

# Modelos multivariados de predicción de caudal mensual utilizando variables macroclimáticas.

## Caso de estudio Río Cauca

Yesid Carvajal Escobar\*  
Juan B. Marco Segura\*\*

\* Ph.D. Profesor - Escuela de Ingeniería de Recursos Naturales y del Ambiente - EIDENAR - Facultad de Ingeniería - Universidad del Valle - Santiago de Cali, Colombia. e-mail: yecarvaj@univalle.edu.co

\*\* Profesor catedrático - Director del Departamento de Hidráulica y Medio Ambiente - Universidad Politécnica de Valencia, España. e-mail: jbmarco@hma.upv.es

### RESUMEN

Se analizaron las teleconexiones lineales entre el fenómeno ENOS y los caudales mensuales del Río Cauca (Valle del Cauca-Colombia), utilizando dos técnicas estadísticas multivariadas: el análisis de funciones ortogonales empíricas (FOES) y el análisis de correlación canónica (ACC). 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 anterior indica que las variables macroclimáticas aportaron información adicional. El FPE (error final de predicción) se

Fecha de recepción: Marzo 2 de 2005  
Fecha de aprobación: Junio 10 de 2005

redujo en un 9.44% en promedio, usando la primera componente principal de las variables macroclimáticas como variable auxiliar. Mientras, el FPE se redujo en un 13.35% en promedio, usando el primer coeficiente canónico de las variables macroclimáticas como variable auxiliar. Los modelos presentaron buen ajuste, por lo que pueden ser utilizados para predicción. Así mismo, los métodos multivariados de análisis de FOES y El ACC demuestran que son una herramienta eficaz en el estudio de la variabilidad climática para entender las relaciones del ENOS o fenómenos similares con la hidrología de la región.

**Palabras Clave:** Funciones ortogonales empíricas (FOES), análisis de componentes principales (ACP), análisis de correlación canónica (ACC), patrones principales de predicción (PPP), fenómeno el Niño, fenómeno de la Niña.

## ABSTRACT

The linear teleconnections between the ENOS phenomenon and the Cauca river located in the Valle del Cauca, Colombia, S. A. monthly interannual flows were studied. Two multivariate statistical techniques were used: the empirical orthogonal functions (EOFs) and the canonical correlation analysis (CCA). ARMAX models were adjusted to predict the monthly flows, using major components and canonical coefficients of the macroclimatic variables as auxiliary variables in prediction models. When including the basic components of the macroclimatic variables as predictor variables in flow models, an improvement in prediction was attained, indicating that they were contributing with additional information. The FPE (final prediction error) was reduced in 9.44% in average, using the first principal component of macroclimatic variables as an auxiliary variable. Meanwhile, the FPE was reduced in 13.35% in average by using the first canonical coefficients of macroclimatic variables as an auxiliary variable. The flow models presented a good adjustment, so they can be used for prediction. Likewise, the multivariate

EOFs and CCA methods proved to be valuable tools in the study of climate variability helping to understand the relationships between the ENOS phenomenon and the region's hydrology.

**Key Words:** Empirical orthogonal functions (EOFs), principal components analysis (PCA), canonical correlation analysis (CCA), principal patterns prediction (PPP), El Niño, La Niña, ENSO phenomenon.

## 1. INTRODUCCIÓN

Colombia es uno de los países con mayor abundancia de recursos hídricos por superficie en el mundo, con un rendimiento anual del escurrimiento ( $59 \text{ l/s/km}^2$ ) seis veces mayor que el promedio de la tierra ( $10 \text{ l/s/km}^2$ ) y tres veces mayor al de América Latina ( $21 \text{ l/s/km}^2$ ). Presenta un régimen hidroclimático determinado localmente, por la presencia de costas sobre dos océanos, la fisiografía, la circulación de la cuenca Amazónica y la vegetación, entre otros; mientras que a nivel global, está determinada por la Oscilación de la Zona de Convergencia Intertropical, que está directamente relacionada con la ocurrencia del evento El Niño - Oscilación del Sur (ENOS). No obstante la capacidad de aprovechamiento, esta relativa abundancia hídrica está determinada por las limitaciones espacio temporales que presenta. Por otro lado, casi el 90% de los desastres naturales que ocurren en el país son ocasionados por fenómenos hidrometeorológicos, siendo el Valle del Cauca una de las regiones más afectadas por su cercanía al Pacífico, zona de gran influencia del fenómeno ENOS. La agricultura es el principal renglón en la economía de la región, destacándose por ser la más tecnificada. No obstante, sufre los efectos de los eventos derivados de la variabilidad climática, que repercute sobre las condiciones socioeconómicas de la región; por lo cual su entendimiento permite mitigar efectos adversos y obtener los beneficios de una mejor predicción de caudales. Lo anterior impone una mirada diferente de la información hidroclimatológica, dadas las

incertidumbres que se presentan. Las decisiones en planificación y operación de sistemas de aprovechamiento hídrico se fundamentan en la información histórica de las variables hidrológicas; información que debe ser entendida y procesada para un buen funcionamiento de los sistemas con el fin de reducir la vulnerabilidad de disponibilidad hídrica.

En Colombia: la fase cálida del ENOS está asociada a sequías, incendios forestales, racionamientos energéticos, disminución en la producción, agrícola, pesquera y pecuaria, (Carvajal et al, 1998), así como al incremento en los casos de malaria y enfermedades endémicas, Poveda et al (1997). La fase fría (La Niña), se asocia con la pérdida de vidas humanas por desastres naturales, derrumbes, deslizamientos, crecientes, inundaciones, así como erosión y un notable incremento en el transporte de sedimentos de los ríos.

## 2. ESTADO DEL ARTE

Varios autores han investigado el impacto de el ENOS sobre el régimen de precipitaciones, caudales, temperaturas, nubosidad, etc. Ropelewski y Halpert (1987,1996) mencionan 4 regiones continentales persistentemente influenciadas por el ENOS, entre las cuales incluyen el sur de América. Liu et al (1998) referencian más de 65 trabajos de estudios realizados en los últimos años, en los cuales se relaciona la predicción de variables hidrológicas, incluyendo variables del ENOS en el pronóstico hidrológico.

En Latinoamérica hay referencias de perturbaciones climáticas asociadas con el ENOS que indican su influencia en diferentes escalas de tiempo, entre las que se cuentan: México (Cavazos y Hastenrath, 1990), Colombia, (Poveda y Penland, 1994); (Montealegre y Pabón, 1992), Venezuela, (Pulwarty et al., 1992), Perú (Waylen y Caviedes, 1986), en la cuenca Amazónica, (Marengo, 1992; Marengo y Hastenrath, 1993;

Obregón y Nobre, 1990), Brasil (Rao y Hada, 1990; Kayano et al, 1988; Chu, 1991), América del Sur (Aceituno, 1988; Aceituno, 1989).

Fernández y Fernández (2003), Utilizaron modelos estocásticos lineales de funciones de transferencia y modelos no lineales auto regresivos de promedio móvil con variables externas polinomiales de segundo orden, NARMAX, para modelar la estructura de dependencia temporal entre las precipitaciones del cono Sur de América y el campo de anomalías estandarizadas de la temperatura superficial del O. Pacífico, encontrando una dependencia clara y homogénea entre ambos fenómenos, corroborando así resultados obtenidos previamente. Montecinos et al. (2000) analizaron la estacionalidad de la relación simultánea entre la TSM en el Océano Pacífico tropical y las precipitaciones en la región subtropical de América del Sur, utilizando a tal efecto diferentes técnicas multivariadas.

Dölling y Varas (2000) indican que la introducción de indicadores de las anomalías de TSM del océano Pacífico ecuatorial es un elemento explicativo de importancia cuya inclusión en el modelo de predicción de caudales del río San Juan, Argentina, mejoró el comportamiento predictivo planteado sobre la base de redes neuronales artificiales. El régimen de caudales, en amplias regiones de EE.UU., es analizado por Kahya y Dracup (1993) identificando zonas que presentan fuertes y consistentes señales en relación con las fases extremas de ENOS. Las anomalías de precipitaciones, temperaturas y presión superficial, en amplios sectores de los océanos Pacífico e Índico para distintos escenarios de eventos cálidos y fríos de la Oscilación del Sur.

## 3. OBJETIVOS

- ☑ Incluir variables macroclimáticas asociadas al ENOS en la modelación hidrológica de los caudales del río Cauca (Colombia) y su

principal afluente aplicando análisis de funciones ortogonales empíricas y análisis de correlación canónica.

- ☑ Obtener los patrones principales de predicción (PPP) de caudal mensual.
- ☑ Desarrollar modelos de predicción de caudal medio mensual en los dos principales ríos de la región, involucrando variables macroclimáticas asociadas al ENOS, con el fin de mejorar la predicción de la disponibilidad hídrica y brindar mayores elementos de juicio a los planificadores del recurso hídrico en la región, para prevenir y mitigar efectos adversos, adoptando medidas oportunas.

## 4. METODOLOGÍA

### 4.1 Localización de la Zona de Estudio.

El Valle del Cauca está localizado en Colombia, entre las cordilleras Occidental y Central de los Andes (ver figura 1), y presenta una gran variedad climática debido a sus pisos térmicos. Además de los factores generales de la circulación global y del cambio en la posición del sol durante el año, la topografía, la convección profunda, la cercanía a las costas y la vegetación, son los principales factores que gobiernan el clima local en los trópicos, debido a que la circulación general es débil en gradientes de presión, temperatura, humedad y velocidad del viento. La zona Andina presenta una precipitación relativamente baja de 1500 mm/año, comparada con la Costa Pacífica, que es conocida como la región más lluviosa del mundo. Esto se debe en parte, a la disminución de la humedad atmosférica proveniente del O. Pacífico al entrar en contacto con el flanco oeste de la cordillera Occidental. Hay dos períodos lluviosos: Marzo-Mayo y Septiembre-Noviembre, mientras que Diciembre- Febrero y Junio-Agosto son menos lluviosos o secos. La estación seca de mayor duración ocurre en los meses de verano en el hemisferio norte, siendo fundamental la influencia de los vientos alisios y el cambio de posición en la ZCIT, Snow (1976).

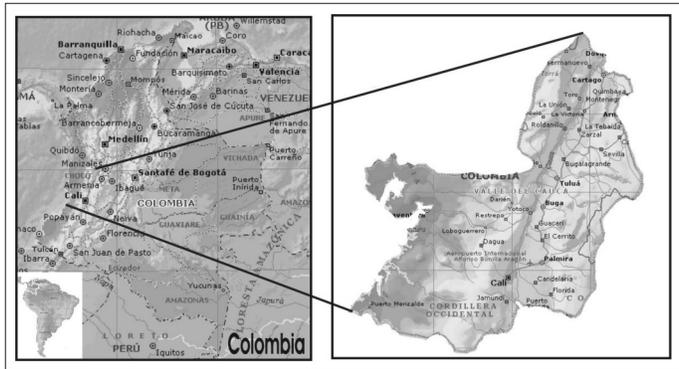


Figura 1. Localización de la zona de estudio

El río Cauca, que atraviesa la región en un recorrido de 340 km, se destaca estratégicamente en el contexto nacional, porque en su cuenca se asienta la industria azucarera, buena parte de la zona cafetera, de desarrollo minero y un sector significativo de la industria manufacturera del país. En el cuadro 1, se observan las características hidrológicas del río Cauca a su paso por el Valle; y la figura 2 presenta la localización de las estaciones hidrométricas que se utilizaron en el presente estudio.

Cuadro 1. Estaciones hidrométricas seleccionadas para el estudio.

Río	Estación	Área (Km <sup>2</sup> )	Caudal mensual medio multianual (m <sup>3</sup> /s)	Caudal mensual máximo multianual (m <sup>3</sup> /s)	Caudal mensual mínimo multianual (m <sup>3</sup> /s)
Cauca	La Balsa	5,156	196	301	111
	Hormiguero	8,018	267	404	151
	Juanchito	8,584	279	434	150
	Mediacanoa	12,186	322	514	171
	Guayabal	16,036	385	625	181
	La Victoria	16,284	386	638	189
La Vieja	Anacaro	17,234	409	673	197
	Cartago	2,852	97	287	188

Fuente: CVC, 1998

Se trabajó con las series mensuales de caudal de 8 estaciones, 7 ubicadas sobre el Río Cauca y 1 sobre el río La Vieja, su principal afluente en la región. Los registros de los afluentes restantes se descartaron por presentar series muy cortas o varios años sin registro. La información histórica de los registros no presentó inconsistencias, pero fue restituida a régimen natural a partir de 1985, cuando entró en operación el embalse de La

Salvajina (1000 hm<sup>3</sup>). Los datos faltantes fueron estimados aplicando un análisis multivariado con una subrutina de componentes principales, y utilizando la información de todo el conjunto de estaciones (ver detalles en Anexo 1). Se utilizó el período comprendido entre 1951-1976 para ajustar los modelos y 1977-2000 como período de validación. Se seleccionaron 9 variables macroclimáticas asociadas al ENOS: Temperatura en las regiones: Niño 1-2, Niño 3, Niño 4, Niño 3-4, Costa oeste de EU, Fidji Hawaii; y el índice de oscilación del sur, el índice multivariado del ENOS, y el índice de la lengua fría. Estos registros se obtuvieron del proyecto COADS (Comprehensive Ocean-Atmosphere Data Set) de la NOAA (National Oceanic and Atmospheric Administration) y de JSAO, instituto adjunto de la NOAA y la universidad de Washington, para el estudio conjunto del océano y la atmósfera.

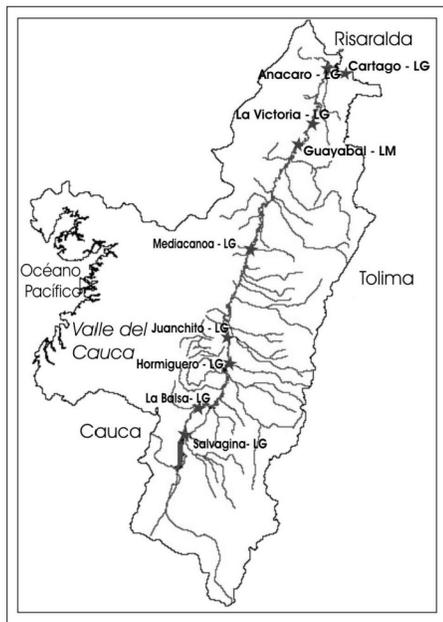


Figura 2 Localización de las estaciones hidrométricas de medición de caudal

#### 4.2 Tratamiento previo de la información.

Inicialmente, se estandarizaron las series, restando la media y dividiendo por la desviación estandar. Como paso previo al ACC, se realizó un análisis de FOES a ambos conjuntos de variables, Preisendorfer y Barnett (1987), con el fin de

compactar la información esencial y eliminar la multicolinealidad. El método ha sido ampliamente utilizado en análisis de variabilidad climática y puede ser usado en predicción, (Zweirs y Von Storch, 1999; Soley y Alfaro, 1999).

#### 4.3 Ajuste de modelos paramétricos.

Se ajustaron modelos paramétricos ARMAX para las series de caudal, usando como variables auxiliares los coeficientes canónicos (CC) y las componentes principales (CP) de las variables macroclimáticas. Para ello, se utilizó el módulo de identificación de sistemas de Matlab, desarrollado por Ljung (1999b). Dicha herramienta, permite construir modelos matemáticos precisos y simplificados de sistemas dinámicos basándose en los datos observados del sistema, y es especial para sistemas complejos de series de tiempo con ruido. La identificación de sistemas se define como el área de la teoría de sistemas, que estudia métodos para la obtención de modelos matemáticos de sistemas dinámicos a partir de mediciones sobre el sistema. En la modelación, es conveniente distinguir entre las señales de entrada (independientes) y las de salida (dependientes); teniendo en cuenta que las salidas son parcialmente determinadas por las entradas. Para este caso, las variables hidrológicas corresponden a las salidas y las CP y CC de las variables macroclimáticas corresponden a las entradas (variables exógenas).

Generalmente se denota la entrada y la salida del sistema en el tiempo  $t$  por  $u(t)$  y  $y(t)$  respectivamente. La relación más básica entre las salidas y las entradas, es la ecuación diferencial lineal:

$$y(t) + a_1 y(t-1) - \dots + a_n y(t-n) = b_1 u(t-1) + \dots + b_m u(t-m) \quad (1)$$

Una forma práctica y útil de ver la ecuación (1) es determinar la próxima salida dada la observación previa.

$$y(t) = -a_1 y(t-1) - \dots - a_n y(t-n) + b_1 u(t-1) + \dots + b_m u(t-m) \quad (2)$$

Para una notación más compacta, se introducen los vectores:

$$\theta = [a_1 \dots a_n \ b_1 \dots b_m]^T \quad (3)$$

$$\varphi(t) = [-y(t-1) \dots -y(t-n) \ u(t-1) \dots u(t-m)]^T$$

Con esto, la ecuación (2) puede escribirse como:

$$y(t) = \varphi^T(t/\theta) \quad (4)$$

Para enfatizar que el cálculo de  $y(t)$  de los datos pasados depende de los parámetros en  $\theta$ , se denota  $\hat{y}(t/\theta)$  se escribe:

$$\hat{y}(t/\theta) = \varphi^T(t)\theta \quad (5)$$

### Método de mínimos cuadrados:

Si se supone que para un sistema dado no se conocen los parámetros en  $\theta$ , pero se conocen las entradas y salidas sobre el intervalo de tiempo  $1 \leq t \leq N$ :

$$Z^N = \{u(1), y(1), \dots, u(N), y(N)\} \quad (6)$$

Una aproximación obvia consiste en seleccionar  $\theta$  en la ecuación (1) a través de la ecuación (4) de tal forma que ajuste los valores calculados  $\hat{y}(T/\theta)$ . Tan bien como sea posible para las salidas observadas, por medio del método de mínimos cuadrados.

$$\min_{\theta} V_n(\theta, Z^N) \quad (7)$$

Donde:

$$V_n(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^N (y(t) - \hat{y}(t/\theta))^2 = \frac{1}{N} \sum_{t=1}^N (y(t) - \varphi^T(t)\theta)^2 \quad (8)$$

Se denota el valor de  $\theta$  que minimiza la ecuación (7) por  $\hat{\theta}_N$

$$\hat{\theta}_N = \arg\text{-min}_{\theta} V_n(\theta, Z^N) \quad (9)$$

Donde  $\arg - \min$  significa la minimización del argumento; el valor de  $\theta$  que minimiza  $V_n$ . Puesto que  $V_n$  es cuadrática en  $\theta$ , se puede encontrar el valor mínimo haciendo la derivada igual a cero:

$$0 = \frac{d}{d\theta} V_n(\theta, Z^N) = \frac{2}{N} \sum_{t=1}^N \varphi(t)(y(t) - \varphi^T(t)\theta) \quad (10)$$

Que resulta:

$$\sum_{t=1}^N \varphi(t)y(t) = \sum_{t=1}^N \varphi(t)\varphi^T(t)\theta \quad (11)$$

ó

$$\theta = \left[ \sum_{t=1}^N \varphi(t)\varphi^T(t) \right]^{-1} \sum_{t=1}^N \varphi(t)y(t) \quad (12)$$

Estructuras de modelos tales como el de la ecuación (5), que son lineales en  $\theta$ , son conocidas en estadística como regresiones lineales. El vector  $\varphi(t)$  es llamado vector de regresión y sus componentes son los regresores, aludiendo al hecho que trata de explicar el  $y(t)$  que viene de atrás para  $\varphi(t)$ . La estructura del modelo de la ecuación (1) es conocida como el modelo ARX de autoregresión estandar con variables exógenas.

### 4.4 Estructura de los modelos.

En esta investigación, se ajustaron modelos autoregresivos multivariados con variables externas tipo ARMAX, autoregresivo de media móvil, con variables externas o exógenas; incluyendo las CP de un análisis de componentes principales realizado a las variables macroclimáticas estandarizadas de 9 variables macroclimáticas asociadas al ENOS: Temperatura en la regiones: Niño 1-2, Niño 3, Niño 4, Niño 3-4, Costa oeste de EU, Fidji Hawaii; y, el índice de oscilación del sur, el índice multivariado del ENOS, y el índice de la lengua fría. Así mismo, se modelaron los caudales tomado como variables exogenas, las primeros coeficientes canonicos de resultantes del ACC entre los dos conjuntos de variables de caudal y variables macroclimáticas asociadas a El ENOS, Ljung (1999 a, 1999b) hace una descripción detallada de tales modelos. Los pasos a seguir para la identificación de modelos se describen en el Anexo 1.

**Modelos ARMAX.** En un modelo ARMAX( $n_a, n_b, n_c, n_k$ ), [ $n_a \ n_b \ n_c \ n_k$ ] corresponden al orden y los rezagos del modelo respectivamente: Los  $y(t)$  corresponden a las salidas, es decir los caudales en este caso; y las  $u(t)$  corresponden a las variables auxiliares, las macroclimáticas para el caso. A, B y C, son polinomios en el operador de cambio (ésta es una forma compacta de escribir

las ecuaciones diferenciales), (Ljung 1999b)

$$A(q) y(t) = B(q) u(t - n_k) + C(q) e(t) \quad (13)$$

Los parámetros  $n_a$ ,  $n_b$  y  $n_c$  son los ordenes del modelo ARMAX, y  $n_k$  es el rezago. Específicamente:

$$\begin{aligned} n_a : A(q) &= 1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a} \\ n_b : B(q) &= b_1 + b_2 q^{-1} + \dots + b_{n_b} q^{-n_b+1} \\ n_c : C(q) &= 1 + c_1 q^{-1} + \dots + c_{n_c} q^{-n_c} \end{aligned} \quad (14)$$

#### 4.5 Preparación de los datos de entrada.

Se ajustaron tres tipos de modelos a las series hidrológicas, con el fin de comparar la calidad de los mismos incluyendo y sin incluir variables exógenas (macroclimáticas):

- ☑ Se ajustaron modelos ARMA a las series univariadas de la variable hidrológica, sin incluir variables exógenas; es decir sin incluir las CP y CC de las variables macroclimáticas asociadas al ENOS.
- ☑ Se ajustaron modelos a las series hidrológicas, incluyendo diferente número de CP de las variables macroclimáticas asociadas al ENOS como variables exógenas.
- ☑ Se ajustaron modelos a las series hidrológicas, incluyendo diferente número de coeficientes canónicos (CC) de las variables macroclimáticas asociadas al ENOS (U). Los coeficientes canónicos fueron obtenidos del ACC entre las CP de las variables hidrológicas y las CP del ENOS.

Se utilizó 1951-1976 como período de ajuste y 1977-2000 como período de validación. La validación incluyó la simulación del modelo para los datos de validación y el cómputo de los residuales del modelo cuando es aplicado a los mismos datos.

#### 4.6 Criterios de selección de orden y estructura de los modelos.

Se pueden tener varios modelos con diferentes órdenes y estructuras, siendo necesario escoger

cuál es el que se considera mejor o más adecuado para el propósito final. No existe un procedimiento absoluto para obtener el mejor modelo, por lo cual es conveniente probar con una variedad de herramientas para evaluar la calidad del mismo. En el Anexo 1 se describe las técnicas utilizadas en dicha evaluación.

Se compararon modelos de diferentes estructuras con estructuras tipo ARX y ARMAX; para ello se determinó la suma cuadrada de los errores de predicción  $V$  (función de pérdida), cuando es aplicada a los datos de validación; para luego seleccionar la estructura que tiene la función de pérdida más pequeña en la validación de los datos. Este procedimiento es conocido como validación cruzada y es una buena forma de aproximación al problema de selección del modelo. También es recomendable, realizar una inspección visual de cómo cambia el ajuste de los datos de validación con el número de parámetros estimados. Finalmente se escogieron modelos ARMAX, por ofrecer los mejores criterios y un menor número de parámetros.

Para los modelos ARMAX, se probaron diferentes estructuras de modelo. Si el modelo es validado sobre el mismo conjunto de datos del que fue estimado; el ajuste siempre mejora, cuando la flexibilidad de la estructura del modelo se incrementa, siendo necesario compensar por este decrecimiento de la función de pérdida. Hay varios métodos para esto, probablemente el más conocido es el criterio del error final de predicción (FPE) de Akaike, (Ljung, 1999a,1999b) y el criterio de información teórica AIC, (Ljung, 1999a,1999b) que es muy parecido y está muy relacionado con el primero. Ambos simulan la situación de validación cruzada, donde el modelo es probado en otro conjunto de datos. Para más detalles consultar las referencias.

Los residuales obtenidos del ajuste del modelo a los datos de validación deben ser idealmente, ruido blanco e independientes de las entradas del modelo. Para evaluar si el modelo describe

correctamente el sistema, se estimaron los residuales (error de predicción) del modelo, aplicados a los datos de validación, y se efectuó el análisis de independencia para las funciones de autocorrelación y de correlación cruzada entre las entradas y los residuales de las salidas, estableciendo unos límites de confianza del 99% para esta variable. Si la función de autocorrelación está significativamente fuera de estos intervalos de confianza, no se acepta el correspondiente modelo, como buen descriptor del sistema. Por último, se validó el modelo de pronóstico de cada análisis para el período independiente comprendido entre 1977-2000.

## 5. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

### 5.1 Ajuste de modelos paramétricos.

La figura 3 muestra la gráfica comparativa del error final de predicción (FPE), ajustando las series de caudal de cada estación a modelos arma (1,1) y armax (1,1,1,1): sin variables auxiliares, con las primeras CP como variables auxiliares y con los primeros coeficientes canónicos (CC) como variables auxiliares; se observa que el error disminuye, cuando se utilizan las variables auxiliares, y que el mejor ajuste se obtiene utilizando coeficientes canónicos (CC) de las variables macroclimáticas. El error final de predicción del caudal a rezago 1 (variables macroclimáticas anticipadas) se redujo en un 9.44% en promedio, utilizando la primera componente principal de las variables macroclimáticas como variable auxiliar. La mayor reducción del FPE a rezago 1 se obtuvo con la serie de caudal del río La Vieja (11.28%). De otro lado, El error final de predicción del caudal a rezago 1 (variables macroclimáticas anticipadas) se redujo en un 13.35% en promedio, utilizando

el primer coeficiente canónico de las variables macroclimáticas como variable auxiliar. La mayor reducción del FPE a rezago 1 se obtuvo con la serie de caudal del río La Vieja (15.1%).

### 5.2 Criterios de ajuste utilizados para evaluar los modelos.

La figura 4 muestra el ajuste obtenido con los modelos al aplicar la ecuación (18) del anexo 1 a los datos de validación, para las diferentes series de caudal: en primer lugar, se comparan: el ajuste del modelo arma (1,1) de la serie sencilla de caudal sin incluir variables exógenas; incluyendo CP de variables macroclimáticas como variables exógenas e incluyendo CC de las variables macroclimáticas, como variables exógenas.

Se observa que los modelos presentan un mejor ajuste cuando se utilizan las variables auxiliares asociadas al ENOS. El ajuste indica el porcentaje de variación de los datos de validación que fue explicado por el modelo. Los mejores ajustes se obtuvieron utilizando las primeras CP y los primeros CC de las variables macroclimáticas, como variables exógenas o auxiliares.

Lo anterior resulta lógico si se tiene en cuenta que las primeras componentes principales y los primeros coeficientes canónicos representan la máxima variabilidad y covariabilidad significativa entre las series de caudal y las variables macroclimáticas, mientras que a medida que aumentan las CP y los CC en orden, representan menor variabilidad y covariabilidad significativa conjunta, por lo cual podrían aportar menos a la predicción. Lo anterior se deduce de la significancia de las correlaciones canónicas. Lo anterior, no solamente se deduce de los resultados, si no de la significancia obtenida con las primeras correlaciones canónicas del análisis.

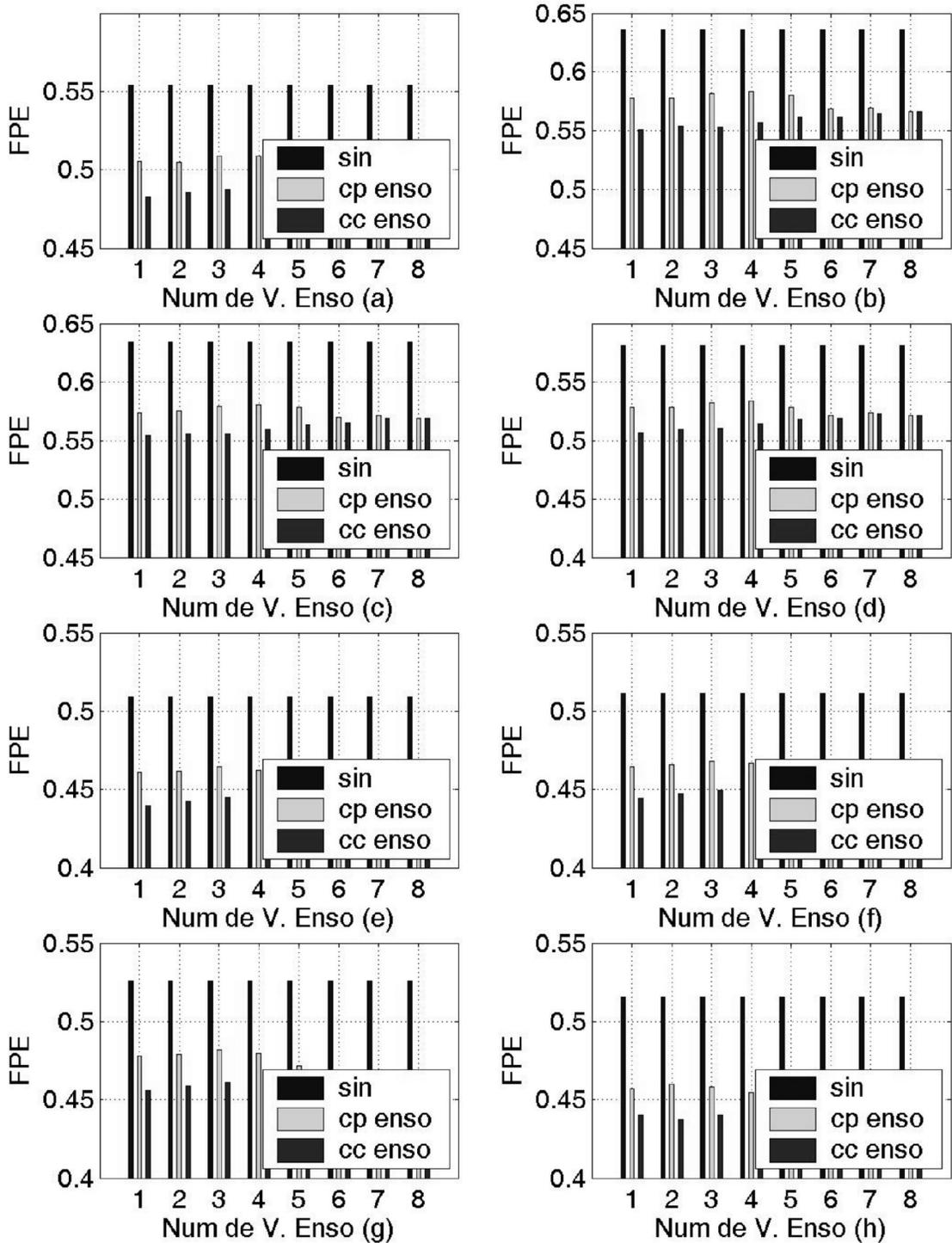


Figura 3. Error final de predicción (FPE) de los modelos ARMAX(1,1,1) de caudal para los diferentes análisis: Sin variables auxiliares, con componentes principales y con coeficientes canónicos de las variables macroclimáticas, como variables auxiliares. a) E. La Balsa, (b) E. Hormiguero (c) E. Juanchito, (d) E. Mediacanoa (e) E. Guayabal, (f) E. La Victoria, (g) E. Anacaro, (h) E. Cartago

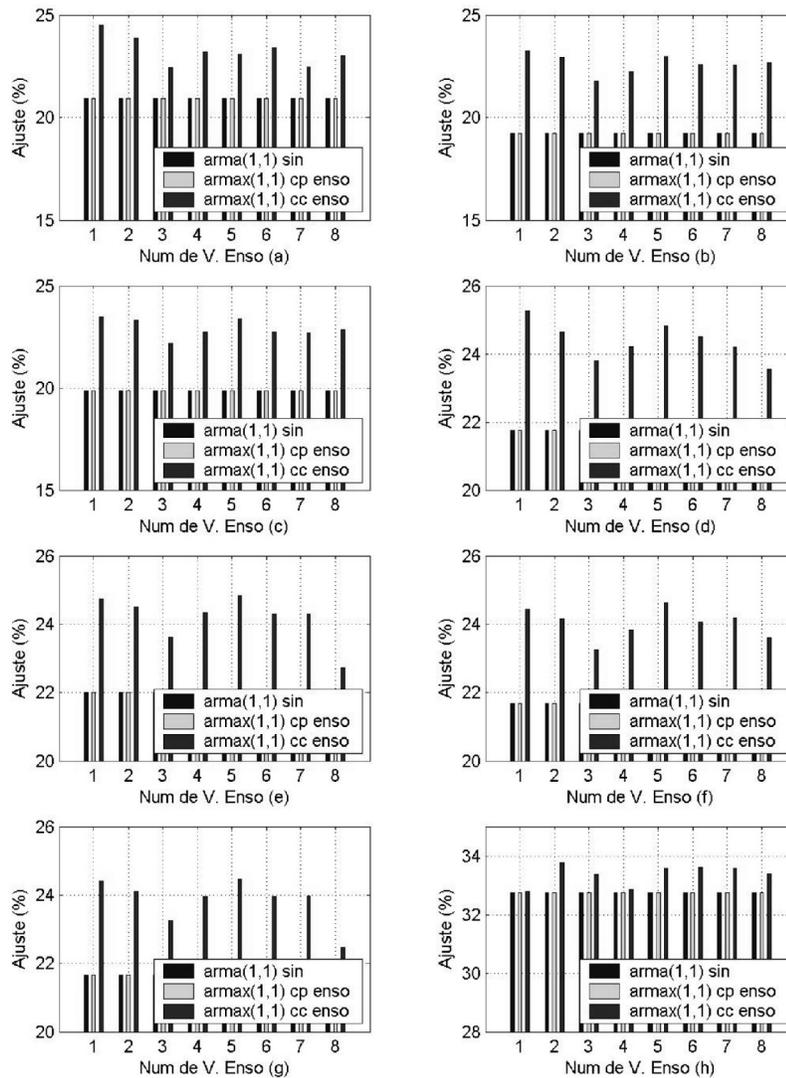


Figura 4. Ajuste de los modelos obtenidos a los datos de validación para las diferentes estaciones. (a) E. La Balsa, (b) E. Hormiguero (c) E. Juanchito, (d) E. Mediacanoa (e) E. Guayabal, (f) E. La Victoria, (g) E. Anacaro, (f) E. Cartago

### 5.3 Función de correlación de los residuos.

Se realizó la comparación gráfica del ajuste de los modelos a los datos de validación, y se revisaron las funciones de autocorrelación y de correlación cruzada entre las entradas  $u(t)$  (variables macroclimáticas) y los residuales de las salidas (caudal)  $e(t)$ . Esto con el fin de valorar la calidad de los modelos. Como caso ilustrativo, la figura 5 (a) presenta la verificación de independencia de los residuos, a partir de la función de autocorrelación obtenida de ajustar los caudales del Hormiguero, utilizando el primer CC de las variables

macroclimáticas como variable auxiliar (modelo ARMAX(1,1,1)), se verificó que los residuos estuvieran dentro de los niveles de confianza del 95 %, por lo cual se consideraron no correlacionados. Resultados similares se obtuvieron para los modelos de las demás estaciones. Así mismo, se verificó la independencia de los residuos de la función de correlación cruzada entre las entradas (primer CC de variables macroclimáticas), y los residuos de caudal a un 95% de confianza, ver La figura 5 (b).

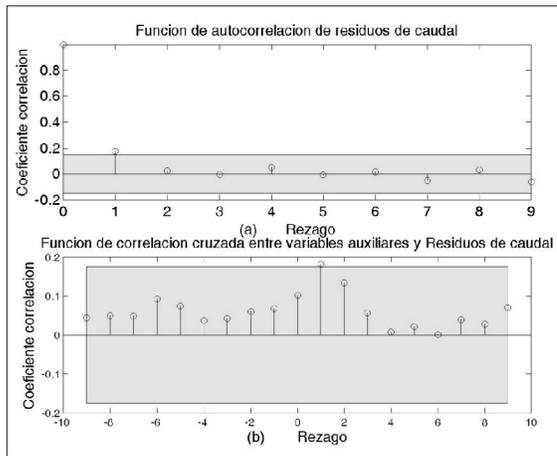


Figura 5. Análisis de residuos, Modelo ARMAX(1,1,1) estación Hormiguero, utilizando el primer CC como variable auxiliar (a) Función de autocorrelación de residuos con límites de independencia al 95%. (b) Función de correlación cruzada entre el primer CC de las variables macroclimáticas y los residuos de caudal con límites de confianza del 95%.

La figura 6 muestra el ajuste de la validación para el modelo obtenido con las series de la estación Hormiguero (río Cauca). Como se mencionó anteriormente, el ajuste en este caso representa el porcentaje de varianza tomada en cuenta por el modelo en las salidas; es decir, que a mayor ajuste, mejor es el modelo. En este caso, se observa un buen ajuste en la validación. Así mismo, se obtuvieron resultados similares para las demás series de caudal.

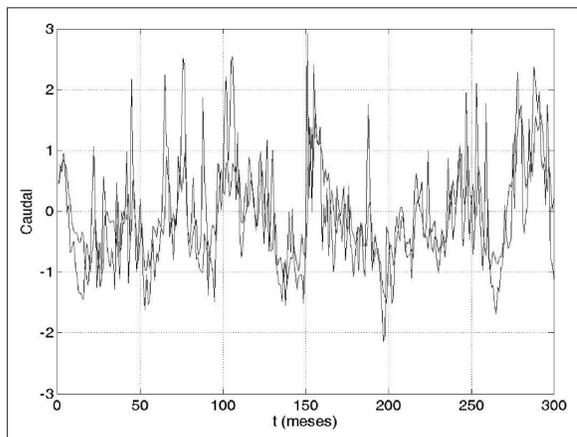


Figura 6. Validación gráfica del modelo. Modelo ARMAX(1,1,1) estación Hormiguero, utilizando el primer CC como variable auxiliar

## 6. CONCLUSIONES

La inclusión de las componentes principales y de los coeficientes canónicos de las variables macroclimáticas como variables auxiliares en los modelos armax de caudal mejoraron el ajuste y la predicción, indicando que aportan información adicional. El error final de predicción de los caudales a rezago 1 se redujo en un 9.44% en promedio, cuando se utilizó la primera componente principal de las variables macroclimáticas como variable auxiliar, y en un 13.35% en promedio, cuando se utilizó el primer coeficiente canónico de las variables macroclimáticas como variable auxiliar. La mayor reducción del FPE se obtuvo con la serie de caudal del río La Vieja (15.1%).

La inclusión de las variables macroclimáticas como variables auxiliares, mejoró la representación de los datos en la validación, al tener en cuenta un mayor porcentaje de varianza en los caudales. Los modelos de caudal presentaron buen ajuste, por lo que pueden ser utilizados para predicción. Así mismo, los métodos multivariados de análisis de FOES y El ACC demuestran que son una herramienta eficaz en el estudio de la variabilidad climática para entender las relaciones del ENOS o fenómenos similares con la hidrología de la región.

Los resultados proporcionan confianza en la aplicabilidad de las FOES y el ACC para entender el ENOS y fenómenos similares bajo perspectivas locales o regionales. Los resultados coinciden bien con observaciones y otros análisis estadísticos de gran escala. Para el Valle del Cauca la fase cálida del ENOS (El Niño) está asociada a la disminución de la precipitación y el caudal, mientras la fase fría (La Niña) está asociada a un incremento en los mismos.

Existe una teleconexión muy importante entre la variabilidad del ENOS y la hidrología de la región, con un rezago anticipado de 2 a 3 meses. Esto abre perspectivas en la utilización de variables macroclimáticas para mejorar la predicción y planificación.

El ENOS tiene un gran impacto socioeconómico en nuestro país, que no ha sido suficientemente estudiado, y que está aún por evaluar; por lo que es importante abordar la temática, incluyendo las variables macroclimáticas asociadas al ENOS en los análisis de planificación hídrica, estudios de producción agrícola, incendios forestales, de salud, etc., de tal forma que sean incorporadas a los procesos de planificación y administración de los recursos naturales, mediante una integración multisectorial y una visión que trascienda y adopte estrategias sostenibles.

Aunque los métodos clásicos ofrecen una explicación cuantitativa de la asociación y proporcionan confianza para entender las variables macroclimáticas, presentan la restricción de linealidad. En estudios futuros, se aplicarán métodos modernos no lineales tales como redes neuronales.

### Agradecimientos

El autor agradece a la Agencia Española de Cooperación Internacional (AECI), al Departamento de Ingeniería Hidráulica y Medio Ambiente de la Universidad Politécnica de Valencia, y a la Escuela de Ingeniería de Recursos Naturales y del Ambiente de la Universidad del Valle, Colombia, que han financiado el desarrollo de la presente investigación.

## 8. REFERENCIAS

- ☑ ACEITUNO, P. 1988. On the functioning of the southern oscillation in the south American sector. Part I: Surface Climate, *Mon. Wea. Rev.*, 116, 505-524.
- ☑ ACEITUNO, P. 1989. On the functioning of the southern oscillation in the south American sector. Part II: upper air circulation; *Journal Climate*, 2, 341-355.
- ☑ CARVAJAL, Y.; JIMENEZ, H. Y MATERÓN, H. 1998. Incidencia del fenómeno El Niño en la hidroclimatología del Valle del río Cauca. *Bulletin de l'Institut D'Francais 'Études Andines* (27) 3, 743-751.

- ☑ CAVAZOS, T. AND HASTENRATH, S. 1990. Convection and rainfall over Mexico and their modulation by the southern oscillation. *Intern. Journ. Climatol.*, 10, 377-386.
- ☑ CHU, P. S. 1991. Brazil's climate anomalies and ENOS, En *Teleconnections linking worldwide anomalies* (Glantz et al, eds) Cambridge University Press, 42-64.
- ☑ CVC. 1998. Estudio de regionalización de caudales máximos para diseño. Corporación Autónoma Regional del Valle del Cauca. Colombia. 220pp
- ☑ DÖLLING, O. y VARAS, E. (2000) Pronóstico de Caudales en Cuencas Nivales Usando Redes Neuronales Artificiales. *Anales XIX Congreso Latinoamericano de Hidráulica, Córdoba, Argentina.*
- ☑ KAHYA, E. y DRACUP, J. A. (1993) U. S. Streamflow Patterns in Relation to the El Niño/Southern Oscillation. *Water Resources Research*, 29, 2491-2503.
- ☑ KAYANO, M. T.; V. B. RAO AND MOURA, A.D. 1988. Tropical circulations and the associated rainfall anomalies during two constranting years. *Journal Climatology*, 8, 477-488.
- ☑ FERNÁNDEZ, B. y H. FERNÁNDEZ, 2002. Influencia de las fases extremas de el Niño Oscilación del sur sobre las precipitaciones en el cono sur de América. *Ingeniería Hidráulica en México, Vol XVII, N3, 5-16.*
- ☑ LIU, Z., VALDÉS, J., ENTEKHABI, D. (1998) Merging and error análisis of regional hidrometeorologic anomaly forecasts conditioned on climate precursors. *Water Resources Research*. Vol 34 No 8pages 1959-1969.
- ☑ LJUNG, LENNART. 1999a. System identification: Theory for the user, 2<sup>nd</sup> Edition, Prentice Hall, Englewood Cliffs, N.J.
- ☑ LJUNG, LENNART. 1999b. System identification toolbox, for use with Matlab, User's Guide Version 4, The MathWorks, Inc, Natick, MA.
- ☑ MARENGO, J., AND HASTENRATH, S. 1993. Case studies of extreme climatic events in the amazons basin. *Journal of climate*. 6,617-627.

- ☑ MARENGO, J. 1992. Interannual variability of surface climate in the Amazon basin. Intern. Jour. Climatol., 12, 853-863.
- ☑ MONTEALEGRE, J. Y PABÓN, J. 1992. Características climáticas relevantes durante la ocurrencia de Fenómenos ENOS en el Noroccidente suramericano. HIMAT. Bogotá, Colombia
- ☑ MONTECINOS, A., DIAZ, A. y ACEITUNO, P. (2000). Seasonal Diagnostic and Predictability of Rainfall in Subtropical South America Based on Tropical Pacific SST. Journal of Climate, 13, 746-758.
- ☑ OBREGÓN, G. O. AND NOBRE, C.A. 1990. Principal component analysis of precipitation fields over the amazon river basin. Climanálise, 5, 35-46.
- ☑ POVEDA, G. Y C. PENLAND, 1994. "Predicción de caudales medios en Colombia usando modelación lineal inversa", memorias del XVI congreso Latinoamericano de Hidráulica e Hidrología, IAHS, Santiago de Chile, Vol. 4, 93-115.
- ☑ POVEDA, G.; BOUMA, M.J.; ROJAS, W.; CHAVASSE, D.; QUINONES, M.; COX, J. AND PATZ, J. 1997. Predicting high risk years for malaria in Colombia using El Niño southern oscillation. Trp. Med. Int. Health. V. (12). 1122-7
- ☑ PREISENDORFER, R. W. AND BARNETT, T.P. 1987. Significance tests for empirical orthogonal functions. Fifth Conference on probability and statistics, American Meteorological Society, 169-172.
- ☑ PULWARTY, R.S.; BARRY, R.G. AND RIEHL, H. 1992. Annual and seasonal patterns of rainfall variability over Venezuela. Erdkunde, 46, 273-289.
- ☑ RAO, V. B. AND HADA, H. 1990. Characteristics of rainfall over Brazil: annual variations and connections with the southern oscillation El Niño. Mon. Wea. Rev., 110
- ☑ ROPELEWSKI, C.F. AND HALPERT, M. S. (1987). Global and regional scale precipitation patterns associated with the EL Niño/Southern Oscillation, Mon. Weather Rev., 115, 1606-1626.
- ☑ ROPELEWSKI, C.F. AND HALPERT, M. S. (1996). Quantifying southern oscillation precipitation relationships, J. Climate, 9, 1043-1059.
- ☑ SOLEY, F. Y ALFARO, E. (1999) Aplicación de análisis multivariado al campo de anomalías de precipitación en Centroamérica. Top. Meteor. Oceanog., 6(2), 71-93.
- ☑ SNOW, J.W. 1976. The climate of northern South American, EN : Climates of central and South America, (W. Schwerdtfeger, ed.), Elsevier, Amsterdam, 295-403.
- ☑ WAYLEN, P. R. AND CAVIEDES, C.N. 1986. El Niño and annual floods on the north Peruvian litoral. Jour. Of Hydrology, 89, 141-156.
- ☑ ZWEIRS, F.W AND VON STORCH, H. 1999. Statistical analysis in climate research. Cambridge University Press. 484 pp.

#### ANEXO 1:

#### Aplicación del método de componentes principales para la estimación de datos faltantes

Frecuentemente, la información presenta datos faltantes, en cuyo caso, se pueden derivar las FOES y CP estimadas de los datos disponibles, con una ligera modificación en el método (Zweirs y Von Storch, 1999). Cada elemento de la matriz de covarianza  $\Sigma$  se estima formando sumas de todos los productos disponibles.

$$\hat{\sigma}_{ij} = \frac{1}{|K_i \cap K_j|} \sum_{k \in K_i \cap K_j} (x_{ki} - \hat{\mu}_i)(x_{kj} - \hat{\mu}_j)^*$$

Donde  $K_i = \{k: \text{componente } i \text{ de } x_k \text{ que está disponible}\}$

y donde  $\hat{\mu}_i = \frac{1}{|K_i|} \sum_{k \in K_i} x_{ki}$ .

Las FOES estimadas son entonces los autovectores  $\hat{e}_i$  de esta matriz de covarianza estimada. El conjunto de  $K_i \cap K_j$  es el conjunto de todos los índices tales que  $x_{ki}$  y  $x_{kj}$  son datos disponibles. Los coeficientes  $\hat{\alpha}_i$  del vector de datos  $\vec{x}$  no pueden obtenerse como el producto entre el vector de datos y las FOES estimadas  $\hat{e}_i$ , pero sí puede obtenerse un estimativo de los

mínimos cuadrados escogiendo  $\hat{\alpha}_i$  para minimizar  $\|\vec{x}-\hat{\alpha}_i\hat{e}_i\|$ . El estimativo de mínimos cuadrados se calcula mediante:

$\hat{\alpha}_i = \frac{\sum_{j \in K} x_j \hat{e}_j^*}{\sum_{j \in K} |\hat{e}_j|^2}$ , donde  $x_j$

y  $\hat{e}_j$  son las  $j$ ésimas componentes de  $x$  y  $\hat{e}$  respectivamente, y donde  $K = \{j: x_j \text{ está disponible}\}$ . La ecuación anterior, se reduce a  $\hat{\alpha}_i = \langle \vec{x}, \hat{e}_i \rangle$  cuando no hay datos faltantes en el conjunto de datos que corresponde al análisis de CP tradicionales.

### Pasos del procedimiento de identificación de los modelos.

El problema básico en la identificación de sistemas dinámicos consiste en estimar el modelo de un sistema basándose en unos datos de entrada (independientes) y unos de salida (dependientes). El procedimiento para determinar el modelo, involucra tres componentes básicos, que son:

1. Los datos de entradas (variables hidrológicas) y salidas (variables macroclimáticas).
2. Un conjunto de posibles modelos (candidatos)
3. Un criterio para seleccionar un modelo en particular del conjunto, basado en la información de los datos (identificación del modelo).

El proceso de identificación tiene en cuenta la repetida selección de una estructura de modelo, computando el mejor modelo y evaluando las propiedades del mismo para determinar si son satisfactorias. El ciclo se resume en los siguientes pasos:

1. Examinar los datos, remover tendencias, outliers y seleccionar el conjunto de datos para el ajuste y el conjunto de datos para la validación.
2. Seleccionar y definir la estructura del modelo, utilizando un conjunto de modelos candidatos que describan el sistema, dentro de los cuales se selecciona un modelo apropiado basándose en los datos disponibles de entrada-salida. Es decir, se selecciona el modelo que mejor se

ajusta a los datos.

3. Computar la estructura del modelo de mejor ajuste de acuerdo a los datos de entrada y un criterio de ajuste dado.
4. Examinar las propiedades del modelo obtenido: residuos, ajuste, independencia, etc.
5. Si el modelo es lo suficientemente bueno, entonces se detiene el proceso; de otra forma se retorna al paso 2 y se prueba con otro conjunto de modelos, o posiblemente, con otros métodos de estimación (Paso 3).

### Criterios utilizados para la selección del modelo de mejor ajuste

#### Criterio del error final de predicción (FPE).

El FPE, está definido como:  $FPE = V * \frac{1 + \frac{d}{N}}{1 - \frac{d}{N}}$  (15)

Donde V es la función de pérdida (ajuste cuadrático) para la estructura en cuestión, d es el número de parámetros y N es el número de datos utilizados en la estimación.

**Criterio de información de Akaike (AIC).** El criterio de información de Akaike, se define como:

$$AIC = \log(V) + \frac{2d}{N} \quad (16)$$

Donde V es la función de pérdida, d es el número de parámetros y N es el número de datos utilizados en la estimación. De acuerdo a la teoría de Akaike, en un conjunto de diferentes modelos, se escoge el que tenga el menor FPE (o AIC). Otro criterio relacionado es el criterio MDL (Minimum Description Length) de Rissanen.

#### Criterio de longitud de mínima descripción (MDL).

El criterio de longitud de mínima descripción (MDL) de Rissanen, donde se selecciona la estructura que conduce a la más corta descripción de los datos. Este criterio ha sido llamado BIC por Akaike, (Ljung, 1991 a, 1999b).

$$V_{mod} = V \left( 1 + \frac{d \log(N)}{N} \right) \quad (17)$$

Luego de examinar diferentes estructuras y ordenes de modelos, se decidió cuál de estos era el mejor y se verificó que la descripción de los datos de validación fuera la adecuada; para ello, se evaluó el ajuste con los datos de validación.

El criterio AIC es un estimador inconsistente del orden del proceso, porque la varianza no decrece con el incremento del tamaño de la muestra. Consecuente con lo anterior, la variabilidad muestral del AIC para determinado orden no decrece con el incremento del tamaño de la muestra. Por consiguiente, el criterio AIC tiende a sobreestimar el orden del proceso. Sin embargo estos problemas no son serios en la práctica. La tendencia a sobreestimar el orden se aligera, con el incremento del tamaño de la muestra, (Zweirs y Von Storch, 1999).

La principal diferencia entre AIC y BIC es que la penalización usando un parámetro extra es mucho mayor en el segundo. Zweirs y Von Storch (1999) mencionan que el BIC es un criterio consistente para determinar el orden, y además tiende a seleccionar modelos más parsimoniosos (con menos parámetros) que AIC, y normalmente identifica el orden del proceso en forma más precisa que AIC, mejorando con el incremento de la muestra. Adicionalmente, tiene más bajo sesgo y variabilidad que AIC, por lo cual se recomienda su uso.

**Criterio gráfico de ajuste del modelo a los datos de validación.** Una buena forma de obtener las características de calidad del modelo consiste en simular los datos de validación (no utilizados en el ajuste del modelo), y compararlos visualmente con los datos observados para inspeccionar la forma como el modelo reproduce los datos y si las características de los mismos fueron capturadas. Lo anterior da una idea de cuáles propiedades del sistema han sido tenidas en cuenta por el modelo, y cuáles no. En este caso, se computó la salida del modelo simulado y se comparó con los

datos de validación, para calcular posteriormente el porcentaje de variación explicado por el modelo, así:

$$Fit = 100(1 - \text{norm}(y_h - y) / \text{norm}(y - \text{mean}(y))) \quad (18)$$

Donde  $y_h$  es la salida obtenida cuando el modelo es simulado con la entrada  $u$ , y  $Fit$ , corresponde al porcentaje de variación de la salida que es explicada por el modelo.

### **Análisis de residuos.**

En el modelo  $y(t) = G(z)U(t) + H(z)e(t)$  los residuos son la parte de las salidas de datos que no se pueden reproducir. Para tener un buen modelo los residuos deben ser independientes de las entradas, es decir: se requiere que la función de correlación cruzada entre las entradas  $u(t)$  y los residuos de las salidas  $e(t)$  estén dentro de la región de confianza de independencia (95%). De otra forma, existe algo de los residuos que se origina desde las entradas y no ha sido descrito adecuadamente por el modelo; por lo tanto, el modelo se rechaza.