

MODELOS DE PREDICCIÓN DE CAUDAL UTILIZANDO VARIABLES MACROCLIMÁTICAS Y TÉCNICAS ESTADÍSTICAS MULTIVARIADAS EN EL VALLE DEL RÍO CAUCA



Yesid Carvajal Escobar, Ph.D.

Profesor Asociado

Escuela de Ingeniería de Recursos naturales y del
Ambiente. EIDENAR- Universidad del Valle
Cali, Colombia
yecarvaj@univalle.edu.co

Guillermo A. Correa, Ing.

Escuela de Ingeniería de Recursos naturales y del
Ambiente. EIDENAR- Universidad del Valle
Cali, Colombia

Flor Matilde Muñoz, Ing.

Escuela de Ingeniería de Recursos naturales y del
Ambiente. EIDENAR- Universidad del Valle
Cali, Colombia

RESUMEN

Esta investigación se enfoca a la revisión del estado del arte de los fenómenos océano-atmosféricos que afectan la hidrología Colombiana especialmente el fenómeno El Niño Oscilación del Sur (ENOS) que causa un impacto socioeconómico de primer orden en nuestro país, por lo tanto es importante abordar la temática, incluyendo las variables macroclimáticas asociadas a ENOS en los análisis de planificación hídrica. Los análisis incluyen revisión de técnicas estadísticas de análisis de consistencia de datos hidrológicos con el objetivo de conformar una base de datos de caudal mensual del río Cauca confiable y homogéneo. Se utilizan métodos estadísticos (análisis de datos multivariados) específicamente el análisis de componentes principales para involucrarlos en el desarrollo de modelos de predicción de caudales medios mensuales en el río Cauca con el enfoque lineal

* Recibido : Enero 15 2007 * Aceptado : Febrero 2 2007

como son los modelos autoregresivos AR, ARX y Armax y el enfoque no lineal redes neuronales artificiales.

PALABRAS CLAVE

Fenómeno ENOS, Variables Macroclimáticas, Redes Neuronales Artificiales.

ABSTRACT

The project this centred in the revision of the state of the art of the ocean-atmospheric phenomena that you affect the Colombian hydrology especially The Phenomenon Enos that causes a socioeconomic impact of first order in our country, it has not been sufficiently studied; therefore it is important to approach the thematic one, including the variable macroclimáticas associated to the Enos in the analyses of water planning. The analyses include revision of statistical techniques of analysis of consistency of hydrological data with the objective of conforming a database of monthly flow of the river reliable and homogeneous Cauca. Statistical methods are used (Analysis of data multivariante) specifically The analysis of principal components to involve them in the development of models of prediction of flows monthly means in the river Cauca involving the Lineal focus as they are the model autoregresivos AR, ARX and Armax and the focus non lineal Net Artificial Network.

KEYWORDS

Enos phenomena, macroclimatic variables, artificial neuronal networks.

1. INTRODUCCIÓN

El fenómeno de El Niño Oscilación del Sur (ENOS) es uno de los eventos de escala global que afectan directamente la hidroclimatología de Colombia, generando pérdidas económicas y pérdidas humanas por lluvias excesivas, que ocasionan destrucción de cosechas y desastres naturales como crecientes, avalanchas, inundaciones y derrumbes, en su fase fría (La Niña). Por el

contrario, en la fase cálida (El Niño) se presentan sequías extremas, causando déficit de suministros de agua potable, producción agrícola y energía eléctrica.

Las fases extremas del ENOS tienen un impacto socioeconómico de primer orden en Colombia que no ha sido suficientemente estudiado; por lo tanto, 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.

El método de Componentes Principales puede ser utilizado como herramienta de análisis para la simulación multivariada de la sequía meteorológica e hidrológica, incluyendo la asociación con variables macroclimáticas; esto es muy importante porque permite simular caudales teniendo en cuenta su asociación con el fenómeno Enos. En el contexto anterior se planteó el siguiente trabajo de investigación para el desarrollo de modelos de predicción de caudales medios mensuales en el río Cauca. Se realizaron ventanas de predicción de mediano plazo de 1, 2 meses empleando los resultados de la predicción hidroclimática (relación con parámetros oceánico-atmosféricos) y la teoría de sistemas; aprovechando toda la información que actualmente está disponible en la región del Valle del Cauca a través de la Corporación regional del Valle del Cauca (CVC).

2. METODOLOGÍA

2.1. DESCRIPCIÓN Y LOCALIZACIÓN DE LA ZONA DE ESTUDIO

La zona de estudio está comprendida entre el embalse de Salvajina en el departamento del Cauca y Cartago en el departamento del Valle del Cauca, aproximadamente entre los 2°00' y 5°00' N y entre los 75°41' y 77°33' W, (Figura 1).

El caudal promedio multianual del río Cauca es de 279 m³/s en la estación Juanchito. El aporte total promedio de los ríos provenientes de las cordilleras Central y Occidental es de 277 m³/s y su recorrido total en el departamento del Valle del Cauca tiene una longitud de 340 km (CVC, 1998).

2.2. ANÁLISIS DE LA INFORMACIÓN HIDROLÓGICA

Para el análisis de la información hidrológica se utilizaron 8 estaciones hidrométricas seleccionados

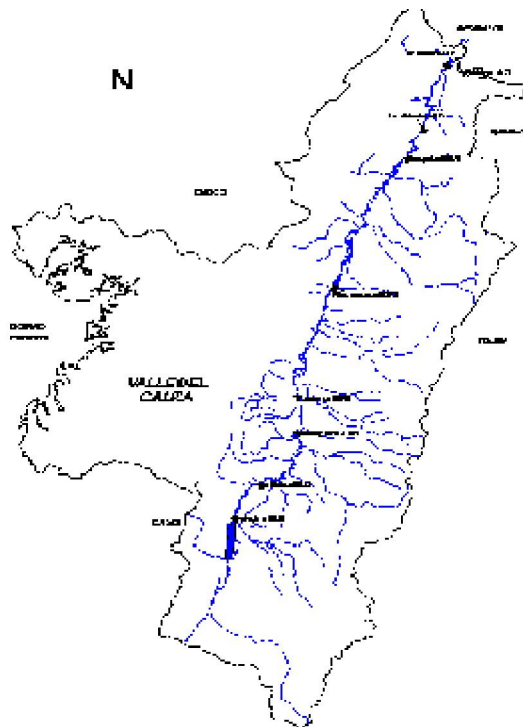


Figura 1 Localización de la zona de estudio

con base en diferentes criterios, tales como: cobertura, periodo de registro y calidad de la información (Carvajal, 2004). Se escogió un periodo común de registros para su análisis, comprendido entre los años 1951 y 2000. En la Tabla 1 se presenta el listado de estaciones hidrométricas seleccionadas para el análisis.

2.3. ANÁLISIS EXPLORATORIO DE DATOS (AED)

El análisis exploratorio de datos fue aplicado a los caudales medios mensuales de las estaciones seleccionadas con el objeto de evaluar su estructura, consis-

tencia, presencia de valores atípicos y relaciones entre las variables analizadas, como incrementar el entendimiento sobre los datos y disminuir los grados de incertidumbre al evaluar la validez de los resultados obtenidos en los modelos de predicción.

.Restitución de los caudales mensuales del río Cauca a régimen natural

El régimen hidrológico del río Cauca ha experimentado desde 1985 una variación, principalmente en sus caudales extremos, cuando entró en operación el embalse de Salvajina. Este cambio en el régimen de caudales y niveles es una consecuencia del efecto regulador del embalse. Se utilizó una técnica clásica de comparación conocida como análisis de las curvas de Doble masas o doble acumulada, con el fin de corregir las tendencias producidas por el embalse Salvajina, para obtener datos consistentes e involucrarlos en el modelo de predicción; los caudales fueron restituidos a régimen natural.

.Método Gráfico

Los métodos gráficos se desarrollaron para las series hidrológicas con el objeto de verificar si tenían cambios

Grafico serie de tiempo estacion

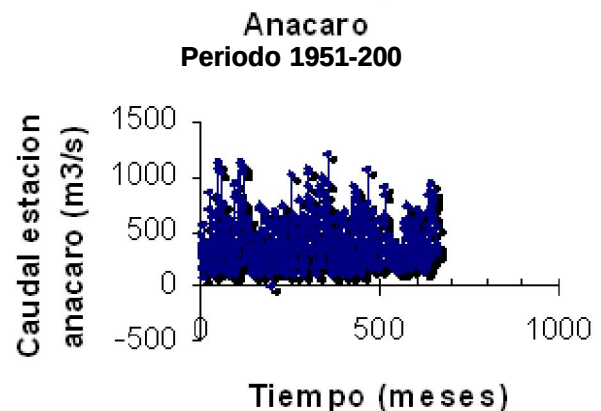


Tabla 1 Estaciones hidrométricas seleccionados y caudales característicos

Río	Estación	Área (Km ²)	Caudal Promedio multianual (m ³ /s)	Caudal Máx. promedio multianual (m ³ /s)	Caudal Min. promedio multianual (m ³ /s)
Cauca	La Balsa	5.156	196	301	111
Cauca	Hormiguero	8.018	267	404	151
Cauca	Juanchito	8.584	279	434	150
Cauca	Mediacanca	12.186	322	514	171
Cauca	Guayabal	16.036	385	625	181
Cauca	La Victoria	16.284	386	638	189
Cauca	Anacaro	17.234	409	673	197
La Vieja	Cartago	2.852	97	287	88

otendencias en la media y la varianza (Gráfica de Series de Tiempo, Masa Simple, Masa Residual, S-S, Q-Q); estas herramientas permiten conocer información pertinente a la concentración y la variación de los datos. En la Figura 2 se presentan a manera de ejemplo las gráficas obtenidas en el método Gráfico en la estación Anacaro.

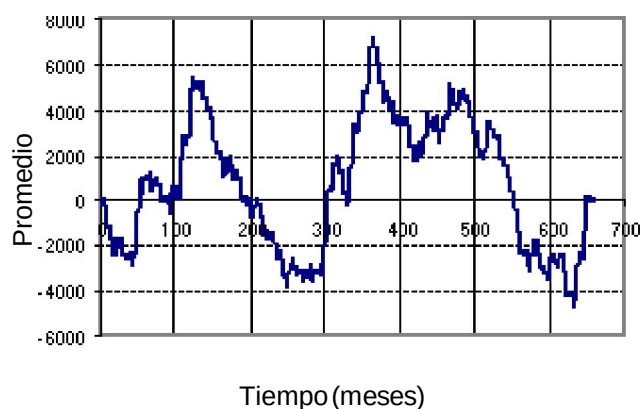


Figura 2. Análisis Exploratorio Método Gráfico

2.4. ANÁLISIS CONFIRMATORIO

Para el estudio de homogeneidad en la media y la varianza se usaron pruebas paramétricas y no paramétricas, pruebas de cambio y de tendencia. La elección de la prueba estadística consistió en definir una medida, d , que permitiera comparar los resultados obtenidos en la muestra con los resultados teóricos según la hipótesis planteada.

2.5. AUTOCORRELACIÓN DE CAUDAL

Se realizó el análisis de autocorrelación de caudal a las series de caudales medios mensuales en cada una de las estaciones de estudio con el objetivo de identificar la estructura que se genera; es decir, cómo influyen las observaciones del pasado en las observaciones del futuro (memoria que guarda la serie), conocida como función de autocorrelación. Los límites de significancia fueron estimados mediante un proceso MA (m) de 1000 desviaciones gaussianas.

2.6. CORRELACIÓN CRUZADA DE CAUDAL Y VARIABLES MACROCLIMÁTICAS

La variabilidad de la hidrología de la región involucra la influencia de diversas variables macroclimáticas, por lo cual es útil cuantificar el grado de dependencia lineal entre las mismas a través de coeficientes de correlación

cruzada a diferentes rezagos en el tiempo. Por tal motivo, se calculó la función de correlación cruzada entre el caudal con respecto a las variables macroclimáticas, estableciendo los límites de independencia mediante un proceso MA (m) de 1000 desviaciones gaussianas.

Se consideraron 7 variables macroclimáticas del océano para analizar las correlaciones cruzadas con la hidrología de la región; dichas variables fueron: el índice de oscilación del sur (SOI) y las temperaturas superficiales del océano Pacífico ($^{\circ}\text{C}$) en las regiones como son, Niño 1-2 (0° - 10° S, 90° - 80° W), Niño 3 (5° N - 5° S, 150° O - 90° W), Niño 4 (5° N - 5° S, 160° E - 150° O), Niño 3-4, Hawaii (THAWAI) y costa oeste de América (TCOAS). El grupo de variables de base para el estudio se caracteriza por presentar correlaciones significativas entre sí, por lo cual es necesario un posterior análisis de Componentes Principales para estudiar su variabilidad conjunta y las relaciones que tienen con la hidrometeorología de la región.

2.8. COMPONENTES PRINCIPALES

La base de datos de las variables macroclimáticas se obtuvo del archivo de datos climáticos de la JISAO (Joint Institute for the study of the atmosphere and ocean), instituto adjunto de la NOAA y la Universidad de Washington para el estudio conjunto del océano y la atmósfera y el Centro de predicción Climática de la NOAA cuyas direcciones electrónicas se presentan a continuación: http://tao.atmos.washington.edu/data_sets/#time_series <http://www.cpc.ncep.noaa.gov/data/indices/>

Se seleccionó un grupo original de 14 variables macroclimáticas medidas en los océanos Pacífico y Atlántico. Según los estudios realizados anteriormente por Carvajal, (2004), se recomienda trabajar con las variables macroclimáticas que mayor relación tienen con la hidrología de la región, para el caso del Valle del Cauca aquellas medidas o estimadas en el océano Pacífico. Este filtro eliminó a 3 variables en el océano Atlántico: las temperaturas superficiales del océano Atlántico en los hemisferio Norte (NATL), Sur (SATL) y en la región tropical del Atlántico y mar Caribe (TROP). Con el grupo de 11 variables restantes se procedió a la aplicación de la técnica de componentes principales, por lo tanto en esta fase se encontró que había variables que presentaban correlaciones significativas (coeficientes de correlación bajos) y uno de los objetivos del ACP es tratar de simplificar la estructura del fenómeno estudiado, proporcionándole una estructura lo más simple posible, que permita una interpretación más fácil;

Tabla 2. Matriz de correlaciones entre las variables macroclimáticas

		Matriz de correlaciones						
		NIÑO12	NIÑO3	NIÑO34	NIÑO4	THAWAI	TCOAS	SOI
Correlación	NIÑO12	1.000	.869	.716	.493	.278	.576	-.394
	NIÑO3	.869	1.000	.941	.728	.369	.511	-.448
	NIÑO34	.716	.941	1.000	.880	.416	.440	-.447
	NIÑO4	.493	.728	.880	1.000	.454	.349	-.390
	THAWAI	.278	.369	.416	.454	1.000	.560	-.682
	TCOAS	.576	.511	.440	.349	.560	1.000	-.673
	SOI	-.394	-.448	-.447	-.390	-.682	-.673	1.000
Sig. (Unilateral)	NIÑO12		.000	.000	.000	.000	.000	.000
	NIÑO3	.000		.000	.000	.000	.000	.000
	NIÑO34	.000	.000		.000	.000	.000	.000
	NIÑO4	.000	.000	.000		.000	.000	.000
	THAWAI	.000	.000	.000	.000		.000	.000
	TCOAS	.000	.000	.000	.000	.000		.000
	SOI	.000	.000	.000	.000	.000	.000	

con base en este criterio fueron eliminadas nuevamente las variables que presentaron correlaciones bajas por estar representadas por variables que aportaban una mayor comprensión del fenómeno estudiado. En esta fase fueron eliminadas las variables de presión atmosférica sobre el nivel del mar medidas en Darwin y Tahití, la Oscilación Decadal del Pacífico (PDO), el Índice de la lengua fría (CTI), y la Radiación de Onda Larga (OLR). Por lo tanto, del conjunto total de variables macroclimáticas fueron escogidas finalmente 7 que son objeto del estudio y se relacionan a continuación: el SOI las temperaturas sobre la superficie del mar de las diferentes regiones en el océano Pacífico

El grupo de variables de base para el estudio se caracteriza por presentar correlaciones significativas entre sí (Tabla 2), por lo cual es necesario un posterior análisis de componentes principales para estudiar su variabilidad conjunta y las relaciones que tienen con la hidrometeorología de la región.

2.8. TEORÍA DE SISTEMAS

. Identificación de Sistemas

La identificación de sistemas puede definirse como el área de la teoría de sistemas que estudia las metodologías para la obtención de modelos matemáticos de sistemas dinámicos a partir de mediciones sobre el sistema (ver Figura 3). En la modelación es conveniente distinguir entre las señales de entrada (independientes) y las de salidas (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) de las variables macroclimáticas corresponden a las entradas (variables exógenas).

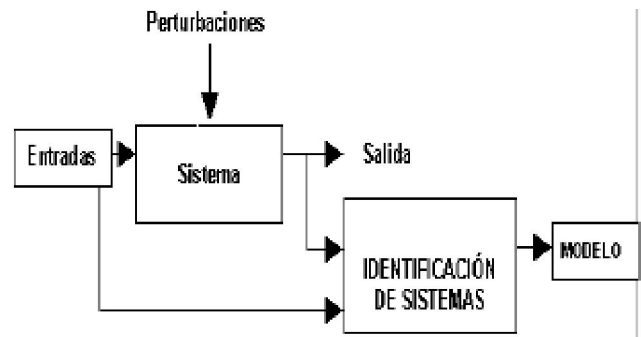


Figura 3. Diagrama de entradas y salidas en la identificación de un sistema

. Estructura de los modelos

Se ajustaron dos tipos de modelos: autoregresivos con variables externas, ARX y ARMAX, autoregresivos de media móvil, con variables externas o exógenas incluyendo las CP de las variables macroclimáticas asociadas a El Enos como variables exógenas (Ljung, 1999 y Box y Jenkins, 1970).

2.9. PREDICCIÓN DE CAUDALES EN EL RÍO CAUCA USANDO REDES NEURONALES

Las series analizadas fueron corridas en las Redes Neuronales Artificiales (RNA). Se usaron como predictores los registros

históricos de caudal y las componentes principales de las variables macroclimáticas. Se utilizó la información mensual de 1951 al 2000 de las estaciones hidrométricas del estudio.

Como un primer paso se elaboró un programa para una Red Neuronal Artificial (trainbr, regularización bayesiana). El número de entradas a la red y el número de neuronas en la capa oculta se determinó basado en la metodología utilizada por Granados, et al, (2004). La capa de entrada de la RNA está conformada por las neuronas que reciben la señal de entrada del exterior y que corresponden a las variables independientes del problema. En este caso se utilizaron como entradas los caudales y las CP de las variables macroclimáticas. La capa de salida entrega la señal procesada por la RNA y corresponde al caudal a predecir.

Luego se creó un archivo de patrones para el entrenamiento y otro para la validación con la información de caudales medios mensuales y las CP de las variables macroclimáticas; es fundamental mencionar que debido a la naturaleza de la información se normalizaron los datos de entrada y salida para que se ajusten en el rango entre 0 y 1, ya que la función de activación (función sigmoidea) arroja resultados entre 0 y 1).

Para normalizar la entrada y la salida a cada dato se le restó la media y se dividió por la desviación estándar. El procedimiento anterior que se conoce como estandarización ó escalamiento de los datos y es necesario para evitar que el proceso de aprendizaje sea muy lento y garantizar una adecuada velocidad de convergencia; la disposición general de los patrones se presenta en la Tabla 3.

Finalmente se corrió la Red; el número de neuronas en la capa de entrada fue de 36; ya que el orden de regresión de los parámetros de entrada es de 12 meses y en la capa oculta se tomaron 5, 7, 12 y 23 neuronas, determinadas a partir de las siguientes recomendaciones: el promedio geométrico de neuronas empleadas entre la capa de entrada y de salida, manejar un nivel de compresión adecuado y tratar de utilizar el menor número de neuronas en la capa oculta (Caicedo, 2005).

Para la predicción del primer mes, el entrenamiento de la red se realizó con un conjunto de datos entrada – salida conocidos (patrones), para que una vez finalizado el entrenamiento, ésta sea capaz de aproximar o calcular las salidas para los nuevos patrones de entrada presentados.

Para realizar el entrenamiento del modelo se empleó el 75% de los patrones generados (1951 a 1984) y el 25% restante (1985 a 2000) se empleó en la validación del mismo; Estos porcentajes son los recomendados ya que permite que la red tenga una mejor capacidad de comprensión de los datos en la Tabla 3 se observa la disposición de los patrones en las matrices de entrenamiento y validación el superíndice representa el número de variable y el subíndice el número de dato en la variable; lo cual, significa que los caudales y las CP de los doce meses anteriores predicen el caudal del mes que sigue.

Por ejemplo los caudales mensuales y las CP de los meses de enero a diciembre de 1951, predicen el caudal mensual de enero de 1952.

	ENTRADAS			SALIDAS
	Estación Hidrométrica (Q^1)	CP1 ($CP1^2$)	CP2 ($CP2^3$)	Estación Hidrométrica (Q)
1	$Q^1_1 \dots Q^1_{12}$	$CP^2_1 \dots CP^2_{12}$	$CP^3_1 \dots CP^3_{12}$	Q_{13}
2	$Q^1_2 \dots Q^1_{13}$	$CP^2_2 \dots CP^2_{13}$	$CP^3_2 \dots CP^3_{13}$	Q_{14}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
npat	$Q^1_{ndat-12} \dots Q^1_{ndat-1}$	$CP^2_{ndat-12} \dots CP^2_{ndat-1}$	$CP^3_{ndat-12} \dots CP^3_{ndat-1}$	Q_{ndat}

Tabla 3. Disposición general de la matriz con el fin de entrenar y validar la red neuronal para la predicción caudales de las estaciones hidrométricas.

3. RESULTADOS Y DISCUSIÓN

3.1. ANÁLISIS CONFIRMATORIO

Las pruebas estadísticas se aplicaron con un nivel de significancia de 0,05. El resumen de los resultados obtenidos al aplicar las pruebas estadísticas se presenta en la Tabla 4; para las series de caudales medios mensuales. En esta tabla, la "A" significa que se acepta la prueba a un nivel de confianza del 95 % y la letra "R" significa que se rechaza la prueba a un nivel de confianza del 95 %.

Aplicando dichas pruebas a cada serie se identificaron posibles puntos de cambio. De las 8 estaciones seleccionadas para realizar el análisis, la estación Hormiguero presentó tendencias significativas en la media, que fueron corregidas; la estación Cartago presenta cambios significativos en la varianza, así lo indican las pruebas en las que se rechazan las hipótesis nulas en la igualdad de varianzas. Las otras estaciones no presentan cambios en la media y en la varianza, así como no presentan tendencias, en general, las 8 estaciones hidrométricas

consideradas para este estudio; 6 (75%) no tienen tendencia significativa en la media y en la varianza 2 (25%) presentaron tendencias significativas en la varianza; por lo cual se concluye que, en general, la información disponible es de buena calidad y puede ser utilizada para el ajuste de los modelos.

.Autocorrelación de Caudal

En la Figura 4 se muestran las funciones de autocorrelación de los caudales medidos en las estaciones hidrométricas de Salvajina, La Balsa, Hormiguero y Juanchito; se presentan coeficientes de autocorrelación (memoria) de orden 1 $\rho_1=0.657$, $\rho_1=0.705$, $\rho_1=0.685$ y $\rho_1=0.687$, respectivamente; estos indican la existencia de una relación significativa y positiva entre una observación y los valores precedentes de caudal del otro mes; es decir, un proceso de almacenaje en los acuíferos de las cuencas.

Tabla 4 Resultados de las pruebas aplicadas a los registros de caudales mensuales para detectar tendencias

Estacion	Hipótesis											
	1	2	3	4	5	6	7	8	9	10	11	12
La Balsa	A	R	A	A	A	A	A	A	A	A	A	A
Hormiguero	A	A	A	A	A	A	R	R	A	A	R	A
Juanchito	A	A	A	A	A	A	A	A	A	A	A	A
Mediacanoa	A	A	A	A	A	A	A	A	A	A	A	A
Guayabal	A	A	A	A	A	A	A	A	A	A	A	A
La Victoria	A	A	A	A	A	A	A	A	A	A	A	A
Anacaro	A	A	A	A	A	A	A	A	A	A	A	A
Cartago	R	R	A	A	A	A	A	A	A	A	A	A

1. Prueba de Levene	7. Prueba Z de Kolmogorov-Smirnov
2. Prueba F para dos varianzas	8. Prueba de Kruskal-Wallis
3. Prueba t para muestras independientes y varianzas iguales	9. Prueba de Friedman
4. Prueba t para muestras independientes y varianzas desiguales	10. Prueba de los signos
5. Prueba t para muestras relacionadas	11. Prueba de U Mann-Whitney
6. Prueba de Wilcoxon	12. Prueba de las medianas

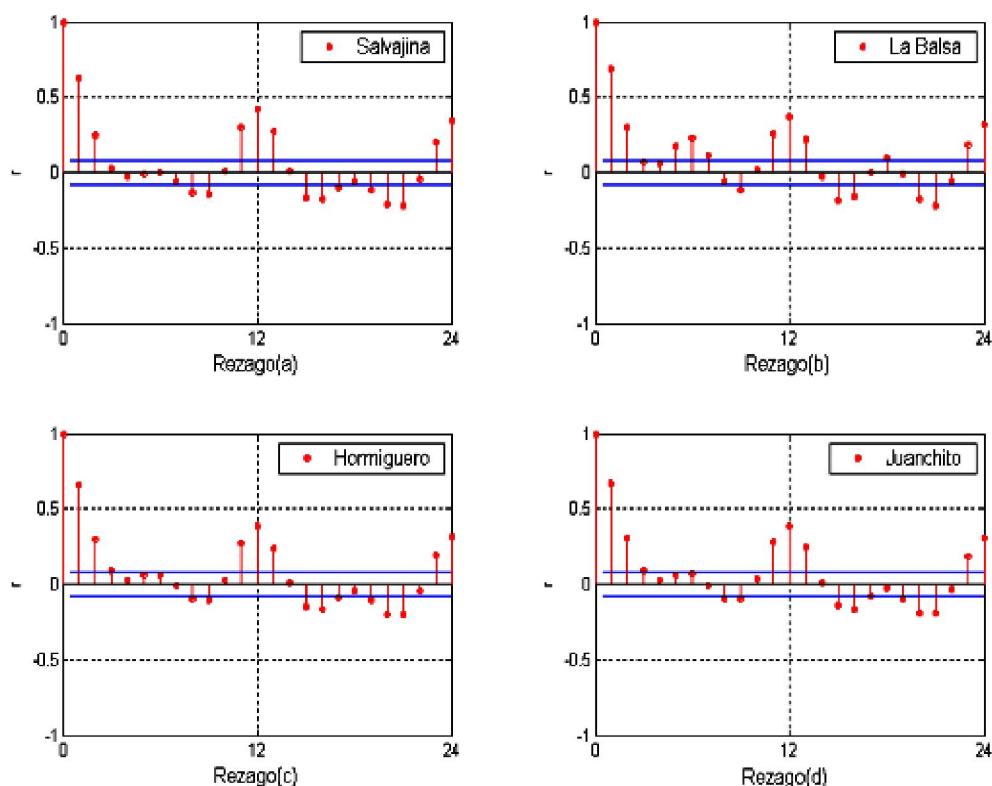


Figura 4 Funciones de autocorrelación: (a) Estación Salvajina, (b) Estación La Balsa, (c) Estación Hormiguero y (d) Estación Juanchito.

. Correlaciones con temperatura superficial de regiones del océano Pacífico

En las Figuras 4 y 5 obtenidos al estimar los coeficientes de correlación cruzada entre el caudal con los registros de temperatura sobre la superficie del océano en las regiones: Hawai, costa oeste de América, Niño 1-2, Niño 3, Niño 4, Niño 3-4; la figura 5 presenta la función de correlación cruzada de caudal con la temperatura superficial del océano en la región Niño 1-2.

Se observa que esta variable presenta una gran correlación entre el caudal y la temperatura superficial del mar en la región.

Se obtuvieron valores muy consistentes con correlaciones cruzadas máximas que oscilaron entre -0.510 y -0.684 a un rezago anticipado de -3 meses con la hidrología de la región; las mayores correlaciones se obtuvieron para las regiones Niño 4 y Niño 3-4 en el océano Pacífico.

Estas variables pueden ser utilizadas como posibles predictoras del caudal en la región porque presentan las correlaciones más altas a rezagos anticipados.

. Componentes Principales

Las correlaciones entre las variables fueron significativas y, por tanto, se justifica buscar otras dimensiones ortogonales que definan aspectos independientes. Las pruebas de adecuación aplicadas se relacionan a continuación:

El nivel de significancia de los conjuntos de variables fue de 0.05 y a este nivel todos pasaron la prueba. En la prueba de esfericidad de Bartlett se obtuvieron valores altos, entre 4765.20 (ver Tabla 5) asociados a una probabilidad de 0.00. En este caso, se rechazó la hipótesis nula, consistente en que las matrices de correlación de los diferentes grupos de variables son matrices identidad, por lo cual se aconseja efectuar el análisis de FOES. Se presentan los resultados obtenidos para el conjunto de variables Macroclimáticas; el índice obtenido en las variables fue considerado bueno o excelente. Los KMO fueron superiores a 0.685 (valor medio), lo que indica que el análisis de FOES es pertinente y puede proporcionar conclusiones satisfactorias.

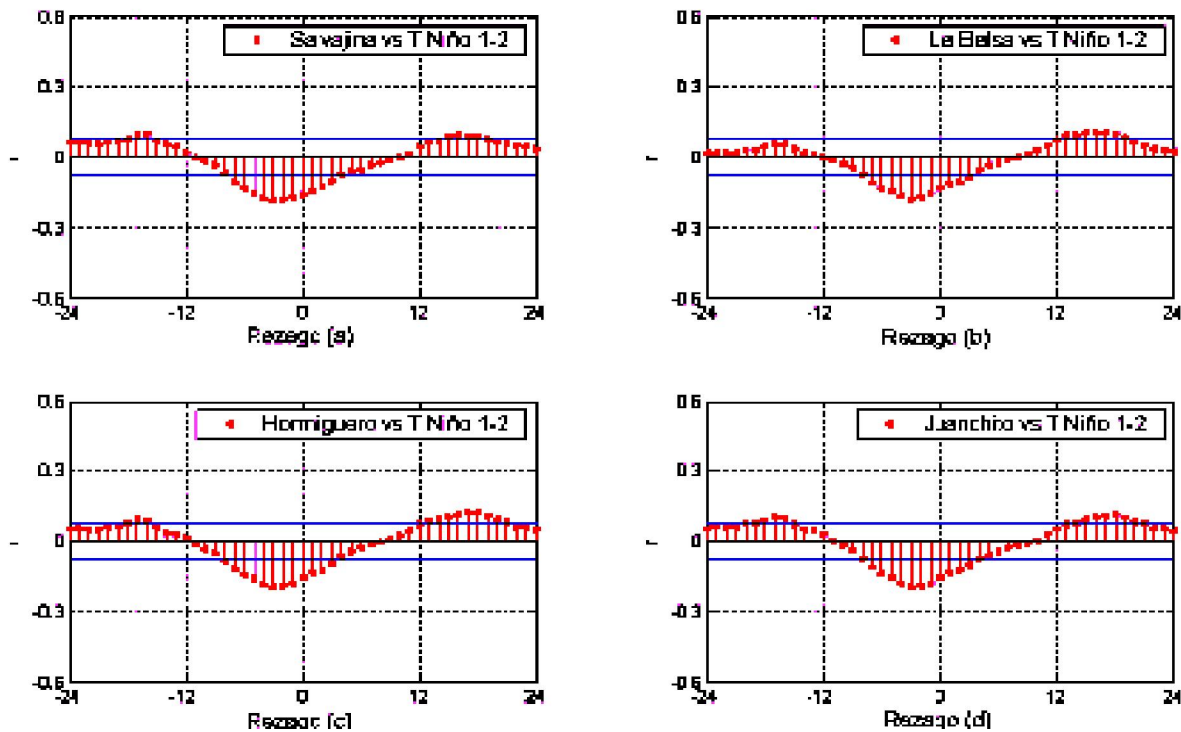


Figura 5. Correlación Cruzada de Caudal y Variables macroclimáticas: (a) Estación Salvajina vs. T Niño 1-2. (b) Estación La Balsa vs. T Niño 1-2. (c) Estación Hormiguero vs. T Niño 1-2. (d) Estación Juanchito vs. T Niño 1-2.

Tabla 5. Análisis de las condiciones de aplicación del Índice de Káiser Meyer Olkin (KMO) y prueba de Bartlett

KMO y prueba de Bartlett			
Medida de adecuación muestral de Káiser-Meyer-Olkin.			.685
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado		4765.205
	gl		21
	Sig.		.000

Análisis de Componentes Principales (ACP)

En la Tabla 6 se presenta la varianza explicada por las CP de caudal. Se presentan los cuadros del porcentaje de varianza explicado por cada CP de los diferentes conjuntos de variables.

Tabla 6. Varianza total explicada por cada CP de Caudal

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	4.362	62.313	62.313	4.362	62.313	62.313
2	1.258	17.970	80.284	1.258	17.970	80.284
3	0.713	10.187	90.471			
4	0.287	4.103	94.574			
5	0.271	3.878	98.452			
6	0.097	1.385	99.838			
7	0.011	0.162	100.000			

Las CP de caudal (Tabla 6) retienen un mayor porcentaje de varianza, la primera CP explica el 62.31% de la variabilidad total del conjunto de datos, mientras que la segunda CP explica el 80.28% de la variabilidad total. Estos resultados coinciden con los obtenidos en las matrices residuales, que indican un buen ajuste, asociado a un alto porcentaje de varianza total representada por la primera componente principal.

3.2. AJUSTE DE MODELOS PARAMÉTRICOS MULTIVARIADOS

Para cada una de las series de caudal se ajustaron

modelos paramétricos AR, ARX y ARMAX. El criterio para establecer el orden de los modelos se basó en los resultados obtenidos por las funciones de autocorrelación presentadas, lo cual indica que este tipo de series ajustan modelos con un rezago de 1 mes. Las ventanas de predicción a largo plazo no dan resultados satisfactorios porque la función de autocorrelación es muy débil. En las Figuras 6 y 7 se presentan la comparación del ajuste de modelos a los datos de validación el error final de predicción (FPE), ajustado a las series de caudal en cada estación: (a) sin variables exógenas, Modelos AR, (b) con la primera CP como variable exógena, Modelos ARX y ARMAX.

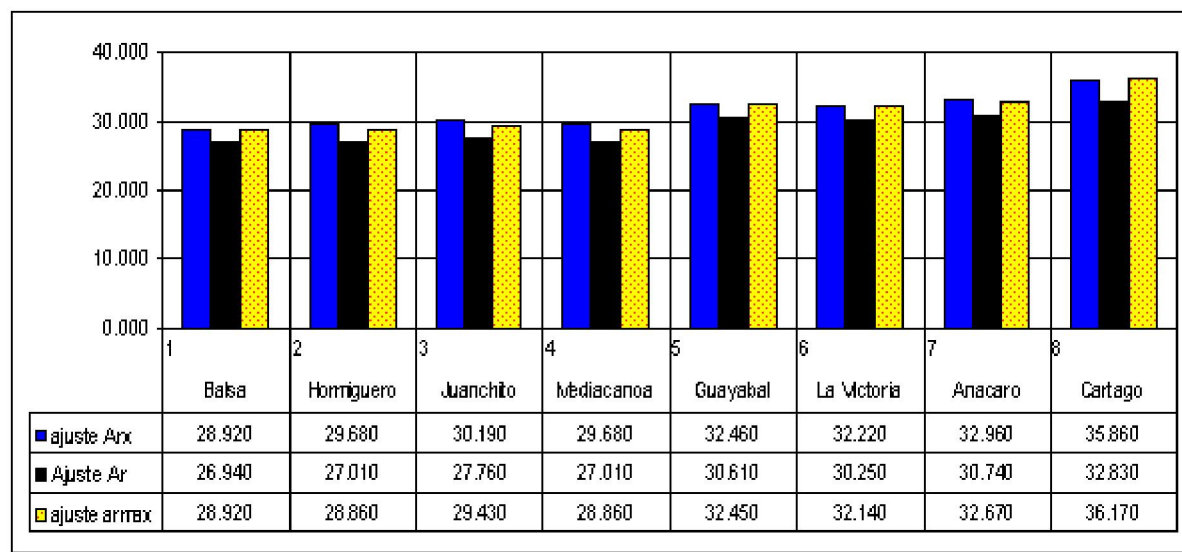


Figura 6 Criterio Gráfico de ajuste de Modelos a los datos de validación

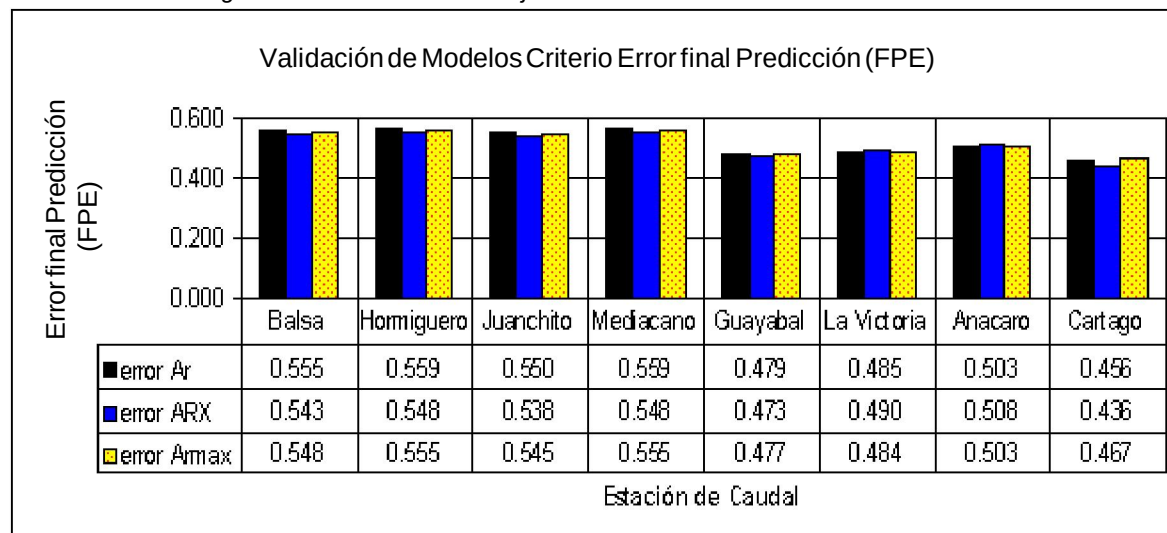


Figura 7 Criterio Error Final de Predicción (FPE)

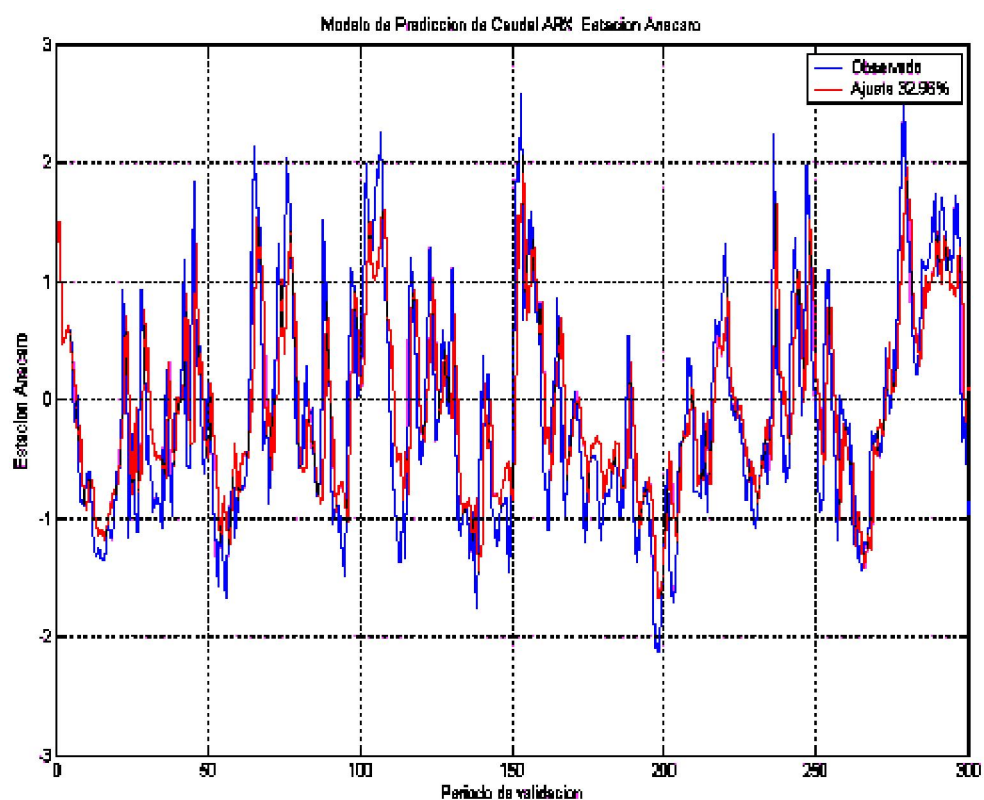


Figura 8 Ajuste y simulación de los datos de validación Modelo ARX (1,1,1) estación Anacaro

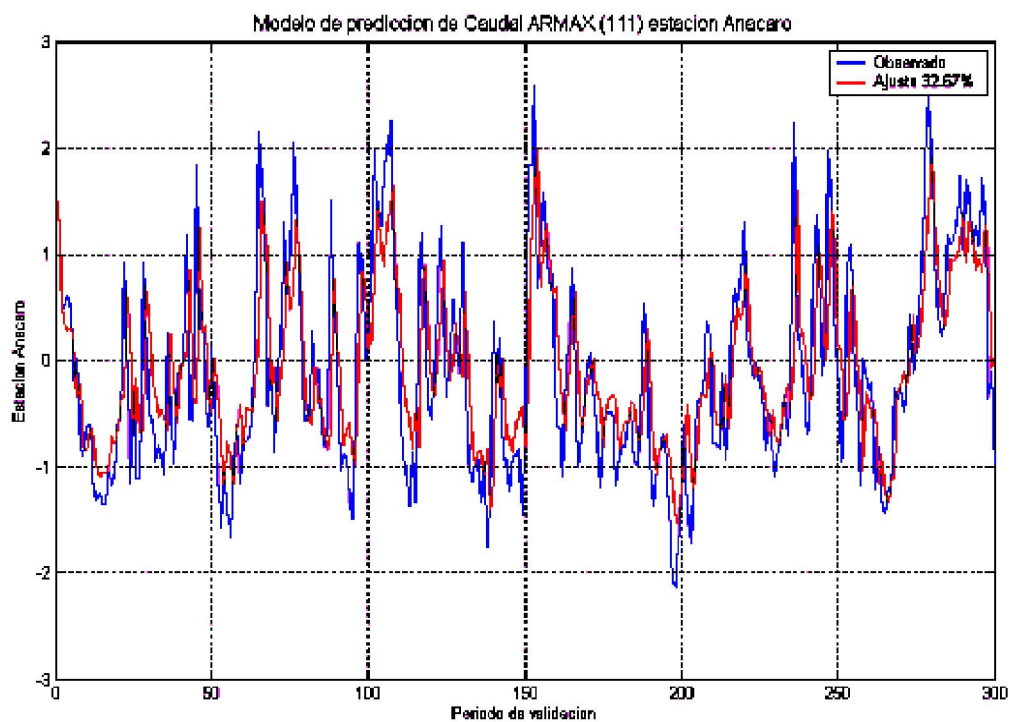


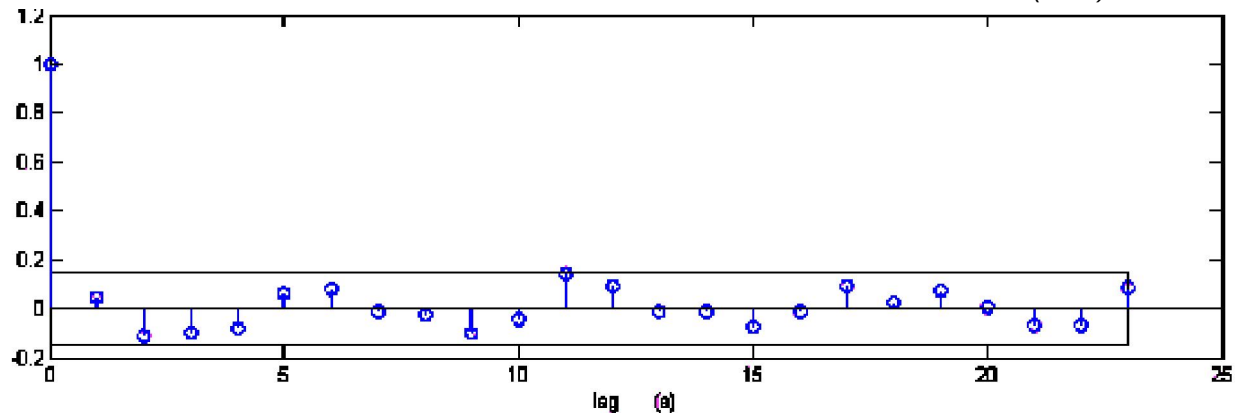
Figura 9 Ajuste y simulación de los datos de validación Modelo ARMAX (1,1,1,1) estación Anacaro.

De acuerdo con la metodología aplicada, los criterios utilizados en la validación de los modelos estimados se resumen en los Figuras 6 y 7, los cuales indican que se obtienen los menores errores de predicción (Criterio del error final de Predicción (-FPE)), cuando se ajustan modelos paramétricos incluyendo como variables auxiliares las primeras CP de las variables macroclimáticas; por lo tanto, permiten obtener modelos con menor error residual y que el porcentaje de variación explicada por el modelo presente mayor ajuste en los modelos que incluyen la primera componente principal (CP) de las variables macroclimáticas ARX; es decir, el sistema identificado mejora la representación de los datos en la validación cuando se aplica un modelo ARX(1,1,1). Lo anterior resulta lógico si se tiene en cuenta que la primera CP representa la máxima covariabilidad significativa entre las series de caudal y las variables macroclimáticas, mientras que a medida que aumentan

los CP en orden, representan menor covariabilidad significativa conjunta, por lo cual aportan menos a la predicción.

En la Figura 10 se presentan la función de autocorrelación de los residuales de los caudales (a) y la función de correlación cruzada (b) entre las entradas (CP de variables macroclimáticas) de la serie de caudal en la estación Anacaro; se observa que en la primera figura (a) los residuales están dentro de los intervalos de confianza del 99% presentados en la función de autocorrelación; esto indica que al realizar el análisis residual del modelo generado se consideran independientes y que el modelo ha recogido e identificado en buena forma las características del sistema. En la figura (b) se puede observar que para verificar esta independencia los rezagos positivos se encuentran dentro del intervalo de confianza; por lo tanto, el modelo ha sido descrito de forma apropiada. Se obtuvieron resultados similares en los demás análisis.

Función de autocorrelación de los residuos de la serie de caudal en la estación Anacaro Modelo ARMAX (1:1.11)



Función de correlación cruzada de CP 1 de la variable Macroclimática de la serie de caudal en la estación Anacaro Modelo ARMAX (1:1.11)

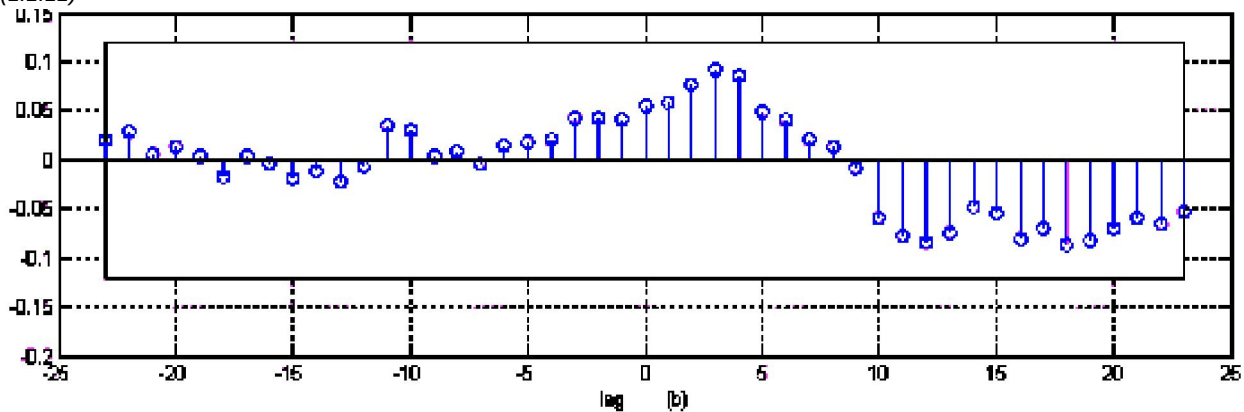


Figura 10 (a) Función de autocorrelación de los residuales de la serie de caudal de la estación Anacaro. (b) Función de correlación cruzada de CP1 de las variables macroclimáticas y los residuales de la serie en la estación Anacaro

3.3. PREDICCIÓN DE CAUDALES EN EL RÍO CAUCA USANDO REDES NEURONALES

Después de haber corrido la red neuronal para los tres meses y después de 50 iteraciones se obtuvieron los resultados registrados en la Tabla 7 para la predicción de caudales de las estaciones hidrométricas. Se puede observar que al trabajar la red con un algoritmo que incluye un término de regularización el error de validación para las predicciones del segundo y tercer mes tiende a bajar; esto puede ocurrir porque los datos de entrenamiento ingresados a la red presentan un ruido no muy bien distribuido y la red escoge los mejores valores para simularlos. Para la predicción del segundo mes se escogen los valores simulados con el menor error para el primer mes e igual para el tercer mes. Por esta razón, los gráficos de validación de las predicciones para el primer y segundo mes muestran una simulación casi perfecta. También se puede observar que al alimentar la red neuronal con variables de caudal y con las componentes principales de las variables macroclimáticas la predicción de caudales es buena.

Tabla 7 Medida del error de la Red para la predicción de caudales de las estaciones hidrométricas sobre el río Cauca.

Estación	Mes	Arquitectura NNE-NNCO- NNS	Número de Iteraciones	MSE (Validación)
LA BALSA	1	36-12-1	10	0.427
	2	36-7-1	10	0.153
	3	36-12-1	10	0.162
HORMIGUERO	1	36-5-1	20	0.415
	2	36-5-1	30	0.123
	3	36-12-1	20	0.128
JUANCHITO	1	36-23-1	10	0.406
	2	36-5-1	10	0.118
	3	36-7-1	20	0.118
MEDIACANOA	1	36-12-1	10	0.390
	2	36-5-1	10	0.122
	3	36-5-1	10	0.115
GUAYABAL	1	36-12-1	10	0.363
	2	36-12-1	10	0.155
	3	36-7-1	20	0.150
LA VICTORIA	1	36-23-1	10	0.360
	2	36-7-1	10	0.158
	3	36-7-1	10	0.149
ANACARO	1	36-12-1	10	0.344
	2	36-12-1	10	0.153
	3	36-5-1	10	0.140

NNE: Número de Neuronas de Entrada
NNS: Número de Neuronas de Salida
NCO: Número de Neuronas Capa Oculta
MSE: Error cuadrático medio

En la figura 11 se presentan los resultados de entrenamiento y validación del esquema de predicción con Redes Neuronales para los caudales en el río Cauca - estación Anacaro.

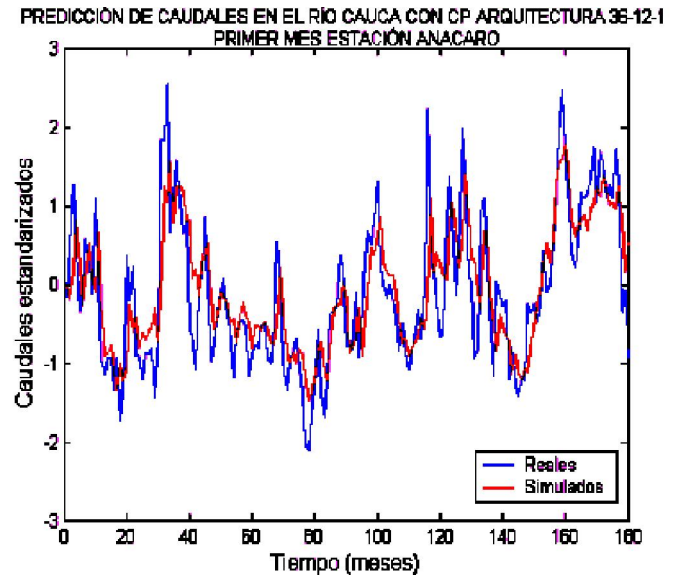


Figura 11 Validación de la Red Neuronal para predicción de caudales en el río Cauca - Estación Anacaro

3.4. COMPARACIÓN DE MÉTODOS LINEALES PARA PREDICCIÓN DE CAUDAL EN EL RÍO CAUCA

Se aplicaron dos metodologías para la elaboración de modelos de predicción de caudal primero utilizando métodos de predicción lineal como son los modelos Autorregresivos (AR, ARX y ARMAX), los cuales aplican relaciones lineales por lo tanto el método de cálculo es analítico.

También se utilizaron métodos de predicción no lineal (RNAS); y el método de cálculo utilizado es iterativo en el cual se empleó una arquitectura que se ajustaba a la estructura de este tipo de datos.

Cada metodología por separado produce modelos que se ajustan al comportamiento de los datos por lo tanto, para tener una mayor aproximación de los modelos estimados fueron sometidos a pruebas de validación

las cuales determinan la calidad y la capacidad de pronóstico de los modelos estimados. Para el caso de los modelos lineales se aplicaron pruebas de validación como el análisis de residuales, los errores finales de predicción y el ajuste; con estos análisis se pretende estudiar esa parte de los modelos que no puede ser explicada por éstos; por ejemplo, los residuales miden las diferencias entre los valores observados y los valores estimados por cada modelo.

En el caso de los errores de predicción, la minimización se realizó utilizando un criterio cuadrático, arrojando resultados satisfactorios. Para las redes neuronales se utilizó un algoritmo Backpropagation - trainbr, el cual minimiza el error de propagación hacia atrás; esta función permite minimizar el número de parámetros utilizados para predecir el comportamiento de los datos empleados en la simulación; por lo tanto, se produce un error mínimo entre las salidas estimadas por el modelo y los datos observados en la realidad; además, con esta clase de algoritmos, la técnica es comparable debido a que ambas utilizan el principio de parsimonia, el cual sugiere construir modelos que incluyan el menor número de parámetros.

Al comparar el error de validación (Tabla 8) de los modelos utilizados para predecir los caudales del río Cauca con las componentes principales de las variables macroclimáticas, se puede observar que los errores más bajos de predicción se obtienen con los modelos no lineales (Redes Neuronales).

Esto puede ocurrir puesto que los modelos AR, ARX y ARMAX se basan en una dinámica lineal y el comportamiento de la naturaleza no obedece a leyes de tipo lineal.

4. CONCLUSIONES

Los métodos clásicos multivariados son una herramienta importante para el análisis y la predicción de caudales; sin embargo, tienen el inconveniente de la linealidad. Existen métodos modernos de predicción hidrológica que permiten mejorar la predicción, teniendo en cuenta la no linealidad; entre estos métodos se tienen las redes neuronales, las cuales ofrecen robustez en el manejo de series de datos ruidosas.

Las técnicas y las pruebas estadísticas aplicadas a las series de caudales medios mensuales del río Cauca sirvieron para detectar señales, inconsistencias y tendencias que al corregirse permitieron conformar una base de datos confiable y homogénea. Esta base se utilizó para ajustar modelos paramétricos y redes neuronales asociándolas con las variables macroclimáticas.

La técnica utilizada en la restitución de caudales mensuales a régimen natural por la curva de doubles masas indica que el efecto regulador del embalse Salvajina no produce cambios considerables en los caudales de las estaciones en estudio; lo corroboran los resultados obtenidos en los análisis exploratorios y confirmatorios que se realizaron a cada una de las estaciones.

De las ocho estaciones hidrométricas consideradas para este estudio 6 (75%) no tienen tendencia significativa en la media y en la varianza, dos (25%) presentaron tendencias significativas en la varianza; por lo cual se concluye que, en general, la información disponible es de buena calidad y puede ser utilizada para el ajuste de los modelos.

Tabla 8 Error de mínimos cuadrados para la predicción de caudales en el río Cauca utilizando modelos lineales y no lineales

Estación	Mes	Error de validación			
		AR	ARX	ARMAX	Redes Neuronales (segundo esquema)
LA BALSA	1	0.555	0.543	0.548	0.427
HORMIGUERO	1	0.559	0.548	0.565	0.415
JUANCHITO	1	0.55	0.538	0.545	0.406
MEDIACANOA	1	0.559	0.548	0.555	0.390
GUAYABAL	1	0.479	0.473	0.477	0.363
LA VICTORIA	1	0.495	0.49	0.494	0.360
ANACARO	1	0.503	0.508	0.503	0.344

Ambos tipos de modelos Lineales y No lineales han sido determinados a partir del mismo conjunto de datos observados; las redes neuronales han podido ajustar y considerar el conjunto de componentes más relevantes, por el hecho de que al determinar su estructura permiten juzgar y analizar componentes no lineales imposibles de considerar en los modelos lineales AR, ARX, ARMAX. Los modelos obtenidos por las dos metodologías propuestas pueden ser utilizados con fines de investigación de pronóstico o de explicación del fenómeno Enos en las estaciones hidrométricas del río Cauca.

La alta variabilidad climática de Colombia hace que la predicción de caudales sea difícil y que este sujeta a una incertidumbre muy grande. El desconocimiento del comportamiento de la variable y su incertidumbre se hacen mayores a medida que se amplía el horizonte de la ventana de predicción, pues ésta viene afectada por los errores de la predicción de los meses anteriores.

El ajuste de los modelos depende de la naturaleza y comportamiento de las variables, ya que si éstas tienen un comportamiento lineal se ajustarán mejor a los modelos lineales AR, ARX y ARMAX; por el contrario, si su comportamiento es no lineal se ajustarán mejor a los modelos no lineales (Redes Neuronales).

5. REFERENCIAS BIBLIOGRÁFICAS

- Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19:716–722.
- Box, G. E. and Jenkins, G. M. (1970). Time series analysis forecasting and control. 3rd ed. 1994. Holden-Day, San Francisco.
- Caicedo, B. E. (2005). Curso: Redes Neuronales Artificiales. Perceptrón y Perceptrón Multicapa (Algoritmo de Aprendizaje Backpropagation). Escuela de Ingeniería Eléctrica y Electrónica. Grupo Percepción y Sistemas Inteligentes. Febrero – Junio de 2005.
- Carvajal, Y. (2004). El uso de funciones ortogonales empíricas y análisis de correlación canónica en el estudio de la variabilidad hidrometeorológica. Aplicación al Valle del Cauca-Colombia. Universidad Politécnica de Valencia. Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos. Departamento de Ingeniería Hidráulica y Medio Ambiente. Valencia, España. Tesis doctoral.
- Cazalac (2004). Guía metodológica para la delimitación de zonas áridas, semiáridas y subhúmedas secas de América latina y el Caribe para realizar los análisis de datos elaborados por el Centro de Agua para zonas Áridas y Semiáridas para América Latina y el caribe (CAZALAC) y el Proyecto Hidrológico Internacional (PHI).
- CVC (1998). Estudio de regionalización de caudales máximos para diseño Corporación Autónoma Regional del Valle del Cauca.
- Dunne, T., L. Leopold. (1978). Water in Environmental Planning: W. H. Freeman et al., San Francisco, p.41.
- Granados, L. Y., Vargas, L. A., Obregón, N. N. (2004). Calibración de una red neuronal artificial en la predicción de caudales medios mensuales a partir de variables hidrometeorológicas. XXI congreso Latinoamericano de Hidráulica São Pedro, estado de São Paulo, Brasil.
- Guerrero, V. M. (1991). Análisis estadístico de series de tiempo económicas. Universidad Autónoma Metropolitana, México.
- Ljung, Lennart. (1999). System identification: Theory for the user, 2nd Edition, Prentice Hall, Englewood Cliffs, N.J.
- Ljung L. (1994). From Data To Model: A Guided Tour Of System Identification. Linköping University. Department of Electrical Engineering..
- Rostlac-Unesco. (1982). Guía metodológica para la elaboración del balance hídrico de América del Sur.
- Smith R. (2000). Campuzano Claudia. Análisis exploratorio para la detección de cambios y tendencias en series hidrológicas. XIV Seminario Nacional de Hidráulica e Hidrología. Villa de Leyva.
- Smith R., Poveda G., Mesa, O., Pérez, C. y Ruiz, D. (1996). En búsqueda de señales de cambio climático en Colombia. IV Congreso Colombiano de Meteorología. Santa Fé de Bogotá..
- Smith, A. R., Vélez, J. I., Velásquez, J. D., Cevallos, A., Correa, P. L., Góez, C. y Salazar, L. F. (2004) Algunos modelos de predicción de caudales para el sector eléctrico Colombiano. XXI Congreso Latinoamericano de hidráulica São Pedro, Estado de São Paulo, Brasil.