

**DESARROLLO DE UN MODELO DE PRONOSTICO DE CAUDALES
SEMANALES ASOCIADO A LA VARIABILIDAD CLIMATICA
INTRANUAL EN COLOMBIA**

Tesis presentada como requisito para optar por el título de
Magíster en Ingeniería - Recursos Hidráulicos

Juan Sebastián Arenas Cárdenas I.C

Director: **Luis Fernando Carvajal Serna. I.C. MSc**

Universidad Nacional de Colombia – Sede Medellín

Facultad de Minas

Escuela de Geociencias y Medio Ambiente

2009

TABLA DE CONTENIDO

1	Introducción	ii
2	Información Empleada.....	1
2.1	Series hidrológicas	1
2.2	Series climáticas.....	2
3	Generalidades del clima en Colombia	3
3.1	ENSO	3
3.2	Chorro Chocó	5
3.3	Vientos Alisios	5
3.4	Complejos Convectivos de Mesoescala	7
3.5	Oscilación Madden Julian	9
3.6	ZCIT	10
3.7	La Influencia de los Océanos Pacífico y Atlántico	11
4	Métodos de predicción de caudales.....	12
4.1	Modelos Estocásticos	13
4.1.1	Modelos Univariados (AR)	14
4.1.2	Modelos Integrados	14
4.1.3	Modelos de Media Móvil	15
4.1.4	Modelos Autorregresivos de Media Móvil	15
4.1.5	Modelos Autorregresivos Integrados de Media Móvil.....	15
4.2	Polinomios de Regresión Multivariados y Adaptivos	16
4.3	Redes Neuronales Artificiales.....	18
4.4	Análisis Espectral.....	22
5	Análisis de la Información	24
5.1	Análisis previo de la información	24

5.1.1	Persistencia hidrológica.....	26
5.1.2	Correlación espacio temporal de las series.....	27
5.2	Análisis de las variables hidrometeorológicas y oceánicas de mayor influencia en la variabilidad intraestacional.....	28
5.2.1	Detección de la banda intraestacional con serie de caudales diarios.....	29
5.2.2	Detección de la banda intraestacional con Advección de humedad.....	31
5.2.3	Oscilación Madden-Julian.....	32
6	Metodología modelo de predicción de caudales.....	33
6.1	Modelo ACP.....	33
6.1.1	Autovalores y longitud de la ventana.....	34
6.1.2	Componentes Principales (CP).....	35
6.1.3	Componentes reconstruidas (CR).....	35
6.1.4	Análisis de un modelo multivariado.....	36
6.1.5	Regresión Lineal Múltiple Semanal.....	37
6.1.6	Predicción usando Análisis por Componentes Principales ACP.....	38
6.1.7	Transformación de datos.....	39
6.1.8	Pasos Análisis Espectral Singular.....	39
6.1.9	Reconstrucción de las componentes principales.....	40
6.1.10	Predicción.....	41
6.2	Redes Neuronales.....	42
6.2.1	Principio de Funcionamiento.....	43
6.2.2	Funciones de transferencia.....	43
6.2.3	Conexión entre neuronas.....	44
6.2.4	Función o regla de activación.....	44
6.2.5	Aprendizaje o entrenamiento.....	45
6.2.6	Estructura de la red – Caso semanal.....	45

7	Resultados	47
7.1	Modelo ACP	47
7.1.1	Autovalores y Componentes principales	47
7.1.2	Autovalores y Componentes Principales – Caso Multivariado	50
7.1.3	Calibración	52
7.1.4	Validación	54
7.2	Redes Neuronales	62
7.3	Evaluación comparativa entre los modelos ACP y RN	64
8	Conclusiones	66
9	Bibliografía	69
Anexo 1		73

LISTA DE TABLAS

Tabla 2.1 Características de las estaciones de caudal seleccionadas para la predicción semanal.....	1
Tabla 2.2 Variables atmosféricas usadas.....	2
Tabla 7.1 Número de aciertos por terciles para cada uno de los ríos.....	61
Tabla 7.2 Porcentaje de caudales semanales con aciertos por terciles para cada índice.....	61
Tabla 7.3 Error promedio y diferencia de errores de los modelos ACP y RN.....	64



LISTA DE FIGURAS

Figura 3.1 Localización y representación esquemática de los principales elementos geográficos que determinan el clima de Colombia.	3
Figura 3.2 Anomalías de la temperatura de la superficie del mar observadas en diciembre de 1997 en el Pacífico Tropical época en la que se desarrolló la fase madura del fenómeno El Niño de 1997-1998.....	4
Figura 3.3 Mapas medios mensuales de viento (flechas) y de presión (isolíneas) sobre el continente suramericano.....	6
Figura 3.4 Foto de un sistema convectivo de meso-escala en el pacífico de Colombia. Ondas del Este.	7
Figura 3.5 Mapa de isobaras en superficie, distribución de la divergencia y convergencia en la superficie, eje de la onda en la superficie y en cúspide a 10000 pies.....	8
Figura 3.6 Sección transversal sobre el ecuador de la propagación de la OMJ desde el océano Índico al Pacífico Oeste. Las flechas rojas indican la dirección del viento y las flechas azules indican los niveles positivos (negativos) de las anomalías de temperatura superficial del mar.	10
Figura 3.7 Banda de dominio de la Zona de convergencia Intertropical.....	11
Figura 4.1 Estructura lógica de una red neuronal.....	19
Figura 4.2 Funciones de Activación o Transferencia.....	19
Figura 5.1 Serie de caudales semanales Río Otún.....	24
Figura 5.2 Serie de caudales semanales Río Cauca.....	25
Figura 5.3 Serie de caudales semanales Río San Eugenio.....	25
Figura 5.4 Serie de caudales semanales Río Risaralda.....	25
Figura 5.5 Serie de caudales semanales Río Mapa.....	26
Figura 5.6 Correlación de caudal de la semana actual con las semanas anteriores	27
Figura 5.7 Relación serie de caudal río Otún con una serie de precipitación Aeropuerto Matecaña en la misma cuenca.	28
Figura 5.8 Relación serie de caudal río San Eugenio con la serie de precipitación de la estación Termales en la misma cuenca.	28
Figura 5.9 Espectros de potencia en onditas para la serie de caudales diarios	30
Figura 5.10 Espectros en ondita y de potencias para las series de caudales diarios filtrados en la banda 30-120 días.....	31
Figura 5.11 Series de advección de humedad por viento zonal y por viento meridional generada por el Chorro del Chocó (parte superior) y espectros de potencias en ondita correspondientes.	32

Figura 5.12	Espectro de potencias en ondita para la serie índice de la Oscilación Madden-Julian	32
Figura 6.1	Aporte a la varianza de los autovalores de la matriz construida con 12 semanas de rezago de caudal.....	41
Figura 6.2	Primeras 4 componentes principales.....	42
Figura 6.3	Funciones de Activación o Transferencia (Fuente MATLAB).....	43
Figura 6.4	Esquema de función de activación en una neurona.....	44
Figura 6.5	Estructura de una red neuronal con una capa oculta teniendo como datos de entrada los caudales con rezagos 1,2 y 3, serie índice OMJ y posibles variables macroclimáticas	46
Figura 7.1	Autovalores obtenidos para una aplicación univariada del modelo ACP.....	47
Figura 7.2	Serie original estandarizada vs. reconstruida con las componentes principales.....	49
Figura 7.3	Selección de autovalores para la reconstrucción de la serie original.	50
Figura 7.4	Autovalores seleccionados (hasta donde indica la línea verde) y gráfico de la serie de componentes principales reconstruida y la original.....	52
Figura 7.5	Selección del período de calibración y validación para cada una de las series de caudales de los cinco ríos.	53
Figura 7.6	Estimación de los coeficientes del modelo a partir de la semana actual	54
Figura 7.7	Comparación entre los caudales históricos y los caudales estimados del río Otún iniciando el pronóstico en la semana 15 y usando una ventana de pronóstico de 17 semanas.....	55
Figura 7.8	Comparación entre los caudales históricos y los caudales estimados del río Cauca iniciando el pronóstico en la semana 15 y usando una ventana de pronóstico de 17 semanas.....	56
Figura 7.9	Comparación entre los caudales históricos y los caudales estimados del río Risaralda iniciando el pronóstico en la semana 15, 30 y 45 y usando una ventana de pronóstico de 17 semanas	56
Figura 7.10	Comparación entre los caudales históricos y los caudales estimados del río San Eugenio iniciando el pronóstico en la semana 15 y usando una ventana de pronóstico de 17 semanas	57
Figura 7.11	Comparación entre los caudales históricos y los caudales estimados del río Mapa iniciando el pronóstico en la semana 15 y usando una ventana de pronóstico de 17 semanas.....	58
Figura 7.12	Errores medios de validación. El eje horizontal representa la semana de inicio del pronóstico y el eje vertical la semana pronosticada.....	59
Figura 7.13	Errores de predicción en el período de validación por Redes Neuronales para los ríos Otún (a), Cauca (b), Risaralda (c), San Eugenio (d) y Mapa (e).....	63

1 Introducción

En Colombia, la variabilidad climática y la posición geográfica son dos de los aspectos que más influyen en las actividades sociales y económicas las cuales provienen del manejo de una variedad de recursos naturales, destacando las actividades agrícolas, de riego, entre muchos otros. Un factor que incide en la relación de estos dos aspectos es el océano Pacífico tropical al considerarse como uno de los moduladores más influyentes en el clima global a diferentes escalas temporales y que relacionado con la atmósfera constituye el fenómeno El Niño-Oscilación del Sur (ENSO); muestra sólida de la variabilidad climática que influye los regímenes climáticos de gran parte del planeta. El segundo factor se debe a que el país está en medio de dos océanos y de la Amazonía, cuya interacción con la topografía local hace que sea tan variable nuestro clima.

Investigaciones recientes reportadas en Poveda (1994); Poveda y Mesa (1996), Poveda y Mesa (2000); Poveda et al. (2001); Waylen y Poveda (2002), Poveda (2004); evidencian como los fenómenos macroclimáticos están fuertemente asociados a la dinámica de los caudales; lo que motiva a estudiar cómo estos, inciden en la climatología de nuestro país. Para ello se han utilizado técnicas lineales y no lineales, con el fin de determinar las teleconexiones existentes entre los fenómenos macroclimáticos asociados a la variabilidad climática.

Actualmente existen diversas metodologías para estimar condiciones hidrológicas futuras, las cuales relacionan variables meteorológicas con hidrológicas o hidrológicas entre sí. Diversas técnicas de pronóstico, tanto técnicas empíricas-estadísticas como los métodos dinámicos son empleadas en el pronóstico climático a niveles regionales y nacionales (OMM, 2003). Estas metodologías incluyen análisis de patrones de circulación atmosférica general; métodos análogos, series de tiempo, correlación, regresión lineal múltiple y análisis de anomalías climáticas asociadas al ENSO, métodos dinámicos, modelos híbridos, entre otros.

Con este trabajo se busca desarrollar una herramienta de predicción de caudales de ríos a escala semanal en un horizonte máximo de 3 meses, que permita mejorar el conocimiento a corto plazo de la evolución y el comportamiento de los caudales durante el mes teniendo como fundamento la incidencia de variables macroclimáticas o parámetros oceánico – atmosféricos. De aquí, que entre los métodos planteados se describen algunos que consideran esta relación para determinar la respuesta hídrica futura.

El informe está organizado de la siguiente manera: el capítulo 2 describe la información climatológica e hidrológica utilizada en los alcances del estudio mencionando la obtención, características de los registros y escala temporal de las series base.

El capítulo 3 contiene las generalidades de la climatología colombiana haciendo énfasis en aquellos procesos asociados a la variabilidad intraestacional, además se describen los factores físicos que controlan la climatología colombiana.

El capítulo 4 comprende la descripción de algunas de las metodologías que se han desarrollado en predicción de caudales, se evalúan las características de Modelos Estocásticos y las múltiples formas de adaptarlos según el problema, se tienen la metodología MARS (Regresión Múltiple Adaptiva por tramos) que permite el modelamiento flexible no paramétrico. Se analizan los modelos asociados a las Redes Neuronales los cuales consideran algoritmos matemático-computacionales, diseñados con el objetivo de aprender a partir de la dinámica de uno o más eventos de la historia que se le suministre del mismo. Las redes, tienen como objetivo aprender acerca de la dinámica del sistema y, con dicho aprendizaje, simular bajo ciertas características el comportamiento futuro del mismo. Finalmente se describen los modelos de Análisis Espectral como el método adoptado para el desarrollo del modelo de predicción semanal debido a que es una buena herramienta que permite identificar las frecuencias de fenómenos físicos periódicos.

En el capítulo 5 se presentan los análisis previos que requieren las series hidrológicas y climáticas, la correlación espacio temporal de las series y un análisis de las variables hidrometeorológicas y oceánicas de mayor influencia en la variabilidad intraestacional de Colombia.

En el capítulo 6 se presenta la metodología de trabajo de este estudio y la descripción del modelo Análisis por Componentes Principales y un modelo de Redes Neuronales desarrollado como comparativo al modelo inicial propuesto.

El capítulo 7 contiene los resultados de la calibración y validación del modelo y un ejercicio de optimización de parámetros por algoritmos genéticos y la evaluación de un segundo modelo de Redes Neuronales con el cual se desea investigar y comparar los resultados de las calibraciones y validaciones de ambos modelos.

El capítulo 8 se basa en las conclusiones y temáticas de estudios posteriores y finalmente el capítulo 9 presenta la bibliografía recopilada en este trabajo.

2 Información Empleada

Para la realización de este trabajo se empleó información de dos clases: Una climática proveniente de satélites del proyecto Reanálisis climático NCEP/NCAR, de la NOAA (National Oceanic and Atmospheric Administration) y otra hidrológica que comprende registros de caudal y precipitación diarios correspondientes a las estaciones de precipitación y caudales del IDEAM. A continuación se describen las principales características de estas bases de datos.

2.1 Series hidrológicas

El caudal es una de las variables más relevantes en este trabajo. La información hidrológica usada consta de series de caudales de estaciones de cinco ríos del eje cafetero, instaladas y operadas por el IDEAM. Esta información hidrológica hace parte de la base de datos del Posgrado de Recursos Hidráulicos de La Universidad Nacional de Colombia Sede Medellín

Para la selección de las cuencas se consideró el análisis de los registros de las series históricas, la longitud de datos faltantes, exploración de datos, y análisis por medio de técnicas estadísticas de anomalías y estructuras ocultas dentro de la información. Al mismo tiempo, análisis que permitan determinar posibles cambios en la estructura de la serie, brincos, tendencia, datos anómalos y datos faltantes e identificar las causas que se asocian a la variación dinámica de las series. Se seleccionó el río Cauca para evaluar el desempeño del modelo con grandes caudales, de igual forma se seleccionó el río San Eugenio para evaluar el desempeño del modelo con caudales pequeños.

De la aplicación estadística en hidrología se ha concluido que las series de tiempo para que sean estadísticamente significativas y confiables deben tener una longitud de registro de 20 años. Se buscaron series que cubrieran este rango de años ya que es importante recalcar que muchos de los fenómenos de gran escala con manifestaciones interanuales influyen dentro de la variabilidad intraestacional y se requiere capturar algunos de estos eventos. La tabla 2.1 presenta las características de las estaciones de caudal seleccionadas.

Tabla 2.1 Características de las estaciones de caudal seleccionadas para la predicción semanal

Estación	Código	Cuenca	Tipo	Elevación (m)	Municipio	Registro años
La Bananera	2613711	Río Otún	LG	1530	Pereira-RI	29
La Virginia	2617703	Río Cauca	LG	900	La Virginia-RI	26
La Reina	2613712	Río San Eugenio	LG	1630	Santa Rosa Cabal-RI	24
Puente Negro	2614714	Río Risaralda	LG	920	Belalcázar-CA	27
Breña	2614713	Río Mapa	LM	1080	Viterbo-CA	25

2.2 Series climáticas

Las series de las variables de influencia en la zona de estudio provienen de la información de libre acceso de diferentes bases de datos. Para el diagnóstico de los ambientes atmosféricos que caracterizan la región de influencia, se usó información del proyecto del Reanálisis Climático del National Center for Environmental Predictions/ National Center for Atmospheric Research, conocido como Reanálisis NCEP/NCAR. El papel del NCEP es desarrollar y evaluar el modelo que se va a usar, mientras que el NCAR se encarga de proporcionar una completa base de datos histórica de las variables atmosféricas que considera el modelo. Del sitio web <http://www.cdc.noaa.gov/data/reanalysis/reanalysis.shtml> se tomaron las series de tiempo de los principales índices de fenómenos de gran escala así como otras variables atmosféricas para caracterizar fenómenos de menor escala como Advección de humedad y Oscilación Madden-Julian. La tabla 2.2 muestra las variables climáticas usadas.

Tabla 2.2 Variables atmosféricas usadas

Variable	Base de datos	Resolución espacial	Resolución temporal	Período registro
Índice Madden-Julian	NCEP/NCAR	2.5° x 2.5°	diaria	1970-2008
Humedad Relativa	Reanalysis	2.5° x 2.5°	diaria	1970-2008
Viento Zonal	Reanalysis	2.5° x 2.5°	diaria	1970-2008
Viento Meridional	Reanalysis	2.5° x 2.5°	diaria	1970-2008
Humedad específica	Reanalysis	2.5° x 2.5°	diaria	1970-2008

3 Generalidades del clima en Colombia

El clima de Colombia es muy variado, tanto a lo largo de su territorio como en su escala temporal. Esta diversidad climática está determinada en gran medida por la ubicación geográfica y por las características fisiográficas del país, ver figura 3.1. El clima, en general está determinado por su ubicación geográfica (en la zona ecuatorial) y por la influencia de algunos factores como la circulación atmosférica, el relieve, la influencia de los procesos del océano Pacífico y del océano Atlántico, el contraste entre el continente y el mar, la influencia de áreas selváticas o boscosas, entre otros. Además la posición geográfica de Colombia en la zona ecuatorial influye directamente en la distribución de la mayor parte de las variables climatológicas (radiación, temperatura del aire, presión atmosférica, vientos, humedad del aire, precipitación). Las variaciones espaciales y temporales del clima juegan un papel determinante en las diferentes formas de actividad humana en el ámbito nacional. La variabilidad climática interanual, como la causada por el ciclo El Niño - La Niña-Oscilación del Sur, es una de las que más afecta considerablemente.

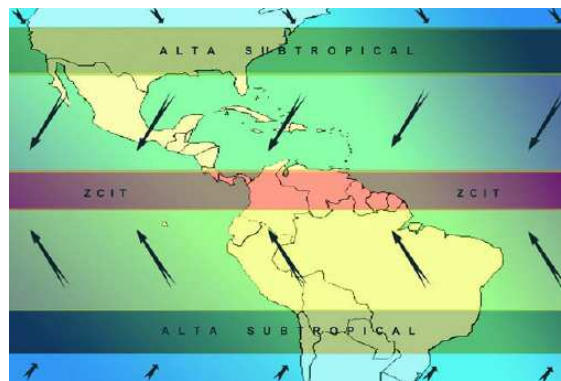


Figura 3.1 Localización y representación esquemática de los principales elementos geográficos que determinan el clima de Colombia.

Fuente: IDEAM 2001

Para tener un detalle de estas características del clima de Colombia, se presenta a continuación una descripción de los aspectos más relevantes que intervienen en la dinámica climática del país y sus variaciones espaciales y temporales en diferentes escalas.

3.1 ENSO

El fenómeno El Niño-Oscilación del Sur (ENSO) es un evento natural que se da como resultado de la interacción entre el océano y la atmósfera en la región del pacífico ecuatorial. Sus características atmosféricas lo clasifican como un fenómeno de gran escala y de implicaciones dominantes en gran

parte de la tierra. Las variaciones atmosféricas y oceánicas ocurren en escalas de tiempo de meses a varios años.

El Niño es el término originalmente usado para describir la aparición, de tiempo en tiempo, de aguas superficiales relativamente más cálidas que lo normal en el Pacífico Tropical central y oriental, frente a las costas del norte de Perú, Ecuador y sur de Colombia (figura 3.2).

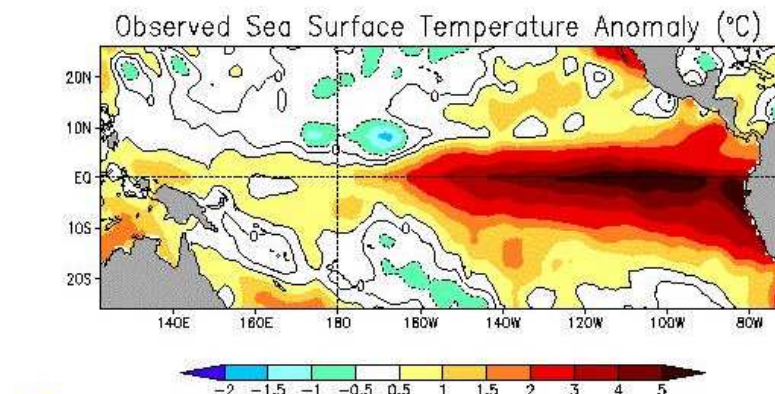


Figura 3.2 Anomalías de la temperatura de la superficie del mar observadas en diciembre de 1997 en el Pacífico Tropical época en la que se desarrolló la fase madura del fenómeno El Niño de 1997-1998.

Fuente: NOAA

El fenómeno ENSO influye sobre la hidroclimatología de Colombia teniendo como base las hipótesis que se han dado en la literatura para explicar las anomalías hidrológicas en la región tropical de Suramérica durante la fase cálida del ENSO.

El Niño es la fase cálida del ENSO, en la que se presenta un aumento de las temperaturas superficiales del mar, en particular sobre el centro y el oriente del pacífico tropical, lo cual conduce a alteraciones en los patrones de circulación de vientos, de presiones atmosféricas superficiales y de precipitación sobre el océano pacífico. Caso contrario en la fase fría se da La Niña donde se presenta una disminución de las temperaturas superficiales del mar sobre el pacífico central del este lo que conduce a incrementar las lluvias en el pacífico colombiano y tropical. La variabilidad del clima de Colombia a la escala interanual asociada con el ENSO tiene implicaciones en la disminución de la precipitación y en los caudales medios mensuales, disminución de la humedad del suelo y en la actividad vegetal (Poveda & Jaramillo, 2000) e incremento en las temperaturas medias del aire. Todas ellas generando efectos en la agricultura y la generación de energía eléctrica entre otros.

3.2 Chorro Chocó

La corriente de vientos en chorro del occidente de Colombia o chorro del Chocó (Poveda, 1998; Poveda & Mesa, 1999 y 2000) transporta grandes cantidades de humedad desde el océano Pacífico hacia el interior del país. Este interactúa con la topografía de los Andes occidentales y con los vientos Alisios del este para favorecer la convección profunda y que contribuye a producir grandes cantidades de precipitación sobre el occidente del país. La dinámica del chorro del Chocó exhibe un marcado ciclo anual, siendo notorios para los meses de septiembre-noviembre y débiles para los meses de febrero-marzo. La corriente en chorro obtiene su energía del gradiente de temperaturas superficiales que existe entre la zona de temperaturas más frías de la llamada “lengua fría” de las aguas del océano Pacífico en Ecuador y Perú, y las temperaturas más cálidas del Pacífico al frente de las costas de Colombia. Con la información del Reanálisis NCEP/NCAR se estudió la dinámica de la circulación de vientos alrededor del trópico de Suramérica y su variación durante las fases del ENSO (Poveda & Mesa, 1999). Se encontró que estas distribuciones de vientos tiene dirección sureste-noroeste en el hemisferio sur y se recurvan al cruzar el ecuador cambiando su dirección a suroeste-noreste en el hemisferio norte. Los vientos Alisios del noreste que proceden del Mar Caribe también se recurvan y adquieren una dirección del oeste. Los vientos del oeste que penetran por la costa del Pacífico colombiano provienen de una corriente de chorro, llamada corriente del Chocó.

3.3 Vientos Alisios

Los vientos Alisios se originan en las altas subtropicales y fluyen hacia el ecuador, algunas veces traspasándolo con dirección del noreste en el hemisferio norte y del sureste en el hemisferio sur (Riehl, 1979). Estos vientos, al converger en una franja cercana al ecuador terrestre, forman la denominada Zona de Convergencia Intertropical (ZCIT), caracterizada en general por ser una zona nubosa y productora de lluvias.

La influencia de los vientos Alisios sobre la región norte de Suramérica durante el invierno boreal se observa en la figura 3.3. Para los meses de enero a marzo se observa una región continua ocupada por los Alisios llegando a extenderse desde la costa occidental de África hasta las latitudes 5-10° S en Suramérica. En esta, se observa también la intensidad que se da hacia el sur sobre la Amazonía incrementando su magnitud. Para los meses de junio a agosto la dirección de los vientos es más definida en dirección este- oeste sobre el océano atlántico en las latitudes cercanas al ecuador. Para los meses de octubre a diciembre han predominado vientos que provienen en dirección este-oeste alrededor de la línea del ecuador incrementando su magnitud hasta comienzos de enero.

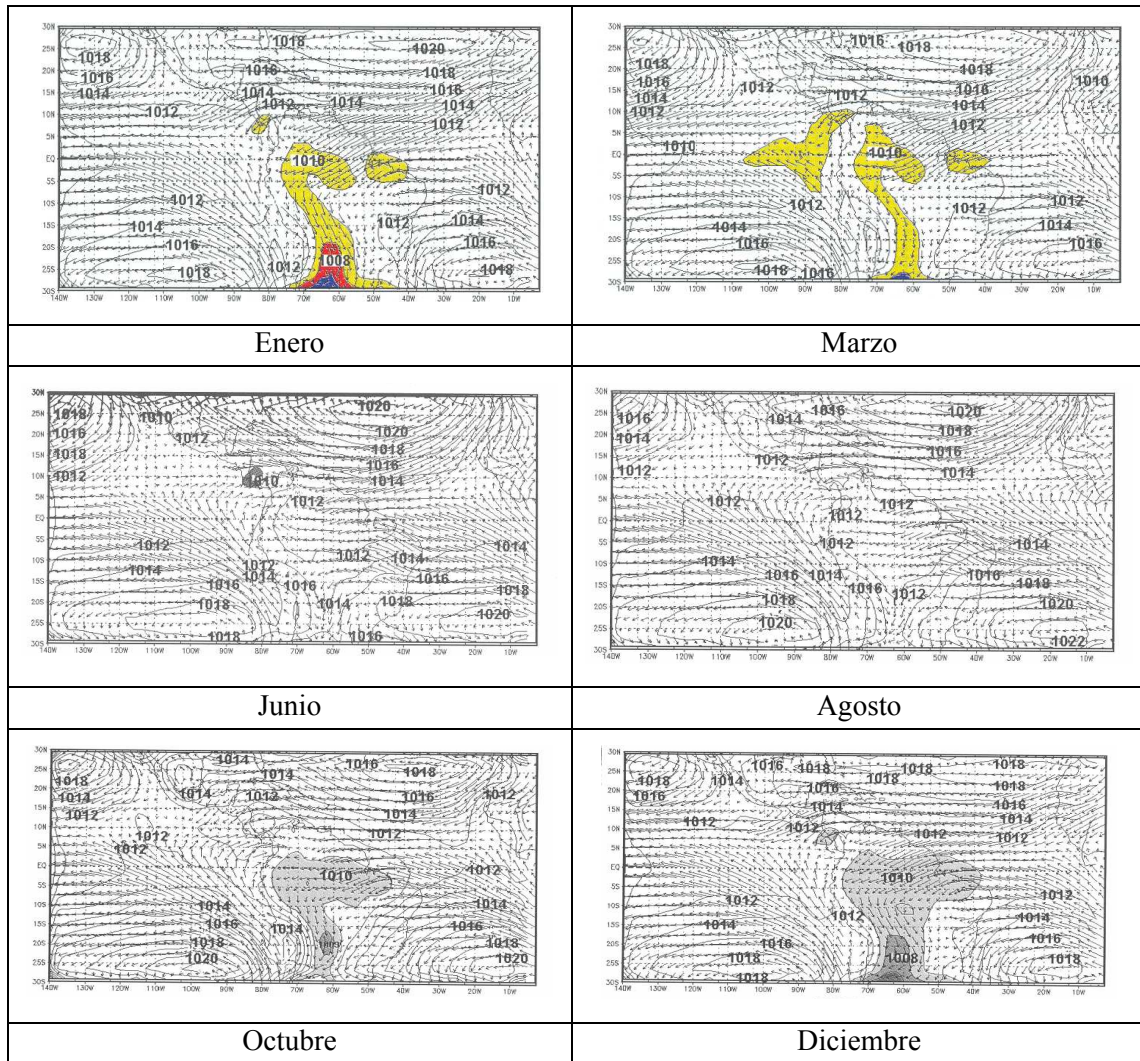


Figura 3.3 Mapas medios mensuales de viento (flechas) y de presión (isolíneas) sobre el continente suramericano

Fuente: Montoya, et al., 2001

Se confirman las características medias de los vientos Alisios y de la circulación tropical analizadas por Hastenrath (1996). La influencia de los vientos Alisios es evidente sobre la región norte de Suramérica, se observa una región continua ocupada por los Alisios del noreste desde diciembre hasta marzo aproximadamente. Esta región llega a extenderse desde la costa occidental africana hasta 5-10° S en Suramérica. Los contornos de presión muestran que a partir de octubre comienza a formarse un sistema de baja presión sobre el continente suramericano hasta marzo.

3.4 Complejos Convectivos de Mesoescala

Los Complejos Convectivos de Mesoescala (MCC, sigla en inglés) son estructuras convectivas de grandes dimensiones que se desarrollan con mayor frecuencia sobre la parte central de Suramérica (ver figura 3.4) afectando con intensas precipitaciones, tormentas eléctricas y en algunos casos con tornados. Las regiones afectadas con mayor frecuencia son: Paraguay, Uruguay, noreste de Argentina, sur de Brasil y en menor frecuencia el Perú. Según Velasco y Fritsch (1987), hay una predominancia de MCC en los subtrópicos en primavera y verano. Cotton y Anthes (1989), mencionan la situación a gran escala que incluye también la existencia de inestabilidad, condicional y convectiva, y una vaguada de onda corta en niveles medios, que actuaría como mecanismo de disparo a nivel sinóptico. Esos factores deben persistir por un tiempo suficiente para que las tormentas iniciales se transformen y se organicen en sistema de mesoescala de mayores proporciones.

En Colombia, las altas precipitaciones que ocurren en el oeste del país no solo se explican por la Advección de humedad proveniente del este por parte de los vientos Alisios. En efecto, la disminución de la diferencia meridional de las temperaturas del océano Pacífico entre Perú y Colombia ocasiona una disminución de la Advección de humedad proveniente del océano Pacífico hacia la costa occidental de Colombia. En la parte noroccidental de América del Sur los vientos Alisios que cruzan el ecuador desde el hemisferio sur adquieren una dirección del oeste penetrando al territorio colombiano (Poveda, Mesa & Carvajal, 1997)

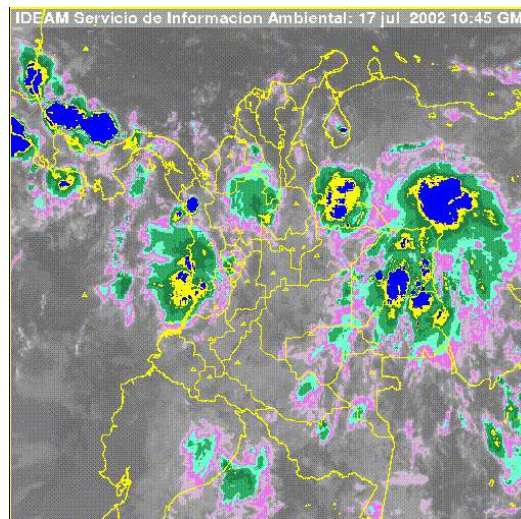


Figura 3.4 Foto de un sistema convectivo de meso-escala en el pacífico de Colombia. Ondas del Este.

Fuente: http://www.emc.ncep.noaa.gov/research/cmb/sst_analysis/images/wkanomv2.png

Las alteraciones más comunes de los vientos Alisios son el régimen de ondas en el este, a la vez que son uno de los factores que se conjuga para explicar las anomalías hidrológicas en Colombia durante las fases extremas del ENSO por medio de la reducción del número de ondas del Este que se generan en la costa occidental de África y que viajan hacia el oeste, desplazándose por el océano Atlántico tropical y el mar Caribe (ver figura 3.5). Diversa literatura resume las características de estas ondas en los últimos 30 años en diferentes regiones, pero es conveniente comenzar con el modelo propuesto por Riehl (1979) que se basa en las observaciones sobre el área del Caribe.

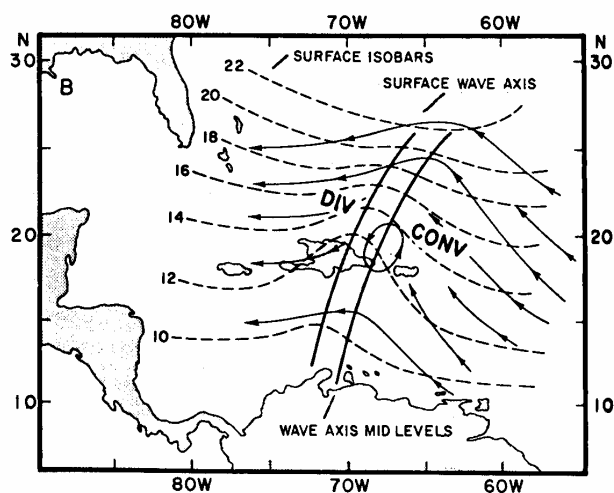


Figura 3.5 Mapa de isobaras en superficie, distribución de la divergencia y convergencia en la superficie, eje de la onda en la superficie y en cúspide a 10000 pies.

Fuente: Carlson & Lee (1978) en Hastenrath, 1991

La existencia de perturbaciones ondulatorias con longitudes de onda entre 3000 y 4000 km, velocidades de propagación entre 8-10 m/s y periodo en el rango de 4-5 días en la ZCIT fue descubierta por Chang (1970) a través del Atlántico tropical y en el análisis de fotografías diarias de satélite. Este modelo, tiene como base el seguimiento de las ondas, que se extiende en un eje norte-noreste a sur-suroeste. La longitud de onda medida de cresta a cresta es de unos 1700 kilómetros. El desplazamiento se da hacia el noroeste, siguiendo en gran medida el flujo base. Esto correspondería a un periodo en el rango de tres días. (Hastenrath, 1991).

El viento predominante en este mar tropical, durante casi todo el año, es desde la región Este (desde el Nordeste hasta el Sudeste, con la mayor frecuencia en la dirección Este) con velocidades entre 12 y 28 km/h (Fuerza 3-4, según la escala Beaufort), el cual llega a alcanzar valores en la parte oriental entre 20 y 38 km/h durante los meses de mayo, junio y julio (NOAA, 1985). Este régimen de

viento genera alturas de olas que van de 0.5 a 1.25 m, con un incremento hasta de 2,5 m en el caso de persistir un número determinado de horas las velocidades del viento entre 29 y 38 km/h.

Las ondas del este y los ciclones tropicales parecen estar estrechamente relacionados. Existe muy buena evidencia acerca de la reducción en el número de tormentas tropicales y huracanes en el mar caribe cuando se presenta el fenómeno del El Niño, lo cual parece indicar una disminución en el número de ondas del este.

3.5 Oscilación Madden Julian

La Oscilación Madden-Julian (MJO) es una perturbación tropical que se propaga hacia el este en las zonas tropicales de todo el mundo con un ciclo en el orden de los 30-60 días. La OMJ tiene un gran impacto en los patrones de las precipitaciones tropicales y extratropicales, en la circulación atmosférica, y la temperatura de la superficie de casi la totalidad de las zonas tropicales y subtropicales del mundo. Hay pruebas de que la MJO influye en el ciclo ENSO. No causa el fenómeno de El Niño o La Niña, pero puede contribuir a la velocidad de desarrollo y la intensidad de ambos.

La Oscilación Madden-Julian se caracteriza por la propagación hacia el este en las regiones de mayor lluvia tropical, principalmente en los océanos Índico y Pacífico. La anomalía de precipitación es la primera evidencia sobre el Océano Índico, y sigue siendo evidente, ya que se propaga hacia el este a lo largo de las cálidas aguas de la zona occidental y central del Pacífico tropical. Cuando las temperaturas de las aguas oceánicas son frías en el Pacífico oriental, el patrón de las lluvias tropicales generalmente es poco, pero a menudo vuelve a aparecer en el Atlántico tropical y África. Estas características se extienden por todo el mundo y no se limitan al hemisferio oriental. Por lo tanto, proporcionan información importante sobre las regiones de movimiento ascendente y descendente asociada a las fases de la oscilación. La figura 3.6 muestra una sección transversal vertical ecuatorial de la MJO mostrando los cambios en la nubosidad, precipitación, velocidad y dirección del viento, y TSM y de como la OMJ se propaga hacia el este en todo el trópico (Adaptado de Madden y Julian, 1971; 1972).

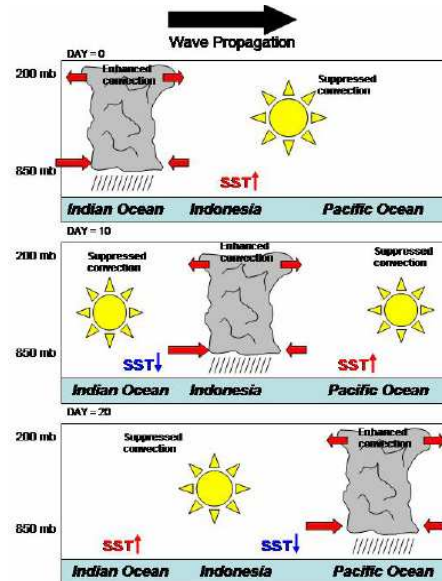


Figura 3.6 Sección transversal sobre el ecuador de la propagación de la OMJ desde el océano Índico al Pacífico Oeste. Las flechas rojas indican la dirección del viento y las flechas azules indican los niveles positivos (negativos) de las anomalías de temperatura superficial del mar.

Fuente: http://www.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/MJO_summary.pdf

Estas características se extienden por todo el pacífico tropical y no solo se limita al hemisferio oriental, por tanto, proporcionan información importante sobre las regiones de movimiento ascendente y descendente asociada a las fases de la oscilación. En estudios realizados sobre el tema, se han utilizado datos de anomalías de viento zonal en varios niveles y radiación de onda larga emergente (OLR) la cual está relacionada con el incremento o disminución de la actividad convectiva de gran escala (Hoyos, 2001).

3.6 ZCIT

La confluencia de los vientos Alisios del Noreste y del Sureste en la ZCIT produce un movimiento ascendente del aire, el transporte de la humedad por la vertical y la consecuente condensación y generación de nubosidad de desarrollo vertical y de la precipitación. La ZCIT se desplaza latitudinalmente siguiendo el movimiento aparente del Sol con respecto a la latitud, con un retraso aproximado de dos meses. Con este desplazamiento, la ZCIT pasa sobre el territorio colombiano en dos ocasiones cada año, generando un comportamiento bimodal (dos máximos) de la precipitación durante el año en el sur y centro, y unimodal (un máximo) en la región Caribe.

Sobre Colombia y áreas vecinas en el sector del Océano Pacífico Oriental, la ZCIT alcanza su posición extrema meridional en los 2° N entre enero y febrero, mientras que en diciembre está un

poco más al Norte. Durante los eventos El Niño la ZCIT puede alcanzar los 5° de latitud Sur. En la parte continental, la ZCIT aparece y se ubica entre los 5° y 10°S. (Ver Figura 3.7)

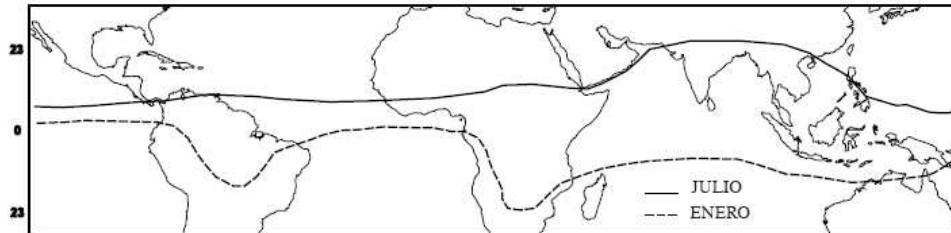


Figura 3.7 Banda de dominio de la Zona de convergencia Intertropical

Fuente: Synoptic and dynamic Climatology – (Roger & Carleton, 2001).

3.7 La Influencia de los Océanos Pacífico y Atlántico

La situación geográfica del territorio continental de Colombia al Noroccidente de Suramérica, condiciona la influencia de los procesos que ocurren en los Océanos Atlántico (incluido el Mar Caribe) y Pacífico. En la costa Pacífica, el Océano, en interacción con una circulación atmosférica de tipo monzónico que se desarrolla en la zona ecuatorial, aporta humedad al Sur de la vertiente Occidental de la cordillera Occidental. De la misma manera, en el sector Norte del Pacífico colombiano, un sistema cuasipermanente dirige la humedad hacia la cordillera. Lo anterior hace que la región Pacífica colombiana sea una de las zonas más lluviosas del mundo. El Océano Atlántico influye en el clima nacional a través de los centros permanentes de alta presión. La alta subtropical de las Azores extiende su influencia con un zona de alta presión sobre el Mar Caribe, la cual afecta estacionalmente el Norte del territorio colombiano. Esta influencia se pone de manifiesto principalmente en la distribución estacional de los vientos (predominio Este y Noreste), la nubosidad (escasa, en el extremo Norte) y la precipitación (regiones relativamente secas). De la misma manera, el clima nacional, particularmente el de las áreas marítimas e insulares del Caribe colombiano y del Norte del territorio continental, se ve frecuentemente influenciado por los ciclones tropicales que nacen sobre el Atlántico tropical. Se ha identificado que en la variabilidad climática interanual del territorio colombiano, los procesos del Pacífico como El Niño-La Niña-Oscilación del Sur, juegan un papel importante. Igualmente, la variabilidad interanual del Atlántico influye en el clima nacional, sin embargo, aún no se ha evaluado plenamente la magnitud de esta influencia en comparación con la del Pacífico.

4 Métodos de predicción de caudales

Desde hace un par de décadas se ha venido investigando la influencia de fenómenos macroclimáticos sobre la hidrológica colombiana. En particular sobre las causas, efectos e implicaciones del fenómeno El Niño/Oscilación del Sur (ENSO), la oscilación del Atlántico Norte, la oscilación cuasi-bienal, entre otras. La predicción de caudales ha sido un elemento importante dentro de ese contexto, por sus múltiples aplicaciones en generación de energía eléctrica, en agricultura, en salud humana, etc. La literatura publicada en ese contexto incluye los trabajos de Poveda y Mesa (1993), Mesa et al. (1994b), Carvajal et al. (1994), Poveda (1994); Poveda y Penland (1994), Poveda y Mesa (1996); Poveda (1998); Poveda y Jaramillo (2000); Poveda et al. (1999). En cuanto a modelos estadísticos, se ha trabajado con técnicas de regresión múltiple, análisis espectral singular (Carvajal et al., 1994), Modelo Lineal Inverso, redes neuronales (Carvajal et al., 1994; Poveda et al., 2002), regresión múltiple adaptiva por tramos (Poveda et al., 2002), modelos autorregresivos dependientes del régimen (Salazar et al., 1994), etc. Para el caso de predicción de caudales en Colombia, los efectos producidos por eventos macroclimáticos -como ENSO, ondas del Este, Madden Julian, entre otros, no son despreciables y la capacidad de predicción cambia de acuerdo a la variable que se quiere interpretar, su rezago y su estimación.

La dinámica de los fenómenos es aleatoria y puede fluctuar indistintamente alrededor de varios estados estables. Esto, de alguna forma justifica el uso de variables secundarias o variables predictoras para la determinación óptima de los caudales. (Poveda, 2004). En hidrología se han utilizado tradicionalmente modelos lineales determinísticos y estocásticos univariados y multivariados en la predicción de series de las variables fundamentales, especialmente de caudales. Es por lo tanto necesario describir algunos antecedentes que fundamentan las ideas de asociar modelos determinísticos y estocásticos con el propósito de mejorar la representación de los procesos y de estudiar, usando experimentación numérica, la capacidad predictiva del modelo determinístico.

El desarrollo de pronósticos para el caudal de los ríos es útil para reducir esta incertidumbre asociada a la gestión y conservación del recurso agua. El pronóstico de caudales es, por tanto, una de las medidas de gestión usadas para lidiar con la incertidumbre del clima. Dichos pronósticos pueden ser de corto plazo, períodos de unas pocas horas o días, y de largo plazo, con tiempos de antelación de hasta nueve meses (Georgakakos y Krzysztofowicz, 2001).

Sin embargo, en una cuenca de corta memoria y bajo almacenamiento, la capacidad de hacer predicciones de caudal de largo plazo depende de la predicción de las condiciones climáticas como

El Niño. Las predicciones de caudal de largo plazo pueden realizarse sobre la base de la estadística estacional local: este método predice el mismo valor todos los años si no se usa un término aleatorio.

Modelos estocásticos que toman en consideración las correlaciones temporales cuando la memoria de la cuenca es corta, estos modelos reproducen los resultados del método anterior.

Modelos empíricos que relacionan, con cierta anticipación, variables climáticas con caudales: en general, este tipo de modelos explica solamente una pequeña parte de la variabilidad del caudal.

Modelos determinísticos climáticos e hidrológicos: estos modelos dependen sustancialmente de la capacidad del modelo climático para predecir la precipitación.

En este capítulo se analizan algunos modelos que se han desarrollado para resolver el problema de la estimación de los caudales a diferentes escalas temporales, resaltando el interés de desarrollar un modelo de predicción semanal asociado a la variabilidad intraestacional en nuestro país donde la incidencia de la dinámica de alta frecuencia dominan e intervienen en los caudales semanales.

4.1 Modelos Estocásticos

Un proceso estocástico, $X(t, w); t \in I, w \in \Omega$ es una función de dos variables t, w . Para un momento $t = t^*$ específico, se tendrá $X(w)$ que es una función solo de w , por tanto se comporta como una variable aleatoria. En el caso de fijar w , $X(t)$ se convierte en una función del tiempo y se denomina una “trayectoria muestral” de un proceso estocástico o serie de tiempo.

En el año de 1970, Box & Jenkins desarrollaron un cuerpo metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series de tiempo en los que las variables de tiempo juegan un papel fundamental. Una parte importante de esta metodología es que la variable de interés indica las características de la estructura probabilística subyacente.

En general, cualquier método de pronóstico estimará el valor futuro de la variable deseada, sin embargo, esta información es puntual y realmente sólo refleja la tendencia de evolución de la observación y no el conjunto de posibles evoluciones que puede sufrir el valor de una observación a partir de su último valor conocido. Al constituir modelos de pronóstico no se pretende que este se ajuste precisamente, sino saber que se está aproximando a una realidad más compleja

En este sentido los modelos que mejor representan los caudales son los ARMA y ARIMA. Su aplicación ha sido ampliamente conocida en series de tiempo resaltando los trabajos realizados por Maqridakis 1983. La forma de estos modelos está basado en la siguiente expresión (Múnera, 1985)

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

Donde

ϕ_j y θ_i son los parámetros del modelo por estimar

ε_t : son las variables aleatorias con distribución normal

Y_t : las variables aleatorias (lluvia o caudal)

p: la cantidad de parámetros autorregresivos

q: número de parámetros de promedio móvil

Las variables aleatorias Y_t y ε_t , corresponden a una distribución de probabilidad normal y tienen media cero; adicionalmente esta última variable es independiente de los ε_t previos y de los Y_{t-k} para $k > 0$.

En la predicción se han aplicado diversas metodologías asociadas a la modelación estocástica donde se destacan:

4.1.1 Modelos Univariados (AR)

Entre éstos se encuentran los Modelos Lineales Univariados AR(p) que se basan en la hipótesis de que las relaciones de dependencia estacionales interanuales son las mismas para todos los períodos. Los procesos autorregresivos suponen una dependencia lineal temporal del proceso consigo mismo; por tanto, el comportamiento de la variable en un momento dado depende de su ocurrencia pasada. El rezago temporal se denota por p, el cual indica el máximo número de períodos precedentes que tienen influencia en el valor presente de la variable. La ecuación general de estos procesos es:

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \phi_2 \bar{z}_{t-2} + \dots + \phi_p \bar{z}_{t-p} + \varepsilon_t \quad (2)$$

donde:

\bar{z}_t : Valor de la variable estandarizada en el instante t .

ε_t : Ruido blanco independiente de z_{t-k} para todo $k \geq 1$.

4.1.2 Modelos Integrados

En su mayoría, los procesos estudiados en Hidrología son no estacionarios, es decir, que la función de distribución probabilística que describe la población, cambia con el tiempo. Una manera

eficiente de garantizar la estacionariedad de una serie consiste en diferenciarla la cantidad (d) de veces que sea necesaria. Por ejemplo, para el caso de diferenciación de orden 1, se tiene:

$$w_t = z_t - z_{t-1} \quad (3)$$

La diferenciación de orden 2 es tal que:

$$y_t = w_t - w_{t-1} = z_t - 2z_{t-1} - z_{t-2} \quad (4)$$

En general, la diferenciación de orden d es de la forma $a_t = z_t - z_{t-d}$

Un valor de d igual a 1 sugiere la presencia de una tendencia lineal en la media, mientras que un $d = i$ indica la existencia de una tendencia polinómica de orden i en la media. En la mayoría de los casos típicos de series analizadas en hidrología se obtiene un orden de diferenciación menor o igual a 2.

4.1.3 Modelos de Media Móvil

En Wold (1983) demostró que un proceso estocástico débilmente estacionario (es decir, con media cero y varianza σ^2 que no contenga componentes determinísticos) puede escribirse como una función lineal de variables aleatorias no correlacionadas:

$$\bar{z}_t = \varepsilon_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (5)$$

donde los ε_i corresponden a variables aleatorias no correlacionadas y \bar{z}_t tiene distribución Normal

4.1.4 Modelos Autorregresivos de Media Móvil

Los modelos ARMA son modelos estocásticos que combinan los modelos autorregresivos (AR) con los modelos de media móvil (MA). La forma general de estos modelos está dada por:

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \dots + \phi_p \bar{z}_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (6)$$

donde los parámetros ϕ_i , θ_i y ε_i tienen el mismo significado que para los modelos AR(p) y MA(q).

4.1.5 Modelos Autorregresivos Integrados de Media Móvil

Los modelos ARMA(p,q) tienen como condición básica la estacionariedad de la serie. Sin embargo, aunque algunos fenómenos pueden presentar este comportamiento, la mayoría de procesos naturales presentan características de no estacionariedad, como es el caso de las señales hidrológicas y climáticas. Por esta razón es necesario eliminar esta característica de la serie por analizar. Para ello se involucra el anteriormente referido parámetro de diferenciación (d) que representa el número de veces que se diferencia la serie. La incorporación de este parámetro da

origen a los denominados modelos autorregresivos integrados de media móvil ARIMA (p,d,q). La expresión general de estos modelos es de la forma:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d z_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (7)$$

donde el operador B se define como:

$$\begin{aligned} Bz_t &= z_{t-1} \\ B^k z_t &= B \dots B z_{t-k}, \quad k \text{ veces} \end{aligned} \quad (8)$$

En predicción de caudales se han empleado varias de estas metodologías de las cuales se destacan los resultados obtenidos por Mohammadi (2006), en la estimación de los parámetros del modelo ARMA para el río Karún en Irán.

Shahjahan (2005), presentó una aplicación donde los resultados demostraron que el modelo PAR puede capturar la variabilidad estacional de los caudales del río Ganges razonablemente bien, tanto en el corto como en el largo plazo con buena significancia estadística.

Rodríguez (1988) presentó un conjunto de modelos determinísticos y probabilísticos, tendientes a resolver los problemas de estimación puntual de caudales (pronóstico) diarios y semanales en varias cuencas del nordeste de Antioquia. En este estudio se concluyó que los modelos estocásticos de pronóstico ARMA y particularmente del ARIMA, resolvieron el problema puntual pero no reflejaron la característica misma del fenómeno.

4.2 Polinomios de Regresión Multivariados y Adaptivos

Otra de las estrategias usadas en la predicción es la modelación adaptiva, la cual tiene como fundamento el mejorar la aproximación de funciones en espacios de alta dimensión y ajustar dinámicamente el comportamiento de la función a ser aproximada. (Friedman, 1991). Una de las bases de la computación adaptiva es el particionamiento recursivo, cuyo objetivo es usar los datos para simultáneamente estimar un buen conjunto de subregiones y parámetros asociados con cada una de las funciones de cada subregión. El particionamiento recursivo tiene la capacidad de detectar la baja dimensionalidad de las funciones de ajuste, es decir, una función puede depender fuertemente de un gran número de parámetros de manera global, y localmente la dependencia es fuerte sólo en unas pocas variables.

Este método no paramétrico y no lineal de regresión se trata de una generalización del particionamiento recursivo y ha sido utilizado en aplicaciones en climatología en la predicción del Índice de Oscilación del Sur (SOI), así como para producir series de tiempo filtradas de pronóstico para este índice de presiones (Keppenne, 1996).

Como antecedentes del método de predicción de MARS se tiene el modelamiento paramétrico lineal que es el modelamiento más común, y ejemplarizado en el ajuste de parámetros por el método de mínimos cuadrados. La expresión general es de la forma:

$$g(x/\{a_j\}_0^p) = a_0 + \sum_{i=1}^p a_i x_i, \quad p \leq n \quad (9)$$

donde g es una función de x , cuyos parámetros son (a_0, a_1, \dots, a_p) .

El modelamiento no paramétrico se caracteriza porque no aproxima una función única en todo el dominio, sino que ajusta una función con varias funciones paramétricas simples, generalmente polinomios de bajo orden, definidas sobre una subregión del dominio (ajuste paramétrico por tramos). El problema de este tipo de modelamiento se presenta en las discontinuidades y cambios bruscos de pendientes en los límites de las funciones, el cual puede ser controlado limitando el número de subregiones posibles y estableciendo el número deseado de derivadas continuas en los límites de las subregiones. El modelamiento no paramétrico presenta limitaciones para aproximar funciones en altas dimensiones.

Los modelos adaptivos resultan adecuados para la expansión de funciones de altas dimensiones y que consiste en la aproximación de una función diferente en cada subregión, basado en el particionamiento (t), cuyas subregiones son divididas. Se optimiza la partición en todas las variables y en todos los puntos usando un buen criterio de ajuste. Las subregiones son entonces recombinadas hasta que se encuentra un óptimo posible.

La función de aproximación puede ser de la forma:

$$\hat{f}(x) = \sum_{m=1}^M a_m B_m(x) \quad (10)$$

donde M es el número de subregiones, a_m son los coeficientes de la función y $B_m(x)$ es una función básica o base, dada por:

$$B_m(x) = \prod_{k=1}^{K_M} H[S_{KM}(x(k, m) - t_{KM})] \quad (11)$$

H es una función de paso que va desde $k = 1$ hasta k_M , el número de divisiones resultantes en la función básica, S_{KM} es igual a ± 1 según sea la división derecha ó izquierda respectivamente, t_{KM} es el nudo o partición de la variable y x es la variable predictora.

Uno de los grandes inconvenientes de este método es que cierto tipo de funciones simples son difíciles de aproximar. Las dificultades aparecen cuando las interacciones dominantes involucran una pequeña fracción del número total de variables. Esta metodología trata de encontrar la mejor relación de dependencia entre la variable a predecir y las variables de apoyo o variables predictoras. Además, intenta superar algunas de las limitaciones de la modelación no paramétrica y el particionamiento recursivo, planteando algunas generalizaciones a los procedimientos, garantizando por ejemplo modelos continuos y derivadas continuas.

En nuestro país se han aplicado los Polinomios de regresión Multivariados y Adaptivos en la predicción de caudales mensuales para el sector energético donde se han incorporado variables macroclimáticas y predicciones de la temperatura superficial del mar en las regiones niño 3-4, sin tener en cuenta las otras variables macroclimáticas. Estudios realizados por la Universidad Nacional de Colombia muestran aplicaciones de esta metodología donde las validaciones en estos no reflejaron mejorías significativas con la predicción de predictores pero si mejoría en los errores de validación con la advección de humedad del chorro del Chocó.

Rendón (1997) en el estudio de beneficios de la predicción hidrológica en el sector eléctrico colombiano, considerando la variabilidad climática por medio de la metodología MARS para predicción de caudales mensuales, Allí es obtuvieron resultados significativos resaltando que con este tipo de modelos no es necesario remover los ciclos de la hidrología debido a que estos modelos se acoplan a la función real satisfactoriamente. Como herramienta de predicción no lineal basada en las variables macroclimáticas y en los caudales previos obtuvo errores aceptables lo que sugiere la búsqueda y exploración con esta herramienta.

4.3 Redes Neuronales Artificiales

La red neuronal artificial (RNA) es probablemente la más exitosa máquina de técnica de aprendizaje flexible con estructura matemática que sea capaz de identificar las complejas relaciones no lineales entre los datos de entrada y salida sin intentar llegar a un entendimiento en cuanto a la naturaleza de los fenómenos.

En la predicción de caudales se ha trabajado esta herramienta por medio de modelos basados en Inteligencia Artificial como las Redes Neuronales Artificiales (RNA). Una red neuronal es un modelo heurístico que copia la estructura de un sistema neuronal biológico con el fin de alcanzar una funcionalidad similar. El procesador elemental o neurona es un dispositivo simple de cálculo que, a partir de un vector de entrada del exterior o de otras neuronas, proporciona una única respuesta o salida. En general una red neuronal consiste en el establecimiento de un conjunto de nodos o neuronas agrupadas en varias capas e interconectadas entre si (capa de entrada, capas ocultas o intermedias y capa de salida). Las RNA son modelos que pueden aprender relaciones altamente no lineales entre las variables independientes y la variable dependiente. Se caracterizan

por un paralelismo masivo de elementos computacionales conectados por pesos, llamados neuronas artificiales. Ver Figura 4.1.

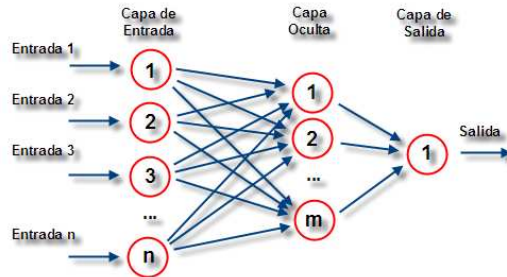


Figura 4.1 Estructura lógica de una red neuronal

Como un principio de funcionamiento en la red, en la etapa de entrenamiento los datos de entrada deben ser estandarizados, de lo contrario la red en muchos casos puede obtener resultados no deseados. Para el caso hidrológico es conveniente estandarizar los datos de acuerdo con la media y la desviación estándar en la escala que se desea hacer predicción. Una vez entrada la información en la red, las neuronas de la capa de entrada transmiten la información a las neuronas de la primera capa oculta a través de las conexiones existentes entre cada neurona de la capa de entrada con cada neurona de la capa oculta. Esta información siempre lleva una dirección de izquierda a derecha como se ve en la Figura 4.2 lo cual muestra que la alimentación de la información es hacia adelante. Cuando la información llega a las neuronas de la capa oculta, ésta es transformada dentro de cada neurona mediante el uso de diferentes funciones de transferencia. Entre éstas funciones existen funciones continuas y no continuas, algunas de las funciones más usadas se muestran en la figura 4.2

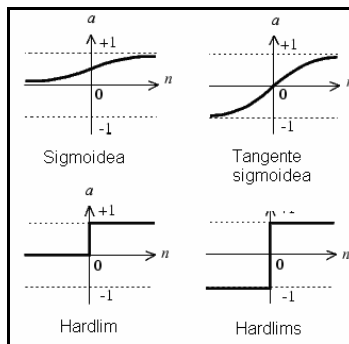


Figura 4.2 Funciones de Activación o Transferencia

Fuente: MATLAB

En hidrología no se deben usar funciones discontinuas, por lo que funciones como “Hardlim” y “Hardlims” no son capaces de capturar ciclicidades de tipo anual. La función sigmoidea y la tangente hiperbólica han sido ampliamente utilizadas para la transferencia de datos en este campo de la ciencia. Las ecuaciones que describen la función sigmoidea y tangente sigmoidea son, respectivamente:

$$f(x) = \frac{1}{(1 + e^{-x})} \quad (12)$$

$$f(x) = \frac{2}{(1 + e^{-2x})} - 1 \quad (13)$$

Una vez se ha entrenado la red, se pasa a realizar la predicción, para lo cual se entra a la red una serie de datos históricos observados, preferiblemente de considerable extensión, entrenando entonces nuevamente la red con los datos disponibles y se hace la predicción a la ventana de tiempo de interés. Se debe tener en cuenta que a medida que la ventana de predicción aumenta disminuye la precisión de la predicción.

Pruebas recientes de la tendencias no estacionarias en series temporales de los recursos hídricos como consecuencia del cambio climático y la misma variabilidad espacio temporal, ha suscitado más interés de mejorar los métodos de modelación para un sistema dinámico no lineal. En este contexto, las redes neuronales han sido ampliamente aplicadas en predicción de caudales en diferentes escalas temporales.

Coulbalya y Baldwin, (2005) construyeron una red neuronal para la predicción de diferentes series de tiempo no estacionarias. El método propuesto selecciona automáticamente la red más óptima de un conjunto de redes para predicción mensual. El rendimiento de la red neuronal es comparado con los resultados obtenidos de predicciones hechas con la aplicación de un Modelo de Regresión Multivariada y Adaptivas. Demostraron que la dinámica de la red neuronal es una alternativa adecuada para la modelación de un sistema dinámico y complejo en un sistema hidrológico con resultados mejores que los obtenidos con la metodología MARS. Esta red neuronal artificial se evaluó con series de tiempo de los volúmenes de almacenamiento del Great Salt Lake en Utah, EE.UU., el río San Lorenzo en Ontario, Canadá y los caudales históricos del río Nilo.

Laki (2001) aplicó un modelo de redes neuronales con algoritmo de retro-propagación para la predicción de caudales medios mensuales del río Huancané (Perú), a partir de datos de caudales, precipitación y evaporación, así como la evaluación de la capacidad de desempeño frente a modelos estocásticos. Los modelos de redes neuronales artificiales mostraron comparativamente mejor desempeño en materia de predicción frente a un modelo autorregresivo periódico de primer orden PAR.

Sahoo y Ray (2005) aplicaron redes neuronales para la predicción de flujo de un arroyo en Hawaii y uno de sus afluentes con el fin de estimar el caudal utilizando curvas de calibración considerando que la exactitud de la predicción se pierde cuando la corriente cambia su comportamiento de flujo. Los resultados mostraron que las redes neuronales pueden mejorar las curvas de calibración de estos afluentes siendo esta una metodología evaluada para la predicción de caudales obtenidos por niveles de flujo.

Dibike y Solomatne (2001) realizaron un estudio de aplicabilidad de redes neuronales RNA para los pronósticos caudales de la corriente en la cuenca del río Apure (Venezuela) bajo la estructura de dos tipos de arquitecturas de RNA, es decir, de varias capas de red perceptrón (MLP) y una función de red de base radial (RBF). Estas últimas se caracterizan por mejorar la aproximación de una función no lineal. Partiendo de una serie de datos afectados por ruido, se realizó una Regresión Funcional, donde fue necesario suponer que una función determinada, que expresara de la mejor forma la relación entre las variables dependientes e independientes. Los resultados obtenidos de estas redes se compararon con los del modelo de lluvia-escorrentía y resultaron ser ligeramente mejores para este río.

A nivel diario y semanal Rodríguez (1988) realizó el estudio de predicción de caudales diarios y semanales mediante modelos probabilísticos ARMA y ARIMA y se propuso un modelo de pronóstico de tipo determinístico y probabilística con el fin de resolver el problema de estimación puntual de caudales (diarios y semanales) para efectos de la operación de sistema de recurso hidráulicos en Colombia. En este estudio se encontró que los modelos estocásticos tipo ARMA y ARIMA resuelven el problema puntual pero no reflejan la característica del mismo fenómeno que se estudia, es decir no permiten pruebas de sensibilidad del tamaño muestral.

Sivakumar, et al (2002). Desarrollan un modelo de predicción de caudales diarios, presentando un paralelo de metodologías entre las redes neuronales artificiales y el uso de la reconstrucción del espacio de fase. Se presentaron mejores resultados con esta última debido a que las redes neuronales cambiaban mucho su entrenamiento al cambiar el número de neuronas en la capa oculta. Otra razón que se evaluó fue la escala temporal, debido a la alta variabilidad que se da en el rango de predicción de 1 a 7 días.

Ludermir y Valença (2001), desarrollaron un modelo de redes neuronales para la predicción de caudales semanales asociados a la operación de embalses en Brasil. Usaron redes neuronales con el fin de compararlas con las metodologías tradicionales de predicción usadas en Brasil como es el modelo PARMA. Se probaron ambas metodologías en cuatro plantas hidroeléctricas situadas en diferentes cuencas hidrográficas. Los resultados obtenidos en la evaluación del desempeño por redes neuronales fueron mejores que los resultados obtenidos con modelos PARMA en todos los cuatro escenarios.

4.4 Análisis Espectral

El análisis espectral Singular (AES) es metodología de análisis de series de tiempo que incorpora la estadística multivariada, sistemas dinámicos y procesamiento de señales con áreas de aplicación en las matemáticas, la física a la economía y la matemática financiera. También en estudios de meteorología, oceanografía y estudios de mercado.

El Análisis Espectral Singular (AES) es usado en estudios climáticos para referirse a la aplicación univariada del Análisis de Componentes Principales (ACP) en el tiempo y es equivalente a aplicar las Funciones Ortogonales Empíricas (FOE), a series de tiempo univariadas. El AES trata de capturar la evolución del sistema en el espacio de fases. La implementación del AES procede de manera análoga a la reconstrucción artificial de la evolución de las trayectorias en el espacio de fases de un sistema no lineal. La reconstrucción artificial (basada en el teorema de Takens, 1981) es un método tradicional en el análisis de procesos que exhiben sensibilidad a las condiciones iniciales o caos determinístico.

Carvajal, et al. (1998) realizó un estudio de predicción hidrológica en Colombia mediante análisis espectral singular y máxima entropía aplicando el método a la predicción de lluvias totales mensuales del periodo 1991-92 en las cuencas de los embalses El Peñol, Troneras y Piedras Blancas del departamento de Antioquia en Colombia. Dada la fuerte influencia del fenómeno El Niño-Oscilación del Sur (ENSO) sobre la hidroclimatología de Colombia, se estimaron las correlaciones entre el índice de Oscilación del Sur (IOS) y las Componentes Principales del campo de precipitación ofreciendo una nueva alternativa para la predicción hidrológica en Colombia.

Smith, et al. (2003) elaboraron un estudio de aplicación de cinco metodologías para la predicción de caudales medios mensuales en Colombia destacando el Análisis Espectral Singular, las Redes Neuronales Artificiales, Redes Adaptativas Neuro-Difusas, y la aplicación de estos modelos en diferentes estaciones de medición de caudal pertenecientes al Sistema de Generación de Energía Eléctrica de Colombia. Los modelos se aplicaron para ventanas de predicción de uno, tres, seis y doce meses.

En Carvajal y Marco (2005), desarrollaron un modelo multivariado de predicción de caudal mensual. Caso de estudio río Cauca, utilizando dos técnicas estadísticas multivariadas: El análisis de funciones ortogonales empíricas (FOES) y el análisis de correlación canónica (ACC). Para este estudio se ajustaron modelos ARMAX, para predecir los caudales, utilizando las primeras componentes principales y coeficientes canónicos de las variables macroclimáticas, como variables auxiliares en los modelos de caudal; obteniéndose mejoras en la predicción Lo que indicó que las variables macroclimáticas aportaron información adicional y optimizaron el ajuste, por lo que los autores propusieron ser utilizados para predicción.

Otra de las metodologías relacionadas fue la propuesta por Hoyos (1999) denominado “PREBEO” que se basa en la descomposición espectral de las series predictoras y la serie de caudales por predecir, usando la llamada “transformada en Onditas” o ‘wavelets” cuyo objeto es descomponer las señales en las distintas bandas frecuenciales características de la variabilidad temporal. La transformada en onditas logra localización temporal y frecuencial simultáneamente, mediante el uso de funciones llamada “ondita madre, de dos parámetros; uno de localización y otro de escalamiento.

Esta herramienta se aplicó para la predicción de caudales a escala temporal de 5 días (o péntada) en Colombia. El método consideró diferentes índices macroclimáticos asociados a fenómenos que afectan la climatología en Colombia en tal escala de tiempo.

Igualmente Poveda, et al (2003) en el estudio de Beneficios de la predicción no lineal de la hidroclimatología en la operación del sistema hidrotérmico colombiano, aplicaron esta técnica no lineal obteniendo y una ganancia importante comparadas con las predicciones obtenidas de los métodos lineales.

5 Análisis de la Información

5.1 Análisis previo de la información

Esta etapa tiene como objeto organizar la información hidrológica en series de datos en péntadas, unificando la presentación de la misma para su adecuado procesamiento. La información base es de tipo diario para cada una de las estaciones seleccionadas. Una vez eliminadas las inconsistencias se tienen entonces series históricas anuales de 365 días en los años regulares y 366 datos para los años bisiestos. Se desea tener igual cantidad de datos por año y se consideran datos en péntadas (5 días), para obtener 73 péntadas por año.

Estación La Bananera: Se tienen 29 años de registro ininterrumpido de caudal, desde enero de 1969, hasta diciembre de 1998.

Estación La Virginia: Se tienen 26 años de registro ininterrumpido de caudal, desde enero de 1969, hasta diciembre de 1994.

Estación La Reina: Se tienen 24 años de registro ininterrumpido de caudal, desde enero de 1972, hasta diciembre de 1995.

Estación Puente Negro: Se tienen 27 años de registro ininterrumpido de caudal, desde enero de 1976, hasta diciembre de 2002.

Estación La Bretaña: Se tienen 25 años de registro ininterrumpido de caudal, desde enero de 1978, hasta diciembre de 2002.

Las series históricas de los caudales semanales de los cinco ríos se presentan en las figuras 5.1, 5.2, 5.3, 5.4 y 5.5.

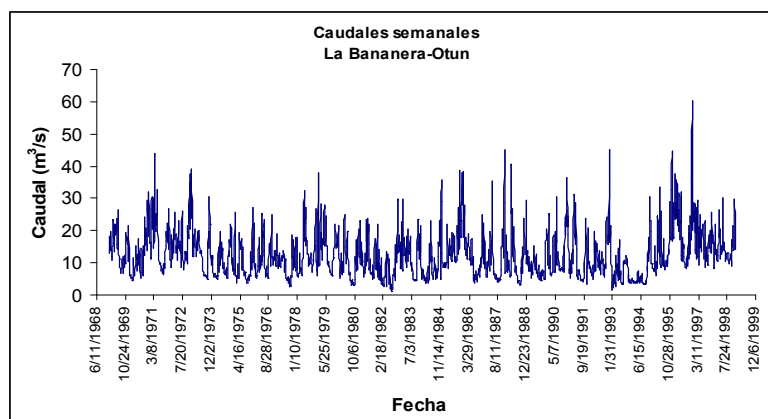


Figura 5.1 Serie de caudales semanales Río Otún

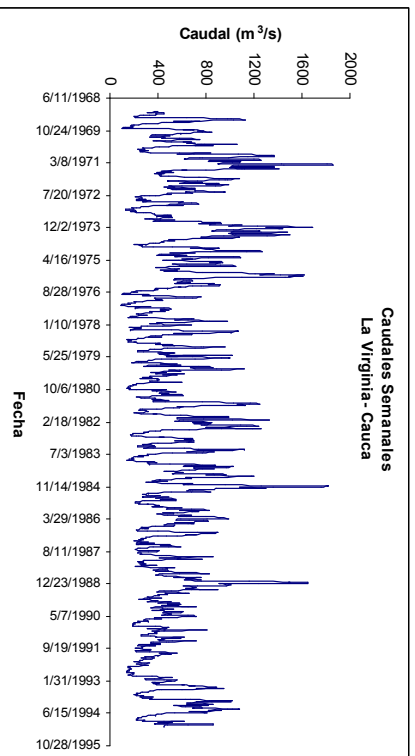


Figura 5.2 Serie de caudales semanales Río Cauca

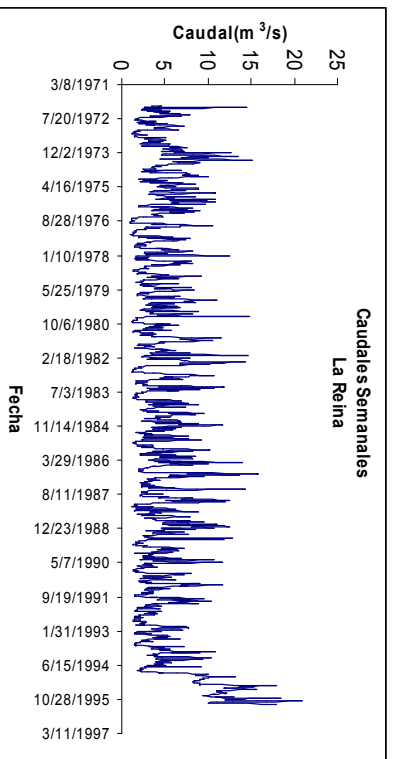


Figura 5.3 Serie de caudales semanales Río San Eugenio

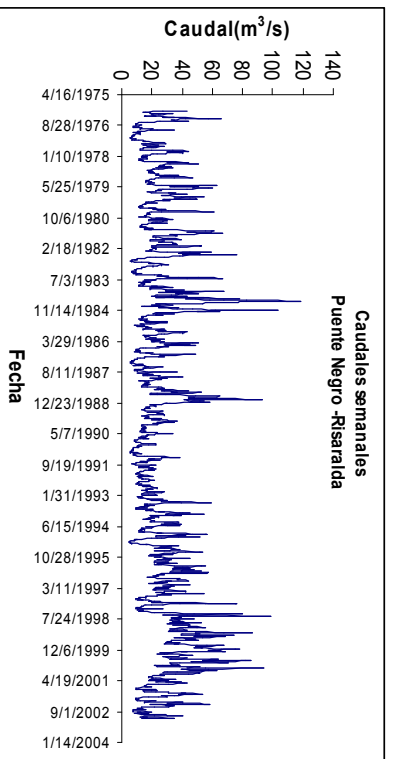


Figura 5.4 Serie de caudales semanales Río Risaralda

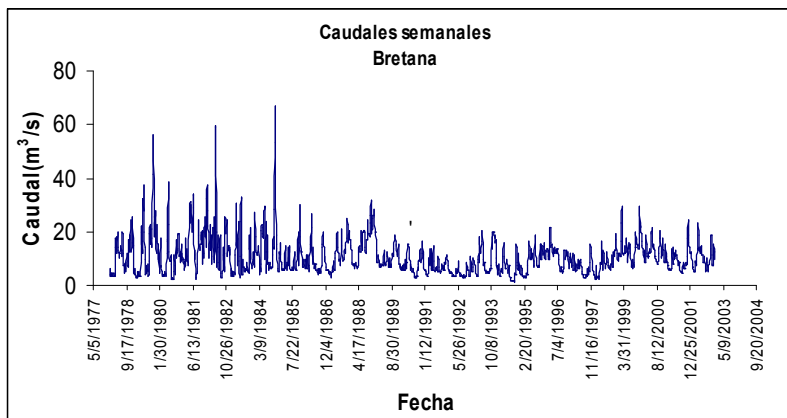
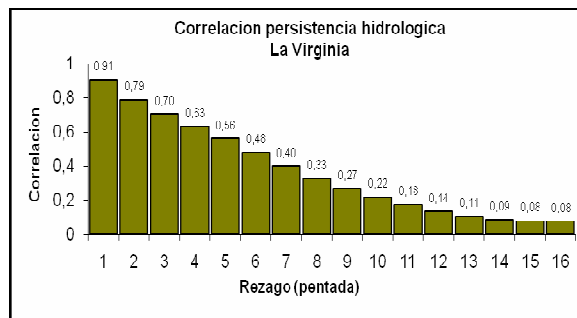
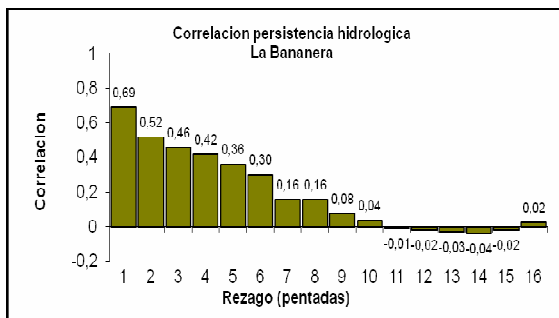


Figura 5.5 Serie de caudales semanales Río Mapa

5.1.1 Persistencia hidrológica

Este método consiste en suponer para cada período t , como predicción del caudal, el valor registrado en el tiempo $t-1$ lo que consiste en repetir la ocurrencia del período inmediatamente anterior. Los caudales tienen memoria y esto se corrobora al correlacionar el caudal de la semana t con el caudal de la semana $t-1$. Por lo tanto se establece como primer parámetro de entrada en el modelo, el caudal o los caudales rezagados un tiempo $t-1$ o $t-2$. La figura 5.6 muestra las correlaciones que tiene el caudal t con los anteriores.



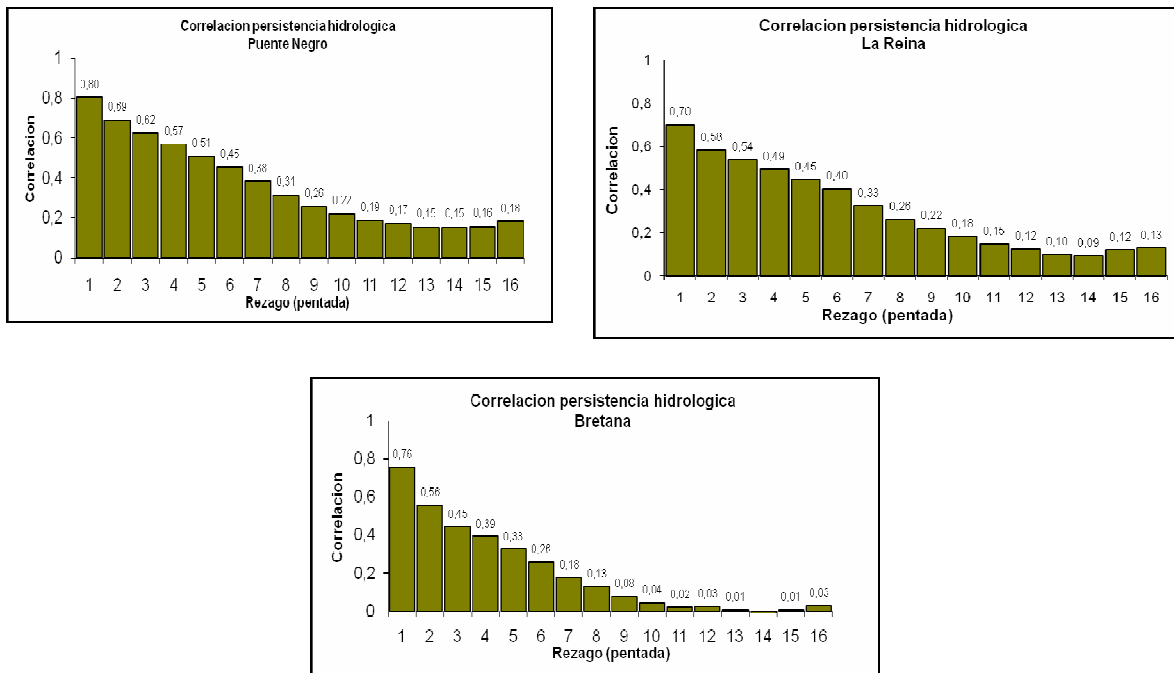


Figura 5.6 Correlación de caudal de la semana actual con las semanas anteriores

Para todos los ríos se estimaron correlaciones altas cercanas a 0.8 en promedio, destacando la correlación del río Cauca-Virginia y Risaralda-Puente Negro donde se tienen correlaciones de 0.91 y 0.83 respectivamente.

Existen también correlaciones con rezago de dos semanas, es notorio que para los mismos ríos anteriores se tengan correlaciones superiores a 0.6 lo que podría considerarse el caudal para un tiempo t-2 como una serie a incorporar dentro del modelo

5.1.2 Correlación espacio temporal de las series

Se realizaron exploraciones que permitieran analizar las series de caudales y posibles series de lluvias, para encontrar correlaciones entre las series de lluvia y caudal. Para obtener predicciones de caudales, es importante involucrar estas series (si existen y se correlacionan) debido a que es importante identificar el patrón de precipitación y si existe dependencia, poder cuantificarla e incorporarla en la predicción. Dada la poca información en los registros de lluvia, y los datos faltantes significativos que se encontraron en otras, se pudieron obtener dos series de esta variable para los ríos Otún y San Eugenio para ser consideradas en el modelo. En la figura 5.7 se observa la serie de tiempo y la correlación con lluvia vs caudal del río Otún y para el río San Eugenio se observa en la figura 5.8, el coeficiente de correlación para el río Otún es de 0.37 lo cual es

significativo si se considera que la cuenca tiene un área de 288 Km². Para el río San Eugenio se obtuvo una correlación de 0.58 para un área de 81 Km².

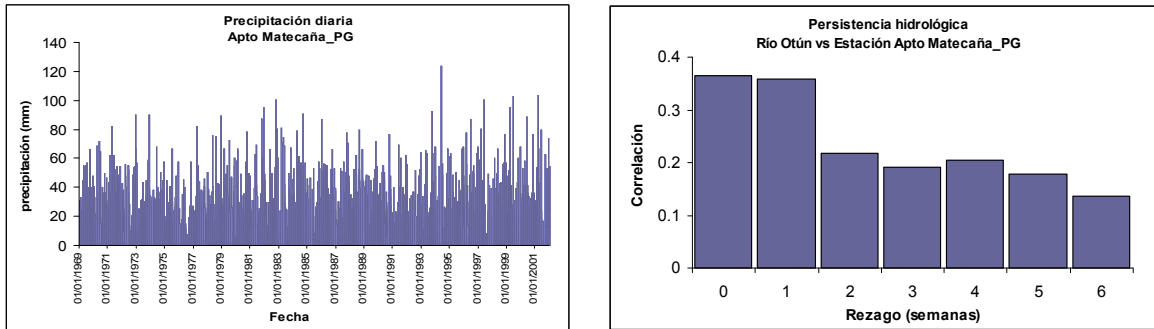


Figura 5.7 Relación serie de caudal río Otún con una serie de precipitación Aeropuerto Matecaña en la misma cuenca.

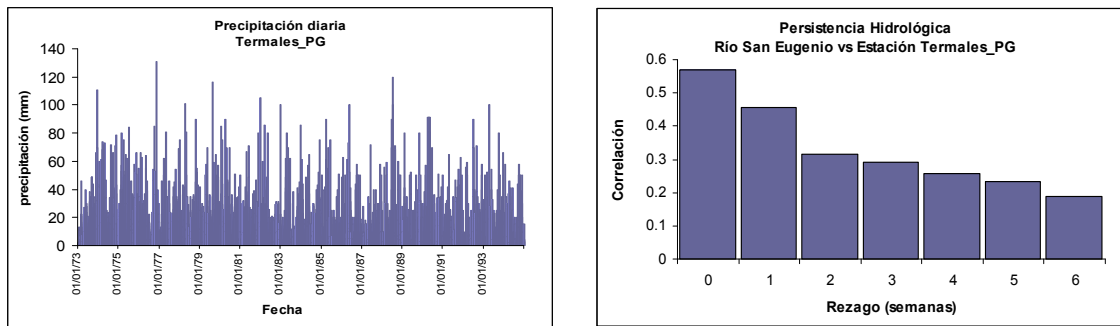


Figura 5.8 Relación serie de caudal río San Eugenio con la serie de precipitación de la estación Termales en la misma cuenca.

5.2 Análisis de las variables hidrometeorológicas y oceánicas de mayor influencia en la variabilidad intraestacional.

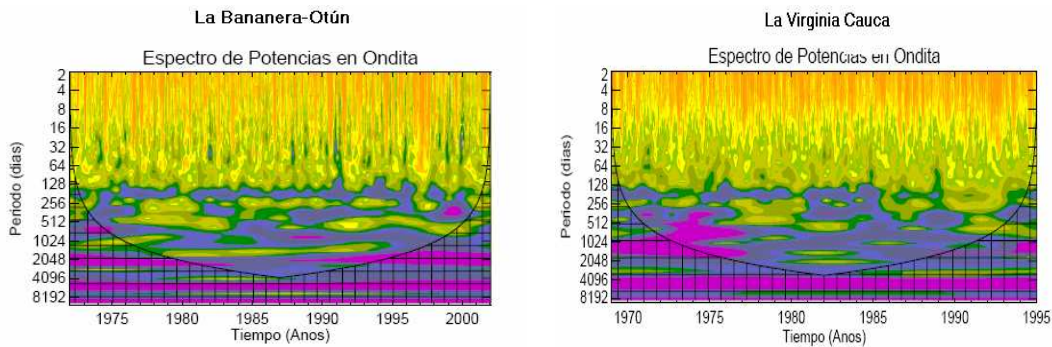
En este numeral se estudia el comportamiento de diferentes variables hidroclimáticas en Colombia que tienen influencia en la variabilidad intraestacional. Además, el comportamiento de señales climáticas provenientes de los océanos Atlántico y Pacífico y su efecto en la climatología de Colombia.

Arias (2005) en el estudio de diagnóstico y predicción de la variabilidad intra-anual de la hidrología colombiana presenta un análisis muy detallado de esta variabilidad y la existencia de actividad intraestacional en registros de precipitación, caudal, humedad del suelo en Colombia, y de

advección de humedad. En este estudio se hace énfasis en los índices climáticos asociados en esta escala y para ello se hace el diagnóstico de actividad en la banda espectral de 30-60 días por medio de los espectros de potencia en onditas para series de tiempo diarias y semanales correspondientes a registros de caudal, advección de humedad por componentes zonal y meridional del viento asociado con el chorro del Chocó y al índice de la Oscilación Madden-Julian. La característica de la función de onditas (wavelet analysis, Lau & Weng 1995) es su habilidad para detectar componentes de la frecuencia intermitentes y transitorias (ondas locales) mientras que el análisis de Fourier muestra las oscilaciones que continúan indefinidamente en el tiempo. En general el análisis de onditas es aplicable para identificar la modulación de una ampliación de una señal o una frecuencia, así como para identificar cambios en una serie de tiempo o en sus características frecuenciales. A continuación se presentan los espectros de potencias promedio y en onditas para los índices climáticos asociados a la variabilidad intraestacional y serie de caudales de los ríos seleccionados.

5.2.1 Detección de la banda intraestacional con serie de caudales diarios

A partir de los caudales diarios se construyen los espectros sin filtrar que muestran una señal en la banda 128-256 días asociados con el ciclo anual y semianual – comportamiento bimodal, para todas las series (figura 5.9). Sin embargo, con igual fuerza aparecen períodos importantes de 512 y 1024 días asociados al fenómeno ENSO. En porcentaje de varianza se pueden observar pequeñas bandas en el rango 30-60 días para unos pocos años.



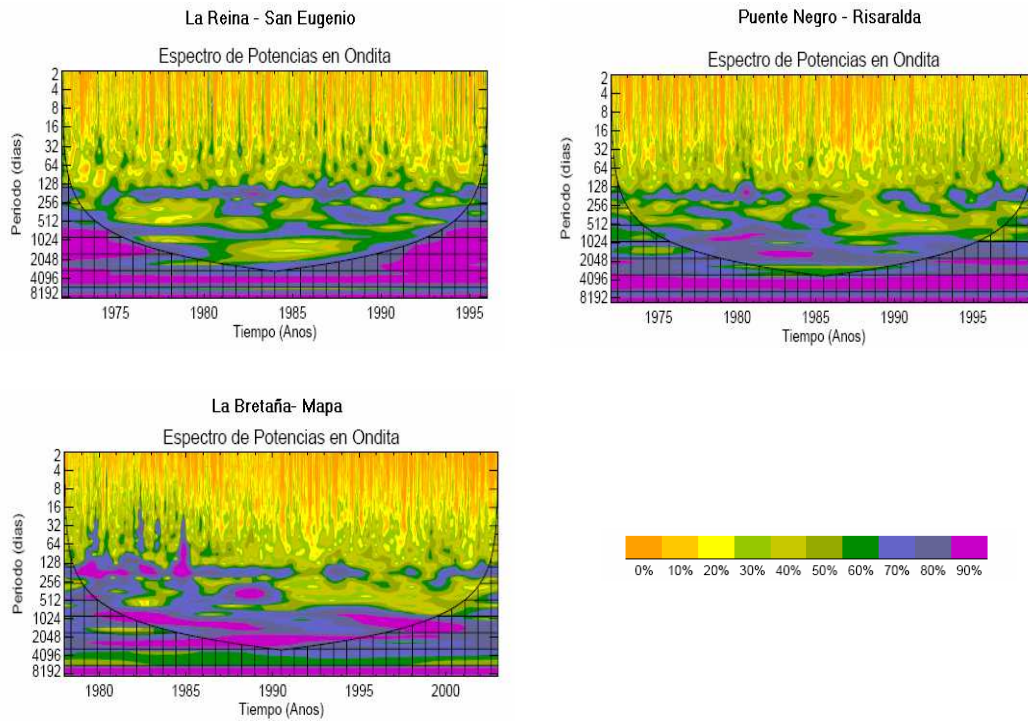


Figura 5.9 Espectros de potencia en onditas para la serie de caudales diarios

Al filtrar las series en la banda 30-120 días se observa la presencia de actividad en la banda 30-60 días en todas las series de caudal consideradas. En la figura 5.10 se muestra el espectro de potencias en ondita correspondiente a las series diarias de caudal. En los espectros de potencia promedio de caudal se observa un pico dentro del rango intraestacional.

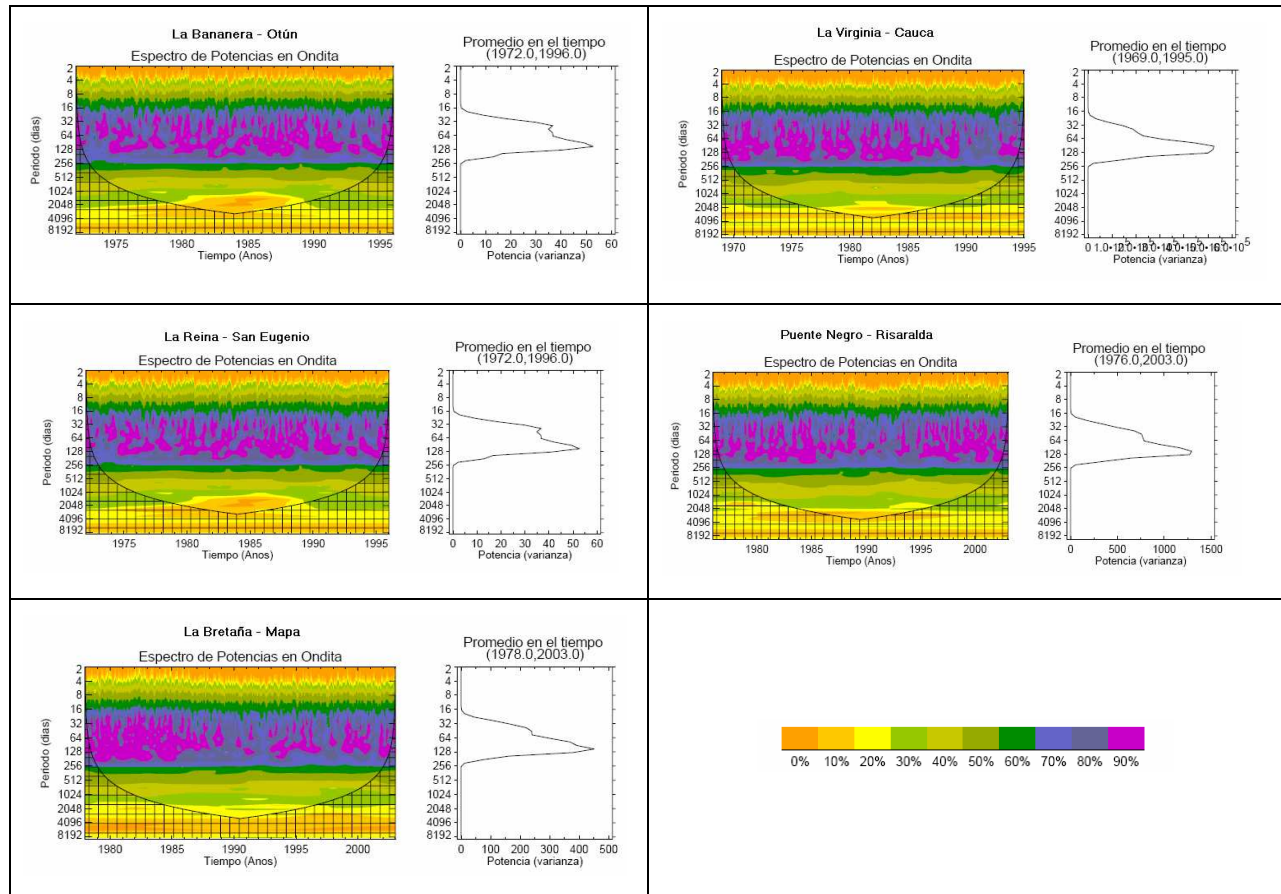


Figura 5.10 Espectros en ondita y de potencias para las series de caudales diarios filtrados en la banda 30-120 días.

5.2.2 Detección de la banda intraestacional con Advección de humedad

Los vientos del oeste que penetran por la costa del Pacífico colombiano se presentan en forma de una corriente de chorro del oeste, muy superficial, en el trópico americano. La corriente de vientos en chorro del CHOCÓ (Poveda, 1998). El transporte de humedad por los vientos zonales a 925 hPa sobre la región es un índice adecuado para representar la actividad y dinámica temporal del Chorro del CHOCÓ (Rendón, 2001). Esta serie hace parte de las variables que se incluirán en el modelo de predicción semanal. Con base en la velocidad del viento y humedad específica se construyeron series de advección de humedad tanto para la componente zonal como la meridional. La figura 5.11 presenta la serie de advección de humedad zonal y meridional con el respectivo espectro en ondita. Se observa la importancia de las bandas frecuenciales asociadas al ciclo anual (256-512 días), al ENSO (1024-2048 días) y para algunos años con menor varianza en las bandas 30-128 días.

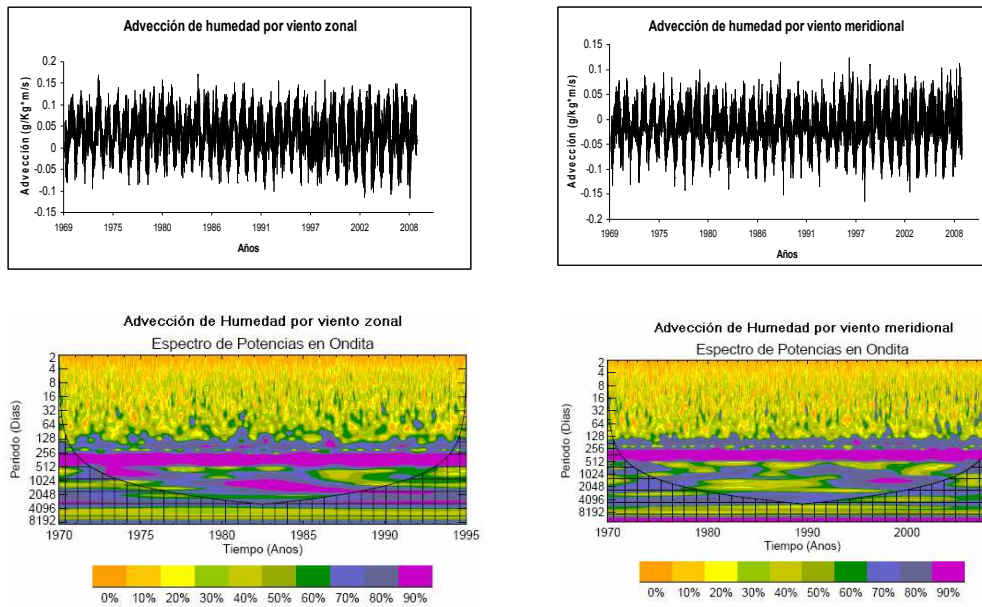
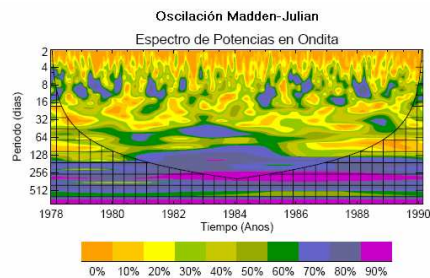


Figura 5.11 Series de advección de humedad por viento zonal y por viento meridional generada por el Chorro del Chocó (parte superior) y espectros de potencias en ondita correspondientes.

5.2.3 Oscilación Madden-Julian

Tomando la serie índice de la Oscilación Madden-Julian (OMJ) se calcularon los espectros en ondita para verificar su incidencia en la variabilidad intraestacional y su periodicidad de aproximadamente 30-60 días. En la figura 5.12 se observa que existen períodos importantes de 4-8 días aproximadamente y otros ciclo importantes con periodicidad entre 30 y 60 días.



----- Varianza de alta frecuencia asociados a ciclos semanales.

=== ciclos con periodicidad entre 30 y 60 días.

Figura 5.12 Espectro de potencias en ondita para la serie índice de la Oscilación Madden-Julian

6 Metodología modelo de predicción de caudales

6.1 Modelo ACP

Se presenta a continuación una metodología para la formulación de un modelo de predicción de caudales semanales asociados con la variabilidad intraestacional en Colombia.

Una de las prioridades que surgen dentro del aprendizaje de la predicción, es la necesidad de que los métodos de predicción puedan tomar ventaja de las componentes deterministas del fenómeno y describir sus principales aspectos no lineales. Dentro de las metodologías que se han aplicado en la predicción de caudales se caracteriza el Análisis Espectral Singular (AES), usado en estudios climáticos para referirse a la aplicación univariada del Análisis de Componentes Principales (ACP) en el tiempo y es equivalente a aplicar las Funciones Ortogonales Empíricas (FOE), a series de tiempo univariadas. El AES trata de capturar la evolución del sistema en el espacio de fases. La implementación del AES procede de manera análoga a la reconstrucción artificial de la evolución de las trayectorias en el espacio de fases de un sistema no lineal. La reconstrucción artificial (basada en el teorema de Takens, 1981) es un método tradicional en el análisis de procesos que exhiben sensibilidad a las condiciones iniciales o caos determinístico.

La metodología se basa en la aplicación de diferentes formas del Análisis de Componentes Principales (ACP) en el dominio del tiempo y del espacio. El AES se basa en el ACP en el espacio de los vectores conformados por versiones rezagadas de una serie de tiempo. En este contexto, la connotación del Análisis Espectral no se refiere al tradicional análisis en el dominio de la frecuencia, sino que se refiere a la expansión de la evolución de un campo espacio-temporal en diferentes modos ortogonales empíricos (obtenidos a partir de los datos). Hassani, (2007)

El análisis clásico de ACP se usa con múltiples series de tiempo y obtiene las direcciones principales de una secuencia de vectores M -dimensionales $(X_i, 1 \leq i \leq N)$, al expandirlos con respecto a una base ortonormal $(E^k, 1 \leq k \leq M)$.

El AES es la expansión de un campo discreto $(X_i, 1 \leq i \leq N)$, en sus componentes principales y Funciones Ortogonales Empíricas considerando una longitud de ventana M .

$$X_{i,j} = \sum_{k=1}^{cp} a_i^k E_j^k \quad 1 \leq j \leq M \quad (14)$$

Los coeficientes de proyección a_i^k son llamados componentes principales (CP) y los vectores E_i^k las funciones Ortogonales empíricas (FOE).

Para un análisis espectral univariado la expansión sería:

$$X_{i+j} = \sum_{k=1}^M a_i^k E_j^k \quad 1 \leq i \leq n; \quad 0 \leq j \leq M-1 \quad (15)$$

Para un análisis espectral multivariado el cual considere L sitios de información

$X_{l,i}$, $1 \leq l \leq L$, $1 \leq i \leq N$, la expansión de las CP y FOE es:

$$X_{l,i+j} = \sum_{k=1}^{L*M} a_i^k E_{l,j}^k \quad 1 \leq l \leq L; \quad 0 \leq j \leq M \quad (16)$$

El problema tiene una longitud de ventana M , pero los autovectores tendrán una dimensión igual a $L*M$.

Las FOE son los autovectores de la matriz Toeplitz, T_x que contiene los coeficientes de covarianza cruzada de los diferentes L vectores para rezagos de 0 a $M-1$.

La matriz de covarianza del proceso x_i , $i=1, \dots, N$, que tiene media cero se define como:

$$T_x = \begin{bmatrix} c(0) & c(1) & \dots & c(M-1) \\ c(1) & \dots & \dots & c(M-2) \\ \dots & \dots & \dots & \dots \\ c(M-1) & \dots & c(1) & c(0) \end{bmatrix} \quad (17)$$

donde $c(j)$; $0 \leq j \leq M-1$, es la covarianza de x en el rezago j .

Para estimar $c(j)$ se tiene la expresión de Yule-Walker:

$$c(j) = \frac{1}{N} \sum_{i=1}^{N-j} X_i X_{i+j} \quad (18)$$

O la expresión:

$$c(j) = \frac{1}{N-j} \sum_{i=1}^{N-j} X_i X_{i+j} \quad (19)$$

6.1.1 Autovalores y longitud de la ventana

T_x es simétrica y no negativa. Sus autovalores λ_k son positivos excepto cuando los datos no tienen ruido y vienen de un sistema dinámico con comportamiento cuasiperiódico. Los autovalores son ordenados en orden decreciente y los correspondientes autovectores están normalizados. La descomposición espectral del proceso se puede presentar como:

$$\sum_{K=1}^M \lambda_k E_h^K E_q^K = T_{x,hq} = c(h-q) \quad 1 \leq h \leq M, \quad 1 \leq q \leq M \quad (20)$$

donde h y p son componentes de cualquier autovector k que corresponde a la componente principal k -ésima.

Un problema importante en el AES es la escogencia de la ventana M , según el número de datos disponibles N , si se considera que la capacidad de resolución del AES es igual a la longitud de la ventana M y se desea reconstruir un atractor extraño, cuyo espectro incluye períodos de longitud arbitraria, el mayor M es el mejor, pero para que los errores estadísticos no afecten los últimos valores de la función de autocovarianza M no debe ser mayor que $N/3$.

6.1.2 Componentes Principales (CP)

La k -ésima CP es el coeficiente de proyección ortogonal de la serie original el k -ésimo vector propio:

$$a_i^k = \sum_{j=1}^M X_{i+j} E_j^k, \quad 0 \leq i \leq N-M \quad (21)$$

6.1.3 Componentes reconstruidas (CR)

Las CP son versiones filtradas de la serie original y se pueden considerar como un promedio móvil de la información en la longitud de ventana M considerada. En la ecuación 5.8 los términos individuales dependen del índice j (variando de 1 a M), por lo tanto hay M formas diferentes de reconstruir los componentes de la señal, que en general no dan los mismos resultados. Otro inconveniente es que las series resultantes son de longitud $N-M+1$ y no de longitud N . Las CR permiten extraer series de longitud N , correspondientes a un conjunto de valores ya seleccionados.

Se busca una serie y de longitud N , $y=R_A x$, tal que la diferencia entre y y x al cuadrado sea mínima, siendo A un subconjunto de K autovalores sobre los cuales se va a realizar la reconstrucción.

La solución de N , $y=R_A x$, en forma óptima de acuerdo al criterio de mínimos cuadrados es:

$$(R_A X)_i = \frac{1}{M} \sum_{j=1}^M \sum_{k \in A} a_{i-j}^k E_j^k, \quad M \leq i \leq N-M+1 \quad (22)$$

$$(R_A X)_i = \frac{1}{i} \sum_{j=1}^i \sum_{k \in A} a_{i-j}^k E_j^k, \quad 1 \leq i \leq M-1 \quad (23)$$

$$(R_A X)_i = \frac{1}{N-i+1} \sum_{j=1-N+M}^M \sum_{k \in A} a_{i-j}^k E_j^k, \quad N-M \leq i \leq N \quad (24)$$

Las CR son las $R_A x$ (denotadas por x^k) y tienen propiedades aditivas, como:

$$(R_A X)_i = \sum_{k \in A} x^k \quad (25)$$

Las series x pueden ser expandidas como la suma de sus CR.

$$x = \sum_{k=1}^M x^k \quad (26)$$

No obstante su aspecto lineal, la transformación de x en y , es no lineal, pues los autovectores E^k , dependen no linealmente de x (Vautard, R., Pascal, Yiou, y Michael Ghil, 1992). Una desventaja de las CR es que son correlacionadas aún en el rezago cero. La ventaja de los CR es que si hay períodos cortos de oscilaciones en la señal, se pueden localizar en forma precisa y que entre x y x^k no se presenta desplazamiento en la fase excepto al final del registro (Vautard et al., 1992).

6.1.4 Análisis de un modelo multivariado

Las mediciones de algunas variables atmosféricas pueden ser incluidas tales como la temperatura, la humedad y la presión del medio y las mediciones de variables hidrológicas como son los caudales y la precipitación (Porporato & Ridolfi, 2001). En general una reconstrucción multivariada para una predicción de caudales puede escribirse así:

$$\xi(t_i) = \left\{ \begin{array}{l} Q(t_i), \dots, Q(t_i - (m_Q - 1)\tau), Q_{uj}(t_i), \dots \\ x Q_{uj}(t_i - (m_{Q_{uj}} - 1)\tau), P_n(t), \dots, \\ x P_n(t_i - (m_{P_n} - 1)\tau), T_l(t_i), \dots, \\ x T_l(t_i - (m_{T_l} - 1)\tau), \vdots w(t_i), \dots, \\ x w(t_i - (m_w - 1)\tau), \end{array} \right\}$$

Donde $\xi(t_i)$ es el punto genérico para el espacio embebido, $Q_{uj}(t_i)$ es el vector de series de tiempo de los caudales. $P_n(t_i)$ y $T_l(t_i)$ son la precipitación y la temperatura medida en diferentes puntos de la cuenca, mientras que $w(t_i)$ resume otras variables hidrológicas que tengan información tales como humedad, evaporación. De esta manera se reconstruye con la totalidad de la información una descripción adecuada del sistema.

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ y_p \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{2,1} & \cdot & x_{n,1} \\ x_{2,1} & x_{2,2} & \cdot & x_{n,2} \\ \cdot & \cdot & \cdot & \cdot \\ x_{1,p} & x_{2,p} & \cdot & x_{n,p} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \cdot \\ C_n \end{bmatrix} \quad (29)$$

ó $Y=XC$, donde Y es un vector $p \times 1$ de observaciones de la variable dependiente, X es una matriz $p \times n$ formada por las p observaciones de cada una de las n variables independientes y C es un vector $n \times 1$ de los parámetros de ajuste desconocidos. Un modelo de ajuste lineal múltiple con intercepto es un caso particular de la ecuación (5.17), donde la variable x_j se debe igualar a 1 para toda observación de y .

El vector de parámetros C puede ser estimado minimizando $\sum e_i^2$, donde $e_i = y_i - \hat{y}_i = y_i - \sum_{j=1}^p \hat{C}_j x_{i,j}$, y \hat{C}_j es el valor estimado de C_j . Finalmente, luego de derivar e igualar a cero, se obtiene la siguiente expresión para estimar C .

$$\hat{C} = (X^T X)^{-1} X^T Y \quad (30)$$

El coeficiente de determinación múltiple R^2 se puede calcular como

$$R^2 = \frac{\hat{C}^T X^T Y - p\bar{Y}^2}{Y^T Y - p\bar{Y}^2} \quad (31)$$

Al combinar el enfoque metodológico planteado con los modelos de regresión lineal múltiple, se obtiene una herramienta de predicción no lineal. El carácter no lineal se adquiere gracias a la existencia de una combinación lineal diferente (ajuste lineal múltiple) para cada semana por predecir. Es decir, en la predicción de semanas diferentes, la estimación mediante regresión lineal del vector C será diferente, permitiendo en cada caso, utilizar las variables de apoyo que verdaderamente influyen en el caudal de una semana determinada.

6.1.6 Predicción usando Análisis por Componentes Principales ACP

La metodología de descomposición espectral permite descomponer la serie original en componentes principales y luego obtener una serie reconstruida apoyada en las variables macroclimáticas. Posteriormente que se establece un modelo lineal auto-regresivo para obtener las predicciones. Dos tercios de los datos se usaran en la calibración y un tercio en la validación como verificación de las predicciones. A continuación se presenta la metodología de este modelo.

6.1.7 Transformación de datos

Se consideran NS sitios y se estandarizan las series por sus respectivas medias y desviaciones estándar mensuales para cada sitio. La componente $x_{i,j}$ representa entonces el caudal en el sitio i , semana i , con longitud de la serie igual a N .

6.1.8 Pasos Análisis Espectral Singular

1. Se construye la matriz de covarianza cruzadas de dimensión $NS*M$, siendo M la longitud de ventana o número de rezagos considerada. Esta matriz tiene en cuenta el espacio y el tiempo. La matriz es cuadrada, donde en la diagonal principal los elementos corresponden a uno y los demás elementos son las autocovarianzas para cada sitio con rezago de 0 a $M-1$ y las covarianzas cruzadas de las series de los NS sitios con rezagos de 0 a $M-1$.

La expresión para calcular la matriz de covarianzas es:

$$c(l^*j, k^*j) = \frac{1}{N-j} \sum_{i=1}^{N-j} X_{l,i} X_{k,i+j}, \quad (32)$$

$$l = 1, \dots, NS; \quad k = 1, \dots, NS; \quad i = 1, \dots, N; \quad -j = 0, \dots, M-1$$

Donde $c(l^*j, k^*j)$ es la covarianza entre los sitios mismos y entre los diferentes sitios con rezagos de 0 a $M-1$.

2. Se encuentran los autovalores λ_k y autovectores $E_{l,j}^k$ de la matriz de covarianzas, con $k=1, \dots, NS*M$. El subíndice k corresponde a las componentes principales y el subíndice l a los diferentes sitios, así que $E_{l,j}^k$ es el vector de la k -ésima componente principal con componentes de 1 a M para el sitio 1, de $M+1$ a $2M$ para el sitio 2 y así sucesivamente, hasta el número total de componentes del autovector que es $NS*M$.

3. Se ordenan los autovalores λ_k en forma decreciente.

4. Se seleccionan aquellos autovalores que permitan capturar una parte importante de la varianza del proceso y que determinan el número de componentes principales a utilizar en el análisis de las series. Como los autovalores se ordenan de menor a mayor, se descartan los últimos autovalores y los últimos autovectores. Estos se eligen al graficar la varianza acumulada, y descartar los que pasan un umbral definido.

5. Cálculo de las componentes principales:

$$a_i^k = \sum_{j=1}^M X_{l,i+j} E_{l,j}^k \quad 0 \leq i \leq N-M; \quad l = 1, \dots, NS \quad (33)$$

Los a_i^k son la proyección de la información en las k autovectores (FOE) considerados significativos, obteniéndose las k componentes principales.

6. Las componentes principales son series de longitud $N-M+1$ y no de longitud N como las series originales.

Cuando se estandarizan los datos mediante el procedimiento descrito, se asume que se remueve el ciclo anual y por tanto, lo que se intenta modelar, es la variación de dicho ciclo durante el periodo de predicción. A diferentes escalas de tiempo la variación del ciclo anual se ve influenciada por fenómenos macro-climáticos a gran escala que poseen unos periodos claramente definidos y que pueden ser capturados fácilmente por el Análisis Espectral. Ciclos de 2.5 a 5 años en las FOE pueden ser asociados a los efectos del fenómeno El Niño/Oscilación del Sur (ENSO), mientras que otros de menor periodo (y por tanto más alta frecuencia) pudieran relacionarse con la Oscilación Madden-Julian (30-60 días, ó alrededor de 2 meses), y otros fenómenos de escala de tiempo intra-anual. Por tanto lo que se desea modelar es la dinámica derivada de aquellos osciladores diferentes del ciclo anual que más aportan a la dinámica no lineal de los caudales, y que están presentes en la serie estandarizada.

6.1.9 Reconstrucción de las componentes principales

Las series de las componentes principales son de longitud $N-M+1$, para obtener las series de longitud N se reconstruye la señal de la siguiente forma:

$$H_A(Y)_l = \sum_{i=0}^{N-M} \sum_{j=1}^M (y_{l,i+j} - \sum_{k \in A} a_i^k E_{l,j}^k)^2, \quad l = 1, \dots, NS; \quad i = 1, \dots, N \quad (34)$$

La solución anterior expresa la diferencia al cuadrado de la señal filtrada y de las componentes principales a_i^k proyectadas en los autovectores $E_{l,j}^k$ siendo A el conjunto de las componentes principales seleccionadas. Resolviendo la ecuación anterior por mínimos cuadrados, se tiene:

$$(R_A X)_{l,i,k} = \frac{1}{M} \sum_{j=1}^M a_{i-j}^k E_{l,j}^k, \quad M \leq i \leq N - M + 1; \quad k \in A \quad (35)$$

$$(R_A X)_{l,i,k} = \frac{1}{i} \sum_{j=1}^i a_{i-j}^k E_{l,j}^k, \quad 1 \leq i \leq M - 1; \quad k \in A \quad (36)$$

$$(R_A X)_{l,i,j} = \frac{1}{N - i + 1} \sum_{j=1-N+M}^M a_{i-j}^k E_{l,j}^k, \quad N-M \leq i \leq N; \quad k \in A \quad (37)$$

Con las ecuaciones anteriores se obtiene un número de componentes reconstruidas igual al número de componentes principales, donde cada serie tiene N datos.

6.1.10 Predicción

A cada una de las componentes reconstruidas, se construye un modelo de Regresión lineal múltiple para hacer la predicción sobre las componentes reconstruidas. Esto hace que cada modelo considere una varianza mucho menor que la varianza total de la señal.

Para obtener la señal filtrada para un sitio se suman las componentes reconstruidas mostradas en la ecuación 5.25. Esta señal filtrada ya tiene una longitud igual a N más el número de predicciones hechas.

$$x_{l,i} = \sum_{k=1}^{na} (R_A x)_{k,l,i} \quad l=1,\dots,NS; \quad i=1,\dots,N \quad (38)$$

Luego se desestandarizan las señales para cada sitio y se obtienen los caudales predichos, correspondientes a los últimos valores de la serie para cada sitio, según el número de predicciones hechas.

Se tomaron de 1 a 12 semanas de rezago de los datos de caudal, por tanto se tienen 12 autovalores, que tienen el siguiente aporte a la varianza total del proceso (figura 6.1)

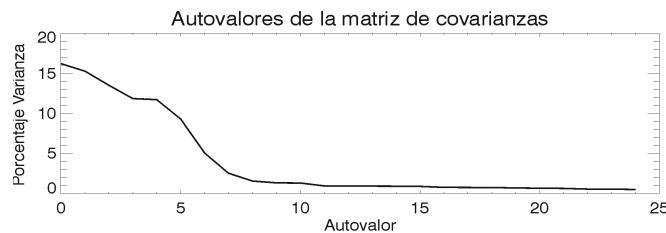


Figura 6.1 Aporte a la varianza de los autovalores de la matriz construida con 12 semanas de rezago de caudal.

Del análisis de los valores propios, se escoge utilizar los primeros 10 valores propios, que explican el 90% de la varianza del proceso. Esto significa que se trabajará sobre las 10 primeras componentes principales.

Las primeras 4 componentes principales se muestran en la figura 6. 2 Allí se puede observar que la componente principal cero, tiene un comportamiento de largo plazo del proceso, luego las otras componentes principales empiezan a tener una menor frecuencia, y relaciones con ciclos de menor escala temporal.

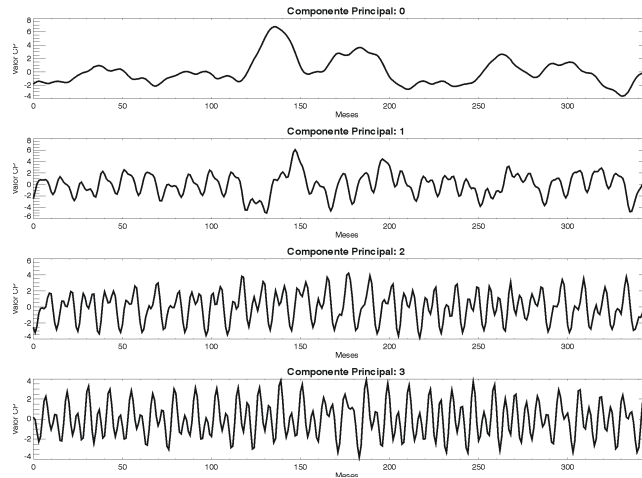


Figura 6.2 Primeras 4 componentes principales

Para poder realizar la predicción sobre las componentes principales, estas se deben reconstruir para que tengan la misma longitud de la serie original.

6.2 Redes Neuronales

Como se describió en el capítulo 4, las redes neuronales básicamente son algoritmos matemático-computacionales, diseñados con el objetivo de aprender a partir de la dinámica de uno o más eventos de la historia que se le suministre del mismo. El objetivo es aprender, de manera artificial, mediante ciertos algoritmos computacionales, acerca de la dinámica del sistema y, con dicho aprendizaje, simular bajo ciertas características el comportamiento futuro del mismo.

Las redes neuronales constan básicamente de tres diferentes tipos de capas: capas de entrada, ocultas y de salida, además dentro de cada una de éstas capas se tiene una cierta cantidad de neuronas. Las neuronas de la capa de entrada son el estímulo inicial que recibe la red, y son los patrones conocidos. Las neuronas que se encuentran en la capa oculta son las unidades de procesamiento de la red, donde se establece una función de suma ponderada de las entradas a cada una de ellas y luego a la suma resultante dentro de cada neurona se le aplica la función de transferencia que se haya elegido para dicha capa o bien neurona.

El algoritmo más comúnmente utilizado para entrenar las redes es llamado de “retropropagación”, que busca básicamente minimizar el error cuadrático medio entre una salida de datos observados y la salida obtenida por la red.

6.2.1 Principio de Funcionamiento

Para un buen funcionamiento en la red, en la etapa de entrenamiento los datos de entrada deben ser estandarizados, de lo contrario la red en muchos casos puede obtener resultados no deseados. Para el caso hidrológico es conveniente estandarizar los datos de acuerdo con la media y la desviación estándar de las semanas.

6.2.2 Funciones de transferencia

Cuando la información llega a las neuronas de la capa oculta, ésta es transformada dentro de cada neurona mediante el uso de diferentes funciones de transferencia que se muestran en la figura 6.3

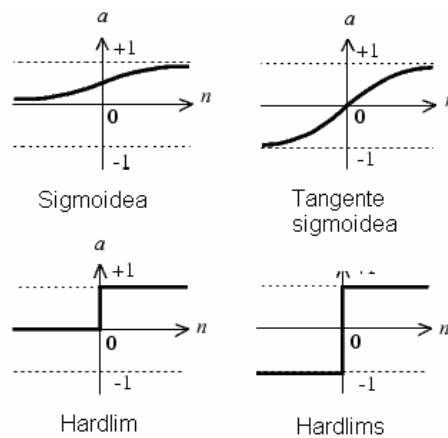


Figura 6.3 Funciones de Activación o Transferencia (Fuente MATLAB).

En hidrología no se deben usar funciones discontinuas, por lo que funciones como “Hardlim” y “Hardlims” no son capaces de capturar ciclicidades de tipo anual. La función sigmoidea y la tangente hiperbólica han sido ampliamente utilizadas para la transferencia de datos en este campo de la ciencia. Las ecuaciones que describen la función sigmoidea y tangente sigmoidea son, respectivamente:

$$f(x) = \frac{1}{(1 + e^{-x})} \quad (39)$$

$$f(x) = \frac{2}{(1 + e^{-2x})} - 1 \quad (40)$$

6.2.3 Conexión entre neuronas

Cada conexión entre la capa de entrada y cada neurona de la capa oculta tiene asociado un peso o valor de ponderación, por lo que la información neta de cada neurona es la suma de cada dato de entrada por su peso de conexión.

Así, los datos de entrada a cada neurona de la capa oculta se representan según,

$$Entrada = \sum_{i=1}^n w_i x_n \quad (41)$$

6.2.4 Función o regla de activación

Así como es necesaria una regla que combine las entradas a una neurona con los pesos de sus conexiones, también se requiere una regla que combine las entradas con el estado actual de la neurona, para producir un nuevo estado de activación. Esta función, F, produce un nuevo estado a partir del estado (a_i) que existía y la combinación de las entradas con los pesos de las conexiones (net_i). Ver figura 6.4.

Dado el estado de activación $a_i(t)$ de las neurona U_i y la entrada total que llega a ella, net_i , el estado de activación siguiente, $a_i(t+1)$ se obtiene aplicando la función de activación F:

$$a_i(t+1) = F(a_i(t), net_i) \quad (42)$$

En la mayor parte de los casos, F es la función identidad por lo que el estado de activación de una neurona en $t+1$ coincide con el Net de la misma en el tiempo t. En este caso, la salida de la neurona i (y_i) será:

$$y_i(t+1) = f(Net_i) = f\left(\sum_{j=1}^N w_{ij} y_j(t)\right) \quad (43)$$

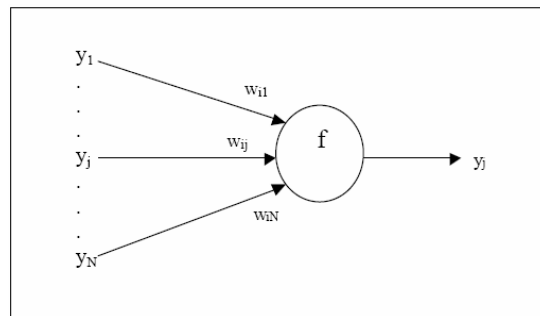


Figura 6.4 Esquema de función de activación en una neurona

Normalmente la función de activación tiene cierto desplazamiento debido a las características internas de la propia neurona. Este desplazamiento se denota por θ_i y representa el umbral de activación de la neurona i .

$$y_i(t+1) = f(Net_i - \theta_i) = f\left(\sum_{j=1}^N w_{ij}(t) - \theta_i\right) \quad (44)$$

6.2.5 Aprendizaje o entrenamiento

Consiste en ajustar los pesos de las conexiones de la red en función de la diferencia entre los valores deseados y los obtenidos en la salida; es decir, en función del error cometido. Una regla o algoritmo simple de aprendizaje por error puede ser:

$$\Delta W_{ji} = \alpha y_i (d_j - y_j) \quad (45)$$

Donde:

ΔW_{ji} : variación en el peso de la conexión entre las neuronas i y j

y_i : valor de salida de la neurona i

d_j : valor de la salida deseado para la neurona j

y_j : valor de salida obtenido de la neurona j

α : factor de aprendizaje ($0 < \alpha < 1$) que regula la velocidad del aprendizaje

Otra forma de entrenar la red es por medio de un algoritmo de retropropagación:

$$E = \frac{\sum_{i=1}^n (Q_i - Q_r)^2}{n} \quad (46)$$

Donde Q_i es el caudal obtenido por la red y Q_r es el caudal observado. El error depende del valor observado y el obtenido por la red, el observado se le entrega a la red como un dato para que ella realice el aprendizaje, a esto se le llama aprendizaje asistido, el cual es el modo de aprendizaje más usado para éste tipo de problemas.

6.2.6 Estructura de la red – Caso semanal

Las primeras redes que se exploraron para predecir caudales fueron las redes estáticas entrenadas iterando el número de capas ocultas y el número de neuronas por capa, teniendo como criterio de selección la red con menor error promedio de validación incluyendo la serie índice Madden-Julian en los arreglos de la información de entrada.

La concepción de las redes como un modelo entrada-salida, que ante un valor de entrada produce un resultado, implica que el entrenamiento haga uso de información histórica, en donde es posible discriminar qué datos corresponden a la “excitación” del modelo, y que datos corresponden a la respuesta deseada. Normalmente, las predicciones han sido calculadas usando como caudales predictores (para una ventana de predicción de cuatro semanas) los cuatro caudales que preceden al periodo de pronóstico, es decir, para predecir el periodo Q_{t+1} , Q_{t+2} , Q_{t+3} y Q_{t+4} se debe hacer uso del Q_t , Q_{t-1} , Q_{t-2} , Q_{t-3} . La figura 6.5 muestra la estructura de una red neuronal con una capa oculta.

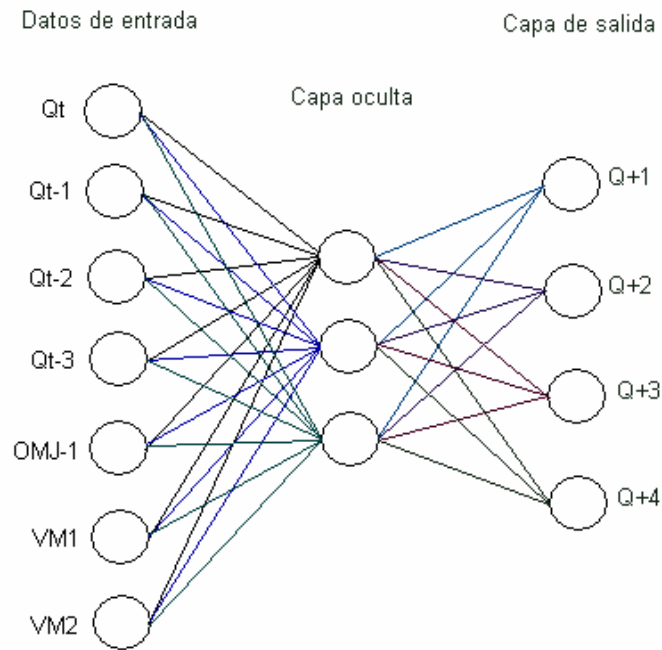


Figura 6.5 Estructura de una red neuronal con una capa oculta teniendo como datos de entrada los caudales con rezagos 1,2 y 3, serie índice OMJ y posibles variables macroclimáticas.

En el ejercicio de la validación se observará qué tanto mejoran las estimaciones de los caudales semanales en comparación con las estimaciones realizadas con el modelo ACP. Además los errores relativos y la representatividad que se da a los picos máximos y mínimos en las series de caudales. Igualmente, los resultados comparativos en el capítulo 7.

7 Resultados

En este capítulo, se presentan los resultados de la operación con el modelo ACP para la predicción de caudales semanales y la correspondiente aplicación de diferentes métricas para la evaluación de la validación. Se presenta además los resultados de la aplicación de Redes Neuronales en los mismos ríos.

7.1 Modelo ACP

7.1.1 Autovalores y Componentes principales

Uno de los principales resultados parciales en el proceso ACP es la forma como se construye la matriz de covarianza de dimensión $M \times M$ siendo M la longitud de la ventana considerados. Para nuestro objetivo donde la escala de los caudales es semanal se tomó una longitud M que varía entre 12 y 17 semanas; la matriz es cuadrada, donde en la diagonal, los elementos corresponden a uno y los demás elementos son las autocovarianzas con rezago de 0 a $M-1$. A dicha matriz de covarianzas T_x se le estimaron los valores propios λ_k y sus correspondientes vectores propios, E_k , conocidos también como Funciones Ortogonales Empíricas (FOEs), Posteriormente se proyectaron las series de datos sobre dichos vectores los cuales conforman una base ortogonal. Luego los autovalores fueron ordenados según su contribución a la varianza total del proceso seleccionando aquellos que permitan capturar la mayor parte de la varianza total del proceso. La figura 7.1 muestra los autovalores estimados de tres estaciones para el caso univariado, esto es, tomando solo la serie de caudales semanales en la descomposición espectral.

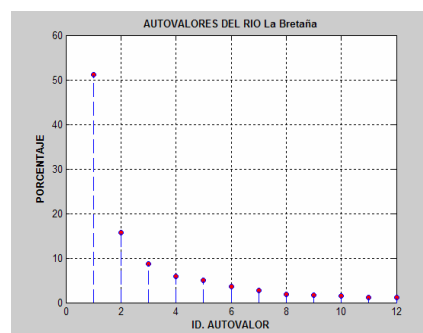
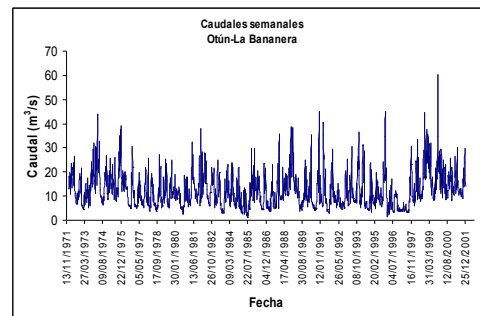
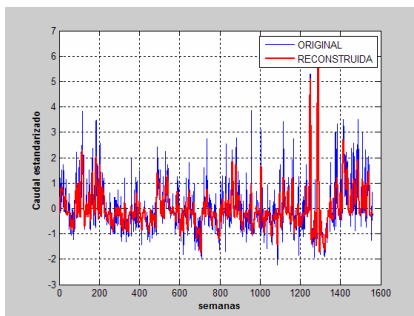


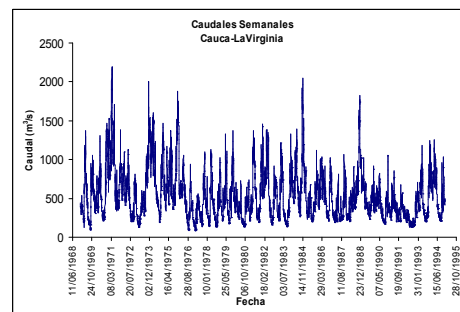
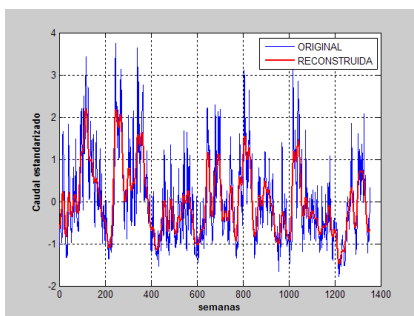
Figura 7.1 Autovalores obtenidos para una aplicación univariada del modelo ACP

Para los gráficos anteriores se tomó solo el primer autovalor que representa más del 60% de la varianza para reconstruir la serie en los ríos Cauca-La Virginia y Risaralda-Puente Negro. Para el río Otún-La Bananera, San Eugenio-La reina y Mapa-Breña se tomaron 2 autovalores que sumados representan el 63%, 66% y 65% de la serie original respectivamente.

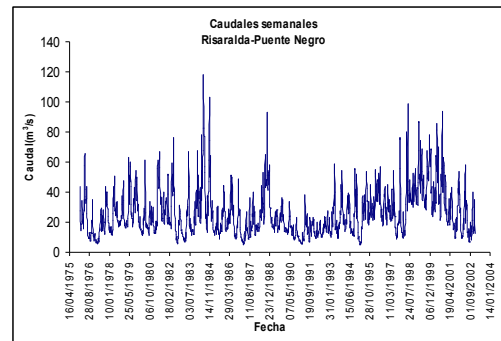
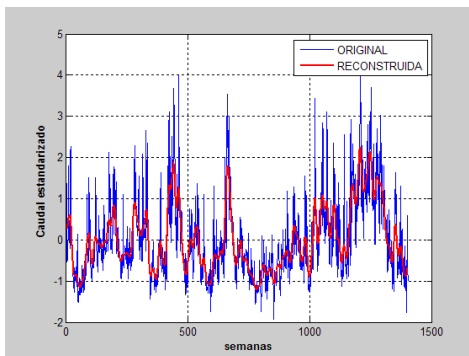
La serie estandarizada se reconstruye mediante análisis espectral singular considerando un tamaño de ventana $M=15$ (que permita capturar variabilidad intraestacional: 60-90 días variando la cantidad de Componentes Principales. En la figura 7.2 se muestra la serie estandarizada vs. la serie reconstruida (filtrada) para las estaciones Cauca-La Virginia, Otún-La Bananera, Risaralda-Puente Negro y San Eugenio-La Reina. La variación de los autovalores determina la estructura de la serie reconstruida, esto explica cómo, para determinado número de componentes se tienen diferentes series reconstruidas.



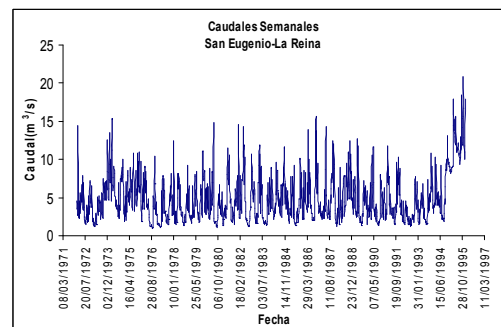
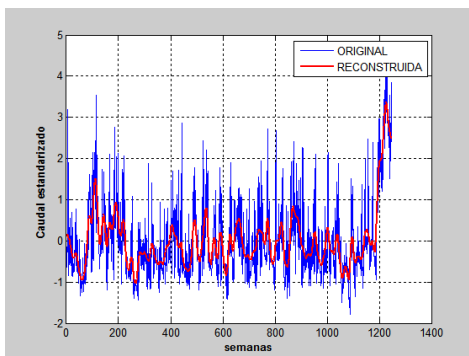
R. Otún -La Bananera



R. Cauca - La Virginia



R. Risaralda - Puente Negro

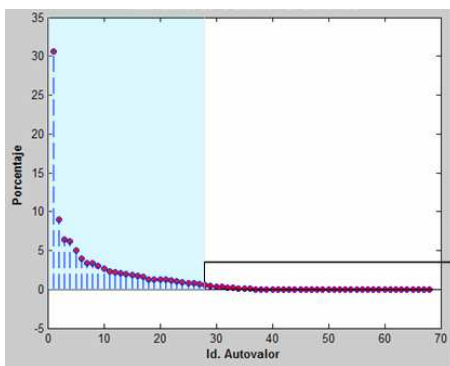


R. San Eugenio -La Reina

Figura 7.2 Serie original estandarizada vs. reconstruida con las componentes principales

Para los cinco ríos se puede observar que la serie reconstruida representa significativamente la serie original de acuerdo al número de componentes seleccionadas. Si se reduce la información a las primeras p componentes principales es posible obtener una serie que omite gran parte del ruido dado que las primeras p componentes describen la mayor parte de la varianza total acumulada conservando la esencia de la serie (Carvajal, 1994). Sin embargo hay que recordar que solo se analizaron caudales y no se consideraron series externas de variables macroclimáticas. Como los primeros auto-valores explican la mayor cantidad de varianza del proceso, estos pueden ser utilizados para reconstruir la serie, luego la señal queda dividida en dos partes, una que aporta al proceso y otra conocida como “ruido”, definida como aquella parte de la señal que no aporta a la varianza total del proceso. Tal calificación puede estimarse en términos del porcentaje de varianza

de cada autovalor y de manera gráfica la cantidad de autovalores que tienen los mayores porcentajes de varianza. Un ejemplo de la selección de autovalores se presenta en la figura 7.3 donde se tiene un caso multivariado con 68 autovalores y se seleccionan 27 autovalores para reconstruir la serie original. Esta selección del número de autovalores es de libre decisión para el modelador. De acuerdo a la distribución de los autovalores de acuerdo al porcentaje de varianza que explican se decide a criterio propio el número final de autovalores con los cuales se reconstruirá la serie. Este análisis es sensible al número de autovalores que se seleccionan para la reconstrucción de la serie original y se entiende como un aspecto de sensibilidad del modelo.



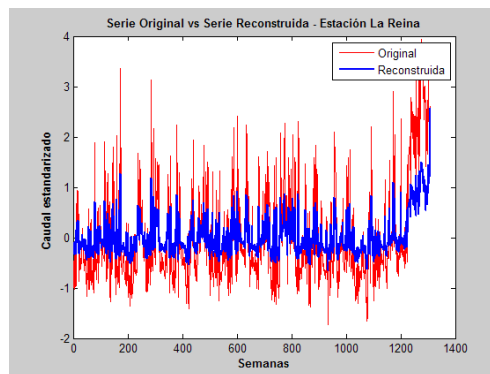
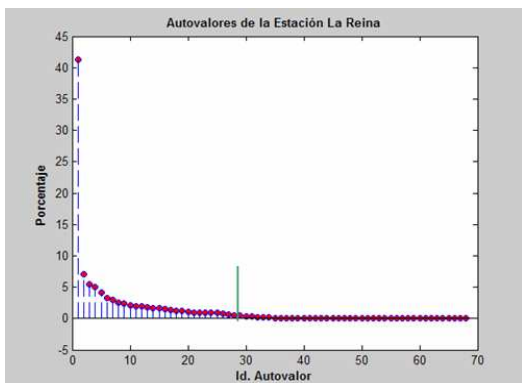
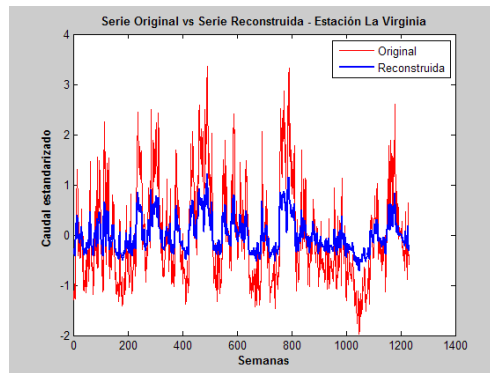
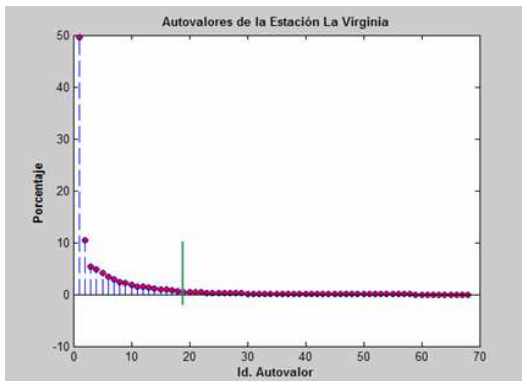
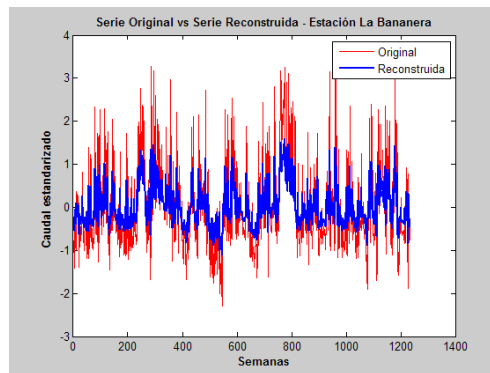
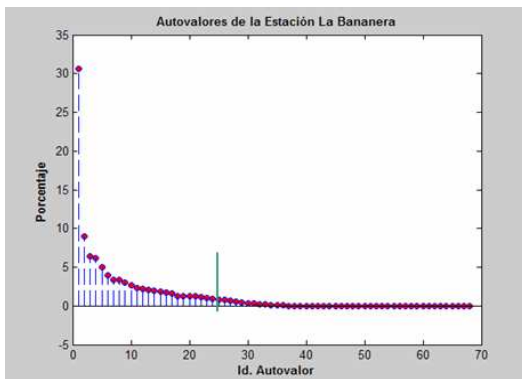
Número de autovalores que determinan el porcentaje de varianza que representarán las componentes principales.

Figura 7.3 Selección de autovalores para la reconstrucción de la serie original.

7.1.2 Autovalores y Componentes Principales – Caso Multivariado

Dentro de las variables analizadas en el modelo, se consideraron caudales rezagados y variables macroclimáticas para la descomposición espectral generando una serie a partir de las componentes principales que representen un porcentaje significativo de la varianza. De acuerdo a las correlaciones y a la asociación con la variabilidad intraestacional se evaluaron los caudales semanales rezagados una y dos semanas, la serie índice de la Oscilación Madden-Julian y la Advección de Humedad por viento zonal rezagado una semana. De igual manera al proceso posterior se involucra la estimación de los parámetros de la regresión lineal multivariada.

En la figura 7.4 se presentan los autovalores seleccionados para cada río que permitieran capturar más del 60% de la varianza total del proceso.



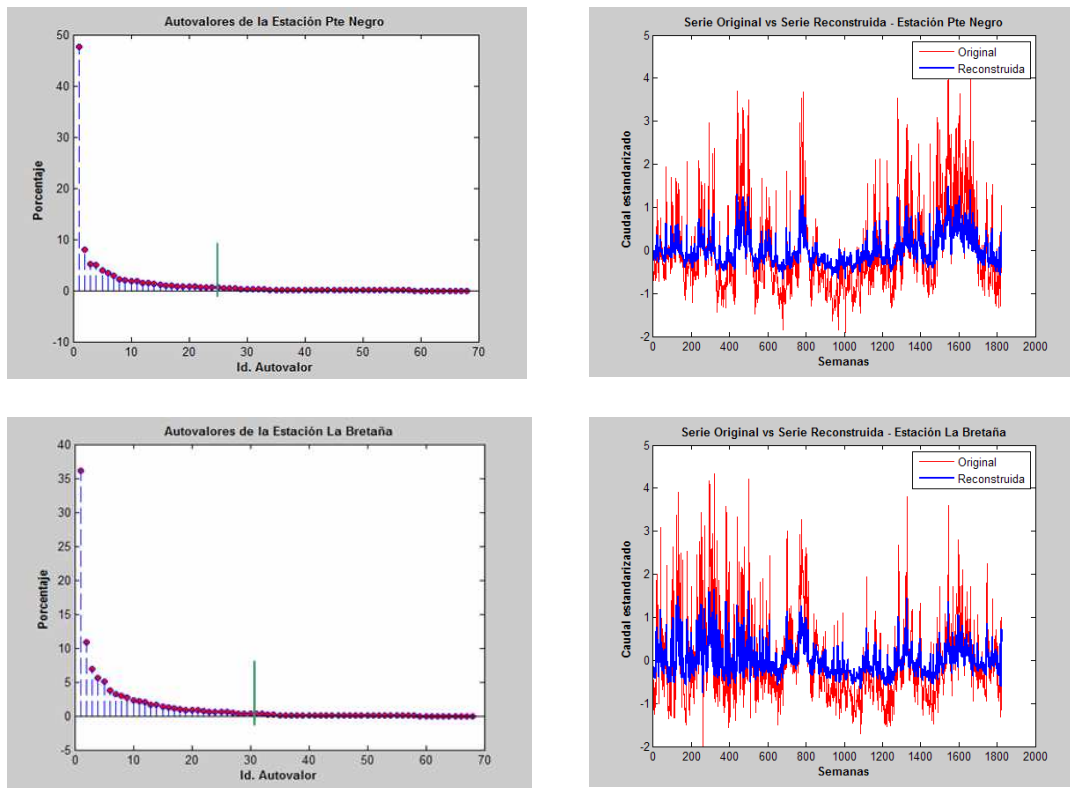


Figura 7.4 Autovalores seleccionados (hasta donde indica la línea verde) y gráfico de la serie de componentes principales reconstruida y la original.

La línea verde muestra el número de autovalores seleccionados para la reconstrucción de la serie original. Como se observa en la gráfica de la derecha de cada río, se toman aquellos valores que representen significativamente la serie original considerando que muchos de los autovalores de menor porcentaje son considerados como ruido de la señal. Sin embargo al tomar un porcentaje alto de la varianza, no se reconstruye muchos de los “picos” de la señal original que están asociados al ruido o factores de alta frecuencia. Hay que destacar que la incertidumbre y la aleatoriedad de la variable se refleja en la señal lo que se convierte en un factor de sensibilidad para el modelo.

7.1.3 Calibración

En la figura 7.5 se muestra la forma cómo se calibra el modelo, en donde cada celda representa datos semanales, primero para la calibración se toman los primeros dos tercios de los datos que se tienen en cada río, el primer año no se tiene en cuenta, se utiliza siempre como punto de partida para poder calcular los autovalores, el tamaño de la ventana M es fijado en un rango de 12 a 15 semanas,

esta se va desplazando por los datos analizados con el fin de obtener los autovalores, los cuales son necesarios para reconstruir la serie y de esta forma eliminar las fuentes de ruido, dejando solamente aquellos datos que realmente aportan información al modelo, después los datos de la serie reconstruida se utilizan para determinar los coeficientes por medio de una regresión lineal múltiple, en este punto el número de los coeficientes son determinados por el usuario.

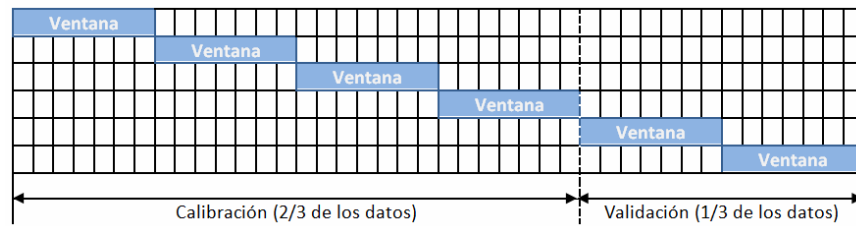


Figura 7.5 Selección del período de calibración y validación para cada una de las series de caudales de los cinco ríos.

Una vez reconstruidas las componentes principales se realizó en cada una de ellas un modelo de regresión lineal múltiple para la predicción, considerando como variables independientes las series de caudales precedentes y las variables macroclimáticas. Para cada río se involucraron diferentes variables, en dos de ellos se incluyó la lluvia.

Si t representa la semana actual, entonces predecir dicha semana utiliza información de las tres semanas anteriores, es decir $t-1$, $t-2$ y $t-3$. Por tanto el modelo multivariado lineal determinístico queda de la forma:

$$y = C_1x_1 + C_2x_2 + \dots + C_nx_n \quad (47)$$

Donde y es la variable dependiente, el conjunto (x_1, x_2, \dots, x_n) contiene las variables independientes y el conjunto (C_1, C_2, \dots, C_n) contiene los parámetros de ajuste que deben estimarse. Los modelos ajustados son periódicos, es decir, para cada semana. Los parámetros C_1, C_2, \dots, C_n son obtenidos mediante regresión lineal múltiple.

En total se organiza la información con 73 valores de caudal para un año, es decir un caudal para cada semana en péntadas ya que la serie de la Oscilación de Madden-Julian originalmente está en péntadas (datos cada 5 días). Para cada semana se tiene un coeficiente que se empleará en la validación de acuerdo a la semana que se desea predecir en función de las anteriores. La nueva serie predicha se desestandariza (se multiplica por la desviación y se le suma la media) para obtener el caudal predicho. El Anexo 1 presenta los coeficientes semanales para río estudiado.

7.1.4 Validación

Para el ejercicio de la validación se utilizó el último un tercio de la serie histórica de cada uno de los cinco ríos y se comparó con la serie estimada por el modelo en el mismo rango temporal. En la validación se tiene en cuenta el número de coeficientes, entonces dependiendo de estos se toman las n semanas anteriores para poder obtener el valor de la semana que se está prediciendo, como se muestra en la figura 7.6

		S_5	S_4	S_3	S_2	S_1	S_A	S_1	S_2	S_3	S_4	S_5	S_6		
		C_5	C_4	C_3	C_2	C_1	D_A								

Figura 7.6 Estimación de los coeficientes del modelo a partir de la semana actual

La semana actual S_A es la semana a la que se quiere hallar la predicción con base en la cantidad de coeficientes determinados, para esto cada coeficiente es evaluado con la misma cantidad de semanas precedentes, es decir el coeficiente C_1 se evalúa sobre la primera semana anterior S_1 , el segundo coeficiente C_2 con la semana S_2 , y así sucesivamente después todos los resultados son procesados para obtener el valor predicho y calcular el error de la predicción con respecto a los datos reales. Debido a que para cada semana en particular se ajusta un modelo diferente no se puede dar un error consolidado de la etapa de validación, es por esto que solo se presentan los resultados por semana en la etapa de validación del modelo.

7.1.4.1 Error relativo

El desempeño del modelo es evaluado para todas las series mediante el error relativo el cual está definido como se muestra a continuación:

$$\text{Error relativo (\%)} = \frac{Q_{\text{observado}} - Q_{\text{predicho}}}{Q_{\text{observado}}} * 100 \tag{48}$$

Si se presenta un valor negativo, indica que la predicción sobre-estimó el caudal real registrado y en caso contrario, es decir, cuando se encuentra un valor positivo se presenta subestimación del caudal.

A continuación se presenta la serie de caudales observados y la serie de caudales predichos para los cinco ríos teniendo como origen la semana indicada (por ejemplo la semana 15, del total de semanas del año). Es decir, al iniciar validaciones en la semana 15, como se comporta el modelo en promedio teniendo como inicio la semana 15.

Validación Río Otún

Para las validaciones del río Otún se obtienen errores con un promedio del 22%, el error mínimo y máximo de 13.5% y 43.5% respectivamente. La validación corresponde a las semanas comprendidas entre 1996-2002. Se puede observar que para el período de validación hay un buen desempeño del modelo en casi su totalidad, los altos valores en los caudales registrados en las semanas 65 y 90 y otros de menor valor entre las semanas 250 y 320 no fueron predichos por el modelo, esto se debe posiblemente a que la dinámica del río Otún es muy variable entre épocas de verano y lluvia, es decir, en verano se dan caudales entre los $7 \text{ m}^3/\text{s}$ y $10 \text{ m}^3/\text{s}$. (ver figura 7.7) En épocas de invierno alcanzan valores promedio de $20 \text{ m}^3/\text{s}$ y máximos puntuales cercanos a $50 \text{ m}^3/\text{s}$, esto es, caudales superiores al ciento por ciento de los valores promedio registrados en verano.

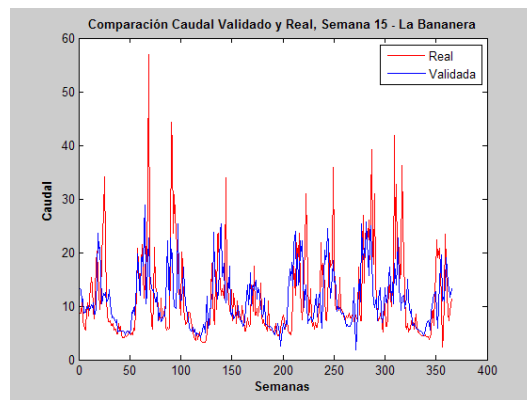


Figura 7.7 Comparación entre los caudales históricos y los caudales estimados del río Otún iniciando el pronóstico en la semana 15 y usando una ventana de pronóstico de 17 semanas.

Validación Río Cauca

Para las validaciones del río Cauca se obtienen errores con un promedio del 24%, el error mínimo y máximo son 11.5% y 41.5% respectivamente. Es importante mencionar que la magnitud de los caudales del río Cauca es alta, del orden de $500 \text{ m}^3/\text{s}$ en promedio, donde la respuesta a muchas de las dinámicas climáticas no se refleja inmediatamente en los caudales. Aun así se observa que la capacidad de predicción es aceptable al considerar los índices de Advección meridional de humedad y de la MJO propuestos dado que representa tanto las magnitudes de los caudales bajos como los altos. Sin embargo no logra reproducir algunos de los eventos extremos (máximos y mínimos). Ver figura 7.8.

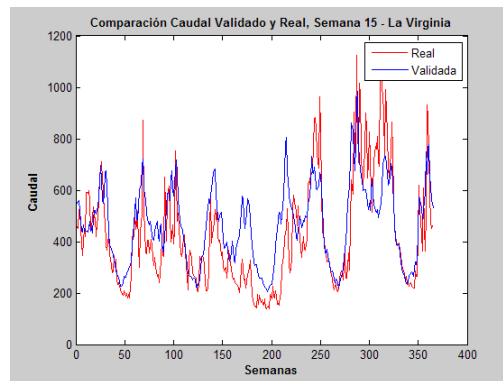


Figura 7.8 Comparación entre los caudales históricos y los caudales estimados del río Cauca iniciando el pronóstico en la semana 15 y usando una ventana de pronóstico de 17 semanas.

Validación Río Risaralda

Para las validaciones del río Risaralda se obtiene errores con un promedio del 26%, el error mínimo y máximo son 8.5% y 44.5%. Dentro de la serie correspondiente a la validación se dieron varios eventos extremos (caudales máximos en este caso) que fueron imposibles de predecir. Una vez más surge la inquietud de la posible incidencia de efectos locales y aleatorios como factores que alteran la dinámica hídrica de la cuenca. En la parte central del período de validación (entre las semanas 200 y 380) el modelo subestimó los caudales semanales mientras que en las semanas 0 y 100 y entre las semanas 380 y 500 dieron relativamente superiores al valor real. Ver figura 7.9.

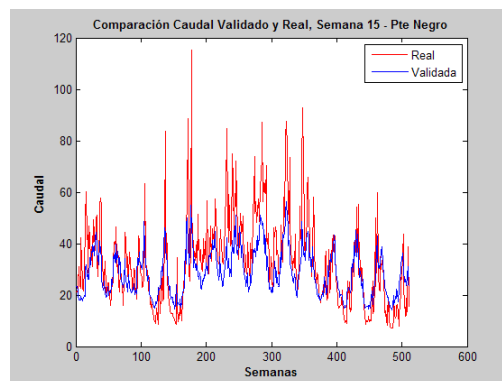


Figura 7.9 Comparación entre los caudales históricos y los caudales estimados del río Risaralda iniciando el pronóstico en la semana 15, 30 y 45 y usando una ventana de pronóstico de 17 semanas

Validación Río San Eugenio

Los caudales semanales asociados al río San Eugenio oscilan entre $5 \text{ m}^3/\text{s}$ y $15 \text{ m}^3/\text{s}$. Para las validaciones se obtuvieron errores con un promedio del 31.2%, el error mínimo y máximo fueron del 12.5% y 38.7% respectivamente. Este río presenta registros de caudales bajos (con respecto a los demás analizados en este trabajo) los cuales en épocas de lluvias tienen un orden de magnitud superior al ciento por ciento de la magnitud promedio que se da en épocas de verano.

Se puede observar que para las primeras 250 semanas del período de validación se tienen estimaciones cercanas a las reales, sin embargo a partir de la semana 282 hay un cambio en la media de la serie lo que se entiende como una alteración en los niveles medios haciendo que esta serie no estacionaria tenga variación estacional. A partir de la semana 300 se observa la no estacionariedad de la serie de caudales semanales. Ver figura 7.10.

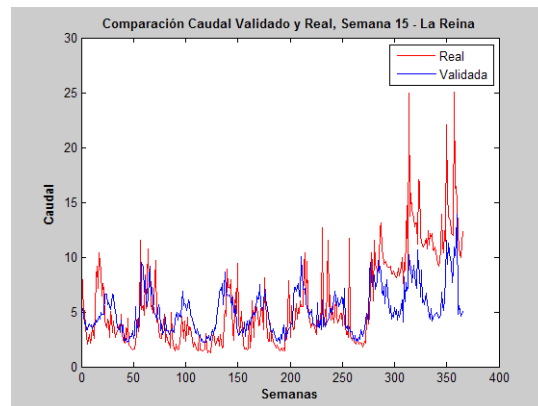


Figura 7.10 Comparación entre los caudales históricos y los caudales estimados del río San Eugenio iniciando el pronóstico en la semana 15 y usando una ventana de pronóstico de 17 semanas

Validación Río Mapa

Para las validaciones del río Mapa se obtiene errores con un promedio del 19%, el error mínimo y máximo son 7.7% y 33.9%. Puede concluirse que la predicción para este río es la que mejor resultado presentó en comparación con los otros ríos predichos.

La inclusión de variables que consideran la oscilación de Madden-Julian y advección conlleva al mejoramiento de la capacidad predictiva del modelo para este río. Se observa que como ocurrió en anteriores series, hay presencia de eventos extremos que son difíciles de predecir. Los eventos medios fueron representados significativamente por el modelo logrando un buen desempeño de validación para el río Mapa. Ver figura 7.11.

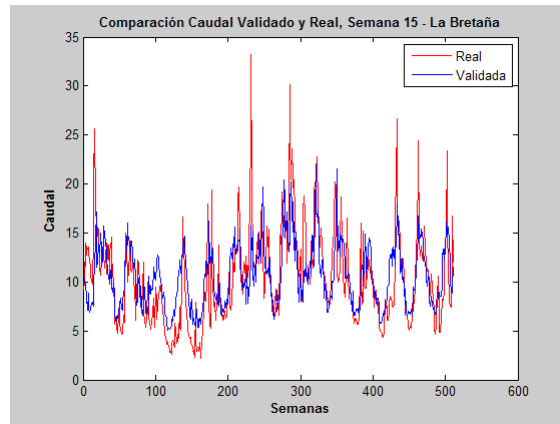
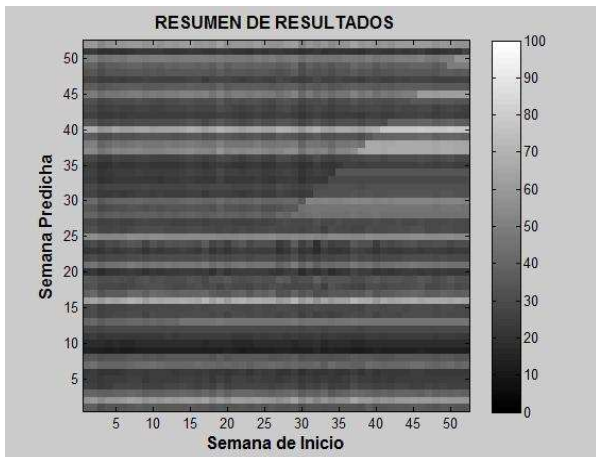
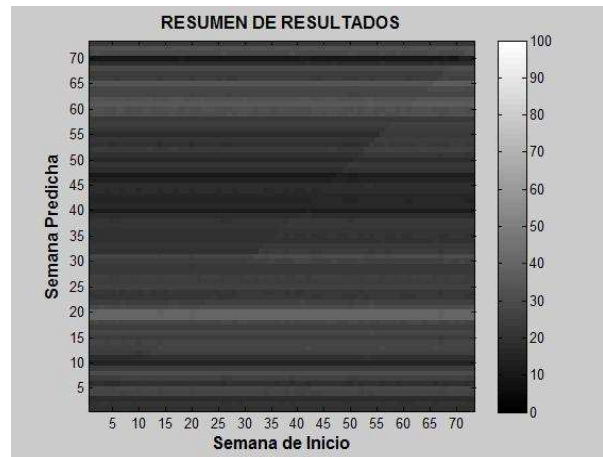


Figura 7.11 Comparación entre los caudales históricos y los caudales estimados del río Mapa iniciando el pronóstico en la semana 15 y usando una ventana de pronóstico de 17 semanas.

En la figura 7.11 se presenta el arreglo matricial de los errores obtenidos en el período de validación del modelo ACP para cada río. En el eje de las abscisas se encuentra la semana en el cual se inician las predicciones, y en el eje de las ordenadas se las semanas predichas. Zonas más oscuras representan menores errores y mayor capacidad de predicción.



Río Otún



Río Cauca

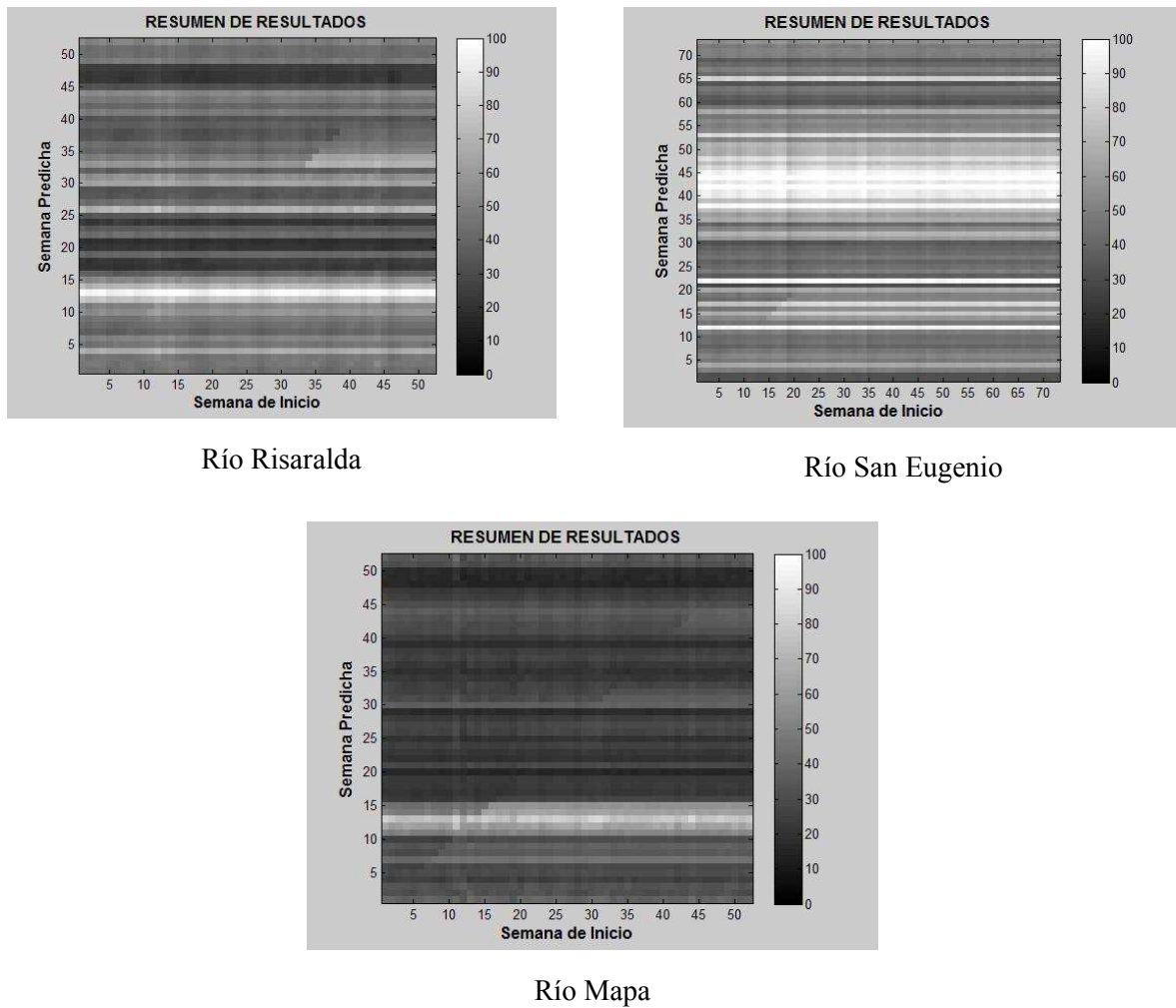


Figura 7.12 Errores medios de validación. El eje horizontal representa la semana de inicio del pronóstico y el eje vertical la semana pronosticada.

Los resultados mostrados en la figura 7.12 corresponden a la evaluación del modelo ACP. Se puede observar el panorama para cada una de las semanas dentro del período de validación. De todos los ríos, el río San Eugenio fue el que tuvo un mayor error de validación, llegando a obtener en algunas semanas errores cercanos al 90%. Caso contrario ocurrió en el río Cauca donde se obtuvieron errores muy promisorios ya que en promedio no superaron el 18%. Los errores obtenidos Los cambios en los ríos Risaralda, Mapa y Otún no fueron significativos y conservaron en promedio el mismo orden de la evaluación sin algoritmos genéticos.

7.1.4.2 Acierto por Terciles

El criterio evalúa la calidad de las predicciones considerando el acierto por terciles de la distribución de probabilidades de la serie de caudal respectiva en el período de validación. Este criterio consiste en calcular los terciles de la distribución de los caudales para la semana en que se realiza la predicción y se desea evaluar, asignando una distribución de probabilidad por el método de “Weibull”, determinando el tercil que le corresponde al caudal predicho y al caudal observado, luego se asigna un índice que valora el acierto dentro de cada tercil.

El procedimiento en detalle consiste en tomar todos los caudales históricos semanales del período de validación, es decir el último tercio del total de la serie de cada río y se ordenan de forma ascendente, asignándoles una posición. Se hace la asignación de probabilidades y se determinan los valores de caudal para la probabilidad de 0.33 y 0.66. Entonces todo caudal menor a la probabilidad de 0.33 se encuentra en el primer tercil, luego de la misma manera los caudales entre 0.33 y 0.66 están en el segundo tercil y caudales mayores a la probabilidad de 0.66 en el tercer tercil.

Los índices son los siguientes:

- 2: Si el caudal observado y el caudal predicho están en el mismo tercil.
- 1: Si el caudal observado y el caudal predicho están en terciles contiguos.
- 0: los demás casos.

Por lo tanto, a mayor error, menor índice.

El índice fue calculado para las predicciones obtenidas en los cinco ríos. Para la evaluación general del desempeño del modelo, se calculó el acierto total como la suma de los aciertos de cada índice para cada semana del periodo de validación. La tabla 7.1 muestra que para el período de validación del río Otún se tiene un total de 366 valores de caudal semanal, de los cuales 154 fueron caudales cuyo índice fue 2 (el caudal observado y el caudal predicho estuvieron en el mismo tercil), 135 caudales con índice 1 (el caudal observado y el predicho estuvieron en terciles contiguos) y 77 en los demás casos. Para el río Mapa se tuvieron 171 caudales con índice 2, 217 valores de caudal con índice 1 y 123 valores con índice 0. De igual forma se leen los aciertos para los demás ríos.

Tabla 7.1 Número de aciertos por terciles para cada uno de los ríos.

Río	Número de aciertos con índice 2	Número de aciertos con índice 1	Número de aciertos con índice 0	Total de caudales semanales período de validación
Otún La Bananera	154	135	77	366
Mapa-La Bretaña	171	217	123	511
San Eugenio-La Reina	105	211	49	365
Cauca-La Virginia	223	95	47	365
Risaralda-Pte Negro	76	238	197	511

Los resultados muestran que en los ríos Otún y Cauca se obtuvieron aciertos dentro del tercil correspondiente al caudal observado. Los ríos Mapa, Risaralda y San Eugenio tuvieron aciertos por terciles con caudales observados y predichos en terciles contiguos. Para los 5 ríos se presentaron menores casos en lo que los caudales predichos estuvieran en el tercil extremo al caudal observado.

Comparativamente se presenta el porcentaje de caudales con acierto por terciles para cada río en la tabla 7.2

Tabla 7.2 Porcentaje de caudales semanales con aciertos por terciles para cada índice

Río	Índice 2	Índice 1	Índice 0
Otún La Bananera	42,1	36,9	21,0
Mapa-La Bretaña	33,5	42,5	24,1
San Eugenio-La Reina	28,8	57,8	13,4
Cauca-La Virginia	61,1	26,0	12,9
Risaralda-Pte Negro	14,9	46,6	38,6

De la totalidad de las semanas que comprenden la validación de acuerdo a la extensión de cada río, se estimó el porcentaje de caudales semanales predichos con aciertos para cada índice. La calificación cuyo índice fue 2, obtuvo el mayor porcentaje en los ríos Otún y Cauca con un 42,1% y 61,1%. Es decir que el 42,1% de los caudales del período de validación del río Otún estuvieron dentro del tercil del caudal observado. Igual para el río Cauca con el 61%. En el río Mapa se dio un mayor porcentaje para el índice 1 con un 42,5%. De igual forma en el río San Eugenio se dio un mayor porcentaje para el índice 1 con un 57,8%. El río Risaralda fue donde se dio el menor porcentaje para el índice 2, con un escaso 14,9% frente a un 46,6% de índice 1 y un porcentaje de

38,6% de índice 0. Es decir que el río Risaralda no tuvo buenas predicciones bajo la aplicación del modelo.

7.2 Redes Neuronales

Para hacer un ejercicio comparativo y de análisis del primer modelo, se estimaron los caudales semanales para el mismo período de validación del primer modelo, por medio de Redes Neuronales. Cada una de las redes calibradas fue usada para predecir las semanas de este período. Los patrones de entrada y los valores de salida son conocidos en la etapa de validación. Se trata de cuantificar la capacidad de predicción de la red neuronal mediante la diferencia entre los caudales medios predichos y los caudales históricos observados, para lo cual se consideró la ecuación del error:

$$Error_{i,j} = \frac{\sqrt{\frac{\sum_{i=1}^k (Q_{observado\ j,i} - Q_{predicho\ j,i})^2}{t}}}{Q_{med\ j,i}} \quad (49)$$

donde k es el número de años de validación y Q_{med} es el valor promedio de la semana predicha en el período de validación.

Bajo el procedimiento descrito por Redes Neuronales los resultados no muestran mejoras, los errores mínimos, medios y máximos se mantienen del mismo orden al modelo inicial en todos los ríos. A partir de estos resultados se puede pensar que es posible revisar el caso de una red con una nueva arquitectura en donde existan más capas ocultas y más neuronas, sin embargo no se espera que los resultados mejoren mucho. Este procedimiento requiere de varios entrenamientos y diferentes estructuras de redes que varíen de acuerdo a los parámetros iniciales, a la cantidad de capas ocultas y al número de neuronas de contenga cada capa oculta. La figura 7.13 se muestran los errores de validación para los cinco ríos. En cada uno se observa la semana de inicio en el eje de las abscisas y semana estimada en el eje de las ordenadas.

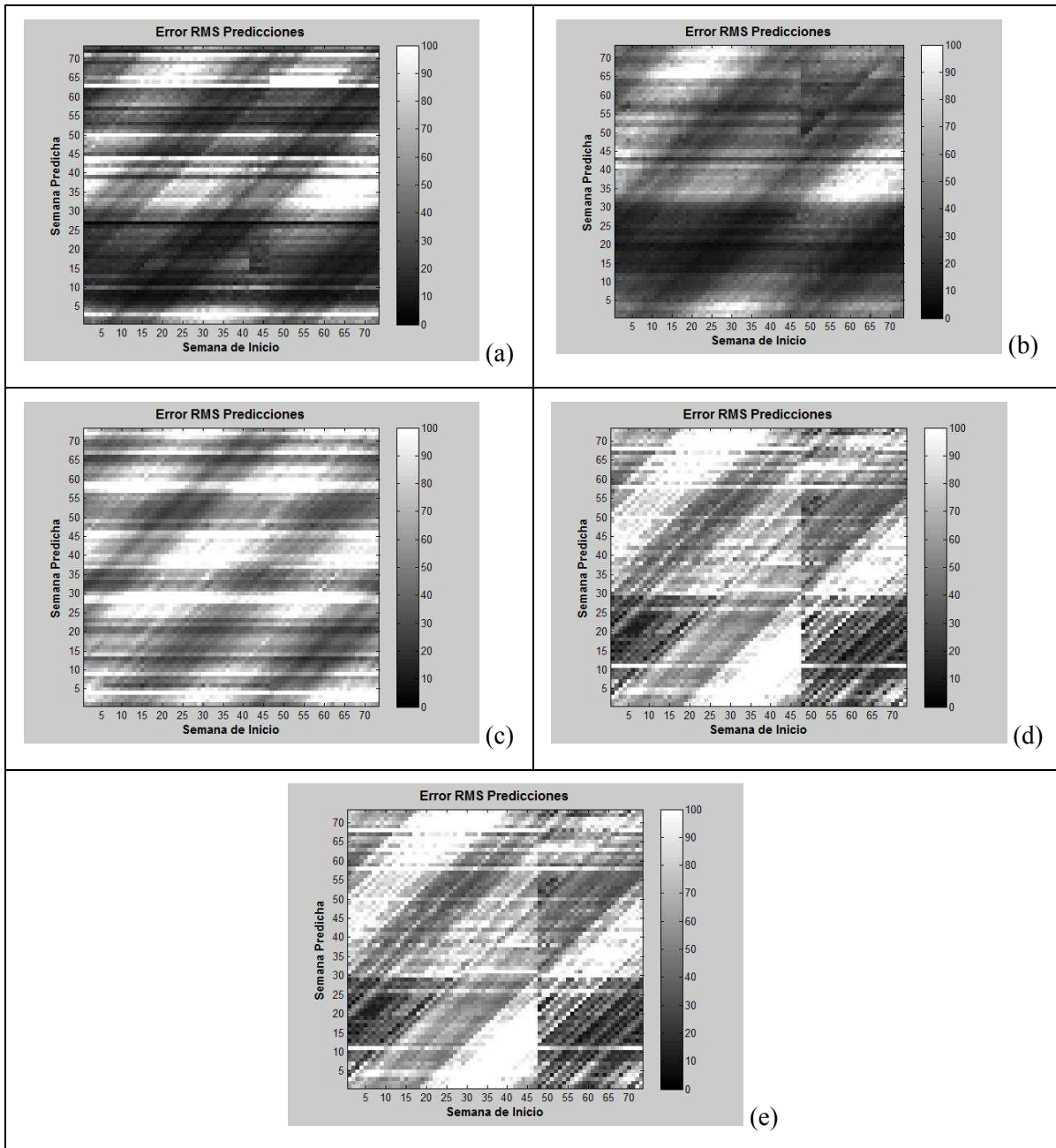


Figura 7.13 Errores de predicción en el período de validación por Redes Neuronales para los ríos Otún (a), Cauca (b), Risaralda (c), San Eugenio (d) y Mapa (e)

En este ejercicio se consideraron redes de una y dos capas ocultas, con una cantidad de neuronas en cada una que varía entre 1 y 10 a un paso de 2, y con una capa de salida que cuenta con 4 neuronas (semanas adelante a la inicial). El número de neuronas en la capa de entrada varía según la

información hidrológica (caudales rezagados una y dos semanas) y la información macroclimática que para este caso fue el índice de la Oscilación Madden-Julian con rezago uno. Para cada río, entonces, se entrenaron diversas redes, producto de las combinaciones de las diferentes estructuras y del uso de 2 combinaciones distintas de índices en los arreglos de entrada. Los algoritmos de entrenamiento y los períodos de validación se conservan iguales que en todos los casos anteriores. De igual forma, el criterio de selección de la mejor red es el mismo el mínimo error promedio de validación.

7.3 Evaluación comparativa entre los modelos ACP y RN

Los resultados estimados por el modelo inicial ACP son comparados con los estimados por Redes Neuronales se resumen en la tabla 7.3. En esta evaluación se miden donde hubo mejor desempeño de cada modelo en los cinco ríos. Cabe mencionar que esta evaluación solo considera las predicciones del período de validación y se hace por medio del error relativo promedio para cada río.

Tabla 7.3 Error promedio y diferencia de errores de los modelos ACP y RN

Error promedio en porcentaje (%)			
Río	ACP	RN	Diferencia de errores
Otún	25.2	32	-6.8
Cauca	23.4	29.9	-6.5
Risaralda	26.8	41.7	-14.9
San Eugenio	32.2	38.3	-6.1
Mapa	22.6	33.1	-10.5

Los errores obtenidos reflejan diferencias importantes que permiten evaluar el desempeño de dos modelos aplicados a los mismos ríos y considerando las misma información de entrada y los mismos rezagos de las variables macroclimáticas.

En todos los ríos el modelo ACP tuvo un mejor desempeño para predecir los caudales en comparación con el modelo de Redes Neuronales. Para el río Otún el modelo ACP tuvo un error promedio de 25.2% comparado con el 32% de las Redes Neuronales. El río Risaralda es el que presenta mayor diferencia (14.9 %) en cuanto a los errores de validación, mientras que el río Cauca presenta menor diferencia entre los errores (6.5%) pero ambos con errores mayores a 30%.

Parte de la diferencia entre los resultados se debe a que la metodología de ambos modelos es totalmente distinta y en las redes neuronales existen diversas estructuras de redes que varían desde el número de neuronas de entrada, número de neuronas de salida y número de capas ocultas, al igual

que el total de neuronas en cada una y las variaciones entre sus pasos. En este caso solo se consideraron redes de una capa oculta, con una cantidad de neuronas que varía entre uno y diez a un paso de dos.

Se concluye que el modelo ACP tuvo mejor desempeño que las Redes Neuronales para los cinco ríos. Es necesario resaltar que para ambos modelos se tuvieron en cuenta series de información climática que corresponden a los fenómenos asociados a la variabilidad intraestacional y que la caracterización de la dinámica asociada a ellos requiere de mayor análisis y detalle para estudios posteriores.

8 Conclusiones

Desde las generalidades del clima de Colombia se pudieron identificar diversas dinámicas atmosféricas que ayudan a entender la variabilidad y evolución intraestacional en Colombia. Recientemente la NOAA pronosticó la transición de las condiciones de La Niña 2008-2009. Las anomalías en las temperaturas negativas en la superficie del mar se debilitaron a través del Océano Pacífico y anomalías positivas se desarrollaron en áreas del Pacífico este. Para el mismo período se tuvo que la oscilación Madden-Julian influenció la circulación atmosférica a través de los trópicos globales, y contribuyó a la fluctuación periódica en la fuerza de los vientos del este en los niveles bajos y los vientos del oeste en los niveles altos a través del Océano Pacífico ecuatorial. Paralelamente, estas anomalías oceánicas y atmosféricas son consistentes con la transición de condiciones ENSO-neutral del segundo semestre del 2009.

En Colombia, prevaleció tiempo lluvioso entre 2008 y 2009 en la mayor parte de las regiones Andina, Pacífica y Amazonía, lo cual se reflejó en lluvias excesivas debido a la Advección de humedad desde el Pacífico oriental y desde la cuenca Amazónica. Este aspecto permitió identificar la dinámica que actualmente se da en el trópico; los índices y estimaciones que actualmente se hacen si corresponden a una realidad climática actual, por lo tanto es necesario considerarlo en estudios posteriores de predicción y diagnóstico.

De forma similar se estimaron correlaciones entre caudales con diferentes rezagos y se obtuvieron valores significativos (0.7-0.9) de la semana actual con respecto a la semana anterior. Estos valores nos permitieron definir el caudal semanal actual como la variable hidrológica principal de entrada al modelo de predicción.

Partiendo de las correlaciones más significativas entre los caudales semanales y las variables macroclimáticas, se evaluó el modelo con la información climática de la serie Madden-Julian, Advección de humedad y caudales rezagados una y dos semanas. Se obtuvieron coeficientes que estimaron caudales semanales cuyos errores obtenidos en el período de validación mostraron una de tantas soluciones dado que estas soluciones no son las óptimas. Existen otras que seguramente estimaran caudales con menores errores, considerando que existe la posibilidad de incrementar la información de entrada al modelo, otras variables macroclimáticas o combinación de ellas con variaciones de rezagos entre la información hidrológica y la información climática.

En la aplicación del modelo ACP, se seleccionó un determinado número de autovalores y se reconstruyó una nueva serie con un porcentaje significativo de la varianza, comparado posteriormente la serie observada con la reconstruida. En esta comparación es importante resaltar que los picos máximos y mínimos de la serie observada que son difíciles de reconstruir se debieron a que existen ciclos de mayor frecuencia dentro del rango semanal que se salen de los alcances de

este estudio. El ciclo diario de la temperatura, el ciclo diario de la lluvia son algunos de los factores que inciden en la representatividad de estas señales y están dentro de una escala temporal menor a la semanal.

Comparativamente, se consideraron series de lluvia semanal en dos de los cinco ríos. La validación mejoró en la significancia de los picos máximos de la serie caudal mejorando el error en la validación. Los resultados dejaron ver la importancia de tener mediciones de la lluvia en diferentes escalas como una variable hidrológica influyente en la predicción de caudales.

El error relativo promedio en el período de validación permitió evaluar la capacidad de predicción del modelo ACP. Los resultados fueron variables entre cada uno de los ríos, teniendo en cuenta que se escogieron ríos con grandes caudales, como ríos de menor caudal. Para el río Otún, y Risaralda se obtuvieron errores entre el 25% y 30% en promedio. El río Cauca y Mapa tuvieron errores cercanos al 23% siendo los más aceptables. El desempeño del modelo ACP en el río San Eugenio fue el más bajo de todos con un error promedio superior al 30%, debiéndose en parte a que los caudales medios de la serie histórica son bajos (del orden de $5.0 \text{ m}^3/\text{s}$). Los ríos con valores bajos tienden a ser sensibles a eventos máximos locales cuya respuesta es inmediata lo que no se permite alcanzar a visualizar en la escala intraestacional.

Los resultados obtenidos se pueden considerar adecuados considerando que se desarrolló un modelo semanal con Análisis Espectral que no consideró muchos de los fenómenos intra-estacionales como las ondas del este y series de precipitación del CMAP. Por lo tanto surgen preguntas sobre como optimizar la capacidad de predicción del modelo y que mejoras se pueden hacer. Esto involucra aspectos como la caracterización adecuada y rigurosa de cada uno de los fenómenos asociados a la variabilidad intraestacional en Colombia.

Una de las métricas usadas para evaluar el modelo ACP fue el acierto por terciles, donde se presentó buena capacidad de predicción para el río Cauca, siguiéndole el río Otún. Esta metodología no mide el valor exacto ni lo compara con el real, sino que lo asocia a un tercil. Este tipo de metodología permitió identificar si las métricas de error relativo son consecuentes con el acierto por terciles. En particular hubo consecuencia con el río Cauca y Otún. De forma similar lo menores aciertos por terciles se dieron para los ríos de mayor error relativo.

El objetivo principal de este trabajo se centró en el desarrollo de un modelo de predicción de caudales semanales asociado con la variabilidad intra-estacional de la climatología colombiana. Los resultados obtenidos evidencian la necesidad de mejorar las predicciones considerando factores tanto temporales como espaciales en el modelo. Factores como la misma variabilidad temporal, espacial, la topografía y la ubicación del país y el hecho de estar rodeada por los océanos Atlántico y Pacífico y la cuenca Amazónica, imponen nuevas interacciones que se traducen en una alta variabilidad climática espacial y temporal además del factor de limitaciones de información

disponible, que inciden notoriamente en el desarrollo y evaluación de este tipo de modelos. También se encontró como la consideración de la información macroclimática mejora el desempeño del modelo, sin embargo, existen más factores que incurren en esta variabilidad y por ende es necesario estudiar más profundamente la interacción de este fenómeno con la climatología colombiana.

Con el modelo ACP se hicieron las predicciones de caudales semanales donde se obtuvieron resultados válidos de acuerdo a las evaluaciones de los errores y a los aciertos por terciles. Se destaca la necesidad de conocer con mayor rigurosidad la dinámica climática asociada a la variabilidad intraestacional. Para estudios futuros se deberá mejorar el entendimiento de la oscilación Madden-Julian, estimar una mejor señal que permita caracterizar con mayor claridad tanto temporal como espacialmente esta oscilación. Es un reto además considerar la incidencia que tienen las ondas tropicales en la hidroclimatología de Colombia.

Se pudo observar que los caudales predichos por ACP evidenciaron significativamente la incidencia de los fenómenos macroclimáticos y el estudio de la variabilidad climática en Colombia como resultado de la interacción de diversos factores macroclimáticos, las condiciones orográficas de Colombia y el hecho de estar rodeada por los océanos Atlántico y Pacífico y la cuenca Amazónica.

Las predicciones usando redes neuronales y referenciadas por diferentes autores hablan de mostrar buenos resultados, debido tanto a los productos numéricos como a la facilidad en el uso de la herramienta. Además, a que la capacidad de automatización de los procesos es alta, al igual que la flexibilidad en la definición de las arquitecturas y el uso de la información. Sin embargo, se requiere ser cuidadoso y consecuente con la mejor estructura de la red que permita estimar óptimos resultados. En este estudio, se realizaron estimaciones de caudales semanales con redes neuronales. La estructura de la red y el número de capas ocultas así como el número de pasos definieron los productos numéricos para los cinco ríos donde los valores obtenidos para el período de validación las validaciones no reflejaron ninguna mejoría en los errores comparados con los obtenidos con el modelo ACP. Esta conclusión nos obliga a explorar el desarrollo de las redes neuronales dada la información que esta metodología requiere y la calidad de la misma en la caracterización de los fenómenos climáticos asociados a la escala de interés que permitan optimizar las aplicaciones de las Redes Neuronales en predicción de caudales.

9 Bibliografía

- ARIAS, P. 2005. Diagnóstico y Predicción de la Variabilidad Intra-Anual de la Hidrología Colombiana. Universidad Nacional de Colombia. Sede Medellín. Facultad de Minas.
- BARRY, R., Carleton, A. Synoptic and Dynamic Climatology. Routledge Editors. United Kingdom. 2001. 619 pp.
- BOX, G.E.P. y G. Jenkins. 1976. Time Series Analysis: Forecasting and Control, Holden-Day.
- CARVAJAL, L.F., SALAZAR J.E., MESA, O.J., y POVEDA, G. 1994. Aplicación del Análisis Espectral Singular a series hidrológicas en Colombia. Memorias del XVI Congreso Latino-Americano de Hidráulica e Hidrología, IAHS, Santiago de Chile, Vol.3, 97-108,
- CARVAJAL, L.F., 1994, Acerca de la predicción no lineal en hidrología. Tesis de Maestría, Universidad Nacional de Colombia, Facultad de Minas, Medellín, Colombia. 136 pp.
- CARVAJAL Y., MARCO B. 2005 Modelos Multivariados de Predicción de Caudal Mensual Utilizando Variables Macroclimáticas: Caso de Estudio Río Cauca. Colombia, Ingeniería Y Competitividad ISSN: 0123-3033, 2005 vol:7 fasc: 1 págs: 18 – 32
- COULIBALY, P., Baldwin, C. 2005. Nonstationary hydrological time series forecasting using nonlinear dynamic methods. Journal of Hydrology 307; 164–174.
- COTTON, W. R., Anthes, R. 1989. Storm and Cloud Dynamics. Academic Press, 883 pgs.
- CUARTAS, L. 2001. Balance Atmosférico de Humedad para Colombia. Tesis Maestría Universidad Nacional de Colombia sede Medellín. Facultad de Minas.
- DIBIKE Y.B. and Solomatine D., 2001, River Flow Forecasting Using Artificial Neural Networks, Journal of Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere, Vol. 26, No.1, pp. 1-8.
- FRIEDMAN, J. H. 1991. Multivariate Adaptive Regression Splines, Stanford University. Stanford, CA.
- GEORGAKAKOS Y KRZYSZTOFOWICZ, 2001. Probabilistic & Ensemble Forecasting. Journal of hydrology. 249.1
- GHIL M., Allen M., Dettinger M., Ide M., Kondrashov D., Mann M., Robertson A., Saunders A., Tian Y. y Varadi F. y Yiou P. 2002. “Advanced Spectral Methods for Climatic Time Series”. Review of Geophysics. American Geophysical Union, Vol. 40, N° 1. pp. 1-41.

- HASSANI, H. 2007. Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*. 5. 239-257
- HASTENRATH, S. 1996. *Climate and Dynamics of the tropics*. 2da edicion. Kluwer Acad. Dordrecht Netherlands.
- HOYOS, C. 1999. Algunas aplicaciones de la transformada de Fourier y la descomposición en ondas a señales hidrológicas y sísmicas. Universidad Nacional de Colombia sede Medellín. Facultad de Minas.
- HOYOS, C. et al. 2001. Predicción no lineal de la hidrometeorología de Colombia caso de estudio: caudales del Rio Nare. Universidad Nacional de Colombia, sede Medellín. XV Seminario Nacional de Hidráulica e Hidrología, Sociedad Colombiana de Ingenieros.
- IDEAM, 2002. Características climáticas y oceánicas del Fenómeno de El Niño 1997-1998.
- KEPPENNE, Christian y Lall, U. Complex Singular Spectrum Analysis and Multivariate Adaptive Regression Splines Applied to Forecasting the Southern Oscillation.
- LUDERMIR, T Valença, M. 2001. Constructive neural networks in forecasting weekly river flow. Conference on Computational Intelligence and Multimedia, Computer Society .
- MEJÍA, J., POVEDA, G. 2005. Ambientes atmosféricos de sistemas convectivos de mesoescala sobre Colombia durante 1998 según la misión TRMM y el Reanálisis NCEP/NCAR. *Revista Ciencias de la Tierra*. Número 113-diciembre de 2005.
- MESA, O. J., Poveda, G., Carvajal, L.F y Salazar, J.E. 1994. Predicción hidrológica usando redes neuronales, *Memorias del XVI Congreso Latinoamericano de Hidráulica e Hidrología, IAHR, Santiago de Chile, Vol. 3, 385-396, 1994.*
- MÚNERA, L. 1983. *Modelos Estocásticos para las series Hidrológicas*, Universidad Nacional de Colombia, Sede Medellín.
- MAKRIDAKIS, Spyros and Winkler, Robert, L. 1983, "Averages of forecasts: Some empirical results," *Management Science*, 29, 987-996. (RLW)
- MOHAMMADI, K., Eslami, H., Kahawita, R. 2006. Parameter estimation of an ARMA model for river flow forecasting using goal programming. *Journal of Hydrology* 331, 293– 299.
- MONTOYA, G., PELKOWSKI, J., ESLAVA, J.A. 2001. Sobre los Alisios del nordeste y la existencia de una corriente en el piedemonte oriental andino. *Rev. Acad. Colomb. Cienc.* 25(96) 363-370.

- MUHAMMAD, A. 2006. Analysis and prediction of flow from local source in a river basin using a Neuro-fuzzy modeling tool. *Journal of Hydrology* 265.
- ORGANIZACIÓN METEOROLÓGICA MUNDIAL (OMM) 2003. *Servicios de Información y Predicción del Clima y Aplicaciones Agrometeorológicas para los países Andinos*. Ginebra- Suiza.
- PORPORATTO, A., Ridolfi, L. 2001. Multivariate nonlinear prediction of river flows. *Journal of Hydrology*. 248. 109-122.
- POVEDA, G., y Mesa, O.J. 1996. Las fases extremas del ENSO - El Niño y La Niña - y su influencia sobre la hidrología de Colombia. *Revista de Ingeniería Hidráulica en México*. Vol. XI, No. 1, 21-37.
- POVEDA, G., MESA, O.J, L.F CARVAJAL, C.D HOYOS, J.F. MEJÍA, L.A CUARTAS Y A. PULGARÍN. 2002. “Predicción de caudales medios mensuales en ríos colombianos usando métodos no lineales”. *Meteorología Colombiana*, 6, 101-110.
- POVEDA, G. 2004. La Hidroclimatología colombiana: Una síntesis desde la escala interdecadal hasta la escala diaria. *Revista Ciencias de la Tierra*. Número 107-junio
- POVEDA, G. El chorro del chocó y su influencia sobre la Hidroclimatología de la costa pacífica y El occidente de Colombia Posgrado en Aprovechamiento de Recursos Hidráulicos. Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia.
- POVEDA, G., y Mesa, O.J. 1993. “Metodologías de predicción de la hidrología Colombiana considerando el evento El Niño-Oscilación del Sur (ENOS)”. *Revista Atmósfera*, 17, Sociedad Colombiana de Meteorología, Bogotá.
- POVEDA, G., y Penland, C. 1994. Predicción de caudales medios en Colombia usando Modelación Lineal Inversa, *Memorias del XVI Congreso Latino-Americano de Hidráulica e Hidrología, IAHR*, Santiago de Chile, Vol. 4, 119-129, 1994.
- POVEDA, G., y O.J. Mesa. 1999. “La corriente de chorro superficial del oeste (“del CHOCO”) y otras dos corrientes de chorro atmosféricas sobre Colombia: Climatología y variabilidad durante las fases de ENSO”. *Rev. Acad. Colomb. Cien.*, 23 (89): 514-528.
- POVEDA, G., Jaramillo, A. 2000. ENSO-related variability of river discharges and soil moisture in Colombia, *Biospheric Aspects of the Hydrologic Cycle*. IGBP. 8. 3-6.
- RENDÓN, 1997 Beneficios de la predicción hidrológica en el sector eléctrico colombiano, considerando la variabilidad climática. Universidad Nacional de Colombia. Sede Medellín. Facultad de Minas.

RENDÓN, A. 2001. Influencia de tres corrientes de chorro sobre la hidrología colombiana. Universidad Nacional de Colombia sede Medellín. Facultad de Minas.

RIEHL, H. 1979. Climate and weather in the tropics. Acad. Press London. 611 pp.

RODRIGUEZ, J. Predicción de Caudales diarios y semanales mediante modelos probabilísticas. Universidad Nacional de Colombia. Sede Medellín. Facultad de Minas. 1988

SAHOO, G., Ray, C. 2006. Flow forecasting for a Hawaii stream using rating curves and neural networks. Journal of Hydrology 317. 63–80.

SIVANANDAM, N., Deepa, N. 2008. Introduction to Genetic Algorithms. Verlag Berlin Heidelberg. Germany

SIVAKUMAR, B. JAYAWARDENA, A.W., T. FERNANDO. 2002. River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. Journal of Hydrology 265. 225–245.

SMITH R.A. et al. 2001. “Modelos de predicción de caudales mensuales para el sector Eléctrico Colombiano”. Avances en recursos Hidráulicos, 11, Facultad de Minas, septiembre de 2004. P 91-102.

VAUTARD, R. et al. 1992. “Singular-spectrum analysis: A toolkit for shot, noisy chaotic signals”. Physica, D 58 (1992): 95-126. North-Holland.

VELASCO, I., Fritsch, J.M .1987. Mesoscale convective complexes in the Americas. J. Geophys. Res., 92, D8, 9591-9613.

WOLD, H. 1938. A Study in the Analysis of Stationary Time Series. Almqvist and Wicksell, Uppsala.

<http://ww.cpc.noaa.gov/products/predictions/experimental/bulletin/Jul96/art18.html>

<http://ww.cpc.noaa.gov/>

<http://ww.ideam.gov.co>

Anexo 1.

	C1	C2	C3
semana 1	0.25	0.01	0.91
semana 2	-0.12	0.58	0.29
semana 3	0.36	-0.11	0.35
semana 4	0.13	-0.17	0.88
semana 5	0.22	-0.59	1.28
semana 6	0.98	-0.89	0.78
semana 7	0.48	-0.19	0.35
semana 8	-0.48	1.07	-0.14
semana 9	0.55	-0.01	0.38
semana 10	0.17	-0.39	1.05
semana 11	-0.06	0.17	0.17
semana 12	-0.26	0.07	0.94
semana 13	-0.08	0.45	0.39
semana 14	1.40	0.23	-0.68
semana 15	-0.60	1.16	0.60
semana 16	-0.40	0.24	0.89
semana 17	0.14	-0.92	1.01
semana 18	-0.80	1.04	0.17
semana 19	0.35	-0.36	0.47
semana 20	0.43	-0.82	1.43
semana 21	-0.54	1.86	-0.82
semana 22	1.32	-0.99	-0.15
semana 23	0.22	-0.04	0.91
semana 24	0.26	0.08	0.44
semana 25	0.37	-0.18	0.53
semana 26	0.23	0.46	0.56
semana 27	0.08	-0.43	0.86
semana 28	1.00	-0.60	0.84
semana 29	-0.02	-0.06	1.02
semana 30	-0.09	-1.06	1.97
semana 31	-0.02	0.35	0.49
semana 32	-0.04	0.89	0.13
semana 33	0.48	0.45	0.01
semana 34	0.69	-0.39	0.61
semana 35	0.20	0.11	0.70
semana 36	0.32	-0.57	1.08
semana 37	-0.26	0.88	0.11
semana 38	0.58	-0.16	0.42
semana 39	0.13	-0.11	0.96
semana 40	-0.11	0.10	0.89
semana 41	-0.33	-0.49	1.79
semana 42	-0.99	1.28	0.64
semana 43	-0.48	0.49	0.79
semana 44	0.33	-0.01	0.58
semana 45	1.15	-0.08	-0.36
semana 46	0.46	0.31	0.24
semana 47	-0.62	-0.16	1.29
semana 48	0.46	-0.17	0.59
semana 49	0.34	-0.44	0.57
semana 50	0.15	-0.37	0.80

	C1	C2	C3
semana 51	-0.14	0.39	0.62
semana 52	0.19	-0.29	0.77
semana 53	-0.16	0.17	0.56
semana 54	0.34	-0.24	0.45
semana 55	-0.26	0.06	0.76
semana 56	0.33	-0.29	0.17
semana 57	0.10	-0.38	0.54
semana 58	0.47	-0.40	0.82
semana 59	0.45	-0.23	0.54
semana 60	-0.04	-0.03	1.00
semana 61	-0.45	0.32	0.45
semana 62	0.99	-0.81	0.64
semana 63	0.32	-0.29	0.52
semana 64	-0.10	-0.34	1.48
semana 65	-0.06	0.07	0.57
semana 66	0.50	-0.63	1.34
semana 67	0.76	-1.15	0.51
semana 68	-0.33	0.43	-0.14
semana 69	0.02	-0.18	0.61
semana 70	0.10	0.04	0.35
semana 71	0.20	-0.34	0.28
semana 72	0.16	-0.20	1.00
semana 73	0.80	-0.07	0.44

Coeficientes Río Otún por el modelo ACP

	C1	C2	C3
semana 1	0.49	-1.39	1.88
semana 2	0.59	-1.20	1.62
semana 3	0.30	-0.99	1.64
semana 4	0.04	-0.50	1.39
semana 5	0.59	-1.50	1.85
semana 6	0.81	-1.77	1.97
semana 7	0.04	-0.17	1.17
semana 8	0.76	-1.56	1.69
semana 9	0.91	-2.01	1.94
semana 10	0.30	-0.88	1.47
semana 11	0.29	-0.80	1.54
semana 12	0.24	-1.26	2.08
semana 13	0.55	-1.53	1.85
semana 14	1.07	-2.42	2.35
semana 15	0.45	-1.21	1.77
semana 16	0.76	-1.39	1.55
semana 17	0.36	-1.23	1.81
semana 18	0.53	-0.85	1.33
semana 19	0.83	-1.04	1.23
semana 20	1.04	-1.63	1.59
semana 21	0.28	-0.48	1.19
semana 22	0.65	-1.45	1.78
semana 23	0.46	-1.20	1.68
semana 24	0.64	-1.68	2.04
semana 25	0.35	-0.70	1.30
semana 26	0.91	-2.12	2.11
semana 27	0.53	-1.29	1.73
semana 28	0.16	-0.64	1.47
semana 29	-0.03	-0.26	1.25
semana 30	0.28	-0.79	1.52
semana 31	0.94	-1.53	1.63
semana 32	0.72	-0.71	0.96
semana 33	1.27	-2.45	2.05
semana 34	0.70	-1.57	1.72
semana 35	0.62	-1.35	1.67
semana 36	0.69	-1.34	1.69
semana 37	0.42	-1.09	1.69
semana 38	0.37	-1.16	1.66
semana 39	0.15	-1.05	1.75
semana 40	1.18	-2.32	2.14
semana 41	0.37	-0.85	1.46
semana 42	0.35	-0.36	1.06
semana 43	1.04	-0.88	0.83
semana 44	-0.04	-0.88	1.82
semana 45	1.50	-2.65	2.19
semana 46	0.50	-1.34	1.75
semana 47	0.68	-1.56	1.78
semana 48	0.55	-1.01	1.42
semana 49	0.89	-1.83	1.90
semana 50	0.19	-0.67	1.40

	C1	C2	C3
semana 51	0.76	-1.65	1.85
semana 52	0.64	-1.29	1.63
semana 53	0.66	-1.23	1.62
semana 54	0.64	-1.45	1.82
semana 55	0.59	-1.58	1.87
semana 56	0.40	-1.60	2.16
semana 57	0.26	-0.69	1.45
semana 58	0.26	0.05	0.65
semana 59	0.95	-1.58	1.57
semana 60	0.99	-1.95	1.95
semana 61	0.57	-1.23	1.59
semana 62	0.43	-1.44	1.91
semana 63	-0.05	-0.40	1.39
semana 64	0.86	-1.85	1.95
semana 65	1.13	-2.33	2.02
semana 66	0.84	-1.74	1.90
semana 67	0.58	-1.14	1.56
semana 68	0.42	-1.27	1.80
semana 69	0.97	-2.12	2.17
semana 70	0.20	-0.75	1.54
semana 71	0.75	-1.60	1.79
semana 72	0.69	-1.34	1.62
semana 73	0.73	-1.47	1.71

Coefficientes Río Cauca por el modelo ACP

	C1	C2	C3
semana 1	0.48	0.00	0.48
semana 2	-0.15	-0.13	1.17
semana 3	-0.09	0.75	0.32
semana 4	-0.46	-0.39	1.83
semana 5	0.69	-0.14	0.53
semana 6	0.80	-0.77	0.96
semana 7	0.45	-0.12	0.47
semana 8	0.54	0.23	0.22
semana 9	0.27	0.31	0.18
semana 10	0.41	0.00	0.34
semana 11	-0.02	0.12	0.86
semana 12	0.18	0.39	0.48
semana 13	-0.96	0.93	0.88
semana 14	0.50	0.04	0.37
semana 15	0.32	0.08	0.59
semana 16	-0.51	0.39	1.15
semana 17	0.64	-0.17	0.60
semana 18	0.84	-0.39	0.59
semana 19	0.12	0.19	0.68
semana 20	-0.15	1.06	-0.02
semana 21	0.16	-0.01	0.76
semana 22	-0.13	0.40	0.55
semana 23	0.45	0.16	0.27
semana 24	0.37	-0.19	0.59
semana 25	0.28	-0.49	1.10
semana 26	0.71	-0.88	0.92
semana 27	0.42	-0.01	0.69
semana 28	0.57	0.56	-0.12
semana 29	0.71	-0.41	0.33
semana 30	0.36	-0.01	0.89
semana 31	-0.19	-0.01	0.93
semana 32	-0.19	0.82	0.25
semana 33	-0.18	-0.06	1.14
semana 34	-0.13	0.43	0.55
semana 35	0.08	0.09	0.72
semana 36	0.97	-0.15	-0.22
semana 37	0.60	0.04	0.36
semana 38	0.34	-0.21	0.87
semana 39	0.26	-0.77	1.08
semana 40	-0.48	0.16	1.27
semana 41	0.07	-0.08	0.80
semana 42	0.32	0.27	0.41
semana 43	0.09	0.29	0.35
semana 44	-0.43	0.65	0.62
semana 45	-0.34	-0.21	1.61
semana 46	0.33	-0.64	1.30
semana 47	-0.27	0.61	0.50
semana 48	-0.07	0.21	0.79
semana 49	0.31	-0.24	0.71
semana 50	0.30	-0.37	0.90
semana 51	0.54	-0.54	0.85
semana 52	-0.02	0.42	0.40
semana 53	0.16	-0.15	0.49
semana 54	-0.07	0.32	0.37
semana 55	0.48	0.26	0.25
semana 56	-0.18	0.48	1.09
semana 57	-0.42	0.33	0.78
semana 58	0.16	-0.21	0.78
semana 59	-0.10	0.14	0.41
semana 60	0.19	0.20	0.52
semana 61	0.10	0.10	0.73
semana 62	0.06	-0.03	0.72
semana 63	0.20	-0.02	0.64
semana 64	0.16	0.23	0.47
semana 65	0.25	0.10	0.41
semana 66	0.43	-0.27	0.70
semana 67	0.27	-0.46	0.93
semana 68	0.08	-0.09	0.82
semana 69	0.81	-0.57	0.54
semana 70	-0.28	0.86	0.31
semana 71	0.34	-0.17	0.64
semana 72	0.10	-0.16	0.99
semana 73	0.78	-0.71	0.83

Coefficientes Río Risaralda por el modelo ACP

	C1	C2	C3
semana 1	0.89	-0.44	0.71
semana 2	-0.18	0.12	0.54
semana 3	-0.53	0.50	0.31
semana 4	0.72	-0.29	0.44
semana 5	0.06	-0.23	0.84
semana 6	0.23	0.21	0.19
semana 7	-0.31	0.35	0.48
semana 8	-0.56	0.84	-0.22
semana 9	0.36	0.14	0.51
semana 10	0.06	-0.52	1.15
semana 11	-0.32	0.18	0.55
semana 12	-0.59	0.30	0.90
semana 13	-0.11	0.43	0.06
semana 14	0.56	-0.12	0.09
semana 15	0.01	-0.24	0.91
semana 16	0.06	0.36	0.32
semana 17	0.63	-0.67	0.76
semana 18	0.08	0.67	0.20
semana 19	0.17	0.04	0.51
semana 20	-0.01	-0.57	0.69
semana 21	-0.12	-0.05	0.51
semana 22	0.04	-0.32	-0.01
semana 23	0.19	-0.34	0.78
semana 24	0.13	-0.52	0.82
semana 25	0.27	-0.21	-0.04
semana 26	0.34	0.08	0.56
semana 27	-0.12	-0.24	0.85
semana 28	0.27	-0.94	1.28
semana 29	0.28	0.13	0.16
semana 30	-0.38	-0.13	1.53
semana 31	-0.11	0.39	0.52
semana 32	-0.14	0.36	0.62
semana 33	-0.20	-0.28	1.24
semana 34	0.59	-0.36	0.44
semana 35	-0.24	0.54	0.27
semana 36	-0.03	0.18	0.29
semana 37	0.50	-0.24	1.07
semana 38	0.05	0.28	0.46
semana 39	0.41	-0.11	1.03
semana 40	0.39	-0.49	1.21
semana 41	-0.54	1.08	0.54
semana 42	0.67	-0.81	0.91
semana 43	-0.67	0.58	0.77
semana 44	0.28	-0.20	0.77
semana 45	0.46	0.46	0.08
semana 46	-0.58	0.65	1.10
semana 47	-0.42	0.53	0.41
semana 48	-0.08	-0.19	0.89
semana 49	0.38	-0.40	0.85
semana 50	0.22	-0.34	0.96
semana 51	0.61	-0.76	1.15
semana 52	-0.08	0.21	0.46
semana 53	-0.44	0.52	0.16
semana 54	0.39	0.34	0.37
semana 55	-0.85	0.00	0.78
semana 56	0.77	-0.19	0.37
semana 57	0.02	-0.22	0.67
semana 58	-0.17	-0.09	0.74
semana 59	0.11	-0.06	0.41
semana 60	0.13	0.00	1.07
semana 61	0.06	0.31	0.24
semana 62	0.65	-0.15	0.24
semana 63	0.07	-0.05	0.87
semana 64	0.78	-0.51	0.45
semana 65	0.22	-0.22	0.96
semana 66	-0.13	0.55	0.24
semana 67	0.09	-0.12	0.20
semana 68	0.41	-0.42	-0.12
semana 69	0.08	0.14	0.45
semana 70	-0.26	-0.06	0.65
semana 71	-0.24	0.28	0.30
semana 72	0.39	-0.32	0.75
semana 73	0.56	0.09	0.46

Coefficientes Río San Eugenio por el modelo ACP

	C1	C2	C3
semana 1	-0.28	-0.15	0.85
semana 2	1.07	-0.88	0.89
semana 3	0.09	-0.18	1.02
semana 4	-0.11	0.73	0.25
semana 5	0.97	-0.55	0.43
semana 6	0.15	-0.17	0.87
semana 7	1.36	-1.14	0.52
semana 8	-0.23	0.69	0.46
semana 9	0.05	-0.18	0.94
semana 10	0.46	0.18	0.32
semana 11	0.18	-0.22	0.63
semana 12	0.03	0.43	0.40
semana 13	0.06	0.05	0.62
semana 14	0.48	0.16	0.04
semana 15	-0.10	0.66	0.51
semana 16	0.10	0.53	0.13
semana 17	-0.69	0.74	0.03
semana 18	0.33	-0.42	0.79
semana 19	0.39	-0.13	0.79
semana 20	0.17	-0.30	1.05
semana 21	-0.06	-0.37	0.88
semana 22	0.20	-0.74	1.19
semana 23	-0.12	0.08	0.72
semana 24	0.66	-0.09	0.38
semana 25	0.26	0.24	0.21
semana 26	0.26	-0.10	0.81
semana 27	0.53	-0.35	0.71
semana 28	0.00	-0.51	1.04
semana 29	-0.69	0.68	0.49
semana 30	-0.01	0.62	0.28
semana 31	-0.22	-0.47	1.12
semana 32	0.19	-0.25	0.97
semana 33	0.26	0.04	0.48
semana 34	-0.06	-0.08	0.86
semana 35	0.10	-0.56	1.10
semana 36	-0.41	0.36	0.81
semana 37	-0.09	-0.46	1.16
semana 38	0.41	-0.78	1.11
semana 39	0.25	-0.04	0.63
semana 40	0.13	-0.28	1.06
semana 41	0.19	0.07	0.22
semana 42	-0.31	0.19	0.93
semana 43	0.00	-0.33	0.77
semana 44	0.24	-0.30	1.07
semana 45	0.04	0.17	0.75
semana 46	-0.53	0.30	1.11
semana 47	-0.08	-0.61	1.41
semana 48	0.20	-0.49	1.17
semana 49	-0.23	-0.01	1.03
semana 50	-0.27	0.28	0.74

	C1	C2	C3
semana 51	0.03	0.03	0.90
semana 52	0.23	0.07	0.61
semana 53	0.37	-0.77	1.27
semana 54	0.41	0.54	-0.09
semana 55	-0.50	0.42	0.84
semana 56	0.31	0.07	0.60
semana 57	0.09	-0.46	1.26
semana 58	0.09	-0.14	0.80
semana 59	0.12	0.19	0.66
semana 60	0.21	-0.43	1.12
semana 61	0.34	-0.68	1.06
semana 62	-0.51	0.57	0.78
semana 63	-0.05	-0.06	1.00
semana 64	-0.55	0.46	0.81
semana 65	0.65	-0.99	1.16
semana 66	0.04	-0.46	1.00
semana 67	-0.01	0.16	0.61
semana 68	-0.14	0.25	0.77
semana 69	0.04	-0.53	1.39
semana 70	0.10	0.53	0.37
semana 71	0.41	-0.18	0.48
semana 72	0.18	0.34	0.45
semana 73	0.34	0.30	0.35

Coefficientes Río Mapa por el modelo ACP