Predicción no lineal de caudales utilizando variables macroclimáticas y análisis espectral singular

• Julián David Rojo-Hernández • Luis Fernando Carvajal-Serna •
Universidad Nacional de Colombia

Resumen

La estrecha relación no lineal entre la hidrología colombiana y los procesos climáticos globales justifican el estudio de modelos no lineales de predicción. Este estudio presenta la dinámica no lineal de los caudales de los ríos de Colombia, utilizando un modelo periódico de predicción basado en el Análisis Espectral Singular (AES). El modelo periódico fue desarrollado y aplicado para ventanas de predicción de tres y seis meses. Se hacen comparaciones para los casos univariado y multivariado; este último incluye variables macroclimáticas. El modelo propuesto se estudia a través del análisis de componentes principales, la estructura de correlación con variables macroclimáticas y parámetros de reconstrucción de las series. Los esquemas desarrollados con base en el AES se aplican a las series de caudales mensuales de los ríos San Carlos, Río Grande II, Guatapé, Magdalena, Guavio y Batá. Los resultados indican que los modelos propuestos reproducen aceptablemente las principales características estadísticas de las series de caudales, obteniéndose predicciones cercanas a las observaciones históricas; además, la inclusión de variables macroclimáticas en el esquema multivariado mejora de forma significativa la capacidad de predicción.

Palabras clave: Análisis Espectral Singular (AES), modelación de series de tiempo, predicción no lineal de caudales, predicción con variables exógenas.

Introducción

Para el sector productivo de un país, la predicción de caudales medios mensuales es un insumo de gran importancia para operar y optimizar el recurso hídrico, y valorar los beneficios económicos derivados de la aplicación de tales predicciones (Smith *et al*., 2004; Mesa *et al*., 1995).

La utilización del Análisis Espectral Singular (AES) en la predicción de caudales medios mensuales en Colombia se remonta al año 1994, cuando Carvajal, con base en el trabajo de Vautard *et al*. (1992), propuso utilizar la descomposición en componentes principales, ajustando a cada componente un modelo auto-regresivo tipo AR para su predicción y luego reconstruir la serie; desde entonces, la predicción con AES ha incorporado el uso de regresiones lineales múltiples sobre componentes principales o componentes reconstruidas, involucrando variables macroclimáticas rezagadas en el tiempo (Carvajal *et al*., 1998; Poveda *et al*., 2002).

El presente trabajo muestra la dinámica no lineal de los caudales de los ríos de Colombia, utilizando un modelo periódico de predicción basado en el Análisis Espectral Singular. Este documento posee una síntesis del AES y el uso de dicha técnica para estudiar los ciclos que influyen sobre la dinámica de las series de caudales mediante el análisis de las componentes principales (ACP). Se utiliza el AES para reconstruir las series de caudales y encontrar relaciones entre las series reconstruidas y las variables macroclimáticas mediante la estimación de los coeficientes de correlación;

con base en las observaciones se desarrolla un esquema periódico de predicción y por último se comentan los resultados.

Datos

A continuación se hace una breve descripción de los datos de caudal y variables macroclimáticas empleadas en este trabajo.

Series de caudales

Se seleccionaron estaciones de caudales con más de cuarenta años de registro de seis ríos de la geografía colombiana: Nare, Río Grande, San Carlos, Magdalena, Guavio y Batá. La figura 1 presenta la localización de las estaciones y los respectivos ciclos anuales de caudales. Las series de caudales han sido estandarizas por su respectiva media y desviación estándar mensual.

Variables macroclimáticas

Algunos de los índices macroclimáticos obtenidos del National Climatic Data Center (http://www.cdc.noaa.gov/data.noaa.ersst.html) de los Estados Unidos son descritos en el cuadro 1. Las correlaciones entre el MEI y las diferentes variables macroclimáticas se presentan en la figura 1.

Análisis Espectral Singular

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 a series de tiempo univariadas. El AES trata de capturar la evolución del sistema en el espacio de fases y la implementación del método procede de manera análoga a la reconstrucción artificial de la evolución de las trayectorias en un sistema no lineal sensible a condiciones iniciales (Ghil *et al*., 2002).

Vautard *et al*. (1992) describieron el Análisis Espectral Singular (AES) para una serie estandarizada de tiempo X_i, donde i varía de 1 hasta N (cantidad total de datos). Si se asume una ventana de visualización M puede formarse una matriz de covarianzas denominada matriz de Toeplitz (T_x), cuyas componentes son los coeficientes de covariancia de la serie con cada uno de los M rezagos a visualizar (Carvajal, 1994), calculada como:

$$T_x = \begin{bmatrix} c(0) & c(1) & \dots & c(M-1) \\ c(1) & c(2) & \dots & c(M-2) \\ \vdots & & \ddots & \vdots \\ c(M-1) & & & c(0) \end{bmatrix} \qquad (1)$$

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

A la matriz cuadrada T_x se le hallan sus valores propios λ_k, resolviendo el sistema ($T_x - \lambda_k I$) $x = 0$; posteriormente se encuentra una base ortogonal, cuyas direcciones corresponden a los vectores propios E^k asociados con los autovalores de la matriz T_x; estos vectores E^k son denominados Funciones Ortogonales Empíricas (FOEs); la proyección de la serie original sobre cada FOE permite encontrar las componentes principales mediante la expresión (2) (Carvajal, 1994):

$$a_i^k = \sum_{j=1}^{M} X_{i+j} E_j^k \qquad 0 \le i \le N-M \qquad (2)$$

Las CP son procesos de longitud $N - M + 1$, considerados como un promedio móvil ponderado del proceso X_i, pero para obtener componentes que posean la misma longitud de la serie original debe hacerse una reconstrucción de las CP utilizando la expresión (3):

$$X_i^k = \sum_{j=1}^{M} a_i^k E_j^k \qquad M \le i \le N-M+1 \qquad (3)$$

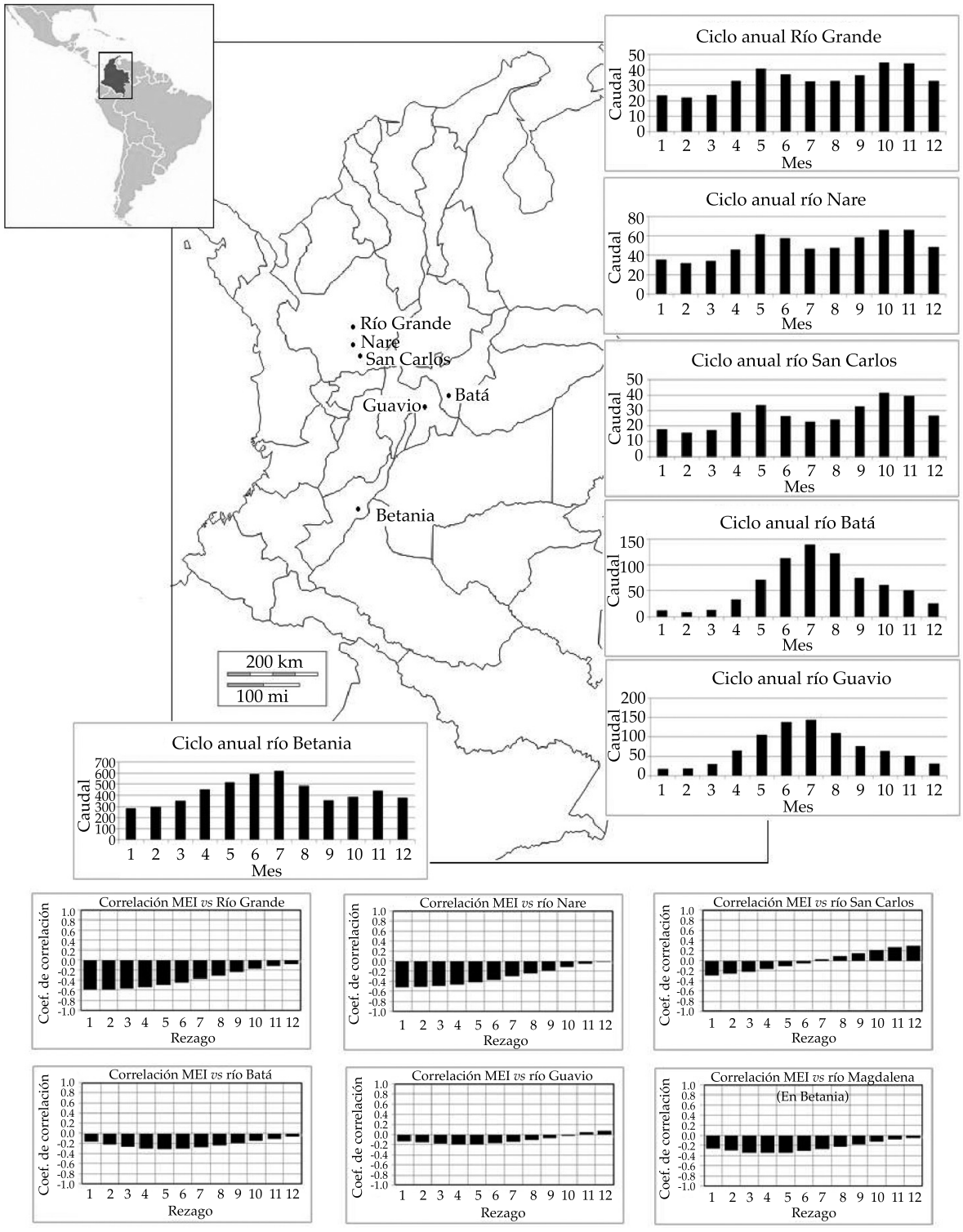

Figura 1. Ubicación aproximada. Ciclo anual en las estaciones de aforamiento (caudales en m^3/s) y correlación con el Índice Multivariado del ENSO (MEI).

Cuadro 1. Descripción de las variables macroclimáticas.

Índice	Descripción
Índice Oceánico del Niño (ONI)	Es el promedio de los últimos tres meses de las anomalías de temperatura superficial del océano (SST) Pacifico en la región Niño 3-4 (5N-5S, 120-170W).
Índice Multivariado del ENSO (MEI)	Puede describirse como un índice promedio de las características principales del ENSO, en relación con las siguientes seis variables: presión a nivel del mar, componentes zonal y meridional de los vientos superficiales, temperatura superficial del mar, temperatura del aire superficial y cantidad total de nubosidad. El MEI se calcula a partir del comportamiento de las seis variables observadas, normalizando los datos y extrayendo el componente principal en la matriz de covarianza de las variables combinadas.
Anomalías en la región Niño 3-4 (AN 3-4)	Definidas como la diferencia entre la temperatura histórica y el promedio sobre la región Niño 3-4 (5N-5S, 120-170W).
Oscilación Decadal del Pacífico (PDO)	Patrón de la variabilidad climática en el océano Pacífico y se define como la primera Componente Principal (CP) de las anomalías de temperaturas en el océano Pacífico norte.
Índice Oscilación del Sur (SOI)	Se estima como la diferencia de presiones atmosféricas entre las islas de Tahití y Darwin en el océano Pacífico, y representa la componente de presiones del ENSO.
La temperatura superficial del Atlántico norte (SST-AN)	Es el promedio de la temperatura entre las latitudes 5°00 N a 20°00 N, y entre las longitudes 60°00 W a 30°00 W. Asociadas con dichas temperaturas también existen índices estandarizados que analizan sus anomalías y se conocen como Anomalías Atlántico Norte y Anomalías Atlántico Sur.

las cuales se conocen como componentes reconstruidas (CR) de la serie. Expresiones para X_i^k con $i < M$ e $i > N - M + 1$ pueden hallarse en Vautard *et al*. (1992). La ortogonalidad en tiempo (covarianza cruzada igual a cero para dos CP en el rezago cero) y el espacio (ortogonalidad de las FOE) implican que λ_k (autovalor k de la matriz de la matriz de Toeplitz) representa la varianza de la k-ésima CP. Si se reduce la información a las primeras p CP podemos obtener una serie que omite gran parte del ruido, dado que las primeras p componentes describen la mayor parte varianza total acumulada, conservando la esencia de la serie (Carvajal, 1994).

Para la implementación del análisis espectral multivariado que considere L sitios (o L variables) con información del campo por estudiar, la expansión de las CP en las FOE está dada por la expresión:

$$X_{l,\,i+j} = \sum_{k=1}^{L\times N} a^k{}_i \, E_{l,j}{}^k \qquad 1 \le l \le L; \qquad 0 \le j \le M \quad (4)$$

Metodología

Las series de caudales representan un sistema dinámico y por tanto pueden ser descritas como una combinación lineal de diferentes osciladores. Mediante una proyección de la serie de caudales sobre sus FOEs se encuentran las componentes principales (CP) (expresión (2)), las cuales, al ser analizadas en el dominio del tiempo y la frecuencia, permiten determinar modos dominantes de oscilación que puedan ser asociados con diversos eventos macroclimáticos. Se trata entonces de utilizar

dicha información para reconstruir las series de caudales con base en componentes de baja frecuencia, con periodos alrededor de cuatro y seis años, además de aquellas cuyos periodos correspondan a ciclos anuales o semi-anuales (Plaut y Vautard, 1994).

Cuando se utiliza el análisis espectral singular con un tamaño de ventana M para reconstruir las series, existen M formas diferentes de reconstruir una señal, por lo que se debe establecer una metodología tendiente a determinar la cantidad de componentes y el tamaño de la ventana (M) necesarios para que la serie reconstruida represente de manera idónea la influencia del macroclima sobre los caudales.

Cabe anotar que para la predicción de la hidrología es vital involucrar la influencia de las diferentes variables macroclimáticas, cuantificando el grado de dependencia lineal entre tales variables a través de coeficientes de correlación a distintos rezagos de tiempo (Mesa *et al*., 1997). Entre mayores sean los coeficientes de correlación, mayor será la capacidad de la serie para representar los cambios debidos al macroclima; luego, se pretende estimar la cantidad de varianza y el valor de M necesarios para reconstruir las series y obtener mayores coeficiente de correlación entre las series de caudales reconstruidas y las variables macroclimáticas.

Una vez ajustados tamaño de ventana y porcentaje de varianza, según el macroclima, la serie reconstruida (filtrada) mediante la aplicación del AES se ordena formando vectores columna para cada mes, y sobre ésta se ajusta un modelo auto-regresivo periódico de orden tres, usando una regresión lineal múltiple que permite definir los parámetros para predecir cada mes. Para predecir el mes t se requieren los meses en t-1, t-2 *y* t-3. Por tanto:

$$Q_t = C_1 Q_{t-1} + C_2 Q_{t-2} + C_3 Q_{t-3} \qquad (5)$$

La serie reconstruida puede utilizar el AES univariado o multivariado; el primero corresponde al análisis hecho solamente a la serie de caudales y el segundo asume que el problema posee L variables que son la serie de caudales y los índices macroclimáticos rezagados un tiempo t. Para este último caso se espera que la serie reconstruida a partir de las componentes multivariadas represente mejor la influencia del macroclima sobre la hidrología.

Una vez concluida la calibración, se cuantifica la capacidad de predicción que tiene el modelo mediante la comparación de los caudales observados o históricos con los caudales pronosticados en un periodo de validación. Dicha evaluación se hace mediante la siguiente expresión:

$$\text{Error}_{i,j} = \frac{\sqrt{\dfrac{\sum_{i=1}^{k} \left(Q_{\text{histórico}\,j,i} - Q_{\text{simulado}\,j,i}\right)^2}{t}}}{Q_{\text{med}\,j,i}} \qquad (6)$$

Donde k es el número de años de validación y Q_{med} es el valor promedio del mes predicho en los años de validación. Esta expresión define el error porcentual de un valor simulado del mes j con respecto al valor real de dicho mes, iniciando las predicciones desde un mes i, en t años de validación. Se construye entonces el cuadro de errores en la validación, con ventanas de tres y seis meses. El promedio de los errores de validación puede ser usado como un índice para cuantificar la capacidad de predicción del modelo.

Análisis de resultados

Análisis de componentes principales

En el cuadro 2 se presentan los resultados del análisis de frecuencia de los diferentes componentes principales de los ríos analizados mediante el espectro de Fourier y la trasformada en onditas (Torrence, 1998).

En términos generales puede notarse que las primeras componentes principales representan la mayor cantidad de la varianza, y sus modos de vibración corresponden a

Cuadro 2. Periodos aproximados de oscilación (en meses) para las diferentes componentes principales en los ríos de estudio utilizando el espectro de Fourier.

	1CP	2CP	3CP	4CP	5CP	6CP	7CP	8CP	9CP	10CP	11CP	12CP
Río San Carlos	128	32	8 - 12	8	4	8 - 3	6	5	4 - 2	4 - 2	3	2
Río Grande	128 - 64	32 - 64	16	8	6	5	4	3	2	3	3	3
Río Nare	128 - 64	32	16	8	8	6	5	4	4	3	3	2
Río Guavio	128 - 64	16	64 - 12	6	5	4	4	2	3	2	2	2
Río Betania	128 - 64	32 - 12	32 - 16	8	6	4	5	8 - 4 - 2	2	4	3	3

periodos grandes entre los 128 y 64 meses (diez y cinco años, aproximadamente) para la primera componente principal; el periodo de la segunda componente varía entre los 64 y 32 meses (entre tres y cinco años), y la tercera varía entre los 32 y 12 meses. Las siguientes tres componentes poseen periodos que oscilan entre los ocho y los cuatro meses; por último, las demás componentes poseen periodos que van entre los cuatro y los dos meses.

A manera de diagnóstico, se pueden relacionar los periodos de las primeras componentes con los ciclos que poseen las variables macroclimáticas cuya influencia es de carácter interanual, como por ejemplo la PDO (diez años), el ENSO (entre tres y cinco años), el ciclo anual y semi-anual; los periodos 12 y 4 meses de las otras tres componentes siguientes pueden estar relacionados con variaciones intranuales, como el paso de la zona de convergencia intertropical (ZCIT) y los diferentes periodos de verano e invierno en el país. Las componentes de alta frecuencia pueden relacionarse con variables como las oscilaciones de Madden-Julian (de uno a tres meses) (Arias, 2005) y las ondas del este (Poveda, 2004).

Cantidad de varianza para reconstruir las series

Para los diferentes tamaños de ventana (*M*) se elaboraron gráficas que muestran la variación de los máximos coeficientes de correlación con la cantidad de varianza utilizada para reconstruir la serie (ver figura 2 para el caso del Río Grande).

Las observaciones permiten afirmar que independiente del tamaño de la ventana, las máximas correlaciones se obtienen a partir de series reconstruidas con base en las primeras componentes principales. Para la mayoría de los índices macroclimáticos, los máximos coeficientes de correlación se obtienen a partir de series reconstruidas con un 50 y 70% de la varianza de la serie de original en los ríos de ciclo bimodal (Río Grande, Nare y San Carlos). Para aquellos ríos unimodales (Magdalena en Betania, Guavio y Batá), los máximos coeficientes de correlación se obtienen a partir de series reconstruidas con un 25 y 45%. En el cuadro 3 se presentan los porcentajes de varianza requeridos para reconstruir las series y obtener las máximas correlaciones con los índices macroclimáticos para los ríos que se analizan en este trabajo.

Sensibilidad al tamaño de la ventana

El valor de la ventana *M* determina la longitud de la periodicidad que es capaz de capturar el análisis espectral singular (Kondrashov y Gil, 2006). Sabemos que los ríos de Colombia poseen un ciclo anual marcado de carácter netamente astronómico, que incluso puede dividirse en periodos trimestrales o semestrales, puesto que existen dos temporadas lluviosas y dos secas o de poca precipitación, definidos como ciclos bimodales; en otras regiones, como el oriente colombiano, la humedad procedente de la Amazonia influye de forma tal que los ríos poseen un ciclo unimodal durante todo el año. Sin embargo, eventos de fenómenos de

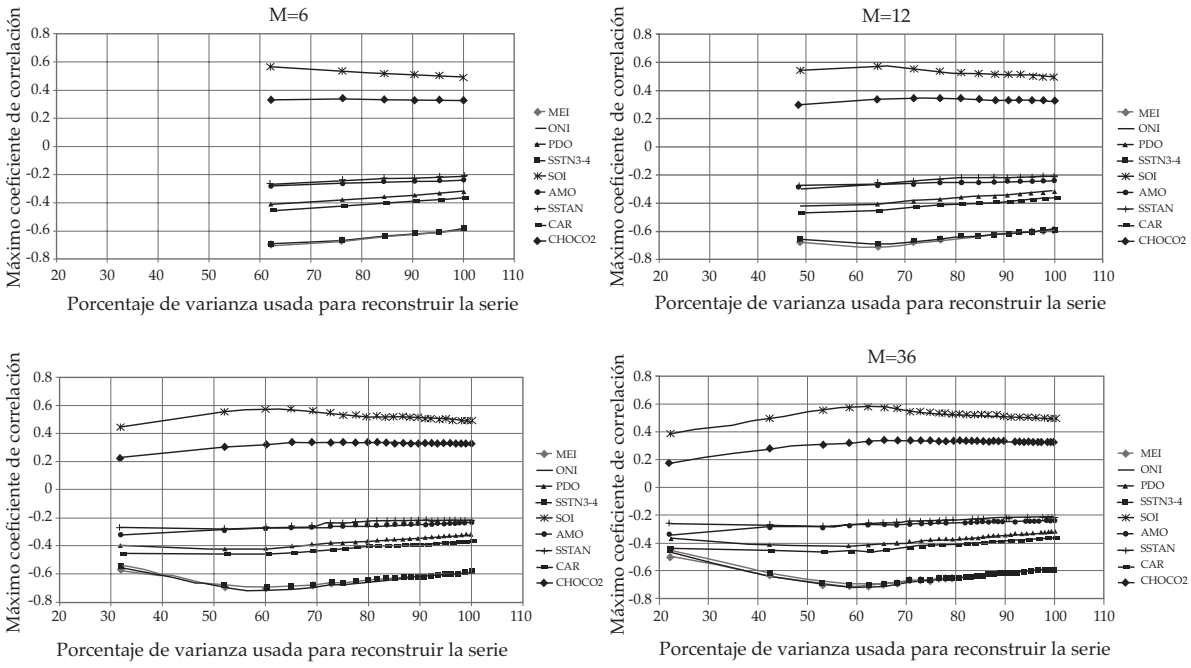

Figura 2. Variación de coeficiente de correlación con los porcentajes de varianza para reconstruir la serie, utilizando tamaños de ventana de 6, 12, 24 y 36 meses para el Río Grande.

gran escala y carácter interanual como el ENSO modulan la hidrología colombiana con una frecuencia de tres, cinco y diez años. Como los datos de la series de caudales son mensuales y teniendo en cuenta las frecuencias encontradas que influyen sobre la climatología colombiana, se han definido los tamaños de ventana *M* de 3, 6, 12, 24, 36 meses, que permitan capturar la variabilidad intra-anual e inter-anual de los fenómenos que modulan las series de caudales.

Los resultados muestran que los máximos coeficientes de correlación aumentan con el tamaño de la ventana de manera asintótica; es decir, inicialmente son bajos, pero en la medida en que aumenta el tamaño de la ventana, el máximo coeficiente de correlación tiende a ser constante, pues no presenta una variación significativa. La figura 3 muestra resultados de la variación de los coeficientes de correlación para algunos de los ríos que se analizan en este documento.

Metodología para reconstruir las series

Teniendo en cuenta los análisis que se incluyen en las secciones "Análisis de componentes principales" y "Cantidad de varianza para reconstruir las series", se propone que para reconstruir las series de forma tal que se obtengan los máximos coeficientes de correlación entre las variables macroclimáticas y las series estandarizadas de caudales se debe utilizar una ventana (*M*) de 12 meses para capturar variaciones relacionadas con eventos de carácter interanual. Para los ríos unimodales, se debe reconstruir, privilegiando entre el 25 y el 45% de la varianza, y para los ríos bimodales entre el 50 y el 70%.

Cuadro 3. Cantidad de componentes (***p***) y porcentaje de la varianza (% VAR) utilizados para obtener los máximos coeficientes de correlaciones en diferentes tamaños.

		Tamaño de la ventana											
		6			**12**			**24**			**36**		
Río	**Índice**	***p***	**% var**	**Máx corr**	***p***	**% var**	**Máx corr**	***p***	**% var**	**Máx corr**	***p***	**% var**	**Máx corr**
Río Grande	MEI	1	62%	-0.70	2	65%	-0.71	3	60%	-0.71	4	58%	-0.71
	SOI	1	62%	0.57	2	65%	0.58	3	60%	0.58	4	58%	0.58
	AN3-4	1	62%	-0.69	2	65%	-0.70	3	60%	-0.70	4	58%	-0.71
	SST AN	1	62%	-0.26	2	65%	-0.26	3	60%	-0.26	4	58%	-0.27
	PDO	1	62%	-0.41	2	65%	-0.41	3	60%	-0.42	4	58%	-0.42
Río Nare	MEI	1	60%	-0.68	2	62%	-0.69	3	58%	-0.69	4	57%	-0.69
	SOI	1	60%	0.57	2	62%	0.58	3	58%	0.58	4	57%	0.59
	AN3-4	1	60%	-0.68	2	62%	-0.70	3	58%	-0.70	4	57%	-0.70
	SST AN	1	60%	-0.28	2	62%	-0.29	3	58%	-0.29	4	57%	-0.29
	PDO	1	60%	-0.42	2	62%	-0.40	3	58%	-0.42	4	57%	-0.43
Río San Carlos	MEI	1	53%	-0.39	2	55%	-0.40	3	50%	-0.41	4	47%	-0.42
	SOI	1	53%	0.38	2	55%	0.40	3	50%	0.40	4	47%	0.40
	AN3-4	1	53%	-0.47	2	55%	-0.49	3	50%	-0.49	4	47%	-0.50
	A ATLAN	1	53%	-0.24	2	55%	-0.24	3	50%	-0.25	4	47%	-0.25
	PDO	1	53%	-0.38	2	55%	-0.37	3	50%	-0.38	4	47%	-0.38
Río Batá	MEI	1	35%	-0.29	2	32%	-0.30	3	29%	-0.31	3	25%	-0.34
	SOI	1	35%	0.21	2	32%	0.21	3	29%	0.23	3	25%	0.26
	AN3-4	1	35%	-0.26	2	32%	-0.28	3	29%	-0.30	3	25%	-0.34
	SST AN	1	35%	-0.10	2	32%	-0.11	3	29%	-0.16	3	25%	-0.17
	PDO	1	35%	-0.15	2	32%	-0.16	3	29%	-0.18	3	25%	-0.17
Río Guaio	MEI	1	32%	-0.31	3	45%	-0.33	4	35%	-0.34	4	25%	-0.33
	SOI	1	32%	0.23	3	45%	0.24	4	35%	0.24	4	25%	0.23
	AN3-4	1	32%	-0.35	3	45%	-0.37	4	35%	-0.37	4	25%	-0.37
	A ATLAN	1	32%	-0.28	3	45%	-0.27	4	35%	-0.28	4	25%	-0.27
	PDO	1	32%	-0.22	3	45%	-0.23	4	35%	-0.23	4	25%	-0.22
Río Betania	MEI	1	39%	-0.52	2	41%	-0.52	3	35%	-0.53	4	33%	-0.53
	SOI	1	39%	0.39	2	41%	0.40	3	35%	0.42	4	33%	0.42
	AN3-4	1	39%	-0.54	2	41%	-0.55	3	35%	-0.56	4	33%	-0.55
	SST AN	1	39%	-0.17	2	41%	-0.17	3	35%	-0.17	4	33%	-0.16
	PDO	1	39%	-0.24	2	41%	-0.24	3	35%	-0.24	4	33%	-0.24

p: cantidad de componentes utilizadas para reconstruir la serie.
% var: porcentaje de varianza acumulada.
Máx corr: máxima correlación obtenida luego de filtrar la serie.
A ATLAN: Anomalías de temperatura en el Atlántico norte.
SOI: Índice de Oscilación del Sur.
AN3-4: Anomalías de temperatura en la región Niño 3-4.

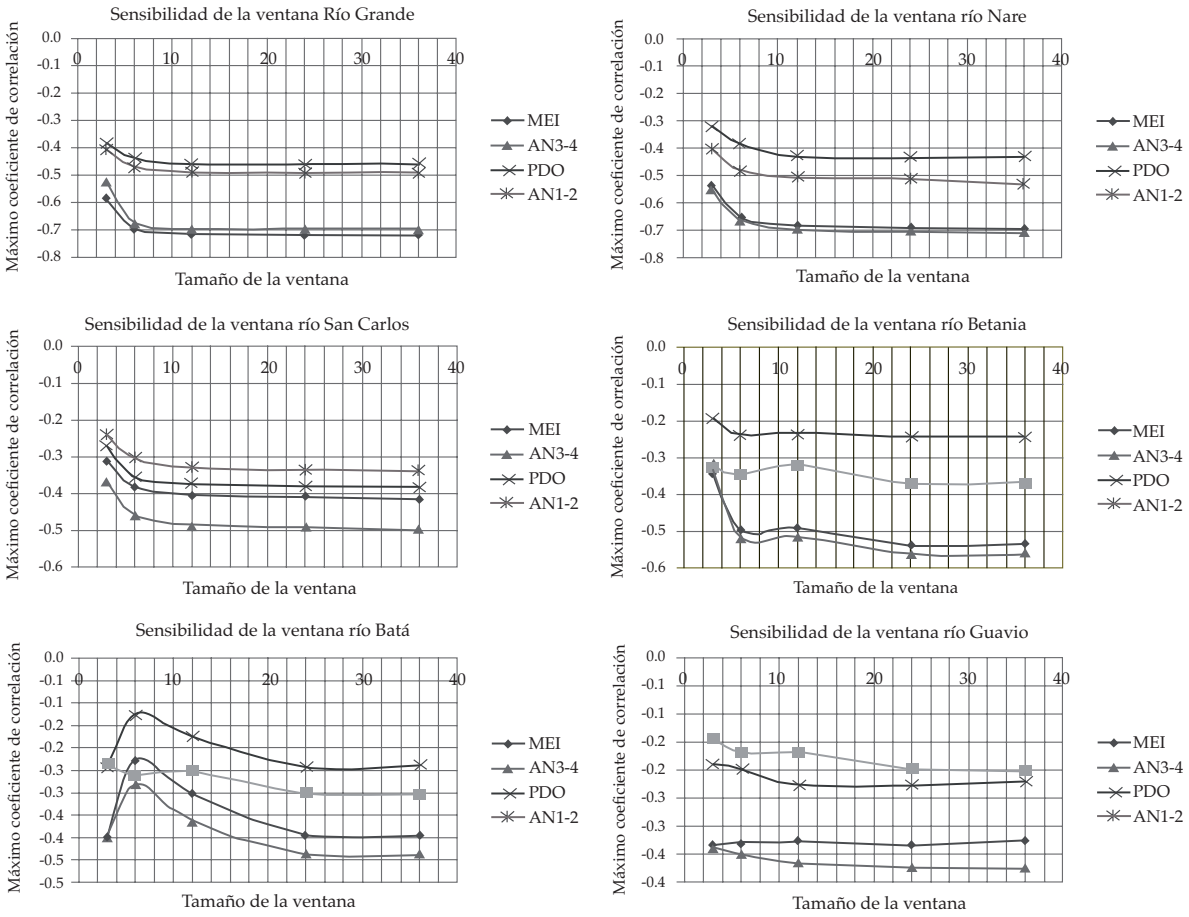

Figura 3. Sensibilidad al tamaño de la ventana: variación del máximo coeficiente de correlación con el tamaño de la ventana.

Cuando se filtra una serie mediante Análisis Espectral Singular (AES), los coeficientes de correlación entre los caudales y los índices de variables macroclimáticas mejoran en porcentajes que pueden ir desde un 20 hasta un 120%, según el grado de influencia de las variables macroclimáticas en relación con las diferentes regiones del país.

Calibración y validación

Cuando se aplica el análisis espectral singular a la serie estandarizada y se reconstruye dicha serie utilizando las primeras dos o tres componentes para un tamaño de ventana de 12 meses (figura 4), la nueva serie filtrada posee propiedades lineales muy fuertes tanto de auto-correlación como con las variables macroclimáticas. Los parámetros de los modelos periódicos obtenidos se presentan en el cuadro 4. El periodo de validación comprende desde enero de 1996 hasta diciembre de 2006. Los esquemas de las series históricas, conjuntamente con las series predichas para ventanas de tres y seis meses, se muestran en la figura 5. Los errores promedio de la validación por mes para todos los ríos se presentan en la figura 6 y en el cuadro 5 se muestra el error promedio total durante el periodo de validación tanto

Cuadro 4. Parámetros de los modelos periódicos para los ríos estudiados.

	Río Grande						Río Nare					
	Univariado			Multivariado			Univariado			Multivariado		
Mes	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**
Ene.	2.49	-2.13	0.63	2.91	-2.88	0.97	2.71	-2.47	0.76	2.68	-2.52	0.83
Feb.	2.62	-2.34	0.71	1.19	0.39	-0.61	0.99	0.89	-0.90	0.94	0.80	-0.78
Mar.	2.64	-2.37	0.73	2.12	-1.24	0.11	1.99	-1.02	0.01	1.94	-0.98	0.01
Abr.	2.78	-2.67	0.89	2.82	-2.69	0.86	2.92	-2.88	0.96	2.87	-2.89	1.02
May.	2.88	-2.86	0.98	2.88	-2.78	0.91	2.83	-2.67	0.84	2.90	-2.94	1.04
Jun.	2.66	-2.43	0.76	2.74	-2.52	0.77	2.55	-2.14	0.58	2.71	-2.56	0.83
Jul.	2.70	-2.49	0.79	2.86	-2.80	0.94	2.67	-2.39	0.72	2.70	-2.59	0.88
Agos.	2.68	-2.47	0.79	2.74	-2.57	0.82	2.69	-2.45	0.75	2.67	-2.50	0.82
Sept.	2.50	-2.12	0.61	2.62	-2.31	0.69	2.66	-2.36	0.70	2.56	-2.26	0.69
Oct.	2.59	-2.31	0.71	2.61	-2.31	0.69	2.71	-2.45	0.74	2.61	-2.34	0.73
Nov.	2.54	-2.17	0.62	2.56	-2.18	0.61	2.49	-2.00	0.51	2.61	-2.28	0.68
Dic.	2.74	-2.60	0.85	2.72	-2.49	0.77	2.64	-2.32	0.68	2.75	-2.58	0.83
Mes	**Río San Carlos**						**Río Batá**					
Ene.	2.41	-2.13	0.71	2.56	-2.29	0.72	2.10	-1.86	0.72	2.76	-2.64	0.87
Feb.	2.39	-2.10	0.69	2.66	-2.46	0.79	1.89	-1.51	0.57	2.48	-2.09	0.60
Mar.	2.51	-2.33	0.79	2.62	-2.45	0.82	2.05	-1.82	0.73	2.72	-2.55	0.82
Abr.	2.53	-2.32	0.77	2.68	-2.49	0.80	2.00	-1.74	0.66	2.85	-2.79	0.94
May.	2.73	-2.72	0.99	2.88	-2.87	0.99	2.06	-1.91	0.75	2.70	-2.51	0.80
Jun.	2.39	-2.03	0.61	2.71	-2.53	0.81	2.07	-1.95	0.78	2.79	-2.71	0.90
Jul.	2.48	-2.22	0.73	2.73	-2.57	0.84	2.05	-1.88	0.71	2.78	-2.68	0.89
Agos.	2.34	-1.97	0.62	2.54	-2.23	0.69	1.96	-1.73	0.66	2.96	-3.01	1.06
Sept.	2.44	-2.07	0.60	2.65	-2.41	0.74	2.04	-1.91	0.79	2.98	-3.06	1.09
Oct.	2.54	-2.30	0.75	2.60	-2.33	0.73	2.08	-1.92	0.76	2.94	-2.97	1.03
Nov.	2.57	-2.35	0.78	2.50	-2.13	0.62	1.90	-1.56	0.53	2.74	-2.59	0.84
Dic.	2.55	-2.37	0.81	2.66	-2.45	0.79	2.08	-1.90	0.75	2.83	-2.76	0.93
Mes	**Río Guavio**						**Río Betania**					
Ene.	2.24	-1.98	0.72	2.73	-2.59	0.86	2.56	-2.20	0.63	2.59	-2.30	0.70
Feb.	2.18	-1.90	0.68	2.60	-2.38	0.77	1.82	-0.79	-0.04	2.62	-2.37	0.74
Mar.	2.30	-2.09	0.75	2.70	-2.58	0.88	2.50	-2.07	0.56	2.36	-1.88	0.50
Abr.	2.09	-1.74	0.59	2.67	-2.53	0.85	2.77	-2.61	0.84	2.63	-2.35	0.71
May.	2.32	-2.21	0.86	2.70	-2.59	0.88	2.75	-2.50	0.74	2.69	-2.42	0.73
Jun.	2.32	-2.22	0.88	2.77	-2.74	0.97	2.12	-1.14	-0.01	2.84	-2.83	0.98
Jul.	2.21	-2.03	0.75	2.69	-2.58	0.87	2.73	-2.58	0.84	2.71	-2.55	0.84
Agos.	2.15	-1.98	0.71	2.70	-2.59	0.86	2.79	-2.57	0.78	2.72	-2.48	0.76
Sept.	2.35	-2.22	0.78	2.66	-2.51	0.83	2.63	-2.29	0.65	2.68	-2.38	0.70
Oct.	2.24	-1.96	0.67	2.67	-2.53	0.84	2.76	-2.58	0.82	2.68	-2.45	0.76
Nov.	2.22	-1.93	0.69	2.71	-2.56	0.84	2.73	-2.51	0.77	2.68	-2.45	0.76
Dic.	2.29	-2.05	0.74	2.81	-2.74	0.93	2.64	-2.36	0.72	2.71	-2.51	0.81

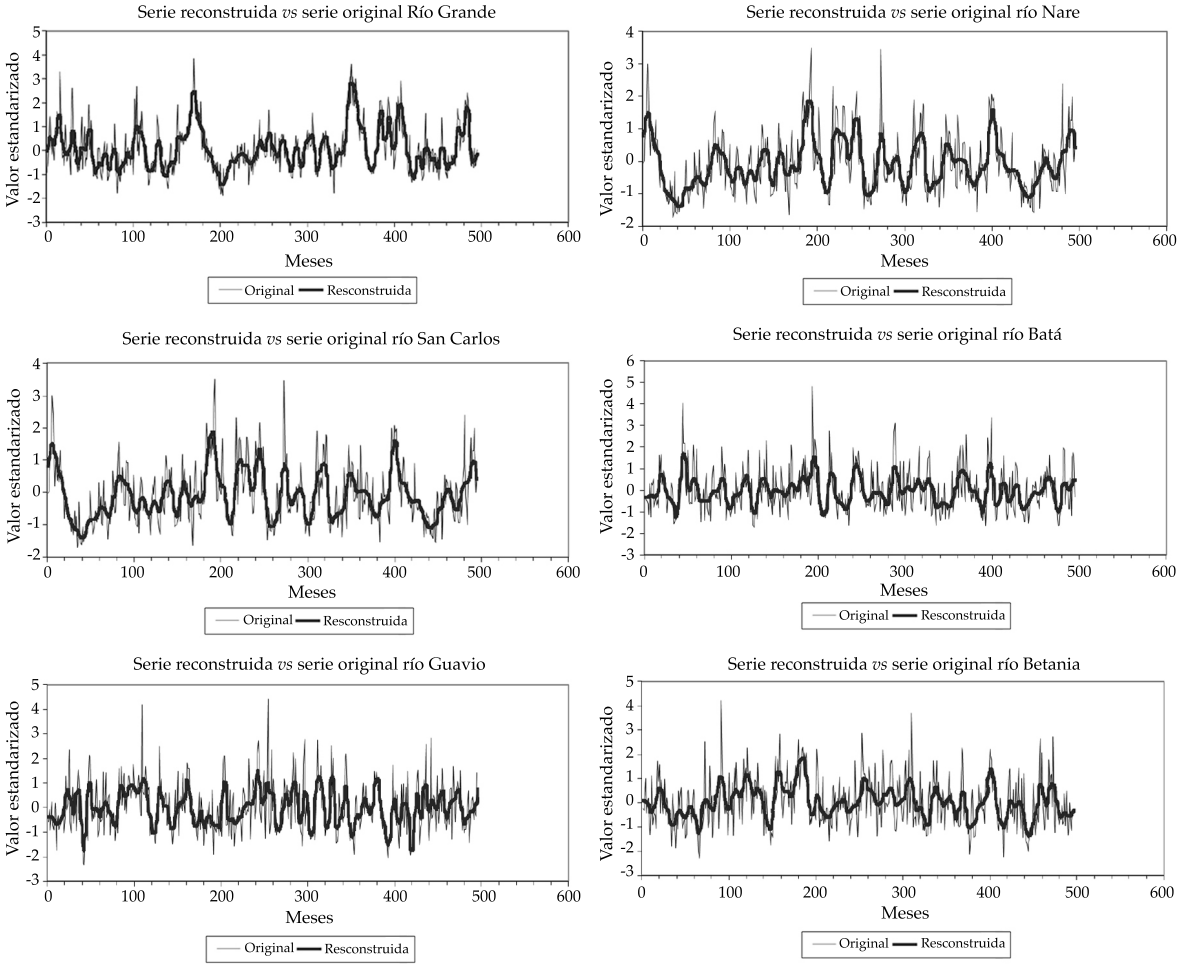

Figura 4. Series reconstruidas para los diferentes ríos.

Cuadro 5. Errores medios de validación (en porcentaje) para los seis ríos. máx: promedio de los máximos; prom: errores promedio; mín: promedio de los mínimos.

Errores de validación, ventana de tres meses							
Río		**San Carlos**	**Río Grande**	**Río Nare**	**Río Betania**	**Río Guavio**	**Río Batá**
Sólo caudal	máx	55.3	33.2	41.3	45.0	57.5	55.3
	prom	25.1	15.3	16.4	19.86	26.3	26.24
	mín	3.3	1.5	2.1	2.3	2.9	3.3
Caudal + var. macroclimáticas	máx	50.6	29.5	37.2	41.5	51.2	50.9
	prom	23.7	13.2	15.1	18.45	25.0	25.11
	mín	2.8	2.3	1.8	2.3	2.9	3.2
Errores de validación, ventana de seis meses							
Río		**San Carlos**	**Río Grande**	**Río Nare**	**Río Betania**	**Río Guavio**	**Río Batá**
Sólo caudal	máx	69.3	39.1	55.2	58.0	61.4	63.6
	prom	28.2	18.6	18.3	22.1	30.7	29.37
	mín	4.3	3.4	2.2	2.8	4.0	3.1
Caudal + var. macroclimáticas	máx	58.9	38.1	50.5	53.0	52.5	60.6
	prom	27.9	16.3	17.2	21.9	29.7	28.3
	mín	4	3.0	2.5	2.7	3.8	3.12

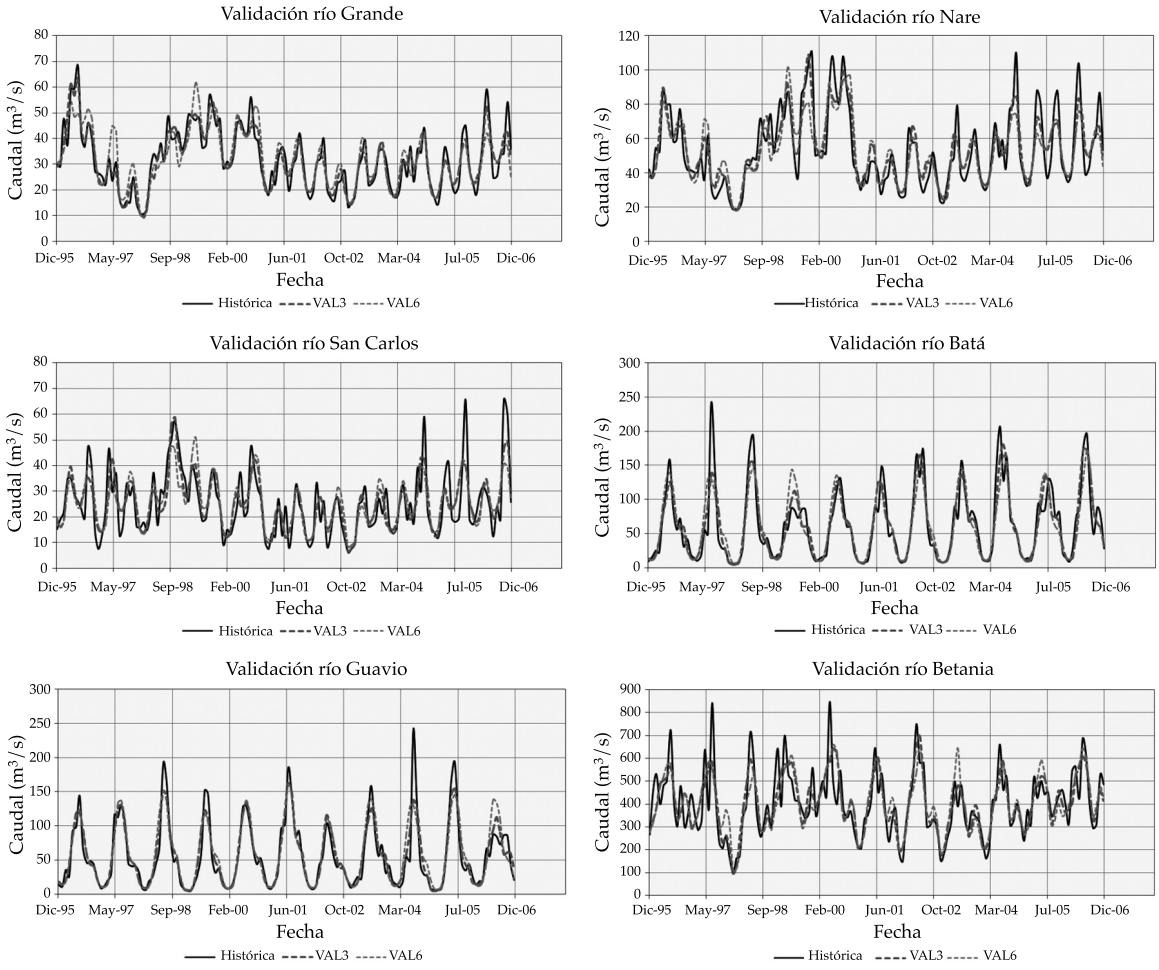

Figura 5. Series históricas y series predichas con ventanas de tres meses (VAL3) y seis meses (VAL6) durante el periodo de validación (de enero de 1996 a diciembre de 2006).

en el caso de AES univariado como multivariado.

Conclusiones

En las series de caudales, cada CP posee una frecuencia asociable con los ciclos de diferentes variables macroclimáticas, siendo las tres primeras componentes aquellas cuyos ciclos pueden relacionarse con la variabilidad interanual de la serie, dado que representan periodos de diez, cinco y un año, respectivamente; por el porcentaje de varianza aportada al proceso, se puede decir que son las que más contribuyen a la dinámica de la serie de caudales; en la mayoría de los casos, el ruido de la serie de caudales lo aportan aquellas componentes cuyos periodos son inferiores a un año asociadas con la variabilidad interanual de la serie.

Cuando las series de caudales se reconstruyen mediante un AES univariado (sólo caudales), tomando las primeras componentes, la serie reconstruida aumenta considerablemente su relación lineal con los índices macroclimáticos de carácter interanual como la PDO, el MEI y las anomalías de temperatura en la región Niño 3-4; por

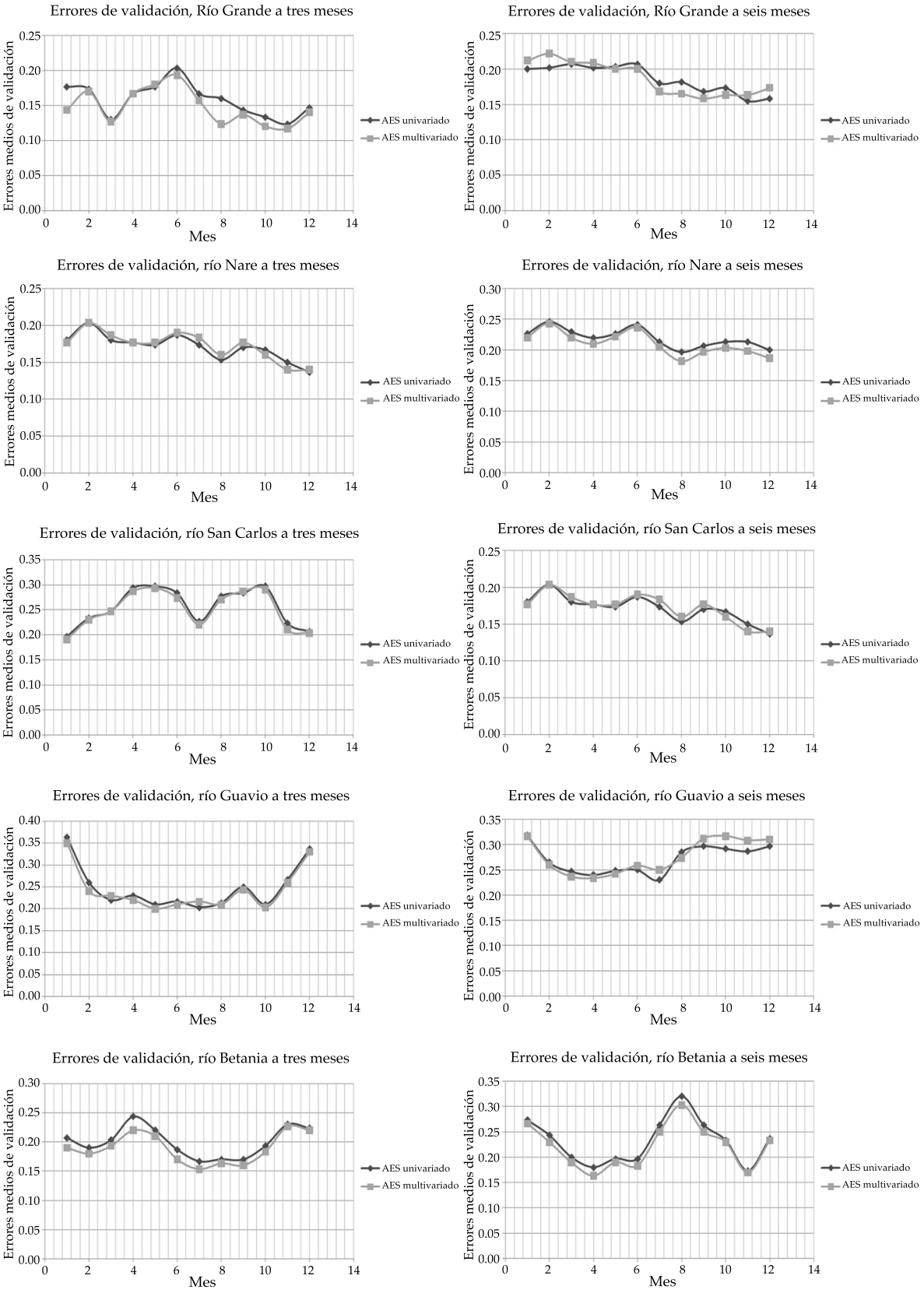

Figura 6. Errores medios de validación para los diferentes meses predichos en los seis ríos.

tanto, la serie de caudales refleja de manera más evidente las variaciones debidas a los fenómenos macroclimáticos de gran escala.

En cuanto a la relación existente entre las series de caudales y las variables macroclimáticas, el tamaño de la ventana se vuelve irrelevante cuando se utilizan tamaños mayores a 12 meses. Los resultados son básicamente los mismos que los obtenidos para 24 y 36 meses, según el análisis de sensibilidad para el tamaño de la ventana y por tanto se aconseja para cualquier análisis de series de caudales utilizar una ventana de 12 meses.

El modelo univariado es una buena aproximación para la predicción, sin embargo en ocasiones no logra representar adecuadamente la variabilidad de la serie y por tanto presenta mayores errores que si se consideraran las variables macroclimáticas. La utilización de variables macroclimáticas mejora la capacidad de predicción entre un 3 y 10%, aproximadamente. Permiten disminuir el error global y los errores mensuales hasta en un 5%, con respecto al error medio mensual. Por tanto, aunque los procedimientos gastan un poco más de tiempo y memoria del computador, los resultados se ajustan mejor en la validación a los datos históricos y los errores de validación disminuyen de forma significativa.

Recibido: 24/04/09
Aprobado: 02/06/10

Referencias

ARIAS, P.A. *Diagnóstico y predicción de la variabilidad intra-anual de la hidrología colombiana*. Tesis de Maestría en Ingeniería-Recursos Hidráulicos. Medellín: Universidad Nacional de Colombia, 2005, pp. 10-207.

CARVAJAL, L.F. *Acerca de la predicción no lineal en hidrología*. Tesis de Maestría. Medellín: Facultad de Minas, Universidad Nacional de Colombia, 1994, 136 pp.

CARVAJAL, L.F., MESA, O. y POVEDA, G. Predicción hidrológica en Colombia mediante análisis espectral singular y máxima entropía. *Ingeniería hidráulica en México*. Vol. XII, núm. 1, enero-abril de 1998, pp. 7-16.

GHIL, M., ALLEN, M., DETTINGER, M., IDE, M., KONDRASHOV, D., MANN, M., ROBERTSON, A., SAUNDERS, A., TIAN, Y., VARADI, F. and YIOU, P. Advanced Spectral Methods for Climatic Time Series. *Review of Geophysics*. American Geophysical Union. Vol. 40, no. 1. 2002, pp. 1-41.

MESA, O., SMITH, R., SALAZAR, J.E. y CARVAJAL, L.F. Modelos de predicción de caudales para el sector eléctrico colombiano. Parte I. *Avances en Recursos Hidráulicos, 3*. Agosto de 1995. Medellín: Universidad Nacional de Colombia, Sede Medellín, Facultad de Minas, 1995.

MESA, O.J., POVEDA, G. y CARVAJAL, L.F. *Introducción al Clima de Colombia*. Bogotá: Imprenta Universidad Nacional de Colombia, 1997, 390 pp.

PLAUT, G. and VAUTARD, R. Spells of low-frequency oscillations and weather regimes in the Northern hemisphere. *Journal of the Atmospheric Sciences*. Vol. 51, no. 2, 1994, pp.210-236.

POVEDA, G. La hidroclimatología de Colombia: una síntesis desde la escala inter-decadal hasta la escala diurna. *Rev. Acad. Colomb. Cien*. Vol. 28, núm. 107, 2004, pp. 201-219.

POVEDA, G., MESA, O.J., CARVAJAL, L.F., HOYOS, C.D., MEJÍA, J.F., CUARTAS, L.A. y PULGARÍN, A. Predicción de caudales medios mensuales en ríos colombianos usando métodos no lineales. *Meteorología Colombiana*. Vol. 6, 2002, pp. 101-110.

SMITH, R.A., VÉLEZ, J.I., VELÁSQUEZ, J.D., CEBALLOS, A., CORREA, P.L., GÓEZ, C., HERNÁNDEZ, O.O., SALAZAR, L.F., y ZAPATA., E.C. Modelos de predicción de caudales mensuales para el sector eléctrico colombiano. *Avances en Recursos Hidráulicos 11*. Facultad de Minas, septiembre de 2004, pp. 91-102.

VAUTARD, R., YIOU and GHIL. Singular-spectrum analysis: A toolkit for shot, noisy chaotic signals. *Physica*. D 58, 1992, pp. 95-126.

Abstract

ROJO-HERNÁNDEZ, J.D. & CARVAJAL-SERNA, L.F. Nonlinear river flow prediction using macroclimatic variables and singular spectral analysis. Water Technology and Sciences, *formerly* Hydraulic engineering in Mexico *(in Spanish). Vol. I, No. 4, October-December, 2010, pp. 59-73.*

The close nonlinear relationship between Colombian hydrology and global climatic processes justifies the study of nonlinear river flow models. This paper shows the nonlinear river flow dynamics in Colombia using a periodic prediction model based on the Singular Spectrum Analysis (SSA). The prediction window length is 3 and 6 months. The prediction includes both univariate and multivariate cases. The multivariate case takes into account macro-climatic variables. The proposed model is studied through the Principal Component Analysis, the correlation structure between macro-climatic variables and river flow time series and the reconstruction of series. This model is applied to the San Carlos, Riogrande II, Guatapé, Magdalena, Guavio, and Batá rivers. The results show an improvement in river flow prediction and indicate that river flow multivariate prediction using macro-climatic variables is better.

Keywords: *Singular Spectral Analysis (SSA), time series modeling, nonlinear river flow prediction, prediction with exogenous variables.*

Dirección institucional de los autores

Ing. Julián David Rojo-Hernández
Ing. Luis Fernando Carvajal-Serna

Universidad Nacional de Colombia
Escuela de Geociencias y Medio Ambiente
Carrera 80 núm. 65-223, bloque M2 Oficina 301
Medellín, Colombia
teléfonos: +57 (4) 5710 147 y 8634 073
fax: +57 (4) 4255 100
jdrojoh@gmail.com, jdrojoh@unal.edu.co
lfcarvaj@unalmed.edu.co, lfcarvaj@bt.unal.edu.co