4-28016	Afluentes a la Laguna del Maule (R.N.)	1952- a la fecha
5	q4-28019	Maule en Bocatoma Central Isla (R.N.)	1958- a la fecha
6	q9-28026	Maule en Armerillo (R.N.)	1915- a la fecha

- Afluentes a la Cuenca Intermedia Isla: los caudales Hoya Intermedia desagüe Laguna Maule menos Bocatoma Central Isla corresponden a la estadística de igual nombre limitada a  $48 \text{ m}^3/\text{s}$ . Dicho caudal se obtiene ponderando el caudal afluente a la Laguna Invernada (R.N.) por el valor 0.023 [Divi82].
- Afluentes al embalse Melado: hasta antes de la puesta en marcha de la central Pehuenche, los afluentes al embalse Melado (R.N.) pueden ser considerados iguales a los medidos en la estación fluviométrica del río Melado en junta con Maule más el caudal captado por el canal Melado (riego). Luego de la puesta en servicio de la central Pehuenche, el caudal afluente al embalse Melado (R.N.) además de considerar el caudal captado por el canal Melado, debe incluir el caudal generado, desaguado y/o regulado por el embalse Melado y, adicionalmente, el caudal aportado por la aducción Maule.
- Claro en San Carlos más Estero Las Garzas: corresponde a la suma de las mediciones directas de ambas series (R.N.)
- Cuenca Intermedia Maule en Bocatoma Pehuenche menos Laguna Invernada menos Laguna Maule: se determina como la suma del caudal afluente al embalse Melado (R.N.) y el caudal aportado por la aducción Maule, a lo cual se le resta el caudal captado por el canal Melado.

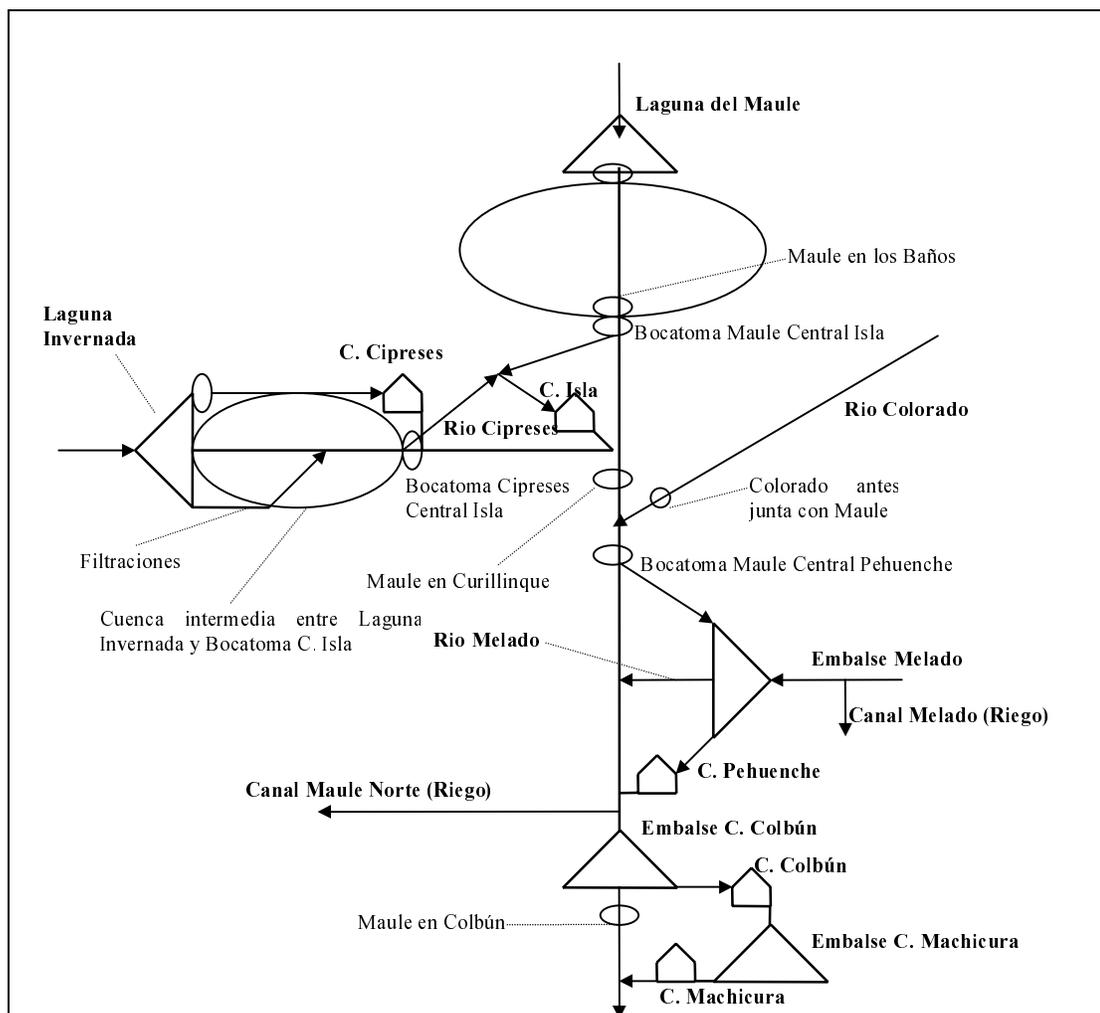


Figura 4.2: Diagrama unilineal resumido de la cuenca del Maule.

- Hoya Intermedia muro Colbún desagües de las lagunas Invernada y Maule: el caudal aportado por dicha hoya intermedia debe ser calculado como la suma del caudal observado en Maule en Colbún y caudal captado por el canal Maule Norte y caudal captado por el canal Melado, a lo que se le debe restar el caudal de filtraciones Laguna Invernada, el caudal generado en la Central Isla, el caudal desaguado Laguna Invernada y, el caudal desaguado Laguna Maule.

- Cuenca Intermedia muro Colbún menos Armerillo: corresponden a la estadística de caudales de igual nombre (R.N.), su medición es directa.

Una revisión exhaustiva de cada una de las series disponibles permite la detección de los períodos de “relleno” de estadísticas, así como la forma en que se han realizado.

Como el modelo OMSIC utiliza matrices de cuarenta años hidrológicos correspondientes al período 41/42 a 80/81, en los lugares donde existe sólo parte de éstos se ha rellenado a través de determinar su correlación con los datos disponibles en otras series relacionadas de la misma cuenca. De acuerdo a lo presentado en la tabla 4.2, los períodos de información realmente medida son bastante dispares entre una serie y otra, se cuenta con series muy antiguas como Maule en Armerillo (1915) y Afluentes a la Laguna Invernada (1941) y otras más contemporáneas (que datan de 1952 y 1958), por lo que si se desea establecer un período de información realista común se debe transar entre los registros de la menor de las series y la importancia relativa de dicha serie.

### iii) Cuenca del Rapel

Interesa realizar un estudio detallado de las series consideradas como información base y compararlas con las utilizadas en la modelación. En esta cuenca se han considerado 6 series de datos de caudales mensuales de años hidrológicos (Abril a Marzo de cada año) de diferentes extensiones. En la tabla 4.3 se presentan los nombres de las series indicadas y los períodos de medición.

En la figura 4.3 se presenta un diagrama unilineal de la cuenca que muestra la relación entre las series y los aprovechamientos.

A partir de los datos presentados en la tabla 4.3 se construyen las series hidrológicas que ocupa el modelo OMSIC y que resumen la información considerada de interés para la operación en la optimización que son las dos que se presentan a continuación (ver [Divi82] y [Epop83]):

- Cachapoal en Bocatoma Sauzal: son los caudales disponibles en la Bocatoma de la central Sauzal, o sea no incluye estero Coya ni el resto de la cuenca intermedia entre Bocatoma y Termas.
- Afluentes al embalse Rapel menos demanda de riego río Teno: se obtiene como la suma del caudal generado por la central Rapel, el caudal evacuado (por rebases superficiales y compuertas de medio fondo) y el caudal regulado por el embalse central Rapel, menos el caudal entregado por el canal Teno-Chimbarongo.

Tabla 4.3: Series hidrológicas registradas en la cuenca del Rapel.

N	Nombre	Lugar de medición	Período
1	q4-25004	Cachapoal en Puente Termas (R.N.)	Ago. 49 - a la fecha
2	q4-25005	Cachapoal en Bocatoma Central Sauzal (R.N.)	Obtenida por correlación con la serie anterior
3	q4-25006	Claro en Campamento	Obtenida por correlación con la serie Cach. en Pte. Termas
4	q4-25101	Coya en Captación Coya (R.N.)	Obtenida como la diferencia entre la serie 1 y la 2
5	q4-27001	Captación Río Teno (R.N.)	1963 - a la fecha Período 41/42 - 62/63 extendido a partir de Teno bajo en la Junta
6	q5-25001	Afluente embalse Central Rapel (R.N.) (Sin Teno)	1963 - a la fecha

Una revisión exhaustiva de cada una de las series disponibles permite la detección de los períodos de “relleno” de estadísticas, así como la forma en que se han realizado.

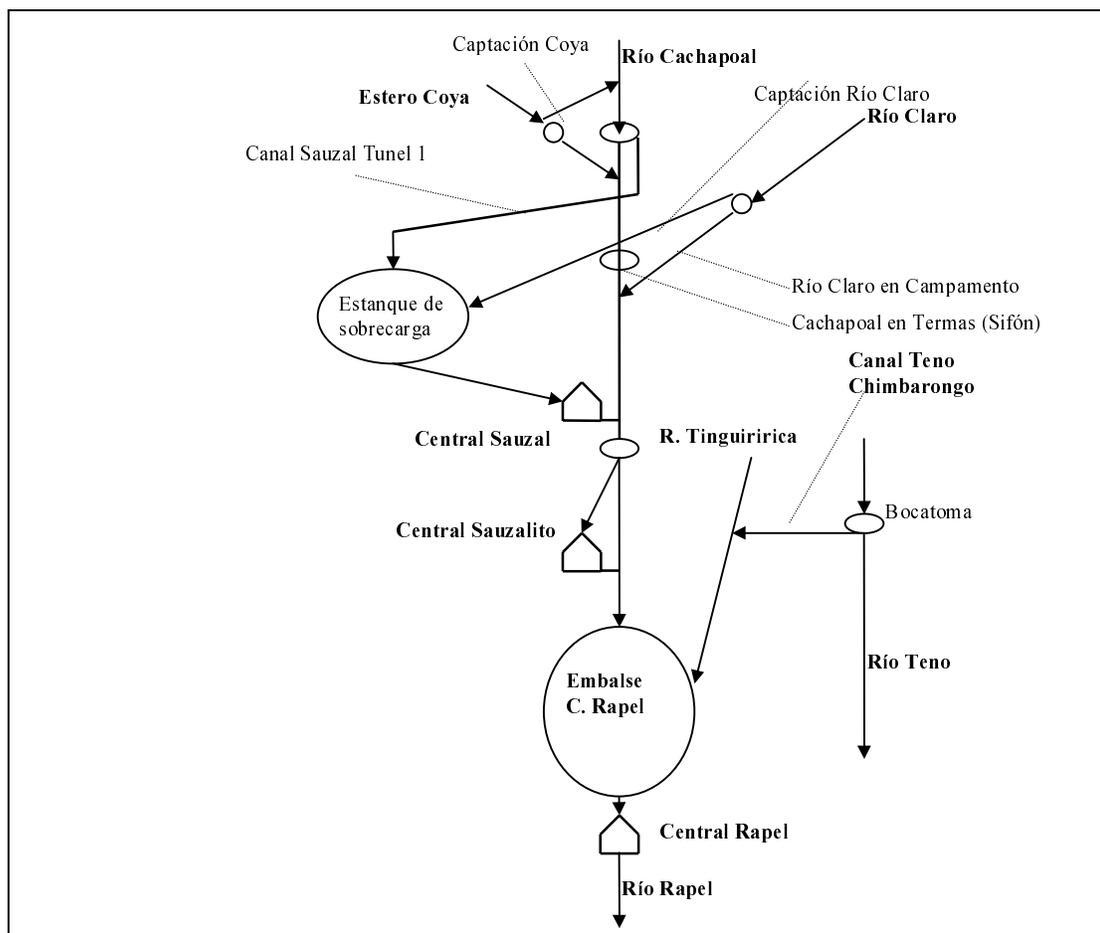


Figura 4.3: Diagrama unilineal resumido de la cuenca del Rapel.

Como el modelo OMSIC utiliza matrices de cuarenta años hidrológicos correspondientes al período 41/42 a 80/81, en los lugares donde existe sólo parte de estos se ha rellenado a través de determinar su correlación con los datos disponibles en otras series relacionadas de la misma cuenca. Es posible apreciar en la tabla 4.3, la cantidad mínima de años a desechar por el “relleno” debe ser 10. Ahora bien, dado que la mayoría de las series se obtienen a partir de Cachapoal en Puente Termas (R.N.) es necesario tener en cuenta la forma en que se han construido las series 2, 3 y 4 de la tabla anterior, es decir, reproduciendo tan sólo momentos de primer orden (promedio).

## iv) Cuenca del Chamiza

Interesa realizar un estudio detallado de las series consideradas como información base y compararlas con las utilizadas en la modelación. En esta cuenca se han considerado 3 series de datos de caudales mensuales de años hidrológicos (Abril a Marzo de cada año) de diferentes extensiones. En la tabla 4.4 se presentan los nombres de las series indicadas y los períodos de medición.

Tabla 4.4: Series hidrológicas registradas en la cuenca del Chamiza.

N	Nombre	Lugar de medición	Período
1	q1-36003	Blanco en Captación (R.N.)	Estimación a partir de precipitaciones
2	q4-36009	Afluentes al Lago Chapo (R.N.)	Estimación a partir de precipitaciones
3	q4-36014	Lenca en Captación (Caudales Captables)	Estimación a partir de precipitaciones

En la figura 4.4 se presenta un diagrama unilineal de la cuenca que muestra la relación entre las series y los aprovechamientos.

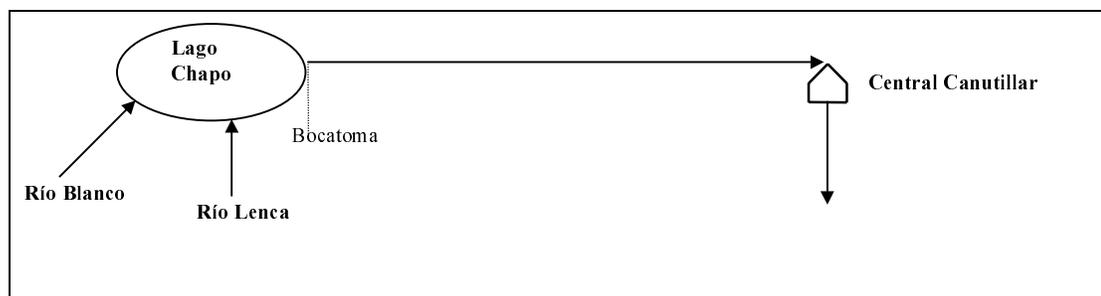


Figura 4.4: Diagrama unilineal resumido de la cuenca del Chamiza.

A partir de los datos presentados en la tabla 4.4 se construyen las series hidrológicas que ocupa el modelo OMSIC y que resumen la información considerada de interés para la operación en la optimización que son las dos que se presentan a continuación (ver [Ende84]):

- Afluentes lago Chapo más Blanco: corresponden a la suma del caudal medido en Blanco en Captación (R.N.) y el caudal de afluentes al lago Chapo (R.N.)
- Afluentes Lenca: corresponde al caudal correspondiente a Lenca en Captación, limitado al caudal captable.

Una revisión exhaustiva de cada una de las series disponibles permite la detección de los períodos de “relleno” de estadísticas, así como la forma en que se han realizado.

En este caso, dado que no se contaba con mediciones de los caudales de interés, para obtener dichos valores se realizó una estimación a partir de las precipitaciones registradas en la zona (ver [Ende84]). Para lo anterior se construyó un mapa de isoyetas de promedios anuales de las precipitaciones en estudio. En aquellas estaciones para las que se cuenta con la precipitación anual para el período 1941/42 a 1981/82, el promedio se obtiene directamente y en las otras estaciones se obtiene a partir de la relación con la precipitación registrada en la estación de Puerto Montt.

Cabe destacar que para el trazado del mapa de isoyetas de la zona, no sólo se han considerado las precipitaciones promedios de las estaciones pluviométricas controladas, sino también se ha tenido en cuenta la esorrentía de las cuencas con control fluviométrico en su salida.

En este caso se debe tener presente la totalidad de la información requerida se ha obtenido a partir de la estimación señalada y tal como se indica en la referencia [Ende84] la duda que se presenta no guarda relación con la magnitud de la precipitación, sino el trazado de las isoyetas que se ha realizado suponiendo que la precipitación aumenta de SW a NW, con un gran incremento en el cordón montañoso que se extiende desde el volcán Calbuco hacia el SE lo que puede no ser realmente cierto. Al no existir antecedentes de las zonas altas, el trazado de las isoyetas ha debido realizarse fijando un criterio que es el antes expuesto.

En este caso, indicar la cantidad de años a utilizar resulta realmente dificultoso por lo antes expuesto.

- v) Otras cuencas menores (Cuenca del Limarí: central Los Molles; cuenca del Maipo: central Alfalfal, Maitenes, Queltehues y Volcán; cuenca del Pilmaiquén: central Pilmaiquén; y cuenca del Huanehue: central Pullinque)

Interesa realizar un estudio detallado de las series consideradas como información base y compararlas con las utilizadas en la modelación. En dichas cuencas se han considerado 8 series de datos de caudales mensuales de años hidrológicos (Abril a Marzo de cada año) de diferentes extensiones. En la tabla 4.5 se presentan los nombres de las series indicadas.

Tabla 4.5: Series hidrológicas registradas en las cuencas del Limarí, Maipo, Pilmaiquén y Huanehue.

N	Cuenca	Lugar de medición
1	Limarí	Bocatoma captación central Los Molles
2	Maipo	Bocatoma río Colorado en capt. más bocatoma río Olivares en capt. central Alfalfal
3	Maipo	Bocatoma capt. río Colorado central Maitenes
4	Maipo	Bocatoma río Maipo en capt. más bocatoma río Volcán en capt. central Queltehues
5	Maipo	Bocatoma capt. río Volcán central Volcán
6	Pilmaiquén	Caudal medido en Pilmaiquén bajo el salto
7	Pilmaiquén	Caudal regulado por el lago Puyehue
8	Huanehue	Caudal del río Huanehue en desagüe lago Calafquén

En las figuras 4.5, 4.6, 4.7 y 4.8 se presentan los diagramas unilineales de la cuenca que muestra la relación entre las series y los aprovechamientos respectivos.

A partir de los datos presentados en la tabla 4.5 se construyen las series hidrológicas que ocupa el modelo OMSIC y que resumen la información considerada de interés para la operación en la optimización que son las 6 que se presentan a continuación (ver [Fern95/2], [Divi82] y [Epop83]):

- moll (corresponde al caudal afluente a la central Los Molles): los datos corresponden a caudales expresados en energía generable. Para el período 41/42 a 44/45 los datos corresponden a relleno y para el período 45/46 a 80/81 estos se obtienen de la generación histórica.

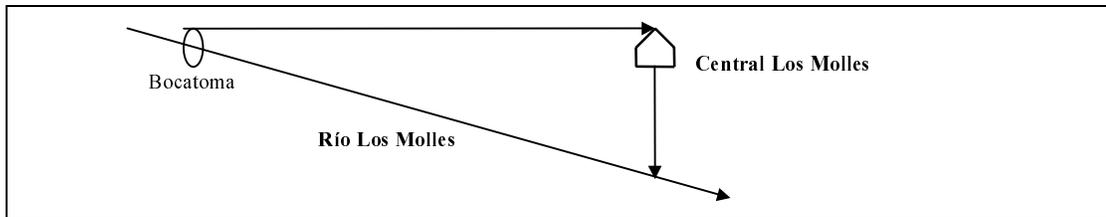


Figura 4.5: Diagrama unilineal resumido de la cuenca del Limarí.

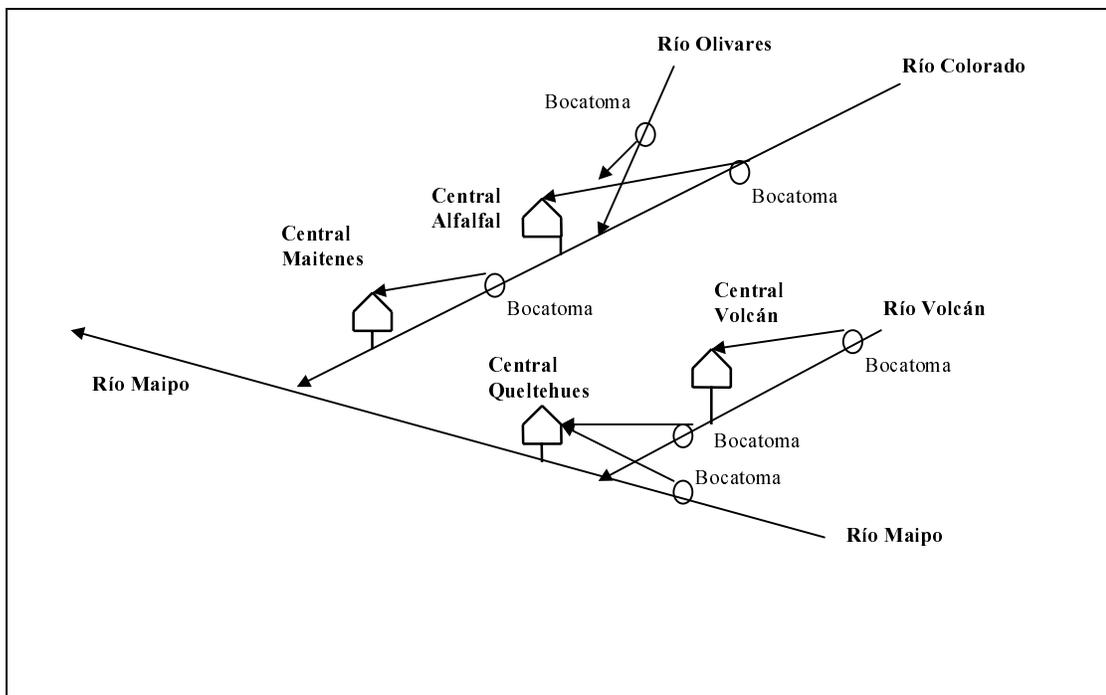


Figura 4.6: Diagrama unilineal resumido de la cuenca del Maipo.

- alfa (corresponde a caudal afluente a la central Alfafal): se cuenta con la siguiente información [Fern95/2]:

- No existen registros de los caudales en las bocatomas anteriores a 1991.

- Las series disponibles corresponden a una proporción de las series de caudales medios mensuales en las estaciones Colorado antes de la junta con Olivares y Olivares antes de la junta con Colorado. Para ello se han considerado coeficientes de ponderación en función de la proporción de áreas y temperaturas medias en las cuencas, propuestos por Motor Columbus (MC) en el estudio hidrológico de la central.

- Las series en las estaciones aludidas tampoco son completas, sino que a su vez se han rellenado en base a otras estaciones del río Maipo, fundamentalmente Colorado en desembocadura.

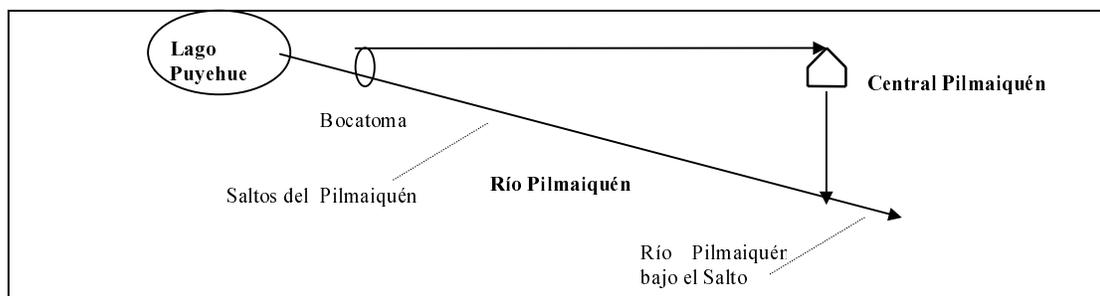


Figura 4.7: Diagrama unilineal resumido de la cuenca del Pilmaiquén.

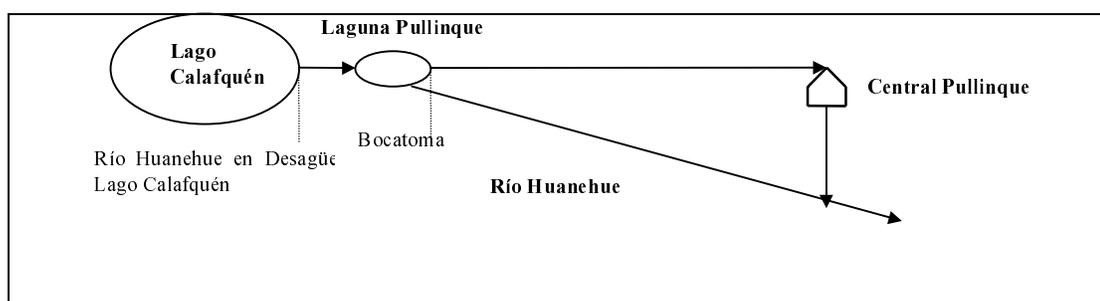


Figura 4.8: Diagrama unilineal resumido de la cuenca del Huanehue.

Tabla 4.6: Registros de información en las estaciones Colorado antes de la junta con Olivares y Olivares antes de la junta con Colorado.

Período	Información base	En los ríos Colorado y Olivares	En bocatomas
41-43	Datos de Colorado en desembocadura rellenos con otras estaciones del Maipo (Maipo en la Obra).	Proporcionales al área de cada cuenca.	Con coeficientes de reparto para cada mes en cada bocatoma.
43-62	Valores registrados de Colorado en desembocadura.	Proporcionales al área de cada cuenca.	Con coeficientes de reparto para cada mes en cada bocatoma.
62-72		Registrados en cada estación, controlados por ENDESA y corregidos por MC.	Con coeficientes de reparto para cada mes en cada bocatoma.
73-76	Valores registrados de Colorado en desembocadura.	Registros con numerosos vacíos corregidos por MC.	Con coeficientes de reparto para cada mes en cada bocatoma.
76-82		Registrados en cada estación, controlados por DGA y corregidos por MC.	Con coeficientes de reparto para cada mes en cada bocatoma.
82-89		Registrados en cada estación, controlados por DGA y corregidos por B. Fernández.	Con coeficientes de reparto para cada mes en cada bocatoma.
91-a la fecha			Mediciones realizadas por Chilgener

- Los registros de ambas estaciones han sido tomados por diferentes instituciones. En la tabla 4.6 se muestra una relación simplificada del origen de las series disponibles para el análisis.

- mait (corresponde a caudal afluente a la central Maitenes): adicionalmente a la nombrada se han considerado las centrales Florida (flor), Sauce Andes (sand), Los Morros (morr), Los Quilos (quil) y Aconcagua (ac12). Hasta

Diciembre de 1956 corresponde a las estadísticas multiplicadas por los rendimientos respectivos de cada central. A partir de Enero de 1957 corresponde a la generación histórica.

- vq93 (corresponde a Volcán más Queltehues): corresponde lo explicado en la serie anterior.
- pilm (corresponde al caudal disponible en bocatoma central Pilmaiquén (R.N.)): este valor corresponde a la suma del caudal medido en Pilmaiquén bajo el Salto y el caudal regulado por el lago Puyehue, estando dicha suma limitada a 150 m<sup>3</sup>/s. Esta estadística queda cortada por supresión de la estación fluviométrica de Pilmaiquén bajo el Salto.
- pull (corresponde al caudal disponible en bocatoma central Pullinque): este valor corresponde al caudal del río Huanehue en desagüe lago Calafquen, limitado a 120 m<sup>3</sup>/s. Debido a que la cuenca intermedia entre el desagüe del lago Calafquén y la bocatoma central Pullinque es muy pequeña, y por no estar regulado el lago Calafquén, los caudales disponibles de la central son los que rebalse el lago. A partir de 1982 se suprime la estación río Huanehue en desagüe lago Calafquén por lo cual la estadística queda cortada.

Una revisión exhaustiva de cada una de las series disponibles permite la detección de los períodos de “relleno” de estadísticas, así como la forma en que se han realizado.

De acuerdo a lo expuesto, los períodos de información realmente medida son bastante dispares entre una serie y otra, aunque por otra parte las centrales agrupadas en esta sección representan una parte poco significativa de las potencias instaladas en el SIC.

#### **4.1.2 Período común recomendado para todas las cuencas del SIC**

De acuerdo a la información presentada en la sección 4.1.1 respecto de las características de los datos presentados y las fechas de data de medición real (libres de relleno) de las series de las distintas cuencas analizadas se ha considerado

desechar los primeros 10 años de la estadística utilizada por CDEC y trabajar con los datos comprendidos entre los años hidrológicos 51/52 y 80/81.

Si bien existen datos para los años que preceden al período antes escogido (81/82 a la fecha), éstos no se han considerado ya que el CDEC no los ha incorporado aún debido a que no se ha logrado un consenso para su inclusión. No obstante lo anterior, dichos datos se utilizaron para el análisis del desempeño de los modelos construidos.

## **4.2 Propiedades hidrológicas de interés**

Para realizar un análisis de las propiedades de interés de las series hidrológicas en estudio se realiza inicialmente una revisión de los aspectos teóricos relacionados con las características de las series hidrológicas para luego presentar ejemplos de lo anterior con series de caudales del SIC.

### **4.2.1 Características de las series hidrológicas**

Las series hidrológicas presentan características distintivas que las diferencian de otras series que pueden encontrarse en economía, ciencias sociales, biología o procesos industriales, en los cuales también tienen amplia aplicación los modelos estocásticos. Las propiedades básicas de cualquier serie cronológica son la deriva, la distribución marginal, la estructura de dependencia y la periodicidad. En hidrología se ha puesto atención además a las propiedades de almacenamiento, corte, intermitencia e irreversibilidad, entre otras. [Fern90] y [Hipe94]

#### **a) Deriva y saltos.**

Una serie de tiempo presenta deriva, llamada también tendencia, si sus propiedades cambian paulatinamente en el tiempo. Si el cambio es brusco se dice que presenta saltos. Estos cambios pueden afectar a la función de distribución, o al valor de alguno de sus parámetros como el promedio, desviación, asimetría, a la estructura de dependencia u otra propiedad. Existen numerosas técnicas que permiten detectar y cuantificar cambios en las series cronológicas, desde las tradicionales dójimas de Student y Fisher [Mood74] para verificar cambios en el

promedio o la dispersión de las series independientes normales, hasta los métodos más elaborados como el denominado análisis de intervención para series dependientes ([Box65]). Si bien los cambios son fácilmente detectables, no existe ninguna recomendación que permita extrapolarlos en el futuro de las series, de manera que al modelar una serie hidrológica una vez detectada una tendencia o un salto, se remueven de la muestra disponible y se modela la serie estacionaria.

b) Periodicidad.

Las series geofísicas presentan ciclos periódicos como consecuencia de los ciclos naturales impuestos en ellas por el movimiento astronómico de la tierra. El período más largo que se ha detectado es de un año y el más corto de utilidad práctica de un día. Si bien el ciclo lunar tiene cierta influencia en algunas series de niveles en general no se considera importante. Como consecuencia de lo anterior, las series que tienen intervalos de discretización iguales o mayores a un año se asumen no periódicas, mientras las series con intervalos menores que un día pueden presentar hasta dos ciclos superpuestos. En [Fern85] se hace un análisis de la periodicidad en series hidrológicas y presenta algunas herramientas para modelarla.

c) Distribución marginal de la variable.

Se han utilizado diversas funciones de distribución para describir el comportamiento estadístico de los datos hidrológicos. En general, la distribución depende de la variable y del intervalo de discretización utilizado. En [Kleme74] se postula que para los datos hidrológicos que físicamente no pueden adoptar valores negativos, como los caudales y las precipitaciones, sólo serían aceptables distribuciones asimétricas con límite inferior. Por otra parte la mayoría de los modelos desarrollados suponen una distribución normal de la variable. Para considerar los efectos de la asimetría se puede recurrir a alguno de los tres procedimientos siguientes: transformar la serie para normalizarla y utilizar modelos normales; determinar las características de la variable independiente de modo de reproducir la asimetría de la serie original; y, finalmente, recurrir a modelos no normales que consideran en su estructura la asimetría.

## d) Dependencia.

La dependencia es uno de los aspectos más importantes de una serie cronológica. De hecho es la responsable de la intención implícita en el proceso a la cual se refiere el apelativo de estocástico. Ella cuantifica en qué medida los valores numéricos que la variable adopta en el instante  $t$  están relacionados con la historia anterior de la serie y refleja la memoria del proceso. Esta dependencia se mide a través de la función de autocovarianza que corresponde a:

$$\text{Cov}(x_t, x_{t-k}) = E[(x_t - u_x)(x_{t-k} - u_x)]$$

donde  $E(\cdot)$  representa el valor esperado. Si el proceso es estacionario esta función no depende del tiempo sino sólo de la separación  $k$  entre los intervalos. Como valor numérico se prefiere utilizar la autocorrelación definida como:

$$\rho_k = \text{Corr}(x_t, x_{t-k}) = E[(x_t - u_x)(x_{t-k} - u_x)] / E[(x_t - u_x)^2]$$

A  $\rho_k$  en función de  $k$  se le denomina correlograma y es una medida de la estructura de dependencia lineal del proceso.

Si el proceso en el instante  $t$  es lineal y depende sólo de lo ocurrido en el instante anterior  $x_{t-1}$ , corresponde al llamado proceso de Markov o autoregresivo de orden uno. En este caso se dice que tiene memoria corta ya que el correlograma decrece exponencialmente con  $k$  y tiende a cero para valores altos. Por otra parte si el correlograma presenta valores importantes para valores altos de  $k$  se dice que presenta memoria larga o persistencia. Una de las peculiaridades interesantes de las series hidrológicas de memoria larga es que exhiben conjuntos agrupados de valores altos y bajos. Este comportamiento fue denominado efecto Josué por [Mande68] como referencia a los bíblicos siete años de abundancia seguidos por siete de escasez.

## e) Almacenamiento y el fenómeno de Hurst.

Muchos de los modelos estocásticos de series de caudales se han propuesto y utilizado para definir o analizar el volumen de embalses de dimensiones importantes en relación con el recurso disponible. En estos casos adquiere relevancia el efecto Josué característico de series de memoria larga. Una propiedad estadística

cuantificable de una serie relacionada con el almacenamiento necesario, es el denominado rango,  $R$ , definido para una serie de longitud  $n$  como la máxima diferencia entre los valores acumulados de la diferencia con el promedio. Ésto corresponde al método tradicional propuesto por Rippl (1883) mediante el cual se determina el volumen necesario para que dada una oferta variable se pueda entregar siempre el valor medio sin que se produzca ningún déficit y, permitiendo que el embalse se vacie sólo una vez.  $R$  es una variable aleatoria cuyo valor esperado depende de la dispersión de la serie, o su desviación típica  $\sigma$ , y su longitud  $n$ , de manera que:

$$E[R_n] = \sigma * n^h$$

Si el proceso de oferta está formado por valores normales independientes estacionarios se puede demostrar que el valor de  $h$  es 0,5. Hurst encontró que para 837 series largas de eventos tales como escurrimientos, precipitaciones, temperatura, anillos de árboles, niveles de ríos y lagos, presión atmosférica, capas de sedimentos lacustres y otros, la constante se distribuye normalmente con valor medio 0,73 y desviación típica 0,092 [Hurs51]. El hecho de que  $h$  fuera diferente de 0,5 o más precisamente su aparente convergencia a un valor constante entre 0,5 y 1,0 en series geofísicas, fue denominado posteriormente por Lloyd [Lloy67] como el “fenómeno de Hurst”.

Intentando una interpretación o explicación física en [Klem78] se describe como denominadores comunes de procesos que exhiben el fenómeno de Hurst el hecho de que la mayoría de esas series geofísicas corresponden al resultado de fenómenos acumulativos en la naturaleza, su propiedad estadística de formar ondas y su tendencia a gatillar cambios bruscos una vez que se ha alcanzado un cierto umbral.

Con el objeto de reproducir el fenómeno de Hurst en la generación de series hidrológicas destinadas a la simulación para analizar el comportamiento de sistemas de embalse importantes, se han propuesto varios modelos que incorporan entre sus parámetros el valor de  $h$ . Mandelbrot [Mand65] propuso un proceso estocástico continuo conocido como ruido Browniano Fraccionado,  $fBn$ , que puede tener valor de  $h$  entre 0 y 1, el cual tiene una “memoria infinita” debido a que se

define como una suma infinita de series aleatorias. Una versión discreta de este modelo conocida como Ruido Gaussiano fraccionado discreto,  $dfGn$ , fue propuesta por el mismo autor quien posteriormente, para evitar algunas de las dificultades matemáticas de este modelo, sugirió el ruido gaussiano fraccionado rápido,  $ffGn$ . Estos modelos han sido criticados porque consideran siempre variables normales y relacionan el fenómeno de Hurst sólo a la dependencia. Además, son muy difíciles de utilizar en series transformadas o periódicas. Con similar objetivo Mejía et al. [Mejí72] desarrollaron el llamado proceso de líneas quebradas (broken lines) para su uso en hidrología. O'Connell [O'Con74] quien investigó el uso de modelos ARMA (1,1) para modelar el fenómeno de Hurst, observando que cuando el coeficiente autoregresivo se aproxima a 1, las series generadas por este tipo de modelos exhiben características similares a las del  $fGn$ . Finalmente, en el último tiempo pareciera haber decrecido el interés por el fenómeno de Hurst en los modelos estocásticos hidrológicos, en parte luego de haberse demostrado que corresponde a un fenómeno asintótico que se hace constante cuando  $n$  tiende a infinito, pero es función del tamaño de las muestras para valores finitos, situación que no corresponde a la realidad de los desarrollos hidráulicos que presentan una vida útil de 50 años o menos.

f) Propiedades de corte. Demandas y sequías.

Otra propiedad que ha sido considerada importante en los modelos estocásticos es su capacidad para reproducir características de corte, o de las sucesiones generadas por procesos de corte [Yevj72]. En muchos casos las series hidrológicas son utilizadas para analizar el comportamiento de sistemas que deben satisfacer una cierta demanda de recursos. La diferencia entre la oferta y la demanda genera sucesiones de valores positivos o negativos que representan superávit o escasez. Éstos pueden caracterizarse mediante propiedades como la longitud de la sucesión, máximo déficit, intensidad y otros similares. Estas propiedades están relacionadas con la distribución marginal de la variable y la dependencia del proceso. Además, cuanto más dependiente es una serie, más largas tienden a ser las sucesiones que con ella se generan. El caso de series independientes normales ha sido ampliamente estudiado, no así el caso de series no normales muy dependientes.

### g) Irreversibilidad o direccionalidad

En la medida que los modelos estocásticos han sido utilizados para el pronóstico se ha hecho necesario reducir el intervalo de tiempo, siendo común en estos casos los modelos de caudales diarios e incluso horarios. En estas series aparece como una característica relevante la irreversibilidad en el tiempo o direccionalidad, definida como una característica de la serie según la cual sus propiedades estadísticas cambian si la dirección del eje del tiempo se invierte. Desde el punto de vista físico esta propiedad se observa en el conocido hecho de que las series de caudales presentan ascensos rápidos y descensos lentos como resultado de las crecidas. Weiss [Weis75] demuestra que un proceso estacionario lineal normal es siempre reversible, por lo tanto, si se desea reproducir direccionalidad es necesario recurrir a modelos que son o no estacionarios, o no lineales, o no gaussianos. Con este objeto Fernández y Salas [Fern86] proponen modelos Gama Autorregresivos Lineales, Multiplicativos y una combinación de ambos. Tong [Tong83] recurre a modelos de umbral según el cual el coeficiente autorregresivo cambia dependiendo del nivel de la variable. Todos estos modelos, si bien exhiben irreversibilidad, no la incorporan en los parámetros, de manera que la reproducción es más bien cualitativa y no cuantitativa.

El método más simple para detectar direccionalidad está basado en la observación de un gráfico de la serie cronológica. Si el tiempo que ocupa el proceso en los ascensos es diferente del tiempo empleado en los descensos, la serie es irreversible.

En hidrología es claro que los caudales y muchas de las series que de ellos se derivan como niveles y algunas concentraciones de elementos descargados en ríos, son irreversibles, mientras las precipitaciones son reversibles. El interés de su reproducción en modelos de caudales para ser usados en pronósticos a corto plazo está en el conocido hecho de que es más simple y confiable pronosticar durante la recesión de una crecida que durante su ascenso.

#### h) Intermitencia

En la medida en que aumenta el interés por modelar el proceso de precipitaciones, o utilizar intervalos de discretización más cortos en series de caudales, es más común encontrarse con procesos intermitentes, en los cuales durante ciertos intervalos el fenómeno no está presente o mantiene valores constantes. En el caso de precipitaciones esto ocurre incluso para series anuales en climas desérticos y es común en series mensuales. En algunas series de caudales en cuencas pequeñas puede aparecer intermitencia como resultado de la periodicidad de las precipitaciones si la cuenca no tiene suficiente capacidad de regulación.

Un proceso intermitente se caracteriza por una distribución marginal de la variable del tipo mixta, en la cual existe una probabilidad no nula de que la variable adopte un valor, por ejemplo cero, y una función de densidad continua para todo el resto de valores posibles. Según Yevjevich [Yevj72] la modelación de procesos intermitentes puede seguir cualquiera de las siguientes tres alternativas: a) caracterización del proceso en su estado activo solamente; b) modelar la secuencia de ceros y valores diferentes considerando el tiempo de inactividad como variable del proceso y c) modelar un proceso continuo con valores en todo el rango pero truncado en el valor estable.

#### **4.2.2 Análisis de series de caudales del SIC**

A continuación se presenta un análisis de las propiedades hidrológicas básicas de un subconjunto de las series utilizadas por el modelo OMSIC. Dichas series reproducen razonablemente variadas situaciones hidrológicas que se presentan en esta zona, incluyendo recursos desde la cuenca del Rapel por el norte hasta el Chapo por el sur. El conjunto de series hidrológicas que intervienen en la modelación del OMSIC son 27. Unas pocas de ellas, que se indican en la tabla 4.7, han sido seleccionadas para ejemplificar la identificación de los modelos, estimación de parámetros, así como la verificación y comparación con el sistema actual. En el ANEXO C se incluyen los datos de cada una de estas series de valores semanales.

Tabla 4.7: Series consideradas para el análisis de las propiedades hidrológicas.

Serie	Corresponde a
Rapel	Afluentes al embalse Rapel - demanda de riego Río Teno
Colbún	Cuenca intermedia muro Colbún Laguna Invernada - Laguna Maule
Melado	Afluentes al embalse Melado
Antuco	Antuco Pasada
Chapo	Afluentes al Lago Chapo más Blanco

a) Propiedades estadísticas

Se trata de series periódicas, que reproducen características estadísticas cada año para los mismos períodos. Estas series han sido formadas para 48 valores pseudosemanales, con meses de cuatro semanas, de manera que la cantidad de días asignados a cada semana fluctúa entre 7 y 8 de acuerdo a una convención adoptada por el CDEC. Los valores están organizados por años hidrológicos, es decir, de abril de un año a marzo del siguiente, siendo la semana 1 la que se inicia el primero de abril.

i) Series anuales y semanales

Las series anuales permiten visualizar el comportamiento general de los recursos en el largo plazo y compararlos entre ellos. La figura 4.9 muestra el comportamiento de la serie anual de Colbún (los valores semanales se adjuntan en el ANEXO C). La tabla 4.8 muestra las principales estadísticas anuales de las cinco series consideradas.

Las series semanales son claramente periódicas, de manera que las propiedades estadísticas son variables de una semana a la otra. Para ellas interesa reproducir los promedios, desviaciones, asimetrías y la dependencia temporal, que queda representada mediante los coeficientes de autocorrelación hasta el orden 15 por ejemplo. Se trata de 48 valores de cada parámetro para cada serie, los que se presentan en el ANEXO C.

Tabla 4.8: Propiedades estadísticas de las series anuales.

Serie	Años	Promedio	Desviación	Asimetría	Autocorrel.
Rapel	41-80	195,8	167,6	1,7	0,785
Colbún	41-80	227,4	110,8	1,4	0,769
Melado	41-80	111,2	57,7	1,5	0,762
Antuco	41-80	57,9	33,6	1,6	0,709
Chapo	41-80	51,3	33,7	1,4	0,312

Tabla 4.9: Valores máximos de las asimetrías semanales de las series en estudio.

Serie	Valor máximo
Rapel	3,502
Colbún	5,086
Melado	4,761
Antuco	4,084
Chapo	4,061

Todas las series presentan valores importantes de asimetría, especialmente durante los períodos de otoño e invierno, épocas en las cuales se presentan las crecidas de mayor magnitud, con excepción de los afluentes de Chapo, que tiene asimetrías importantes en verano. Conviene recordar que los valores normalmente distribuidos tienen una asimetría de valor cero. Ésto obliga a transformar las series semanales antes de identificar y estimar parámetros de modelos que se basan en series normales. La tabla 4.9 muestra los valores máximos de asimetría detectada en las series semanales.

Las figuras 4.10, 4.11 y 4.12 muestran el comportamiento de los promedios, desviaciones típicas y asimetrías semanales de la serie Colbún. Las otras series presentan comportamientos similares como puede apreciarse en las figuras correspondientes agregadas en el ANEXO C.

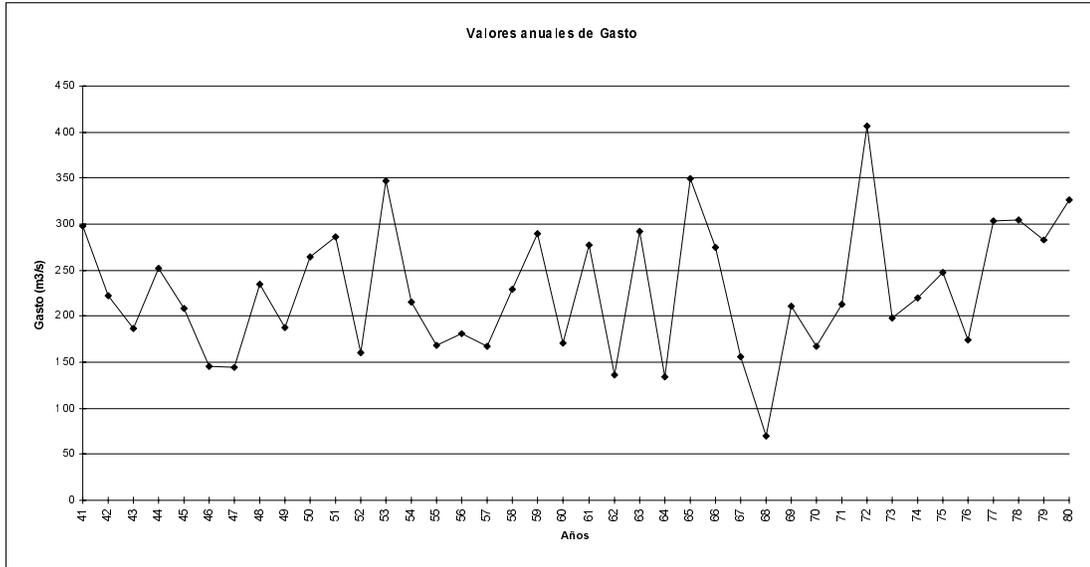


Figura 4.9: Serie de valores anuales de Colbún.

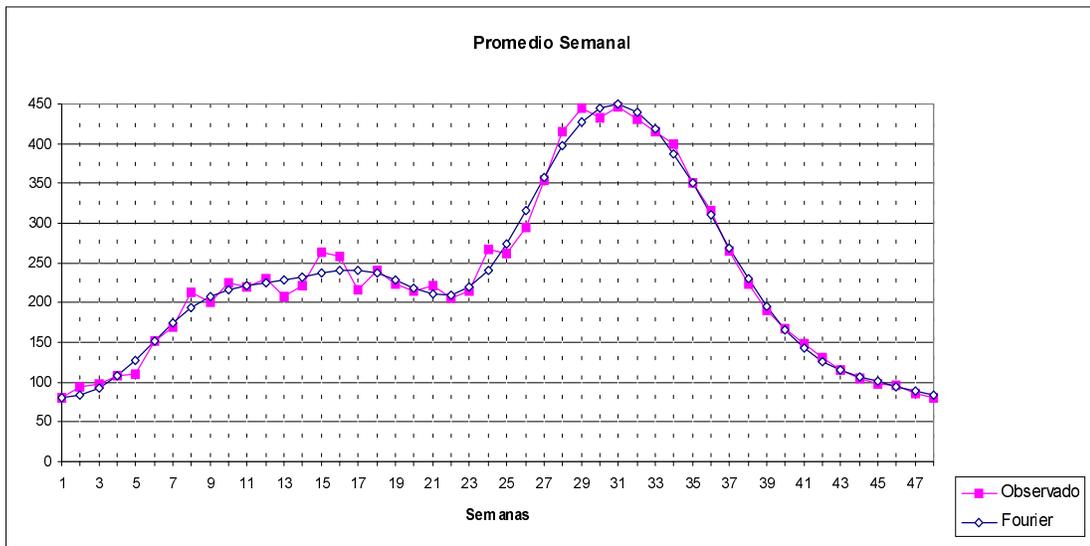


Figura 4.10: Promedios semanales de la serie Colbún. Se agrega la representación de los valores periódicos mediante series de Fourier.

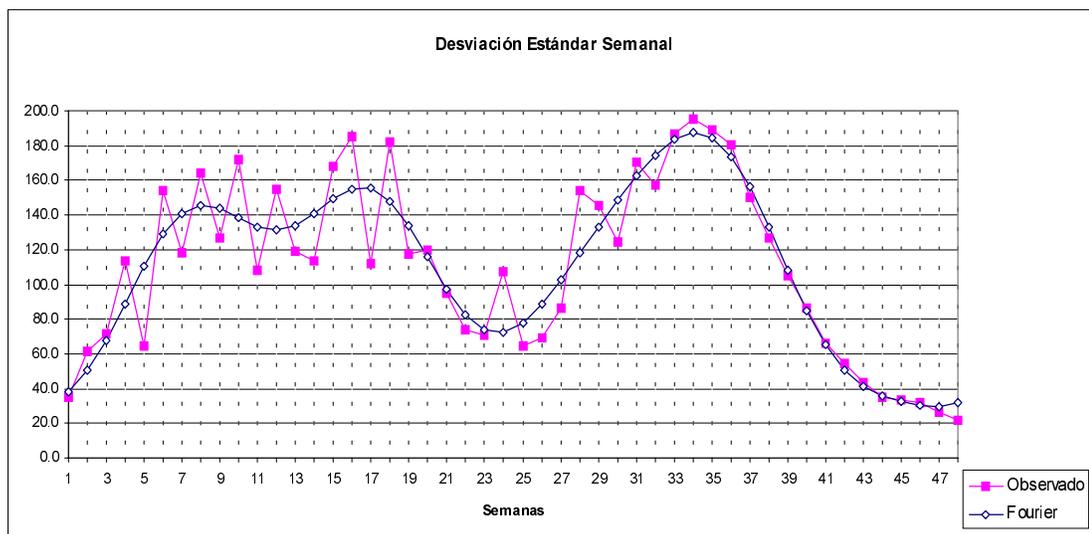


Figura 4.11: Desviaciones típicas semanales de la serie Colbún. Se agrega la representación de los valores periódicos mediante series de Fourier.

## ii) Periodicidad y representación de Fourier

La variación de las propiedades estadísticas para cada semana se debe en parte a la periodicidad propia del fenómeno hidrológico que representan, pero también incluye el efecto de muestreo de las series. No es razonable un cambio demasiado alto de las propiedades para semanas consecutivas y menos aún cambios variables entre semanas consecutivas. Además, la representación mediante modelos de propiedades diferentes para cada semana obliga a una gran cantidad de parámetros. Para solucionar ambos problemas se puede recurrir a la representación paramétrica de estas propiedades, siendo las series de Fourier el mecanismo más utilizado. Ésto permite representar una propiedad variable mediante pocos parámetros. En estos ejemplos si bien se han reproducido todas las propiedades estadísticas semanales de todas las series de acuerdo a una representación mediante series de Fourier, esta representación ha sido empleada sólo para el caso de la asimetría con el objeto de decidir sobre las transformaciones. Para el resto de los parámetros se han considerado los valores estimados de las series de datos.

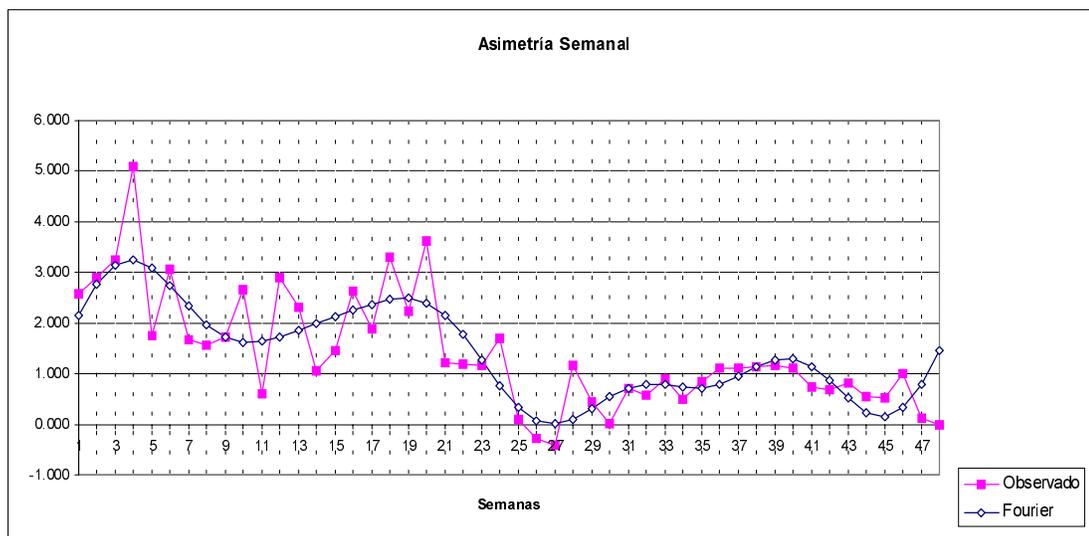


Figura 4.12: Asimetrías semanales de la serie Colbún. Se agrega la representación de los valores periódicos mediante series de Fourier.

En la figura 4.12 se muestran los valores estimados de la asimetría semanal en el caso de la serie Colbún y su presentación mediante series de Fourier. En el ANEXO C se muestran las representaciones de los demás parámetros de todas las series empleadas en este estudio. En estas figuras se aprecia la variabilidad muestral de estos parámetros, al comparar los valores de una semana cualquiera con los de las semanas adyacentes. La variabilidad más importante es la del coeficiente de asimetría, seguida por la desviación típica. El promedio presenta poca variabilidad. En todos los parámetros se aprecia mayor consistencia en los valores del verano.

### iii) Asimetría y transformaciones

El pronóstico se completa con bandas de confianza que supone conocer el comportamiento estadístico de los errores o residuos del modelo. Para poder estimar numéricamente el ancho de estas bandas se supone que estos errores son un ruido blanco, es decir, valores independientes y normalmente distribuidos. Para que esto sea así es necesario que los valores de las series a pronosticar sean normales. Como los caudales semanales no lo son, de acuerdo a los valores de la asimetría

comentados anteriormente, es necesario transformar las series, estimar los modelos y realizar el pronóstico en el espacio transformado, y efectuar, posteriormente, una transformación inversa para expresarlos en términos de los valores reales.

Adicionalmente los modelos que se emplean tanto para la simulación como para el pronóstico usan como supuesto fundamental para la identificación y estimación un comportamiento normal de las series. Es por ello que es recomendable realizar la transformación de los datos previamente a la estimación de los modelos, y no solamente transformar los errores del pronóstico para calcular las bandas de estimación, que es otro procedimiento alternativo.

Para el caso de los valores normalmente distribuidos el coeficiente de asimetría es también normal de promedio nulo y desviación típica  $(6/N)^{1/2}$ , que para series de 40 valores es 0,39. La transformación se basa en reducir la asimetría para cada una de las semanas de la serie a un valor que pueda considerarse normal. Sin embargo, cuando se tienen valores altos de asimetría ello se debe a que está muy influenciada por los valores más altos que se presentan en los datos, algunos de los cuales pueden corresponder a los que se denominan *outliers*. Para considerar este efecto, en este análisis se ha introducido el siguiente método para la transformación: se elimina el máximo valor de cada semana, el cual puede determinar la estimación de la asimetría y se ajusta una serie de Fourier para las asimetrías semanales así calculadas. Si en esta representación el coeficiente de asimetría es superior a 0.6, o inferior a -0.6, se aplica una transformación a la serie de esa semana. Se procede de igual manera para todas las semanas hasta que la representación de Fourier de la asimetría es menor que 0.6 en todos los casos.

La transformación aplicada es del siguiente tipo:

$$y = \begin{cases} \ln X, & \text{si } g > 0.6 \\ X, & \text{si } -0.6 \leq g \leq 0.6 \\ X^2, & \text{si } g < -0.6 \end{cases}$$

donde  $g$  es el valor de la representación de Fourier de la asimetría de esa semana.

Estas transformaciones pueden aplicarse consecutivamente hasta que se logra reducir la asimetría estimada de cada período a valores inferiores a 0,60. Como máximo se consideró aceptable dos transformaciones aunque no siempre se consiguió para todas las semanas reducir la asimetría a valores menores que 0,60; Ésto ocurre sin embargo en muy pocas oportunidades. Por el contrario en la gran mayoría de los casos sólo se requiere la primera transformación y en varios ninguna.

Como resultado de estas transformaciones las series pueden ser consideradas normales. La modelación requiere que para cada serie en cada semana se indique la cantidad y tipo de transformaciones aplicadas hasta lograr su normalización.

Una vez que los valores semanales están transformados, es decir, se pueden considerar normales, se estandarizan de manera de trabajar con series de distribución normal, promedio cero y varianza unitaria. Ésto es, se trabaja con

variables  $z$  dadas por: 
$$z_{m,n} = \frac{y_{m,n} - \bar{y}_m}{\sigma_{y_m}}$$

donde  $y_{m,n}$  es el valor transformado de la semana  $m$  del año  $n$ ,  $\bar{y}$  el promedio de esa semana y  $\sigma_{y_m}$  la desviación típica. La modelación se efectúa entonces con los  $z_{m,n}$  que son todos normales, o aproximadamente normales, de promedio nulo y desviación unitaria.

### 4.3 Observaciones acerca de las series disponibles en el SIC

Basándose en el análisis efectuado se han detectado algunas propiedades que deben considerarse en la modelación de las series cronológicas de caudales de los ríos de la zona central de Chile [Fern95/1]:

- a) Existe una dependencia espacial importante entre cuencas vecinas. La dependencia temporal también es alta aunque variable a lo largo del año, presentándose como hecho destacado una alta correlación entre períodos

sucesivos de la temporada de deshielo en cuencas nivales. La dependencia temporal durante el invierno y otoño es menor.

- b) Las propiedades de interés para la modelación son las siguientes: promedios, desviaciones típicas, asimetrías, coeficientes de autocorrelación, valores mínimos y máximos. Dichas propiedades presentan una importante periodicidad en las series semanales.
- c) 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, 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 a 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 relleno 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 del modelo este persistirá en el mismo error. Ésto es posible apreciarlo en los valores que se presentan para las series Laja en Tucapel y Estero el Toro en Captación, donde los 10 primeros años de la segunda serie, que se presenta en la figura 4.13, corresponden a la misma forma de los análogos de la serie Laja en Tucapel que se presenta en la figura 4.14.
- d) Las series de caudales de las cuencas nacionales presentan altas asimetrías, lo que hace necesario la aplicación de transformaciones de manera de cumplir con los supuestos de normalidad que requiere la construcción de los modelos del tipo que se han empleado.

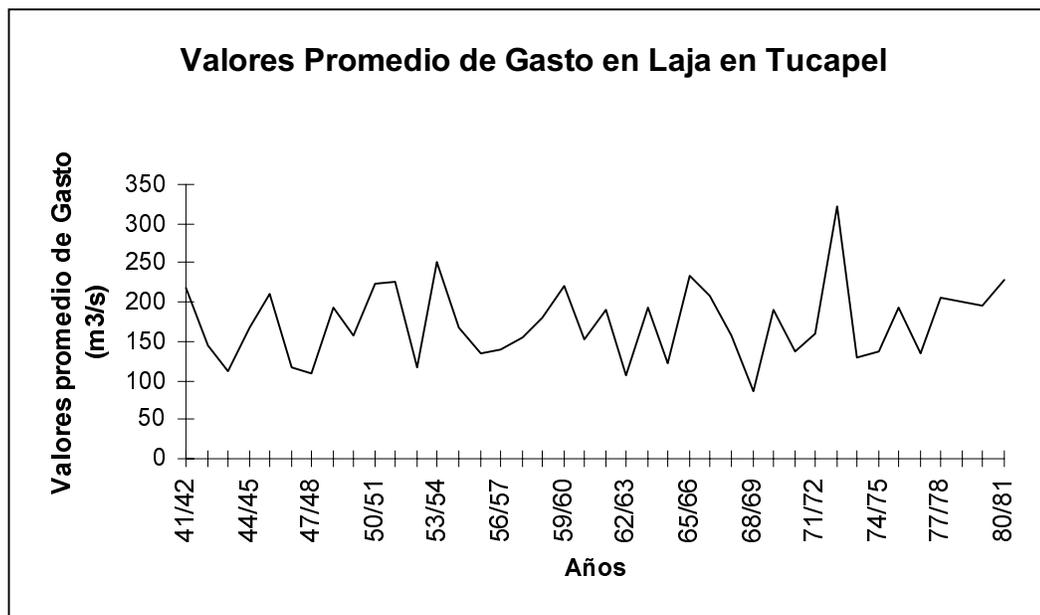


Figura 4.13: Valores promedio de gasto en Laja en Tucapel.

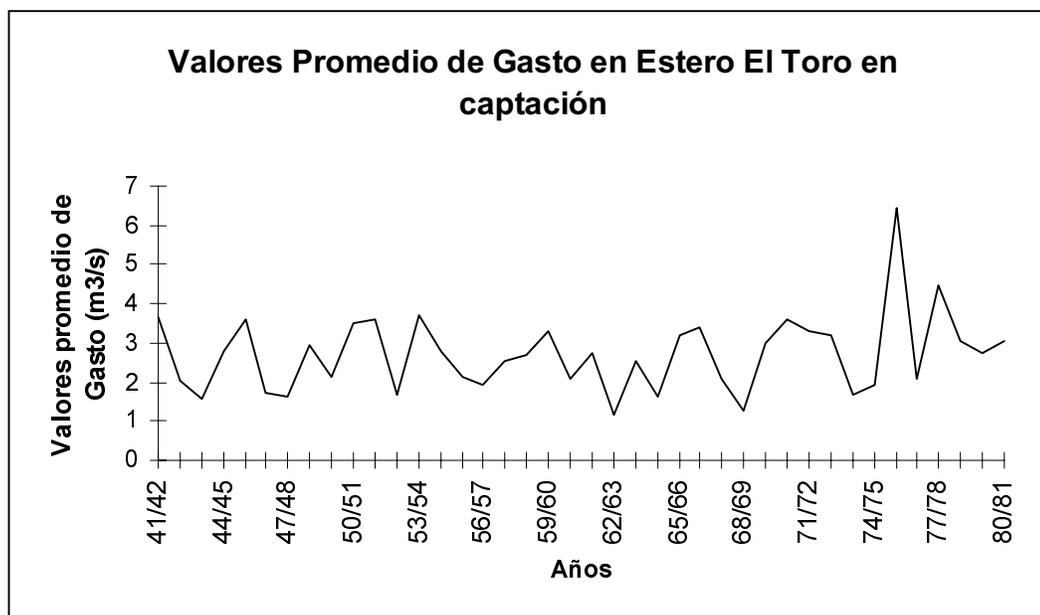


Figura 4.14: Valores promedio de gasto en Estero El Toro en captación.

## V. MODELACIÓN HIDROLÓGICA DE LOS RECURSOS DISPONIBLES

### 5.1 Construcción de un modelo

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

Dado que las posibilidades están restringidas a los modelos conocidos (cuyo desarrollo puede ser encontrado en las referencias [Sala85] y [Hipe94]), en la mayor parte de las situaciones se supone un cierto modelo y se verifica si es el adecuado. En este sentido, no se trata de encontrar un modelo que corresponda al proceso que ha generado la serie en cuestión, sino más bien *elegir por comparación*, entre los posibles modelos, el que mejor represente las características consideradas relevantes.

Para el caso de los modelos lineales existe una metodología de modelación propuesta por Box y Jenkins que es ampliamente aceptada. Ésta consiste en tres etapas que son: identificación del orden, estimación de parámetros y validación. A esto Salas agrega dos etapas previas: composición del modelo y selección del tipo.

Conviene destacar que la selección del tipo de modelo debe apoyarse fuertemente en el comportamiento del fenómeno, la forma de las no linealidades involucradas y la importancia de ellas en el comportamiento de la serie.

Como esquema general de modelación, en esta tesis se abordan los aspectos mencionados en la sección 3.1, que corresponden a la selección del tipo de modelo, identificación del orden, estimación de parámetros y verificación de las condiciones supuestas. La modelación general es complementada con el uso del modelo, ya sea en simulación o pronóstico.

En el caso de la simulación, ésta se realiza mediante la generación de series sintéticas (matrices hidrológicas rellenas con los datos obtenidos a partir de la utilización de los modelos ajustados), a partir de los modelos construidos, siendo

éstas utilizadas como datos de entrada para el modelo OMSIC. Ésto permite obtener distintos valores de costo marginal para cada una de las hidrologías generadas a partir de los modelos.

Así también, es posible utilizar los modelos para realizar pronósticos en tiempo real, obteniéndose tanto los valores predecidos como una banda que define un intervalo de confianza para el pronóstico realizado. Los valores de pronóstico se comparan con los pronósticos para esas mismas situaciones realizados por el sistema actualmente en uso por el CDEC, calculando para ello indicadores basados en los errores de ambos tipos de pronóstico.

## **5.2 Principales modelos utilizados por el CDEC**

Las necesidades de modelación para la operación del SIC de las empresas generadoras, el CDEC y la CNE son satisfechas, actualmente, mediante la utilización de los modelos GOL (Gestión Óptima del Laja) y OMSIC (Operación Mensual del Sistema Interconectado Central). Más recientemente, se ha buscado incorporar modelos que representen con mayor precisión la realidad (multinodales y multiembalses). En este sentido, se ha desarrollado un modelo de operación económica, usando Programación Dinámica Dual Estocástica (SDDP), no obstante no se ha realizado su incorporación definitiva. Adicionalmente, en la actualidad el CDEC se encuentra desarrollando un modelo de características similares al antes descrito. A continuación se presenta una breve descripción del funcionamiento de los modelos antes nombrados:

### **a) Modelo GOL (Gestión Óptima del Laja) [Cne93]**

El Modelo GOL fue concebido como un modelo de programación dinámica estocástica, donde el consumo está centrado en una barra única, optimiza extracciones del Lago Laja, minimiza el costo actualizado de generación termoeléctrica y falla en todo horizonte de planificación. Utiliza criterio azar-decisión y considera etapas trimestrales. Se deja de usar por su prolongado tiempo de reacción ante cambios en las reservas del Lago Laja, ya que se actualizaba cada tres meses; no se representaba la estacionalidad de los consumos dentro de los tres meses y no consideraba la capacidad de regulación de los embalses menores.

En la actualidad, la generación en el SIC es principalmente de origen hidroeléctrico, complementada con generación proveniente de centrales a carbón y petróleo. Dentro de las centrales hidroeléctricas, las hay de pasada, de embalse con pequeña capacidad de regulación y finalmente aquellas asociadas al embalse del Lago Laja, cuya capacidad de regulación es interanual. Resulta entonces evidente que la gestión de este embalse condiciona la generación termoeléctrica y en consecuencia los costos de operación del sistema.

En líneas gruesas, la operación del sistema a costo mínimo significa determinar qué volumen del Lago Laja debe ser generado en un período determinado de tiempo, dada su incidencia en los costos de operación presente y futura. Naturalmente, lo anterior supone una gestión adecuada del resto de las centrales: máxima utilización posible de las centrales de pasada, traspasos óptimos en las centrales con menor capacidad de regulación y utilización de las centrales térmicas según costos marginales de operación crecientes. En el caso de Chile, la hidrología juega un papel de mucha importancia y debe tomarse en consideración su carácter aleatorio.

b) Modelo OMSIC (Operación Mensual del Sistema Interconectado Central)  
[Cdec91]

El modelo OMSIC es un modelo de operación del SIC con etapas mensuales que utiliza programación dinámica y que optimiza las extracciones del Lago Laja y la generación de las centrales térmicas, considerando las centrales de pasada. En este modelo la hidrología se considera independiente entre meses en el período de invierno (Abril-Septiembre) y dependiente en el período de verano (Octubre-Marzo, meses de deshielo), y las decisiones se toman en la modalidad azar-decisión. Adicionalmente, existe una serie de modelos para las centrales que cuentan con embalses (modelo Rapel, de la Laguna Invernada, del embalse Melado, de los embalses Colbún-Machicura y Canutillar. Estos modelos reciben una señal de costo marginal de producción del modelo OMSIC y a partir de ella entregan la energía que económicamente resulta óptima generar a ese precio. Con esa energía se verifica si se satisface la demanda. De ser posible el modelo ha convergido y ese es el costo marginal a utilizar. De no ser así, se inicia una nueva iteración del modelo. A la

interacción entre los modelos de embalses y el modelo OMSIC se le denomina proceso de convergencia.

Este modelo considera dos etapas, una de optimización y otra de simulación.

En la fase de optimización se busca minimizar los costos presentes y futuros de operación del sistema, mediante la utilización de programación dinámica.

Luego, en la fase de simulación del modelo, se realiza un pronóstico de las próximas cuatro semanas. El sistema de pronóstico empleado por el CDEC consiste en la ponderación de los datos de caudales en los 7 días anteriores al pronóstico [Cdec93]. El esquema de ponderación utilizado para el caudal de la semana siguiente considera el del último día con información ponderado por un factor y el promedio de los seis días anteriores con el factor complementario, es decir:

$$Q_1 = A(q_{j-1}) + (1-A) (q_{j-2} + q_{j-3} \dots + q_{j-7})/6$$

donde A toma valor 0,3 ó 0,5 de acuerdo a si se trata del período de verano o invierno, respectivamente. Este valor es igual para todas las cuencas. Como puede apreciarse se trata de un modelo del tipo autorregresivo de los caudales diarios, con coeficientes de ponderación seleccionados arbitrariamente.

- c) Modelo de despacho hidrotérmico con restricciones de transmisión SDDP-PSRI (Programación Dinámica Dual Estocástica desarrollado por Power Systems Research Inc.) [Powe94]

El modelo para el tratamiento de la hidrología que se presenta en el modelo SDDP es uno del tipo PAR (Periódico Autorregresivo), el cual se basa en lo desarrollado por Box y Jenkins [Box76].

El desarrollo teórico presentado asume varias hipótesis respecto de las series hidrológicas, entre ellas de normalidad e independencia. En el caso de la

normalidad las series hidrológicas chilenas (y de ríos en general) presentan comportamientos no normales.

Dado que los modelos PAR asumen normalidad en las series resulta necesario aplicar una transformación a los datos para poder dejarlos normales y luego se ajusta un modelo a dichos datos sin violar los supuestos.

En el modelo SDDP se normalizan los residuos (los residuos son los valores obtenidos después de remover la dependencia temporal) en vez de los datos lo que puede llevar a estimar inadecuadamente los modelos.

El modelo SDDP se compone de dos módulos principales:

- i) cálculo de la política de operación hidrotérmica: determina la política operativa más económica para los embalses, teniendo en cuenta las incertidumbres en las afluencias futuras y las restricciones en la red de transmisión.
- ii) simulación de la operación: simula la operación del sistema a lo largo del período de planificación, para distintos escenarios de secuencias hidrológicas. Calcula índices de desempeño tales como el promedio de los costos operativos, los costos marginales por barra y por bloque de carga, y los intercambios óptimos entre empresas.

### **5.3 Ejemplo de modelación hidrológica propuesto**

Una vez construido el modelo, es decir, determinada su estructura y los parámetros, se puede utilizar para simulación y pronóstico. A continuación se explica brevemente la forma en que se abordan estos usos. No está de más recalcar que 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.

#### **5.3.1 Uso del modelo en simulación de la operación**

En el corto plazo, la influencia de la dependencia espacial resulta poco significativa por lo que la generación de series de caudales sintéticos es realizada en

base a la utilización de la componente temporal de los modelos desarrollados. Por lo antes expuesto, basta con analizar el caso multivariado, ya que el univariado es un subconjunto, en cuanto a requerimientos y complejidad, del primero.

a) Modelación temporal de cada serie

Con la modelación temporal se busca remover la dependencia que tiene la serie consigo misma, de tal manera de obtener ruidos independientes temporalmente, pero dependientes espacialmente.

i) Normalización de las series seleccionadas

La normalización de la serie consiste en hacer una transformación de tal modo que la serie sea lo más simétrica posible, es decir, el valor del coeficiente de asimetría sea cercano a cero. Ésto resulta necesario en caso que el valor del coeficiente de asimetría de la serie original sea significativamente distinto de cero. Se debe considerar que la varianza de la muestra de la serie, según el Test de Anderson, está dada por:

$$\sigma_g = \sqrt{\left(\frac{6}{N} + 29.3\chi^2 + 12\chi^4\right)}$$

donde  $\chi$  es la desviación estándar de la población; y si la serie es simétrica, implica que  $\chi \approx 0$ , por lo que

$$\sigma_g = \sqrt{\left(\frac{6}{N}\right)} = \pm 0.447$$

es considerado cero estadístico. Ésto se debe hacer para cada mes de cada una de las 5 series en estudio.

La serie obtenida se asume que distribuye Normal de media  $\mu$  y desviación estándar  $\sigma$ .

ii) Estandarización de las series seleccionadas

Una vez que la normalización ha sido llevada a cabo se debe estandarizar la serie. La estandarización consiste en restar la media  $\mu$  y dividir por la desviación estándar  $\sigma$ , obteniéndose una serie que debe distribuir Normal estándar ( $N(0,1)$ ).

iii) Orden del modelo

La dependencia temporal se analiza a través del coeficiente de correlación con distintos desfases o retardos (en este caso se consideraron desde 1 al 15). El orden del modelo está dado por el número de parámetros que es necesario estimar para realizar una adecuada predicción. Dichos parámetros deben ser guardados para la construcción de las series sintéticas. Para discriminar entre los distintos órdenes de un modelo del tipo AR se utilizan criterios de comparación; en este caso se utilizó el Criterio de Akaike (este criterio analiza un compromiso entre el número de parámetros estimados y la disminución de la varianza obtenida).

iv) Periodicidad y modelación de la dependencia temporal, utilización de un modelo PAR

Los modelos autorregresivos con parámetros periódicos son aquellos en que todos los parámetros o algunos de ellos varían dentro de un año o tienen un determinado período.

De esta manera, el caudal para el mes  $\tau$  del año  $v$  en un modelo autorregresivo, está dado por la siguiente expresión:

$$Z_{v,\tau} = \phi_{1,\tau} Z_{v,\tau-1} + \phi_{2,\tau} Z_{v,\tau-2} + \dots + \phi_{p,\tau} Z_{v,\tau-p} + \varepsilon_{v,\tau}$$

donde dicho caudal es función de los  $p$  caudales anteriores, siendo  $p$  el orden del modelo que fue identificado anteriormente. Los  $\phi_{p,\tau}$  son los parámetros del modelo a estimar y  $\varepsilon_{v,\tau}$  es la variable aleatoria independiente de distribución normal con media cero y desviación típica  $\sigma_{\varepsilon\tau}$ .

- Estimación de parámetros del modelo para cada serie

Los parámetros  $\phi_{p,\tau}$  del modelo son estimados a través de la ecuación de Yule-Walker que se presenta a continuación:

$$\rho_j = \phi_{1,\tau}(k)\rho_{j-1} + \phi_{2,\tau}(k)\rho_{j-2} + \dots + \phi_{k,\tau}(k)\rho_{j-k}; j = 1, \dots, k.$$

donde  $\rho_j$  es el coeficiente de correlación con desfase  $j$ , los cuales se obtienen al aplicárselo a la matriz normaliza. A partir de los coeficientes de correlación  $\rho_j$  y de los parámetros  $\phi_{p,\tau}$  obtenidos se calcula la varianza de los ruidos  $\sigma_\varepsilon^2$ .

- Estimación de los ruidos

La estimación de los ruidos se hace a partir de la ecuación anterior de donde se despeja el ruido independiente, según se presenta:

$$\varepsilon_{v,\tau} = z_{v,\tau} - \phi_{1,\tau}z_{v,\tau-1} - \phi_{2,\tau}z_{v,\tau-2} - \dots - \phi_{p,\tau}z_{v,\tau-p}$$

donde el  $\varepsilon_{v,\tau}$  distribuye Normal de media cero, desviación típica  $\sigma_{\varepsilon\tau}$ . De aquí se genera la matriz de ruidos que es temporalmente independiente, manteniéndose la dependencia espacial entre las distintas series.

- Verificación de las hipótesis de los ruidos

Se asume que los ruidos son independientes y normales, por lo que dichos supuestos deben ser verificados. Si dichos supuestos no son satisfechos satisfactoriamente la forma del modelo e incluso el tipo de modelo deberá ser cambiado y será necesario repetir esta revisión hasta que el modelo adecuado es encontrado.

Existen varios tests para verificar la independencia temporal. Entre otros los tests de Anderson, Porte Manteau y el del periodograma acumulativo (cumulative periodogram).

El test de Anderson consiste en verificar si los valores de correlación de la serie para desfases distintos de cero se encuentran entre los límites entregados por la expresión indicada más abajo para un cierto nivel de probabilidad:

$$r_k((1-\alpha)\%) = \frac{-1 \pm t_{\infty, (1-\alpha)} \sqrt{N-k-1}}{N-k}$$

Por su parte, el test de Porte Manteau permite saber si la serie  $\varepsilon_t$  es independiente y consiste en comparar el valor de Q con la distribución ChiCuadrado con L-p-q grados de libertad y cierto nivel de significancia. El valor de Q se calcula como:

$$Q = (N-d) \sum_{k=1}^L r_k^2(\varepsilon)$$

donde  $r_k(\varepsilon)$  es el correlograma de los ruidos y L es el desfase máximo considerado. Si Q es menor que el valor de la distribución ChiCuadrado con L-p-q grados de libertad y un nivel de significancia dado, la serie de ruidos es independiente y, por lo tanto, el modelo es adecuado, en caso contrario no lo es.

Por otra parte, la normalidad de las series puede ser verificada a través de varios tests estadísticos, entre otros el Chi-Cuadrado y el de la asimetría. El test de Asimetría se basa en el hecho que el coeficiente de asimetría de una distribución normal es cero. Si la serie distribuye normal estándar, el estimador de la asimetría deberá distribuir normal de media cero y varianza  $6/N$ , lo que permite construir un intervalo de confianza para verificar si el valor calculado de asimetría para la serie pertenece a dicho intervalo. Ésto se utiliza para muestras grandes ( $N > 150$ ), en cambio para más pequeñas se puede calcular directamente el coeficiente de asimetría de la muestra y compararlo con valores tabulados. Finalmente, conviene destacar la importancia de que los residuos sean normales, ya que ellos serán utilizados en la construcción de la estructura espacial. Si la serie resulta ser no normal se puede aplicar una transformación que la convierta en ese tipo de distribución.

b) Modelación espacial del conjunto de series

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.

Considerando los valores de los ruidos a los que ya ha sido removida la dependencia temporal (de acuerdo al tratamiento antes explicado), se construye la matriz de coeficientes espaciales.

Para ello primero se calculan las correlaciones de desfase cero entre cada una de las series consideradas, dando lugar a la matriz  $D^0$ , que como se indicara anteriormente es la matriz de covarianzas. El coeficiente  $b_{ij}$  representa la correlación existente entre la serie  $i$  y la serie  $j$ .

Realizado lo anterior se procede a “guardar” la información de correlación espacial contenida en las series a través de aplicar la descomposición de Cholesky a la matriz  $D^0$ , donde se cumple que:

$$D^0 = MM^T$$

siendo  $M$  una matriz triangular inferior.

Entonces si se dispone de  $k$  valores IIN(0,1):  $w_t(1), w_t(2), \dots, w_t(k)$ , se pueden generar residuos dependientes temporalmente que reproduzcan la dependencia espacial encontrada mediante:

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

La expresión anterior representa el modelo espacial contemporáneo para los ruidos.

### c) Generación de series de tiempo sintéticas

A partir de esta etapa en adelante lo que hay que realizar es deshacer lo hecho en las etapas anteriores, verificando el cumplimiento de las hipótesis en cada paso.

Para la generación de series sintéticas, lo primero es generar un conjunto de números IIN(0,1), a partir de valores U(0,1) y aplicando a ellos la transformación de Box y Miller. Llamando a ellos  $wt(1)$ ,  $wt(2)$ , ...,  $wt(k)$  y usando el modelo espacial se reconstruyen las series  $e_t(i)$ , con  $i=1,2,\dots,k$ .

Utilizando los parámetros del modelo y considerando cinco años para inicializar los datos se reconstruye la serie de valores que son normales de media cero y desviación estándar uno. Si se le aplicó una transformación para normalizar la serie (hacer la asimetría cercana a cero) se debe desnormalizar dicha serie aplicando la transformación inversa. A casi todas las series se les debió aplicar transformación para normalizarlas (se utilizó la transformación de Box-Cox). Una vez que las series están desnormalizadas se debe desestandarizarlas, es decir, multiplicar por la desviación estándar y sumar la media (cabe destacar que todos estos parámetros deben ser almacenados en la fase inicial).

Lo que resulta es una serie de tiempo sintética para cada uno de los lugares escogidos. Finalmente, es fundamental verificar la reproducción de los estadísticos relevantes en cada etapa.

### 5.3.2 Uso del modelo en pronóstico

Como se indica anteriormente el pronóstico de interés para la operación del SIC consiste en estimar el valor del caudal medio 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 del error que se espera cometer.

a) Esquema de pronóstico

Para el pronóstico de los valores semanales se dispone de los valores diarios registrados hasta un cierto día de la semana arbitrario y se desea conocer una estimación del valor medio para la semana siguiente, es decir, los 7 u 8 días siguientes. Así planteado el inicio de la semana a pronosticar no es fijo sino que depende del día seleccionado como inicial. Desde una perspectiva de pronóstico semanal la información diaria disponible se puede agrupar en conjuntos de 7 u 8 días y promediar para disponer de antecedentes semanales. Sin embargo, hay que tener en cuenta que los modelos han sido construidos con semanas que representan días fijos del año calendario, de manera que no comienzan arbitrariamente en el día que se desea realizar el pronóstico.

Para considerar el efecto del día arbitrario de inicio de la semana a pronosticar, y utilizar de todas maneras los modelos de pronóstico semanal, se propone emplear un pronóstico ponderado usando dos resultados, cada uno de los cuales corresponde al empleo de un modelo diferente considerando las dos semanas contiguas al día en cuestión (la semana inmediatamente anterior y la posterior).

El procedimiento propuesto puede aclararse si se supone que se desea hacer un pronóstico para las  $L$  semanas siguientes a partir de un día arbitrario  $j$  hasta el cual se dispone de información diaria. En primer lugar, se estiman los valores semanales de información disponible promediando los días anteriores, así se dispone de los valores de las dos o tres semanas anteriores:

$$Q_0 = \frac{1}{7} \sum_{i=0}^6 Q_{j-i}$$

$$Q_1 = \frac{1}{7} \sum_{i=7}^{13} Q_{j-i}$$

$$Q_2 = \frac{1}{7} \sum_{i=14}^{20} Q_{j-i}$$

donde  $Q_0$ ,  $Q_1$  y  $Q_2$  son los valores semanales registrados de las tres semanas anteriores al día del pronóstico.  $Q_{j-i}$  son datos diarios disponibles al momento de efectuar el pronóstico.

Posteriormente se seleccionan dos modelos para el pronóstico semanal, es decir, se eligen el orden  $p_m$  del modelo PAR y los parámetros correspondientes, de acuerdo a la semana modelada para la serie en cuestión. Ésto se debe a que los modelos semanales fueron establecidos considerando un día de inicio fijo para cada una, en cambio el pronóstico se efectúa para días arbitrarios. Esta selección se hace asignando a cada semana un día de inicio según la partición definida por el CDEC. Así la semana  $S_m$  comienza el día  $I_m$  contado desde el inicio del año hidrológico el primero de abril, y tiene una cantidad de días  $C_m$ , que son 7 ú 8, y a la cual corresponde un modelo  $AR(p_m)$ .

Dado el valor  $j$  del último día con información, se seleccionan los modelos de las semanas  $m$  y  $m-1$  tales que:

$$I_{m-1} < j < I_{m-1} + C_{m-1}$$

La última semana con información tiene entonces  $(j - I_{m-1})$  días de la semana  $m$  y  $(I_{m-1} + C_{m-1} - j)$  días de la semana  $m-1$ . Ésto permite asignar una ponderación a los resultados de cada uno de los modelos en proporción a los días que le corresponde de la última semana con datos.

Una vez seleccionados los modelos se efectúan dos pronósticos, cada uno con los mismos datos ( $Q_0, Q_1, Q_2, \dots$ ), pero con orden y parámetros diferentes que corresponden a los de la semana  $m$  y la semana  $m-1$ , los que se ponderan en proporción a los días involucrados de cada una para entregar el pronóstico de las semanas siguientes.

#### b) Ejemplo de pronóstico con modelos semanales

Para efectuar los pronósticos y comparar los resultados se han programado macros en VISUAL BASIC de EXCEL para PC que trabaja con los

datos de las series disponibles para este estudio. Este programa emplea como información los datos sobre las transformaciones y los parámetros de los modelos semanales ajustados a cada serie obtenidos del período usado para la calibración de los modelos (años 1941-42 al 1980-81), y se le ingresan adicionalmente los datos diarios registrados durante el período de prueba para efectuar pronósticos a partir de cualquier día seleccionable por el usuario.

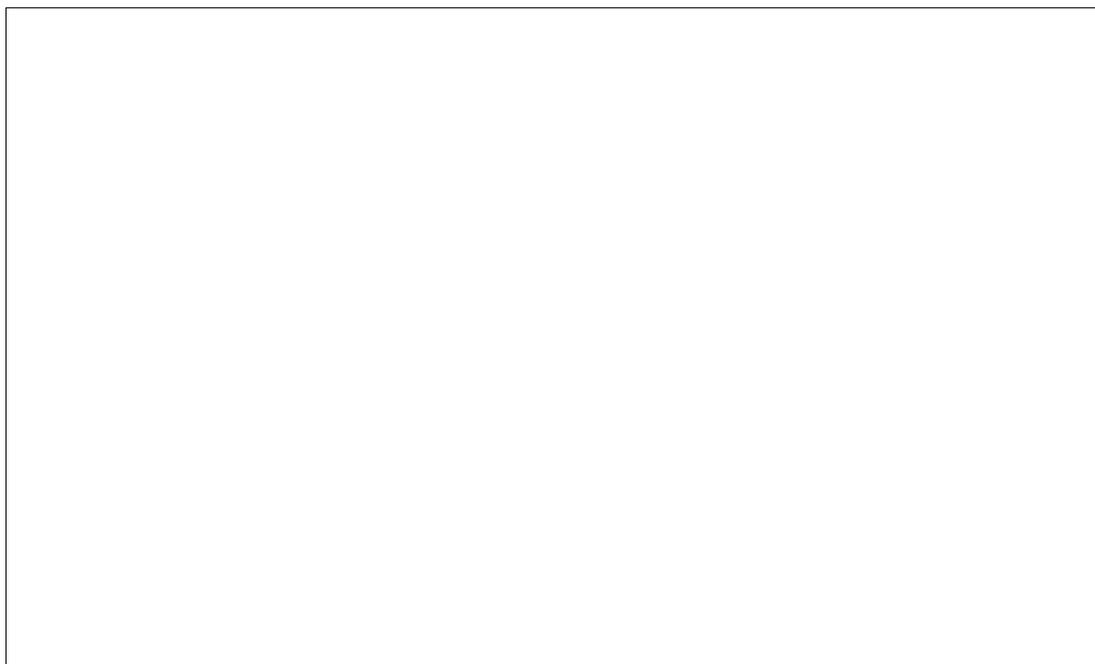


Figura 5.1: Pronóstico para las semanas siguientes al día 200, utilizando el modelo de la semana 26. El pronóstico se realiza con los datos de la semana 4.

Las figuras 5.1, 5.2 y 5.3 muestran el pronóstico efectuado a partir del día  $j=200$  del año 1991-92, de la serie Colbún, empleando los modelos de las semanas 26 y 27, que comienzan en los días 198 y 206 respectivamente. En las figuras se incluyen los valores registrados en las semanas anteriores al día 200, los valores pronosticados para las 7 semanas siguientes y las bandas de confianza con niveles del 50% y 95%. También se agregan los valores efectivamente ocurridos de manera de compararlos con los pronosticados. Las figuras 5.1 y 5.2 muestran los resultados de los modelos 26 y 27 mientras la figura 5.3 corresponde al resultado ponderado, que es

el definitivamente empleado. En esta última figura también se indica el pronóstico de la primera semana según el esquema del CDEC.

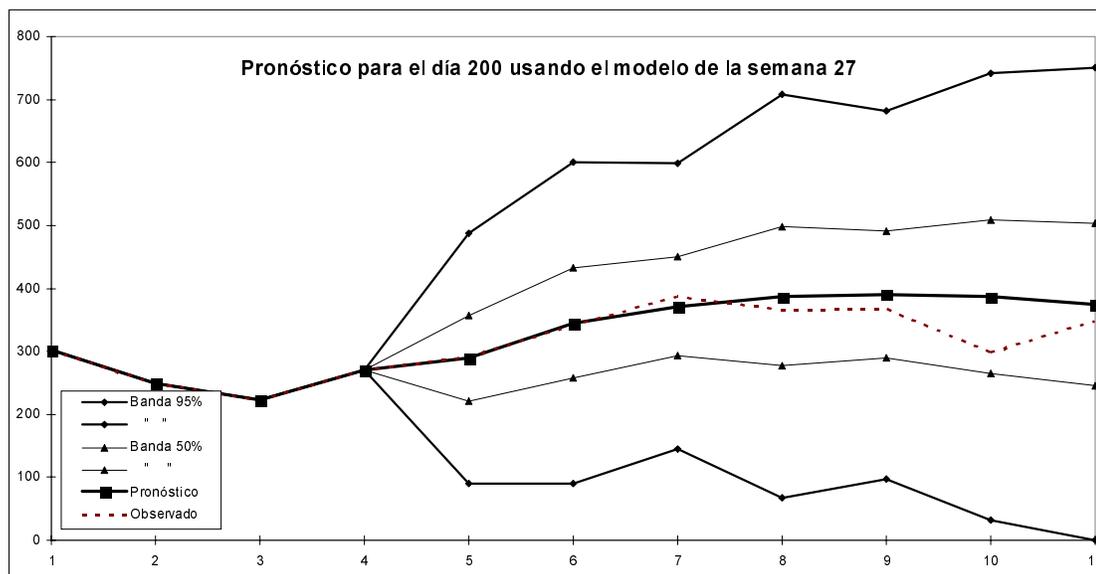


Figura 5.2: Pronóstico para las semanas siguientes al día 200, utilizando el modelo de la semana 27. El pronóstico se realiza con los datos de la semana 4.

La comparación de los resultados del pronóstico se basa en el comportamiento estadístico de los errores, es decir, de la diferencia entre el valor pronosticado para una semana y el valor realmente ocurrido. Se analiza separadamente el comportamiento en verano y en invierno y se considera sólo la primera semana.

La comparación se ha hecho realizando un pronóstico a partir de cada uno de los días de los años empleados para la prueba de los modelos que se indican en la tabla 5.1 para cada serie. El error se ha calculado basándose en el pronóstico de la primera semana.

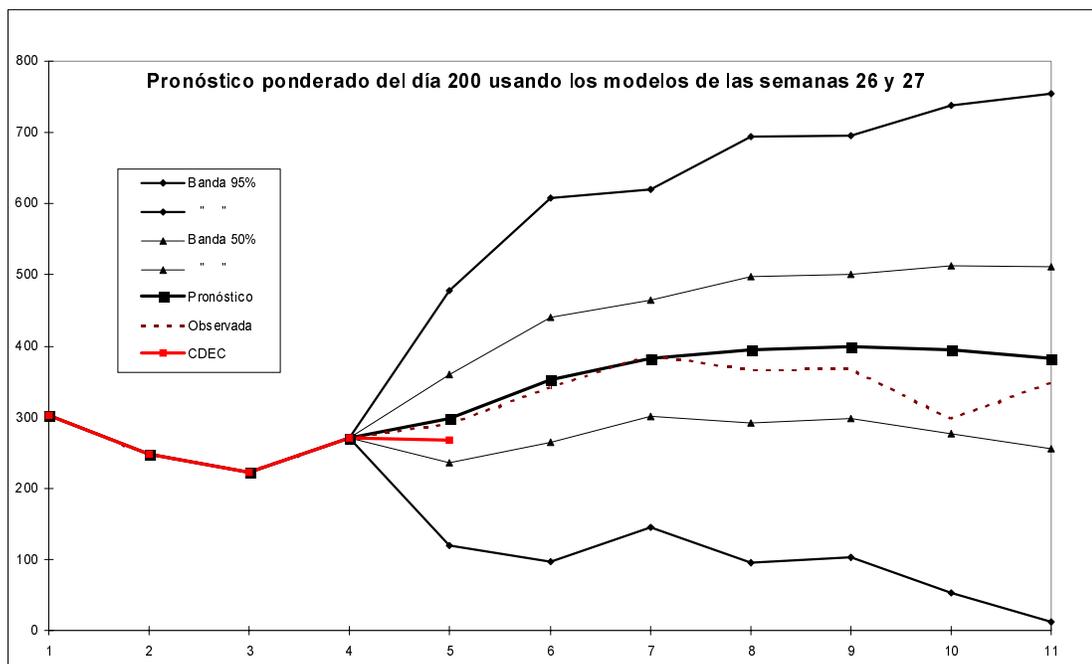


Figura 5.3: Pronóstico para las semanas siguientes al día 200, utilizando los modelos de las semanas 26 y 27 ponderados. El pronóstico se realiza la semana 4. Se incluye pronóstico del CDEC para la semana 5.

Tabla 5.1: Años empleados para la comparación en cada serie.

Nombre	Cantidad de pronósticos	Prueba modelo
Rapel	730	1994-1995
Colbún	3285	1987-1995
Melado	730	1994-1995
Antuco	730	1994-1995
Chapo	730	1994-1995

Suponiendo que la cantidad de pronósticos efectuado en cada serie es  $N$  y que el error en cada caso es  $e_i$ , calculado como:  $e_i = Q_i^{\text{real}} - Q_i^{\text{modelos}}$  de manera que los estadísticos usados para la comparación son los siguientes:

Promedio de los errores: 
$$\text{Pr } mE = \frac{1}{N} \sum_{i=1}^N e_i$$

Valores positivos indican que el modelo subestima y negativos que sobrestima. Este es un indicador del sesgo del pronóstico. Lo ideal es que sea nulo. Para que la comparación sea razonable debe tenerse un valor relativamente alto de N.

Desviación típica del error 
$$\sigma_E = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (e_i - \text{Pr } mE)^2}$$

Este valor debe ser el mínimo posible. Entre dos esquemas de pronóstico es mejor el que presenta la menor desviación típica del error.

Error cuadrático medio 
$$ECM = \frac{1}{N} \sum_{i=1}^N e_i^2$$

El ECM engloba los indicadores anteriores. Si el pronóstico es no sesgado el ECM es igual a la varianza.

Tabla 5.2: Comparación de errores de pronósticos para invierno.

Serie	Modelos PAR			CDEC		
	PrmE	ECM	Desv. error	PrmE	ECM	Desv. error
Rapel	-6,3	29426	164,1	-5,0	25134	155,3
Colbún	7,2	24473	124,6	4,0	19990	115,2
Melado	44,6	9296	84,6	2,2	9343	96,5
Antuco	9,8	4209	61,6	0,7	5028	68,5
Chapo	11,2	6574	67,7	-2,5	5433	72,2

Los valores de estos estadísticos para cada una de las series empleadas se indican en las tablas 5.2 y 5.3 para los pronósticos de invierno y de verano respectivamente.

Tabla 5.3: Comparación de errores de pronósticos para verano.

Serie	Modelos PAR			CDEC		
	PrmE	ECM	Desv. error	PrmE	ECM	Desv. error
Rapel	2,8	1557	38,5	-3,3	1381	36,3
Colbún	-5,0	2602	47,6	-5,5	2776	48,2
Melado	7,9	2374	47,1	-5,0	1219	33,2
Antuco	-0,9	63	7,8	-2,1	52	6,5
Chapo	7,5	960	29,8	-0,9	1323	36,4

El comportamiento de los indicadores comparados está muy influido por los errores ligados a la ocurrencia de crecidas, en las cuales los modelos empleados no pueden anticipar la subida brusca del caudal, por basarse sólo en los caudales registrados recientemente. Las figuras 5.4 y 5.5 muestran los errores de ambos modelos para el período de invierno y verano respectivamente, para un año de la serie Colbún. Se aprecia que el comportamiento de ambos modelos es similar, con errores menores en el período recesivo del verano.

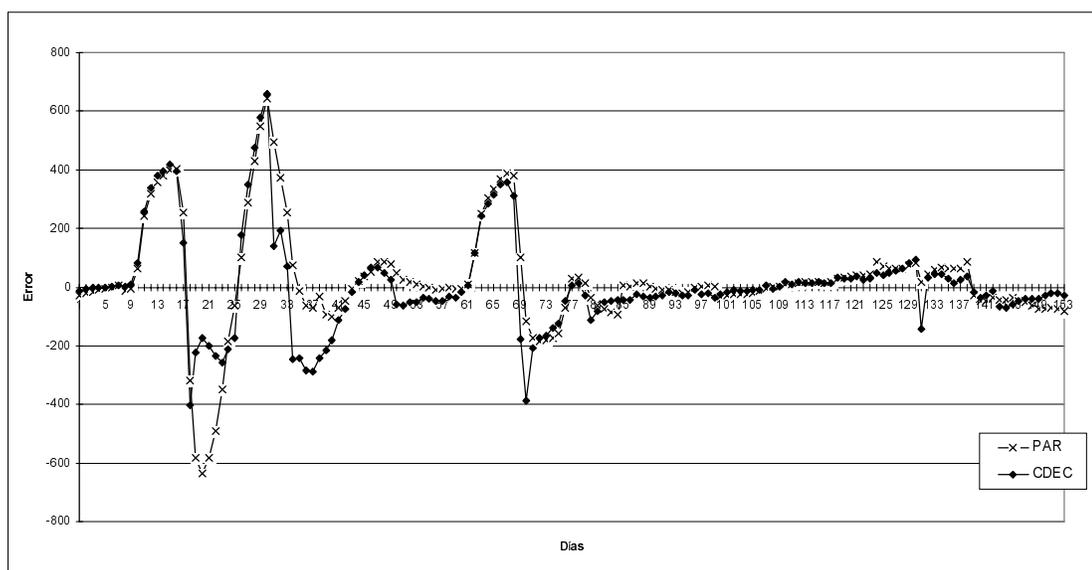


Figura 5.4: Errores de los pronósticos semanales con modelos PAR y CDEC. Serie Colbún en invierno para el año 1988-89.

Al comparar los indicadores mostrados en las tablas 5.2 y 5.3 se observa que los errores de ambos modelos son similares, a pesar de que el modelo del CDEC ocupa datos diarios y el otro semanales. En relación con los resultados de los pronósticos de invierno el sesgo es mayor en los modelos propuestos pero en general tiene menor varianza, especialmente en el caso de Melado, Antuco y Chapo. El gran sesgo que presenta el pronóstico de Melado con los modelos PAR se debe a valores ligados a una crecida, a pesar de lo cual el ECM es menor. Los pronósticos de verano son mejores en ambos casos, con resultados similares con relación a la comparación entre los modelos. El caso de Antuco es la serie que presenta tanto en invierno como en verano el mejor comportamiento. Es curioso observar que en los pronósticos de verano el modelo CDEC presenta siempre sesgos negativos, para todas las series.

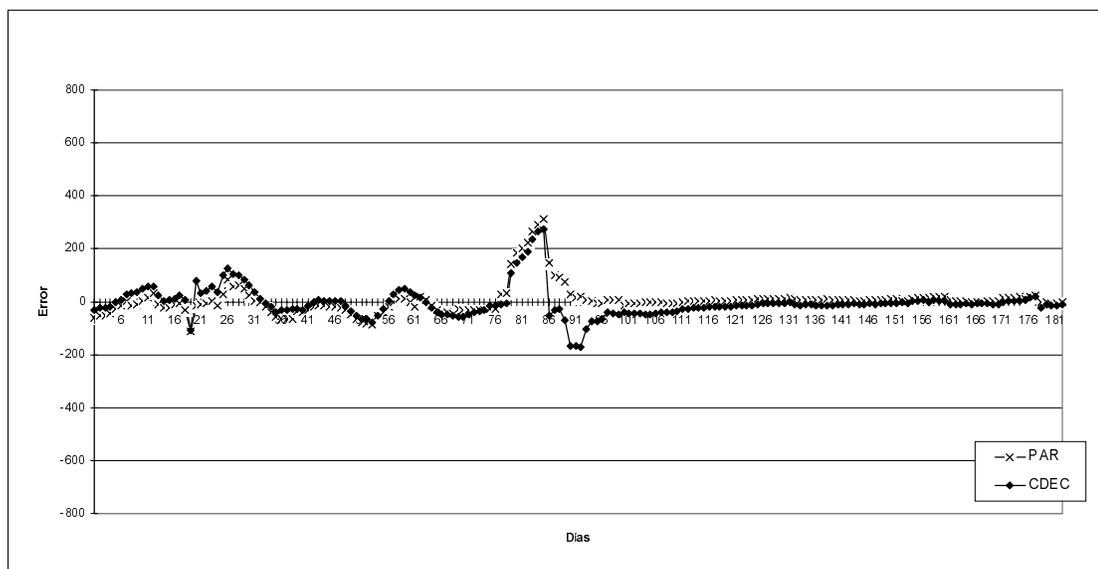


Figura 5.5: Errores de los pronósticos semanales con modelos PAR y CDEC. Serie Colbún en verano para el año 1988-89.

### c) Pronóstico con valores diarios

Tal como se ha mencionado anteriormente los pronósticos de valores medios semanales efectuados comparan el modelo actualmente en uso en el CDEC,

que utiliza información diaria, con un modelo clásico PAR que emplea valores semanales. En estas condiciones los resultados de los errores de pronóstico a una semana plazo son similares al comparar los valores medios. No es posible comparar las bandas de confianza debido a que el modelo del CDEC no las proporciona. A continuación se hace una exploración de los resultados de pronóstico de modelos PAR, pero empleando información diaria para pronosticar los valores medios de la semana siguiente. En este caso se ajustan modelos diarios, con datos también diarios, con los cuales se efectúa un pronóstico de los valores esperados de los siguientes 7 días. El pronóstico semanal corresponde simplemente al promedio de los 7 días pronosticados. Los errores de los valores semanales pronosticados comparado con los registrados se comparan en relación a los resultados que entrega el modelo del CDEC.

Tanto la modelación como la comparación se efectúa solamente para la serie Colbún, debido a que sólo para esta serie se dispone de información diaria suficiente, consistente en ocho años de datos. Para la identificación y estimación de parámetros de los modelos se emplearon 6 años de valores diarios de la serie Colbún, desde 1987 a 1992. Para la comparación de resultados se emplearon los valores de los años 1993, 1994 y 1995.

Es necesario hacer notar que la cantidad de años con información es reducida para estimar los parámetros estadísticos de los modelos diarios, ya que para cada día se dispone de sólo una muestra de seis valores. Esto obliga a incluir los valores de días vecinos en la estimación de parámetros.

Se han empleado dos modelos. Uno del tipo PAR similar al semanal y otro PARMA(1,1), para contrastar la inclusión de promedios móviles. Los parámetros del modelo varían de un día a otro para considerar el cambio en la dependencia temporal de los valores de caudal entre las diferentes épocas del año hidrológico. Para obtener una representación parsimoniosa de los parámetros del modelo se ha recurrido a series de Fourier, que permiten reducir 365 parámetros para un año a 6 si se consideran tres armónicas, normalmente suficientes para seguir las variaciones de los parámetros y eliminar la variabilidad muestra.

La modelación y el pronóstico se ha efectuado después de una transformación logarítmica de los datos para aproximarlos a una distribución normal. Es decir, dados los valores de caudales diarios  $Q_t$ ,  $Q_{t-1}$ ,  $Q_{t-2}$ ,  $Q_{t-3}$ , ... ,se han calculado los valores transformados:

$$Y_t = \text{Log}(Q_t)$$

De estos valores se ha estimado el promedio y la varianza diaria. Como sólo se dispone de una muestra limitada de 6 años con la cual se tendrían sólo 6 valores para la estimación del promedio y la varianza de cada día, lo que puede entregar una alta variabilidad muestral, se han estimado promedios y varianzas móviles alrededor del día en cuestión, considerando los 20 días más cercanos. A estos promedios y varianzas móviles se les ha ajustado una representación de Fourier y con ellos se han estandarizado los valores diarios:

$$Z_t = \frac{Y_t - \bar{Y}_t}{\sigma_t}$$

A estos valores se les ha ajustado un modelo PAR y un modelo PARMA(1,1), ambos periódicos. El primero es similar al empleado con valores semanales. El segundo se expresa como:

$$Z_t = \phi_{1,t} Z_{t-1} - \theta_{1,t} \varepsilon_{t-1} + \varepsilon_t$$

Los parámetros se han estimado con el método de momentos en base a los coeficientes de autocorrelación de cada día con los anteriores.

En este caso sólo se ha efectuado un pronóstico de los 7 días siguientes a la fecha en que se dispone de información y con ellos se ha estimado el valor pronosticado para la semana como el promedio de esos 7 días. Con estos modelos se ha efectuado el pronóstico a una semana plazo para cada día durante los años 1993, 94 y 95, no empleados para la construcción de los modelos, y en cada caso se ha calculado el error del pronóstico y los indicadores correspondientes similares a los empleados en los modelos semanales. Los resultados de estas aplicación de modelos en base a valores diarios se muestra en la tabla 5.4 para el ECM y en la tabla 5.5 para la desviación típica del error.

Tabla 5.4: Comparación de modelos PAR, ARMA y CDEC. ECM de los errores.

AÑO	Invierno			Verano		
	PAR	ARMA	CDEC	PAR	ARMA	CDEC
1993	23795	24045	37653	2090	2244	3867
1994	22309	22748	41125	1006	994	2089
1995	8815	9088	9527	992	979	1401
Promedio	18306	18627	29435	1363	1406	2452

Tabla 5.5: Comparación de modelos PAR, ARMA y CDEC. Desviación típica de los errores.

AÑO	Invierno			Verano		
	PAR	ARMA	CDEC	PAR	ARMA	CDEC
1993	151	152	194	46	47	62
1994	148	150	203	31	31	46
1995	92	94	98	32	31	38
Promedio	130	132	165	36	36	49

Las figuras 5.6 y 5.7 comparan los errores del pronóstico semanal en invierno y verano entre los modelos del CDEC y el PAR.

Se puede apreciar que para todos los años comparados, tanto en época de invierno como en verano, los modelos PAR y PARMA resultan similares y siempre superiores en el pronóstico al modelo actual del CDEC, ya sea que se considere el ECM o la desviación del error. El EMC es en promedio un 60% superior en invierno y un 75% superior en verano. Si se compara la desviación típica del error, ésta es en promedio un 25% mayor en el modelo actual que en cualquiera de los clásicos en época de invierno y un 36% superior en verano.

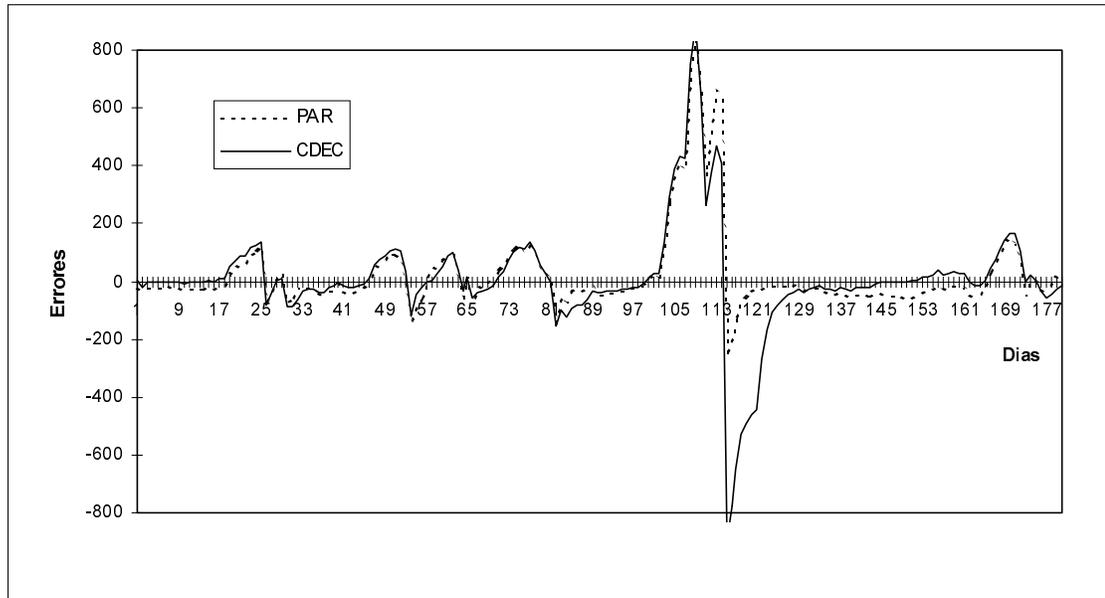


Figura 5.6: Errores de los pronósticos semanales con modelos PAR y CDEC.  
Serie Colbún en invierno para el año 1988-89.

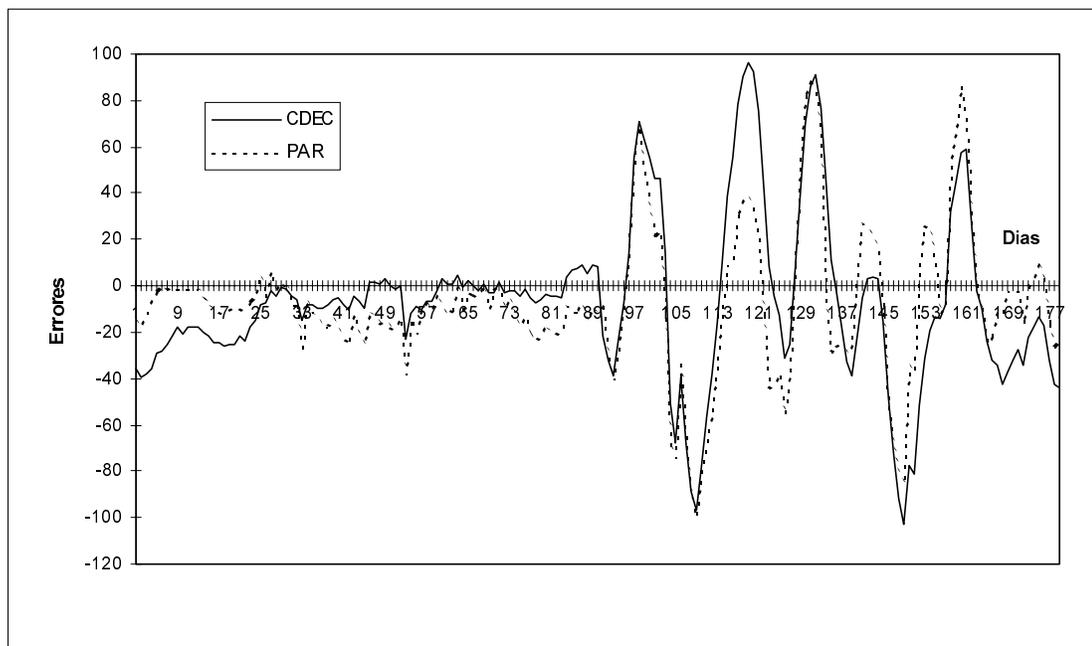


Figura 5.7: Errores de los pronósticos semanales con modelos PAR y CDEC.  
Serie Colbún en verano para el año 1988-89.

#### **5.4 Aplicación del ejemplo de modelación hidrológico propuesto al modelo OMSIC**

Mediante la utilización de los modelos ajustados de acuerdo a lo explicado en el 5.3.1 es posible generar valores distintos de los reales, pero teniendo la certeza que ellos son igualmente válidos (equiprobables). A partir de dichos valores es posible construir matrices hidrológicas sintéticas, y en particular, de las características de las que utiliza el modelo OMSIC, es decir, 40 años hidrológicos correspondientes a los períodos 41/42 al 80/81, que son alimentados al modelo a través del archivo de datos de entrada OMSHID.

La principal ventaja que presenta la utilización de las matrices sintéticas es la posibilidad de generar más de una matriz hidrológica, donde sólo existe una serie de tiempo. Esto posibilita la obtención de un perfil de costos marginales dependientes de la serie hidrológica generada de acuerdo a los modelos.

Para ejemplificar lo antes descrito, se utiliza una corrida del modelo OMSIC realizada con las matrices de datos reales y se comparan con los resultados obtenidos de realizar diez corridas casos con matrices sintéticas del mismo modelo. Las series utilizadas para esta ejemplificación son todas las correspondientes a la Cuenca del Laja (Pasada Abanico, Alto Polcura, Pasada Antuco, Afluente Laja más Alto Polcura y Laja en Tucapel). El resultado obtenido se presenta en la tabla 5.6, donde se puede apreciar que los valores de costo marginal de la energía obtenidos como resultados de las corridas del modelo OMSIC son levemente inferiores, alrededor de 2,00 millsUS\$/KWh en promedio, dependiendo del período considerado, para el caso de las matrices sintéticas y la desviación estándar promedio es alrededor de 1,00 millsUS\$/KWh. El detalle de los resultados de las 10 corridas se presenta en el ANEXO D.

La razón de la diferencia que se presenta entre los valores entregados por el modelo con la estadística original y los entregados con los valores sintéticos se puede explicar fundamentalmente por dos posibles efectos. El primero de ellos es la consideración de una muestra de 30 años hidrológicos (período 1951-1980) debido a la existencia de rellenos, en lugar de los 40 años (período 1951-1980) que actualmente se utilizan.

Tabla 5.6: Resultados de las corridas realizadas al modelo OMSIC.

Período	Caso referencia [millsUS\$/KWh]	Promedio corridas [millsUS\$/KWh]	Desv. Est. corridas [millsUS\$/KWh]
Semana 1	22,98	22,30	0,29
Semana 2	25,41	23,76	0,88
Semana 3	26,99	26,09	0,47
Semana 4	30,01	29,02	0,52
Mes 2	37,43	34,17	1,88
Mes 3	31,46	28,55	1,88
Mes 4	32,11	30,09	1,62
Mes 5	27,17	26,25	0,56
Mes 6	23,17	21,92	0,78
Mes 7	16,86	15,36	1,03
Mes 8	26,04	24,08	0,78
Mes 9	36,96	36,20	0,94
Mes 10	47,55	46,72	1,57
Mes 11	54,68	51,29	1,82
Mes 12	52,82	44,18	2,62
Mes 13	53,32	50,70	1,93
Mes 14	52,71	52,08	1,88
Mes 15	53,18	50,61	1,52
Mes 16	32,85	32,64	0,88
Mes 17	31,79	30,55	0,94
Mes 18	30,07	26,20	2,62
Mes 19	25,11	23,66	1,52
Mes 20	30,32	28,30	1,15
Mes 21	44,96	43,03	1,78
Mes 22	44,09	44,53	2.21

El segundo efecto es que al generar estos valores aparecen algunos datos menores que cero, lo cual no tiene sentido en términos de caudales, por lo que dichos valores son reemplazados por cero. La razón por la que se producen estos inconvenientes es por la alta asimetría de algunas series, la cual a pesar de las transformaciones aplicadas mantiene algún efecto sobre los modelos. Esto se da sólo en algunas series y, en particular, en aquellas con gastos cercanos a cero (menores que 10 m<sup>3</sup>/s). En este caso en particular, se da en la serie Pasada Abanico, cuyo valor promedio es de alrededor de 5 m<sup>3</sup>/s, siendo dos órdenes de magnitud menor que las principales series de la cuenca (Afluentes Laja más Alto Polcura, Pasada

Abanico y Laja en Tucapel). Al llevar los valores a cero, la matriz modificada contiene más agua que la real, debido a la sustitución de los valores negativos por cero, por lo tanto la señal de costo marginal obtenida del modelo OMSIC resulta menor. Este segundo efecto tiene un menor impacto que el primero considerado, por las razones explicadas.

## **VI. CONCLUSIONES Y RECOMENDACIONES**

### **6.1 Consideraciones acerca de los datos**

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 manera relevante de los datos disponibles. Es, por lo tanto, importante que esa información sea revisada cuidadosamente para asegurar que no existen valores rellenados que no aportan información o que la distorsionan. Por otra parte, la construcción del modelo debiera ser una oportunidad única para la revisión de la información. Este tipo de modelos trata de reproducir los valores esperados, las varianzas, las covarianzas o dependencias temporales en una misma serie y las espaciales entre series cercanas.

Es sabido por otra parte que la estadística de caudales en Chile es relativamente corta y que sólo se dispone de unas pocas series completas para el período de análisis, de manera que muchas de ellas contienen datos rellenados 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. Estos dos aspectos son esenciales en el modelo estocástico de caudales planteado para ser empleados en la operación del sistema. Si ésto no se revisa cuidadosamente el modelo reproducirá estas distorsiones. Es por ello importante que para la construcción del modelo se verifique y separe los datos reales de los datos rellenados y completados, y se debe asegurar que las estimaciones de las varianzas y las dependencias son las correctas.

En el análisis de las series escogidas para probar modelos se detectaron problemas en la forma en que se realizaron los “rellenos” de datos para completar las series que no tenían la antigüedad suficiente, ya que el modelo OMSIC requiere que todas las series de caudales estén completas desde el año hidrológico 1941-42. Dichos rellenos se efectuaron reproduciendo sólo los momentos de primer orden (promedios), lo que tiene como consecuencia que las series simuladas tienen menor varianza que el comportamiento real. De acuerdo al análisis presentado al menos los primeros diez años de estadística utilizada en el OMSIC (años hidrológicos 41-42 al

50-51) no debieran ser considerados en la modelación, ya que la gran mayoría de las series son rellenadas.

Adicionalmente, es altamente recomendable agregar datos de mediciones más recientes si se busca ampliar la muestra, usar 50 años hidrológicos, o bien reemplazar los primeros 10 o 15 años de la muestra considerada en la actualidad por los datos de los años 81-82 al 90-91 o al 95-96. La ventaja fundamental de realizar esto, es trabajar con una muestra de datos efectivamente medidos y evitar los problemas antes explicados.

## **6.2 Comentarios del modelo PRSI-SDDP**

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 propiedades del modelo y estimar sus parámetros que aseguren que se ha logrado el modelo más adecuado. En este sentido, la tarea de modelación es muy diferente a la de simulación, debiendo el modelo en sí ser único si los datos lo son. Sólo se podría pensar en modificarlo si se cuenta con nuevos antecedentes. Desde esta perspectiva disponer de un programa que construya modelos a gusto del usuario en la selección del orden y la estimación de los parámetros cada vez que hace una simulación no parece muy razonable.

Los procesos involucrados en la construcción de un modelo requieren adoptar decisiones en varias etapas intermedias del proceso que difícilmente pueden ser meramente automáticas, menos si no siempre se tiene claro las consecuencias de seleccionar alguna de las opciones planteadas. Ésto es especialmente válido en la selección del orden del modelo, la necesidad de normalizar las series, 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.

Las series de caudales mensuales o semanales de los aportes al Sistema Interconectado Central presenta como características estadísticas relevantes 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 varianzas son periódicas con grandes diferencias entre los valores de una estación a otra en un año. Todo ésto hace necesario que el modelo sea del tipo multivariado, periódico y con dependencia temporal. En este sentido, el modelo planteado del tipo PAR con dependencia espacial contemporánea parece adecuado, siendo uno de los más sencillos que puede plantearse.

Otra propiedad destacada de los valores de caudales mensuales o semanales de los ríos de esta zona es que presentan una gran asimetría, la que además es muy variable de un período a otro en una misma serie. Los estimadores de la asimetría son adicionalmente muy ineficientes. Todo ésto hace que sea absolutamente necesario poner atención en la normalización de las series. Esta tiene que ser realizada con cuidado para evitar que el modelo genere valores irreales (por ejemplo valores negativos). En general, se acepta que los valores de caudales medios anuales son del tipo logNormal. Sin embargo, los valores mensuales y semanales presentan asimetrías mucho mayores que logNormal, de manera que no basta aplicar una transformación logarítmica para convertirlos en normales. Esta transformación se puede hacer a los datos o a los residuos del modelo. Si bien el modelo PSRI lo propone para los residuos, sería más conveniente que ello se efectuara a los datos; más aún, si se emplea el método de momentos para la estimación de parámetros, el cual es más eficiente si los valores son normales. Ésto permitiría además tener un mayor control sobre la calidad de los datos al poder detectar en las primeras etapas del proceso la existencia de outliers o información anómala, la que al ser removida al principio no altera la estimación de parámetros ni la identificación de los modelos.

La periodicidad de las series mensuales y semanales genera una gran cantidad de parámetros si éstos se estiman directamente de los datos de cada período (mes o semana). Además estas estimaciones tienen una varianza de muestreo importante, de manera que es recomendable emplear una representación paramétrica de las características de la serie o de los parámetros del modelo. Para ello se pueden emplear representaciones mediante series de Fourier como se propone en el modelo de PSRI. En la realización práctica del modelo es importante que ello se haga, tanto

para disminuir la cantidad de parámetros como para suavizar las estimaciones de los valores periódicos y eliminar el ruido del muestreo.

En la estimación de los parámetros del modelo se recurre a las ecuaciones de Yule-Walker, o el método de los momentos. Puede ser más conveniente recurrir a un método de mínima varianza de los residuos junto a un criterio de selección basado en el Akaike Information Criteria, el cual puede dar mejores resultados globales si las series son no normales o si el modelo se empleará para el pronóstico.

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 comúnmente 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 están lejos de satisfacer estas propiedades, y ello se hace evidente sólo al observar la serie final generada. 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. No debe exagerarse este descarte ya que ello restaría al proceso una aleatoriedad razonable.

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 simulación. Ésto se refiere a que los estimadores de la varianza, correlación y asimetría son sesgados para muestras de tamaño reducido y más aún si estas muestras son dependientes. Para utilizar estas propiedades como valores numéricos del modelo es necesario corregirlas de manera de que, cuando el modelo se emplee en simulación, sea capaz de reproducir características con un valor esperado igual al de la muestra. Existen en la literatura procedimientos establecidos para corregir el sesgo de las principales propiedades de modelos autorregresivos.

### **6.3            Acerca del uso del modelo en pronóstico**

La operación de las centrales del SIC requiere el pronóstico de caudales semanales afluentes a los embalses. Actualmente este pronóstico se realiza en base a un modelo elaborado especialmente con este objeto, el cual considera valores diarios de la semana inmediatamente anterior al día en que se realiza el pronóstico. Este modelo es similar para todos los afluentes y para todas las semanas, de manera que no considera las particularidades de cada serie ni su variación semanal, diferenciando sólo entre condiciones de invierno y verano. Si bien los modelos se pueden considerar del tipo autorregresivo por cuanto el pronóstico se basa en una ponderación de los caudales de los días anteriores, sus coeficientes han sido establecidos de manera arbitraria. Debe hacerse notar que como el modelo pondera los caudales anteriores, en promedio tiende a proponer para la semana siguiente un valor igual al promedio semanal de la semana anterior.

El sistema de las centrales hidroeléctricas del SIC disponen de estadísticas de caudales afluentes a los embalses y centrales de una extensión y calidad suficiente para elaborar modelos de pronóstico de caudales semanales basados en el comportamiento estadístico observado de las series. En base a estas series en este estudio se ha analizado la posibilidad de desarrollar modelos estadísticos clásicos, empleando para ello técnicas conocidas y disponibles en la literatura, de manera de comparar los pronósticos con los actualmente en uso para cinco series de caudales afluentes al SIC.

El pronóstico efectuado actualmente en el SIC sólo entrega valores esperados para las semanas pronosticadas y en base a él no es posible conocer la confianza de los valores pronosticados, por cuanto no se conoce el comportamiento estadístico de los errores. El pronóstico propuesto aquí basándose en modelos PAR entrega límites de confianza que permiten tener una idea precisa de la confianza del pronóstico.

Ambos tipos de modelos elaboran sus pronósticos sobre la base de datos pasados, es decir, son del tipo autorregresivo, en el sentido de que no consideran información adicional a la contenida en los caudales registrados hasta el momento del pronóstico, por lo que la predicción de crecidas no resulta factible, es decir, no

pueden prever el aumento súbito de caudal provocado por una crecida antes que comience el incremento de caudal.

El sistema en uso realiza pronósticos de valores medios semanales, empleando como información los caudales diarios anteriores. Ésto es, utiliza del orden de 7 veces mayor información que los modelos PAR comparados en este estudio. Lo ideal si se dispone de datos diarios es efectuar pronósticos de valores diarios, por lo menos para la primera semana, para obtener a partir de ellos los valores semanales. El uso actual de la información diaria en el modelo CDEC no es eficiente en este sentido. Los modelos PAR propuesto emplean información semanal, aunque podrían extenderse para usar información diaria.

La comparación del comportamiento estadístico de los errores del pronóstico de ambos modelos para las cinco series estudiadas muestra que los resultados de ambos esquemas son similares, a pesar del empleo de diferente cantidad de información.

Adicionalmente se han comparado los resultados que se obtienen para el pronóstico semanal empleando modelos clásicos del tipo PAR y PARMA con datos diarios. Ésto se ha hecho sólo para la serie Colbún, que es de la que se dispone mayor cantidad de datos diarios, usando 6 años para la identificación y estimación de los modelos y tres años para la comparación de los resultados. La comparación muestra que en todos los casos los modelos PAR y PARMA son similares y muy superiores a los resultados que se obtienen del CDEC. Con ellos se obtiene un incremento significativo en la confianza del pronóstico semanal, mejorando las propiedades de los errores, disminuyendo tanto el ECM como la desviación típica de los errores del pronóstico en porcentajes importantes. En efecto, para el período de verano la desviación del error es un 36% superior cuando se usa el modelo del CDEC, y de un 25% en invierno.

Debe hacerse notar que los modelos diarios contenidos han empleado sólo unos pocos años de datos, lo que hace que la estimación de parámetros de los modelos no sea muy eficiente. Se puede afirmar que si se dispone de mayor cantidad de información, por ejemplo 10 a 15 años, en cada estación, la estimación de parámetros puede mejorar significativamente, y en consecuencia los pronósticos.

#### **6.4      Acerca del uso del modelo en simulación de la operación**

Al utilizar matrices sintéticas se puede generar más de una matriz hidrológica para cada serie de tiempo teniendo la certeza que ellas son igualmente válidas (equiprobables), por la manera en que se ha realizado su construcción. La ventaja que presenta este hecho es que, dejando el resto de las variables constantes se puede obtener información acerca de la sensibilidad del modelo OMSIC respecto de las posibles condiciones hidrológicas, a través de la obtención de un perfil de costos marginales. Cabe destacar que esta utilización en la simulación de la operación resulta imposible de realizar con las matrices hidrológicas reales, o con la reordenación de ellas.

Por otra parte, la generación de series sintéticas presenta algunos inconvenientes como la aparición de valores negativos lo que se puede deber, entre otras razones, a la no normalidad de las series consideradas. Resulta muy importante revisar las series sintéticas de modo de detectar dicho problema y corregirlo mediante el ajuste de otro tipo de modelo, o bien, la consideración de otros órdenes para el mismo tipo de modelo.

Es importante destacar que las series hidrológicas presentan valores de asimetría bastante altos, lo que hace necesario la aplicación de transformaciones del tipo Box-Cox. Este procedimiento es mejor que si se realizan las transformaciones a los residuos en vez de hacerlo a los datos y resulta aún más inconveniente en el caso que se estimen los parámetros por el método de los momentos, el cual es más eficiente si los valores son normales.

Finalmente, es posible indicar que existe una diferencia entre las señales de costo marginal obtenidas de las corridas del modelo OMSIC utilizando valores reales y valores sintéticos. Ésto, como se explica anteriormente, se debe fundamentalmente a la utilización de una menor muestra para construir los modelos (30 años). Una diferencia de 2 millsUS\$/KWh puede significar una diferencia significativa en términos de ingresos para las empresas generadoras, por lo que estos temas deben ser estudiados por el CDEC.

## BIBLIOGRAFIA

- [Alar93] ALARCON, I. (1993) **Planificación de la operación y tarificación en el sistema interconectado central chileno**. Chilgener S.A..
- [Box65] BOX, G. E. P. y TIAO, G. C. (1974) **Intervention analysis with application to economic and enviromental problems**. I. Am. Statis, 70, 70-79. Holken-Day.
- [Box76] BOX, G. y JENKINS, G. (1976) **Time Series Analysis, Forecasting and Control**. Revised Edition.
- [Cdec91] CDEC-SIC (1991) **Modelo OMSIC**. CDEC-SIC.
- [Cdec93] CDEC-SIC (1993) **Manual de procedimientos del reglamento interno. Punto IV: Definición de los caudales a utilizar para planificar la operación y calcular el costo marginal**. CDEC-SIC.
- [Cdec95] CDEC-SIC (1995) **Estadísticas de operaciones 1986-1995**. CDEC-SIC.
- [Cne93] CNE (1993) **El sector energía en Chile**. Comisión Nacional de Energía.
- [Divi75] División de Estudios Hidrológicos, ENDESA (1975) **Estudio hidrológico cuenca del río Maule**.
- [Divi82] División de Estudios Hidrológicos, ENDESA (1982) **Recursos hídricos de las centrales hidroeléctricas de Chile**. Volúmenes I y II.
- [Elec96] (1996) **Electricidad & Telecomunicaciones, Año 5 N°23 - Mayo de 1996. "Proyección de la CNE para 1996: Demanda del SING crecerá 24%"**. Editec Ltda.
- [Ende69] ENDESA (1969) **Estudio de los Recursos Hidraulicos del Río Laja**.
- [Ende84] ENDESA (1984) **Estudio de los Recursos Hídricos. Proyecto Canutillar**.

- [Epop83] Sección Planificación de la Operación, Depto. Operaciones, ENDESA (1983) **Información básica de las Centrales Hidroeléctricas del Sistema Interconectado Central**. Informe EPOP - N° 2/83
- [Fern85] FERNÁNDEZ, B. (1985) **Análisis de la periodicidad en series hidrológicas**. Anales de la Universidad de Chile. Quinta serie, N°8, Agosto de 1985, páginas 635-653.
- [Fern86] FERNÁNDEZ, B. Y SALAS, J. (1986) **Periodic Gamma Autoregressive Processes for Operational Hydrology**. Water Resources Research, vol. 22, N°10, 1385-1396.
- [Fern90] FERNÁNDEZ, B. (1990) **Modelos estocásticos en hidrología. Características generales**. Revista de la Sociedad Chilena de Ingeniería Hidráulica. Vol. 5, Núm. 2, Agosto de 1990.
- [Fern95/1] FERNÁNDEZ, B. (1995) **Estudio del modelo estocástico de caudales afluentes a las centrales del Sistema Interconectado Central propuesto en el modelo SDDP-CDEC**. Departamento de Ingeniería Hidráulica y Ambiental de la Pontificia Universidad Católica de Chile.
- [Fern95/2] FERNÁNDEZ, B. (1995) **Desarrollo de una metodología de pronóstico de caudales y concentraciones de sedimentos para la central Alfalfal**. Departamento de Ingeniería Hidráulica y Ambiental de la Pontificia Universidad Católica de Chile.
- [Fern96] FERNÁNDEZ, B. (1995) **Pronóstico de caudales afluentes a los embalses del SIC**. Departamento de Ingeniería Hidráulica y Ambiental de la Pontificia Universidad Católica de Chile.
- [Gati96] GATICA, P. y SKOKNIC, E.(1996) **Marcos regulatorios en el sector eléctrico sudamericano**. Comité chileno de la CIER, CHICIER; subcomité de planificación y medio ambiente- SUPLAMA.

- [Hipe94] HIPEL, K. y MCLEOD, A.. (1994) **Time Series Modelling of Water Resources and Environmental Systems**. Elsevier; Amsterdam, London, New York, Tokyo.
- [Hurs51] HURST, H. E. (1951) **Long term storage capacity of reservoirs**. Trans. Amer. Soc. Civil Eng. 116, pp. 776-808.
- [Kleme74] KLEMES, V. Y BORUVKA, L. (1974) **Simulation of gamma-distributed first order Markov chain**. Water Resources Research, vol. 10, N°1, 87-91.
- [Klem78] KLEMES, V. (1978) **Physically based stochastic Hydrologic Analysis**. En Advances in Hydrosociences. V. T. Chow Editor, Volumen 11, pp. 285-356.
- [Lloy67] LLOYD, E. H. (1967) **Stochastic reservoir theory**. Adv. Hydroscience 4, pp. 281-339.
- [Mand65] MANDELROT, B. B. (1965) **Une classe de processus stochastique homotetique a soi: Application a la loi climatologique de H. E. Hurst**. C. R. Acad. Sci. 260, pp. 3274-3277.
- [Mand68] MANDELROT, B. B. Y WALLIS, J.R. (1968) **Noah, Joseph, and Operational Hidrology**. Water Resources Research, vol. 4, N°5, 909-918.
- [Mejí72] MEJÍAS, J. M.; RODRÍGUEZ-ITURBE, I.; y DOWDY, D.R. (1972) **Streamflow simulatin 2. The broken line process as a potencial model for hydrologic simulation**. Water Resources Research, 8 (4): 931-941.
- [Mind95] MINDER, E. (1995) **Informe de práctica pre-profesional II**. Realizada en la Empresa Nacional de Electricidad S.A. (ENDESA) de acuerdo a los requerimientos de la Escuela de Ingeniería de la Pontificia Universidad Católica de Chile.

- [Mind96] MINDER, E. (1996) **Informe de práctica pre-profesional II**. Segunda práctica preprofesional II, realizada en la Empresa Nacional de Electricidad S.A. (ENDESA).
- [Mood74] MOOD, A. M.; GRAYBILL, F. A. y BOES, D. C. (1974) **Introduction to the theory of statistics**. McGraw Hill.
- [O'Con74] O'CONNELL, P. E. (1974) **Stochastic modeling of long-term persistence in stream flow sequence**. Ph. D. Tesis, Imperial College, University of London.
- [Powe94] Power Systems Research Inc. (1994) **Modelo de despacho hidrotérmico con restricciones de transmisión- SDDP/CDEC**. Power Systems Research Inc.
- [Sala85] SALAS, J.; DELLEUR, J.; YEVJEVICH, V.; y LANE, W. (1985) **Applied Modeling of Hydrologic Time Series**. Water Resources Publications.
- [Sala94] SALAS, S. (1994) **Simulación de la operación del Sistema Interconectado Central**. Memoria de Título, Departamento de Ingeniería Industrial, Universidad de Chile.
- [Tong83] TONG, H. (1983) **Threshold models in non-linear time series analysis**. Lecture Notes in Statistics 21, Springer Verlag.
- [Vara95] VARAS, E. (1995) **Alternativas para la generación de energía eléctrica**. Estudio encargado por la Empresa Eléctrica Pangué S.A. al Departamento de Ingeniería Hidráulica y Ambiental de la Pontificia Universidad Católica de Chile.
- [Weis75] WEISS, G. (1975) **Time reversibility of linear stochastic process**. J. Appl. Prob. Vol. 12, pp. 831-836.
- [Yevj72] YEVJEVICH, V. (1974) **Stochastic process in hydrology**. Water Resources Publications, Littleton, Colorado.

[Yevj84] YEVJEVICH, V. (1974) **Structure of daily hidrologic series**. Water Resources Publications, Littleton, Colorado.

**A N E X O S**

### ANEXO A: CENTRALES QUE COMPOEN EL SING Y EL SIC Y MAPA DE LA ZONA ESTUDIADA

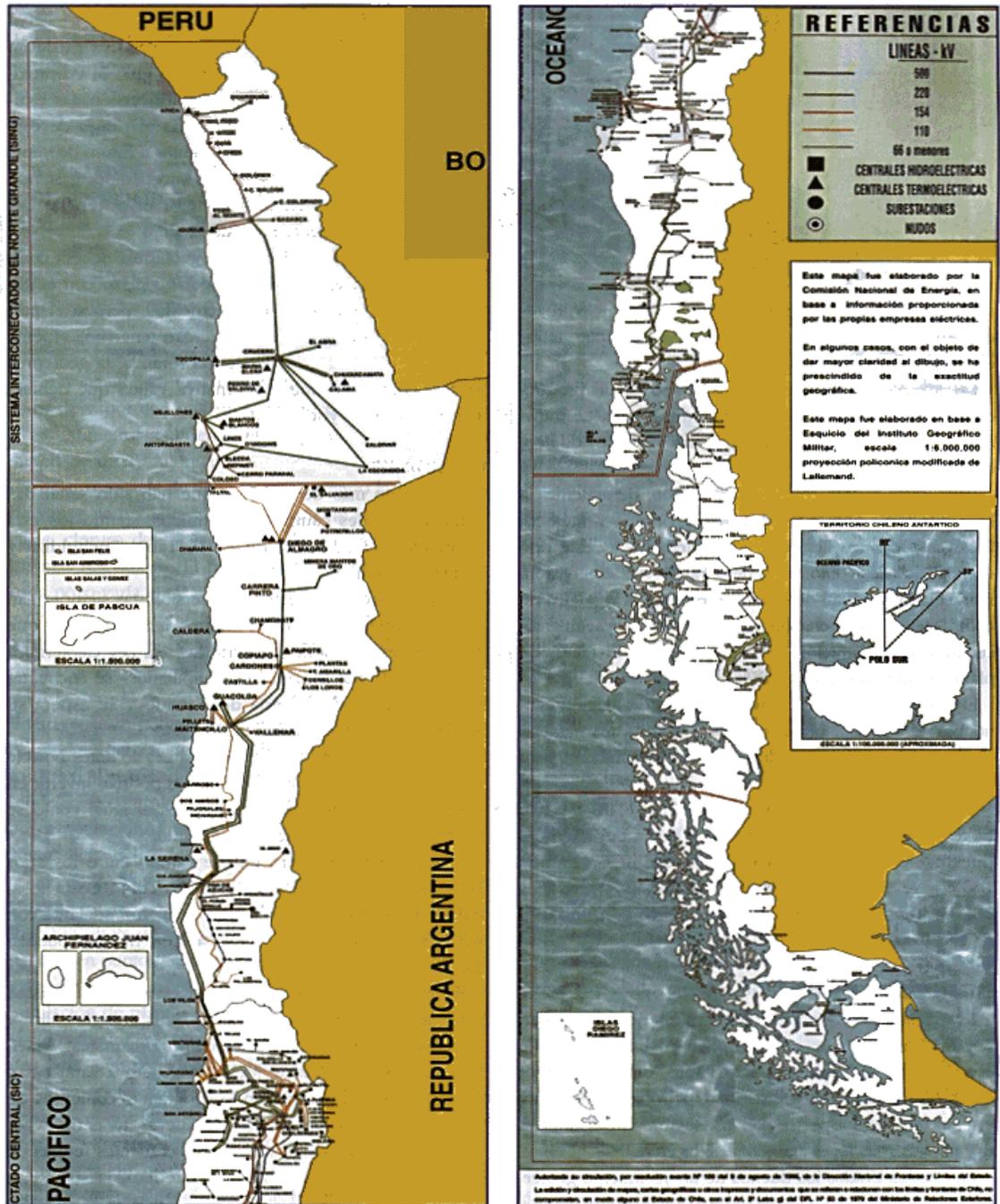


Figura A.1: Sistemas eléctricos chilenos [Cdec95]

Tabla A.1: Potencia bruta instalada en el SING. [Elec96]

Generadores	Potencia Instalada (MW)
Edelnor	230
Endesa	90
Tocopilla S.A.	638
Norgener	133
Mantos Blancos	28
TOTAL	1119

Tabla A.2: Centrales termoeléctricas del SIC. [Cede95]

Nombre	Potencia Inst. (MW)	Unids.	Propietario	Tipos de turbinas
Laguna Verde	54,7	2	Chilgener	Vapor-carbón
Renca	100,0	2	Chilgener	Vapor-carbón
Ventanas	338,0	2	Chilgener	Vapor-carbón
El Indio TG	18,8	1	Chilgener	Gas-diesel
D. de Almagro	23,8	1	Endesa	Gas-diesel
Huasco	16,0	2	Endesa	Vapor-carbón
Huasco	64,2	3	Endesa	Gas-IFO 180
Bocamina	125,0	1	Endesa	Vapor-carbón
Guacolda	150,0	1	Guacolda	Vapor-carbón
Laja	8,7	1	E. Verde S.A.	Vapor-des. Forest.
Constitución	8,7	1	E. Verde S.A.	Vapor-des. Forest.
TOTAL	907,9			

Tabla A.3: Centrales hidroeléctricas del SIC. [Cdec95]

Nombre	Potencia Inst. (MW)	Unids.	Propietario	Tipos de central
Alfalfal	160,0	2	Chilgener	Pasada
Maitenes	30,8	5	Chilgener	Pasada
Queltehues	41,0	3	Chilgener	Pasada
Volcán	13,0	1	Chilgener	Pasada
Colbún	400,0	2	Colbún	Embalse
Machicura	90,0	2	Colbún	Embalse
Pehuenche	500,0	2	Pehuenche	Embalse
Curillinque	85,0	1	Pehuenche	Pasada
Los Molles	16,0	2	Endesa	Pasada
Rapel	350,0	5	Endesa	Embalse
Sauzal	76,8	3	Endesa	Pasada
Sauzalito	9,5	1	Endesa	Pasada
Cipreses	101,4	3	Endesa	Embalse
Isla	68,0	2	Endesa	Pasada
Antuco	300,0	2	Endesa	Embalse
El Toro	400,0	4	Endesa	Embalse
Abanico	136,0	6	Endesa	Pasada
Canutillar	145,0	2	Endesa	Embalse
Pullinque	48,6	3	Pullinque	Pasada
Pilmaiquén	35,0	5	Pilmaiquén	Pasada
Florida	28,0	5	S.C. del Maipo	Pasada
Los Quilos	39,3	3	H.G. Vieja, M.V.	Pasada
Aconcagua	72,9	2	Aconcagua	Pasada
Capullo	10,7	1	E. E. Capullo	Pasada
Sauce Andes	1,2	4	Gen. S. Andes	Pasada
Carbomet	10,9	4	Carbomet	Pasada
TOTAL	3169,1			

## **ANEXO B: REGLAMENTO DEL FUNCIONAMIENTO DEL CDEC**

El Reglamento rige la coordinación de la operación de centrales generadoras y líneas de transporte pertenecientes a distintos propietarios que funcionen interconectados entre sí, y que cumplan simultáneamente las siguientes condiciones, (Artículo 1°):

- i) Operar en sistemas eléctricos de tamaño superior a 100.000 Kilowatt en capacidad instalada de generación, en los cuales se genere más de un 10% de electricidad destinada al servicio público de distribución.
- ii) Pertener a alguna de las entidades que se señalan a continuación:
  - Empresas Eléctricas cuya capacidad instalada de generación sea superior al 2% de la capacidad total que el sistema eléctrico tenía a la fecha de constituirse en dicho sistema, el Comité de Operación que se menciona en el artículo 2° de este Reglamento. Se entenderá por empresas eléctricas aquellas entidades cuyo giro principal es la comercialización de energía eléctrica.
  - Autoproductores cuya capacidad instalada de generación en condiciones normales sea superior a la suma de su demanda máxima anual y de un 2% de la capacidad total que el sistema eléctrico tenía a la fecha de constituirse el mencionado Comité de Operación. Se entenderá por autoproductores entidades cuyo giro principal es distinto de la comercialización de energía eléctrica.

En cuanto a la capacidad instalada de generación de las empresas productoras que les hace obligatorio integrar el CDEC, el reglamento establece que ésta debe ser superior al 2% de la capacidad total que tenía el sistema a la fecha de constitución del respectivo CDEC. En el caso del CDEC-SIC esta capacidad mínima quedó por lo tanto establecida en 61.148 KW (1985).

Por otra parte, el Artículo 25° señala que la valorización de las transferencias de energía y los correspondientes pagos entre integrantes serán contabilizados por el CDEC de acuerdo al siguiente procedimiento:

- i) En las barras de las subestaciones se efectuarán las medidas necesarias para determinar las inyecciones y retiros horarios netos de energía de cada integrante involucrado, las que serán valorizadas multiplicándolas por el costo marginal instantáneo correspondiente.
- ii) Para cada integrante, se sumarán algebraicamente todas las inyecciones y retiros netos valorizados a que se refiere el punto 1, ocurridos en todo el sistema durante el mes. Las inyecciones se consideran con signo positivo y los retiros, con signo negativo. El valor resultante, con su signo, constituirá el saldo neto mensual a favor de cada integrante.

Las inyecciones se consideran provenientes ya sea de centrales o bien de líneas de transporte. Los retiros se consideran destinados a clientes o bien a ser transmitidos por otras líneas de transporte.

El Artículo 26° señala que cada integrante con saldo neto mensual negativo pagará dicha cantidad a todos los integrantes que tengan saldo positivo, en la proporción en que cada uno de estos últimos participe del saldo positivo total del mes, considerando en todo caso los contratos suscritos a precio libremente convenido.

Por otra parte, el Título V trata sobre Pagos entre integrantes por el uso de sistemas de transmisión. En este título en el Artículo 32° se señala que el uso de sistemas de transmisión pertenecientes a un integrante, por parte de otros, dará origen al pago de servidumbre por parte de estos últimos, si dichos sistemas de transmisión han sido establecidos mediante concesión, o bien, si utilizan bienes nacionales de uso público en su trazado. El uso del sistema de transmisión ocurre cuando entre los puntos de inyección y retiro de energía por parte de un integrante, no existe un sistema de transmisión de su propiedad, de capacidad suficiente, y conectado eléctricamente con dichos puntos. El Artículo 33° indica que en el pago de peajes deberá descontarse aquella fracción que ha sido cancelada a través de la valorización

a costo marginal de las transferencias de electricidad en las distintas subestaciones del sistema eléctrico. Esta fracción podrá ser determinada por el integrante que sea propietario de los sistemas de transmisión involucrados comparando, en condiciones típicas de operación del sistema eléctrico, los ingresos anuales que obtiene por la aplicación de los costos marginales a las inyecciones y retiros totales de electricidad que se produzcan, con los costos anuales de capital, mantención y operación de tales sistemas de transmisión.

### **ANEXO C: DATOS DE LAS SERIES UTILIZADAS EN LOS EJEMPLOS Y ESTADÍGRAFOS CALCULADOS**

En este anexo se presentan para cada una de las 5 series consideradas (Afluentes al embalse Rapel - demanda de riego Río Teno; Cuenca intermedia muro Colbún Laguna Invernada - Laguna Maule; Afluentes al embalse Melado; Antuco Pasada y Afluentes al Lago Chapo más Blanco) detalles y gráficos de los valores semanales de gasto, promedio, desviación estándar, asimetría y valores anuales.



















































**ANEXO D: DETALLE MATRICES SINTÉTICAS Y COSTOS MARGINALES OBTENIDOS DE CORRIDAS DEL MODELO OMSIC**

En este anexo se presenta el detalle de los valores de costo marginal obtenido como resultado de correr el caso convergido para el horizonte Junio95-Marzo97 y se compara con los valores obtenidos al utilizar las matrices sintéticas.