

MODELAMIENTO DE SERIES DE CAUDAL USANDO ANFIS

Elizabeth Catalina Zapata, Juan David Velásquez y Ricardo A. Smith

Escuela de Sistemas - Facultad de Minas

Universidad Nacional de Colombia, Sede Medellín

eczapata@unalmed.edu.co

Recibido para evaluación: 07 de Julio de 2004 / Aceptación: 03 de Septiembre de 2004 / Recibida versión final: 20 de Septiembre de 2004

RESUMEN

Este trabajo presenta ANFIS (Adaptative Network Based Fuzzy Inference System) como una técnica de modelamiento no lineal de series de tiempo, que en su formulación puede ser analizada como una generalización no lineal de los modelos autorregresivos, y a diferencia de otras metodologías permite incorporar conocimiento experto sobre la serie de tiempo. Recopila una metodología para la formulación de modelos ANFIS y presenta una comparación entre preprocesamientos de la información utilizados en series de tiempo y series hidrológicas, para definir cual es el más adecuado en este tipo de modelos, concluyendo que los mejores resultados se obtienen con la estandarización mensual utilizada en los modelos tradicionales para series de caudal. Por último compara los resultados obtenidos con los modelos ANFIS y los modelos tradicionales autorregresivos, encontrando que los primeros capturan mejor la estructura de la serie. Para la aplicación se utilizó la serie de caudales de la estación Embalse Betania ubicada en el departamento del Huila.

PALABRAS CLAVES: ANFIS, Series de Tiempo, Series de Caudal, Embalse Betania.

ABSTRACT

This work shows ANFIS (Adaptative Network Based Fuzzy Inference System) like a no linear modelling technique for time series, that in its formulation can be analysed like a no linear generalization of autoregressive models, and that, compared with other methodologies, lets incorporate expert knowledge about the time series. This work compiles a methodology for the formulation of ANFIS models and shows a comparison between pre-processing information used in time series and hydrological time series, for found which one is the most suitable in this kind of models, concluding that best results are obtained by monthly standardization used by traditional models for hydrological time series. Finally, it compares the results obtained by ANFIS models with traditional autoregressive models, founding that first one captures better the time series's behavior. For the application, it was used the flow time series of Betania's reservoir, located in Huila deparment.

KEY WORDS: ANFIS, Time Series, Hydrologic Time Series, Betania.

1. INTRODUCCIÓN

La hidrología juega un papel importante en distintos sectores de la sociedad, la economía y el medio ambiente, tales como el eléctrico, agrícola, ganadero, vías y transporte, calidad de agua y suministro de agua para consumo humano. Ello ha ocasionado que se considere el aprovechamiento adecuado del agua como un caso de estudio desde hace varias décadas (Svanidze, 1980). Se han presentado muchos avances, tanto cualitativos como cuantitativos, en esta área, basándose en la componente aleatoria del caudal, el uso de nuevos métodos matemáticos y los avances en la computación.

En Colombia se evidenció la necesidad de planear adecuadamente el uso de los recursos hídricos a partir de 1992, cuando la planificación inadecuada, la subadministración y el deterioro económico del sistema hidroeléctrico colombiano de generación de electricidad quedaron al descubierto con la depresión de los embalses debido al fenómeno de El Niño que se presentaba ese año, y obligaron al país a entrar en un racionamiento que puso en crisis todo el sector eléctrico. Desde entonces se evidenció la necesidad de desarrollar modelos más precisos de pronóstico para predecir situaciones hidrológicas críticas en el futuro. En este contexto se incluyen los trabajos de Poveda y Mesa (1993); Mesa y Poveda (1993); Carvajal (1994); Salazar, Mesa, Poveda y Carvajal (1994); Salazar y Mesa (1994); Poveda y Mesa (1994); Rendón (1997); Poveda, Mesa, Carvajal, Hoyos, Mejía, Cuartas y Pulgarín (2002); y Prada y Obregón (2004).

Se ha encontrado que la variabilidad del caudal en los ríos se ve influenciada por diferentes variables macroclimáticas. En general, las variables que más relación tienen con la hidrología colombiana son aquellas medidas y/o estimadas sobre el Océano Pacífico, como el Índice de Oscilación del Sur, Temperatura Superficial del Mar en las regiones Niño-3 y Niño-4 y Presiones en la Isla de Pascua (Mesa, Poveda y Carvajal, 1997).

Desde la década del sesenta, se han usado modelos lineales en hidrología y recursos hidráulicos. La familia de modelos propuesta en Box y Jenkins (1997), ha tenido múltiples y muy importantes aplicaciones en el desarrollo de los recursos hidráulicos. Sin embargo, la relación entre el clima y los procesos climáticos globales que muestran un comportamiento no lineal (Mesa et al., 1997), han

incentivado al estudio de otros modelos alternativos que reproduzcan algunas características no lineales en los datos. Se presentan entonces trabajos con técnicas no lineales como Análisis Espectral Singular y Redes Neuronales presentadas por Carvajal (1994), Modelo de Ecuación Diferencial Estocástica no lineal y Procesos Autorregresivos Dependientes del Régimen por Salazar et al. (1994), Predicción por Bandas Espectrales Usando Onditas (PREBEO) y Regresión Lineal Múltiple por Poveda et al. (2002).

El objetivo general de este trabajo es explorar a ANFIS (Jang, 1993) como una técnica de modelamiento de series de tiempo que ha sido poco usada en este campo y por consiguiente en hidrología, pero que en su formulación puede ser analizada como una generalización no lineal de los modelos autorregresivos, que a diferencia de otras metodologías permite incorporar conocimiento experto sobre la serie de tiempo (Velásquez, Dyner y Souza, 2004). Mientras que específicamente se desea desarrollar un sistema de predicción para la serie de caudales medios mensuales de la estación Embalse Betania ubicada en municipio de Palermo en el departamento del Huila.

La sección 2 introduce el modelamiento tradicional de series de tiempo, la sección 3 presenta el modelo ANFIS, la sección 4 contiene una descripción de los datos, la sección 5 contiene los resultados y la sección 6 presenta las conclusiones.

2. MODELAMIENTO TRADICIONAL DE SERIES DE CAUDALES

Los modelos autorregresivos han sido utilizados extensamente en hidrología desde la década de los sesenta, en la generación de series hidrológicas temporales (Svanidze, 1980). Sus principales ventajas están representadas en la forma intuitivamente autorregresiva de las series de tiempo, donde el valor de la variable en el presente depende de los valores previos en el tiempo; y la simplicidad de estos modelos (Salas, Delleur, Yevjevich y Lane, 1980). Sin embargo, el desarrollo matemático de estos modelos se unificó formalmente a partir de 1970 con los trabajos de Box y Jenkins (1970). En ellos, se propone que una serie de tiempo, de caudales en este caso, puede ser modelada a través de la función lineal paramétrica $f(\cdot)$ que en el caso de los modelos autorregresivos toma la forma de:

$$\hat{f}(x_t) = g(x_t | \{\alpha_j\}_0^U) = a_0 + \sum_{j=1}^U a_j \times x_t^{(j)} \quad (1)$$

donde los coeficientes $\{\alpha_j\}_0^U$ son estimados usando mínimos cuadrados:

$$\{\alpha\}_0^P = \arg \min_{\{\alpha\}_{t=0}^P} \sum_{t=1}^T \left[q_t - g(x_t | \{\alpha_j\}_0^P) \right]^2 \quad (2)$$

En (1) el vector de regresores x_t toma diferentes formas de acuerdo al tipo de modelo considerado. En el caso general, x_t está formada por valores rezagados de la serie real y/o los rezagos de otras variables causales, dependiendo de si los modelos son univariados, causales o multivariados.

Salas et al.(1980), indica que para la especificación de un modelo preliminar de la serie de tiempo es necesario hacer una transformación preliminar de la información que asegure su homogeneidad en el sentido estadístico, mediante transformaciones no lineales y de procesos de diferenciación y diferenciación estacional. Posteriormente, se seleccionan como regresores aquellos rezagos de la serie de tiempo que presenten correlaciones significativamente diferentes de cero.

3. ANFIS : ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM

Las redes adaptables neuro-difusas combinan los sistemas de inferencia difusos y las redes neuronales, aprovechando las características sobresalientes de cada modelo. Los sistemas de inferencia difusos proveen un mecanismo intuitivo y de alto nivel para representar el conocimiento experto mediante la utilización de reglas heurísticas *IF/THEN*. Por otro lado, las redes neuronales poseen un alto grado de adaptabilidad y capacidad de aprendizaje y generalización.

3.1. Arquitectura

Para describir la arquitectura de ANFIS se usará por simplicidad un sistema de inferencia difusa con dos entradas x y y y una salida z . Bajo el supuesto que la base de reglas contiene reglas de tipo Takagi-Sugeno (Takagi-Sugeno, 1982):

$$\begin{aligned} & \text{if } x \in A_1 \wedge y \in B_1 \rightarrow z = p_1 x + q_1 y + r_1 \\ & \text{if } x \in A_1 \wedge y \in B_2 \rightarrow z = p_2 x + q_2 y + r_2 \\ & \text{if } x \in A_2 \wedge y \in B_1 \rightarrow z = p_3 x + q_3 y + r_3 \\ & \text{if } x \in A_2 \wedge y \in B_2 \rightarrow z = p_4 x + q_4 y + r_4 \end{aligned} \quad (3)$$

La partición del espacio del dominio dada por las reglas difusas está ilustrada en la Figura 1(a) y la arquitectura de ANFIS es ilustrada en la Figura 1(b). Las funciones en cada nodo de una capa son las mismas y se describen a continuación:

- *Capa 1*: Cada nodo i en esta capa tiene la siguiente función:

$$O_i^1 = \mu_{A_i}(x) \quad (4)$$

donde O_i^1 es la función de pertenencia de A_i y especifica el grado en el cual x satisface A_i . Usualmente, se utilizan funciones de pertenencia normalizadas $\mu_{A_i}(x)$ donde el máximo valor es 1 y el mínimo valor 0. En este caso se utilizaron la función sigmoidea:

$$\mu_{A_i}(x) = \frac{1}{1 + \exp(-\gamma_i \times (x - c_i))} \quad (5)$$

y la función Z :

$$\mu_{A_i}(x) = 1 - \frac{1}{1 + \exp(-\gamma_i \times (x - c_i))} \quad (6)$$

donde $\{\gamma_i, c_i\}$ son llamados *parámetros de la premisa*, y varían dando formas diferentes a las funciones de pertenencia en A_i .

- *Capa 2*: Los nodos de esta capa entregan como salida el producto de las señales entrantes:

$$\omega_i = \mu_{A_i}(x) \times \mu_{A_i}(x), \quad i = 1, 2. \quad (7)$$

- *Capa 3*: El resultado de esta capa establece el porcentaje en que cada regla contribuye con la solución final y se calcula como:

$$\bar{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2}, \quad i = 1, 2. \quad (8)$$

y representan el proceso de agregación de los consecuentes de la regla.

- *Capa 4:* Los nodos de esta capa tienen la siguiente función:

$$O_i^4 = \bar{\omega}_i f_i = \bar{\omega}_i (p_i x + q_i y + r_i) \quad (9)$$

donde $\bar{\omega}_i$ es la salida de la capa 3 y $\{p_i, q_i, r_i\}$ constituyen una serie de parámetros llamados *parámetros del consecuente*.

- *Capa 5:* El único nodo calcula la salida final como la suma de todas las señales entrantes:

$$O_i^5 = \sum \bar{\omega}_i f_i = \frac{\sum_i \bar{\omega}_i f_i}{\sum_i \bar{\omega}_i} \quad (10)$$

La definición de las reglas de Takagi-Sugeno implica que las variables de entrada del antecedente sean las mismas que las variables de entrada en el consecuente, y el dominio de cada variable es cubierto por al menos dos conjuntos difusos. Usualmente, la función especificada para el consecuente es una combinación lineal de sus variables de entrada, sin embargo pueden usarse funciones más complejas, aunque en la práctica no han encontrados sistemas con otras reglas.

La capacidad de ANFIS de aproximar funciones no lineales está determinada por el número de conjuntos difusos asignados a cada variable independiente; esto implica que el número de parámetros en el modelo crece de manera exponencial cada vez que una nueva variable independiente se añade al modelo, requiriendo entonces, un número muy alto de patrones para realizar la afinación del sistema de inferencia.

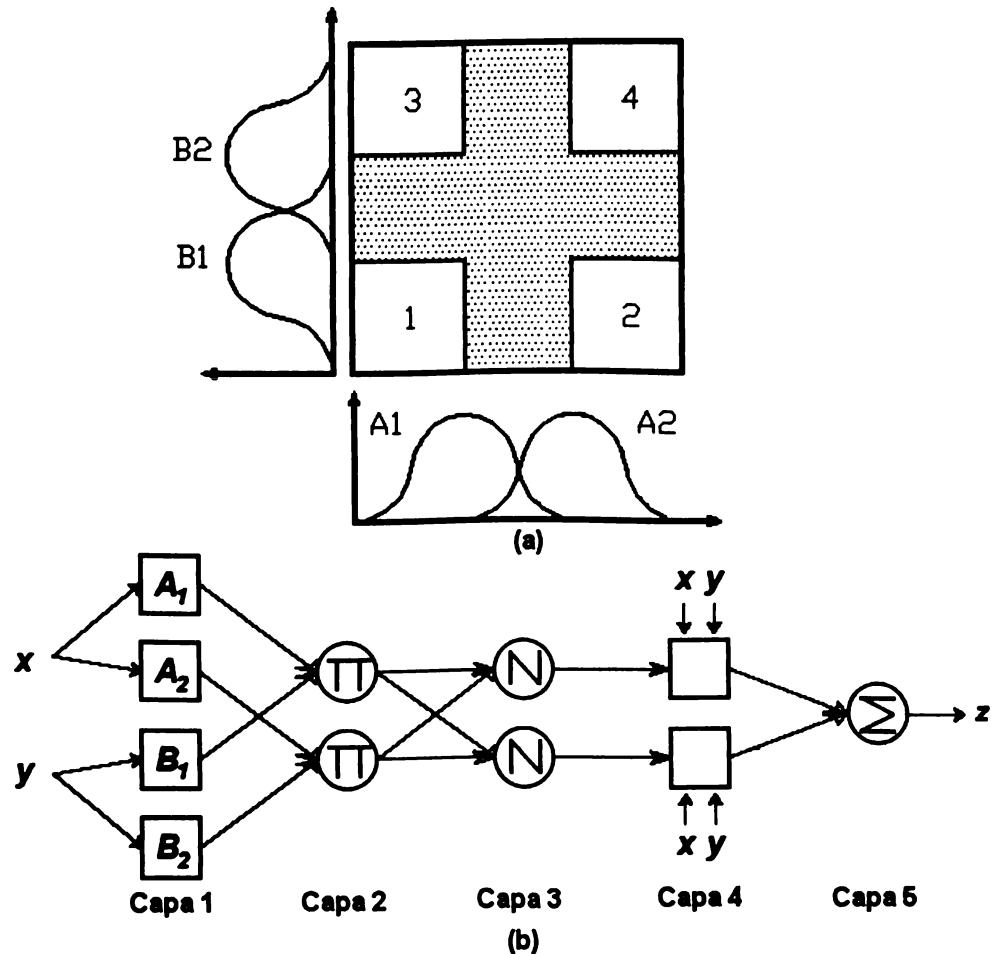


FIGURA 1. (a) Particionamiento en el espacio del dominio dado por los conjuntos difusos. (b) Arquitectura típica de ANFIS.

Para la especificación inicial de ANFIS, Jang (1994) propone el uso del algoritmo CART (Árbol de Clasificación y Regresión) para seleccionar las variables que conformarán las reglas difusas y el número de conjuntos borrosos asignados a cada variable; y propone afinar los parámetros usando un algoritmo de optimización

3.2. ANFIS como un modelo de series de tiempo

Una generalización de la familia de modelos especificada en la Ecuación (1), es permitir que el dominio del regresor

$x_t^{(i)}$ sea particionado en K regiones disyuntas $\{R_k\}_1^K$, dentro de las cuales la serie de tiempo sigue un comportamiento lineal, y asignar un modelo lineal a cada una de ellas (Tong, 1990):

$$\text{If } x_t^{(i)} \in R_k \Rightarrow y_t = g_k(x_t | \{a_j^k\}_0^U) + v_t^{(k)} \quad (11)$$

estos modelos reciben el nombre de SETAR (Self-Exciting Threshold AutoRegressive) o SETARMA (Self Exciting Threshold AutoRegressive Moving Average) de acuerdo con el tipo de modelo con que se implemente el consecuente de la regla.

En el contexto de series de tiempo, ANFIS puede ser interpretado como una generalización no lineal de modelos autorregresivos (Velásquez et al., 2004). De acuerdo a la Ecuación (11), si se permite que las regiones se traslapen permitiendo que sobre la intersección tengan influencia los modelos autorregresivos de cada una de ellas, representando cada región por un conjunto difuso, entonces se obtiene un sistema de reglas de la siguiente forma:

$$\text{If } x_t^{(i)} \in A_1 \Rightarrow y_t = g_k(x_t | \{a_j^k\}_0^U) + v_t^{(k)} \quad (12)$$

equivalentes a un caso particular de ANFIS.

En el caso general, los antecedentes pueden contener más de una variable, e inducen una partición del dominio de la serie de tiempo. Esta representación de ANFIS es flexible en cuanto a la conformación del modelo, es decir, es posible que las variables del antecedente no sean las mismas del consecuente.

Una característica fundamental de ANFIS es que existe un modelo autorregresivo lineal por cada punto del dominio $x_t^{(i)}$, debido a que la pertenencia a cada conjunto

es multiplicada a las reglas del consecuente en cada uno de dichos puntos, mientras que para los modelos SETAR y SETARMA, existe un modelo por cada región en que se partitiona el conjunto de regresores (Velásquez et al., 2004).

3.3. Formulación del Modelo

La estrategia general para la formulación del modelo es propuesta en Velásquez et al. (2004) como una metodología sistemática para la formulación de modelos de series de tiempo, y se resume a continuación:

Primer Paso - Identificación: Se realiza un análisis exploratorio de la serie de tiempo con el fin de seleccionar los posibles regresores x , usados en la construcción de un modelo, y determinar si es posible hacer una transformación de la serie de tiempo que permita al modelo capturar la serie de mejor manera. En la parte final del proceso de identificación se especifica el modelo.

Segundo Paso - Estimación de Parámetros: ANFIS, gracias a la interpretación natural del experto para la definición de los conjuntos difusos en los cuales se dividen los regresores del antecedente, permite al modelo partir de parámetros iniciales cercanos al óptimo; de manera que para el refinamiento de parámetros se puede usar un algoritmo híbrido donde los parámetros de los consecuentes son estimados óptimamente como una función de los parámetros de las premisas.

Tercer Paso - Diagnóstico: Este paso consiste en evaluar los residuos v del modelo con el ánimo de examinar si efectivamente constituyen ruido blanco. Realizando básicamente una inspección visual de diferentes gráficas de los residuos como los residuos en el tiempo, los residuales estandarizados, los residuales acumulados y estas mismas para el cuadrado de los residuales; y el resultado de pruebas estadísticas aplicadas a los mismos con el fin de detectar inconsistencias, como la presencia de componentes estructurales o desviaciones en la hipótesis de normalidad, que no son detectables a la vista, entre estas pruebas se sugieren las autocorrelaciones y autocorrelaciones parciales.

Cuarto Paso - Metadiagnóstico: El fin de este paso es determinar la precisión del pronóstico de la serie real, usando los residuos v del modelo. Usualmente se usan como medida el porcentaje de la varianza explicada :

$$R^2 = 1,0 - \sum_{t=1}^T \frac{v_t^2}{y_t^2} \times 100\% \quad (13)$$

el predictor de la varianza:

$$\sigma_{v_t}^2 \sim \frac{1}{T} \sum_{t=1}^T v_t^2 \quad (14)$$

y la desviación media absoluta (MAD):

$$MAD = \frac{1}{T} \sum_{t=1}^T |v_t| \quad (15)$$

En este paso también se pueden considerar la creación de modelos alternativos con la adición o sustracción de variables del modelo, para llegar al mejor modelo con el mínimo número de regresores en el antecedente. Para evaluar estos modelos se pueden usar criterios de medida como el criterio de Akaike, Bayes o Schwartz (Salas, 1980).

Quinto Paso - Construcción del Pronóstico e Intervalos de Confianza: Despues de aceptar un modelo, este es usado para hacer pronósticos de los valores futuros de la serie real y estimar sus intervalos de confianza. El primer paso después de pronosticar, y_{t+1} , es calcular el intervalo de confianza usando:

$$y_t = \hat{f}(x_t) + v_t \quad (16)$$

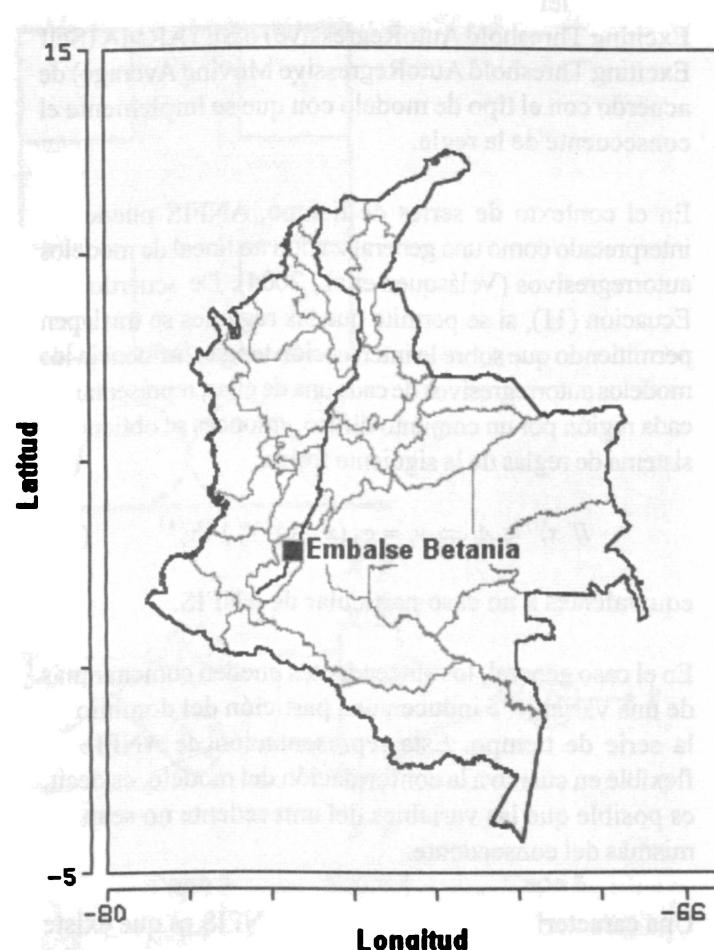
siguiendo una distribución de probabilidad normal con media cero y varianza $\sigma_{v_t}^2$. Cuando se requiere de más pronósticos, y x_t contiene uno o más rezagos de la serie real y_t , entonces es necesario calcular los intervalos de confianza usando simulación de Montecarlo, ya que ANFIS realiza una transformación no lineal de la distribución de probabilidad normal.

4. DESCRIPCIÓN DE LOS DATOS

Los datos utilizados fueron recolectados en la estación Embalse Betania, cuya ubicación geográfica puede verse en la Figura 2, perteneciente al Sistema Hidroeléctrico Betania conformado por una central y un embalse con el mismo nombre que constituye uno de los principales sistemas de la parte centro del país. Esta estación fue instalada en 1987 y continúa funcionando hasta la fecha.

Se obtuvo información de los registros mensuales de caudales medios en la estación desde enero de 1961 hasta diciembre de 1997 sin datos faltantes, con media $433.84 m^3/s$ y desviación estándar $163.51 m^3/s$, esta serie es presentada en la Figura 3(a).

El preprocessamiento realizado a esta serie consiste en la aplicación de una transformación Box y Cox con $l=0.37889$, para aplicar posteriormente una estandarización mensual de los datos, la serie de tiempo resultante es presentada en la Figura 3(b), y su propiedades estadísticas en la Figura 4(I); y una diferenciación de doce meses cuya serie resultante es presentada en la Figura 3(c), y sus propiedades estadísticas en la Figura 4(II). La figura de propiedades estadísticas contiene el autocorrelograma parcial, el espectro y el histograma normalizado. El patrón capturado por la estandarización mensual puede verse en la Figura 3 (d) y el de la diferenciación a doce meses en la Figura 3(e).



Los datos entre los años 1961 y 1982 fueron usados para la estimación de los parámetros del modelo, mientras que los años restantes se usaron para validar el modelo.

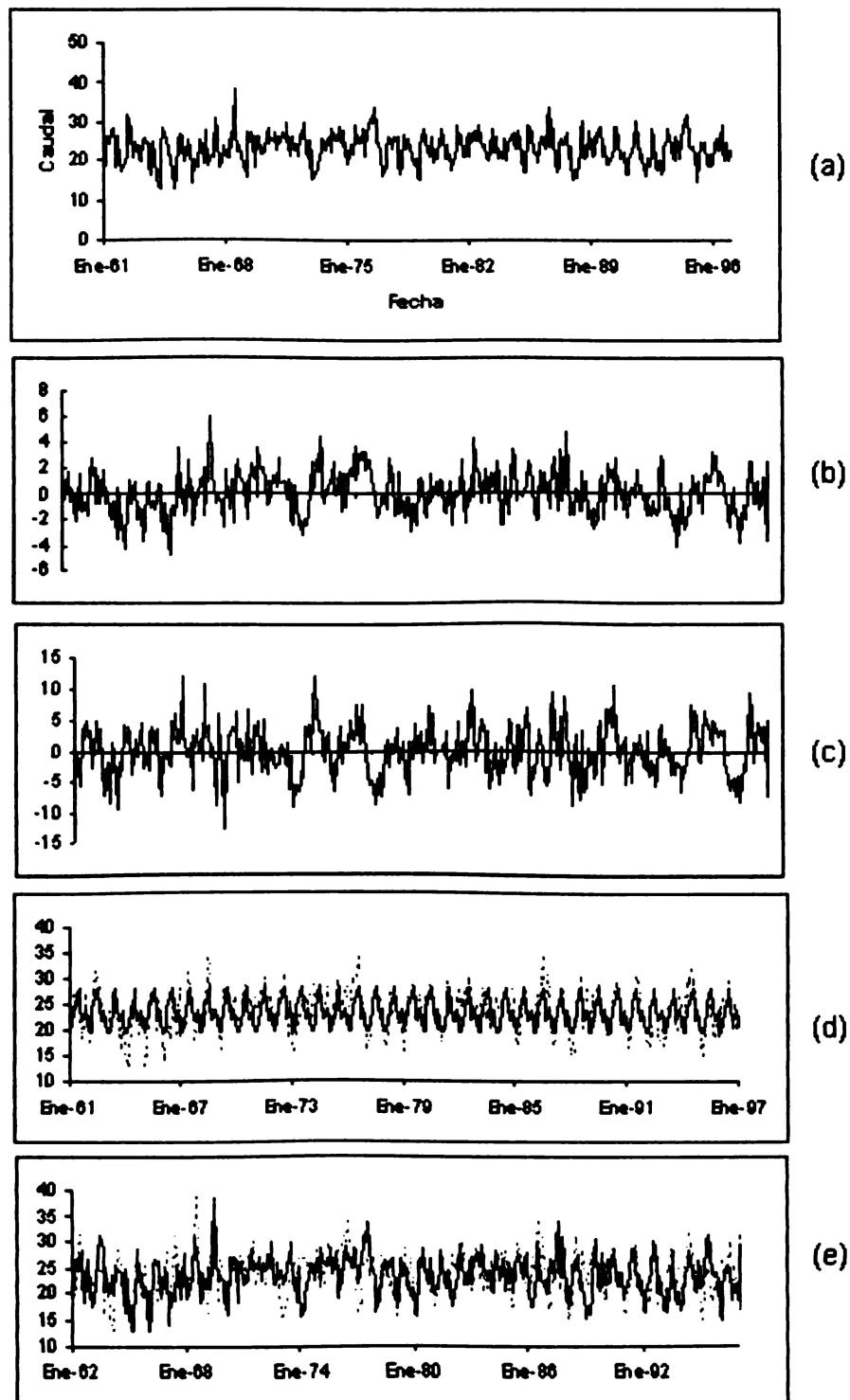


FIGURA 3. Figuras de la Serie Betania

(a) Serie de Caudal Real. (b) Serie Normalizada. (c) Serie Diferenciada. (d) Patrón Capturado por la Normalización. (e) Patrón Capturado por la Diferenciación.

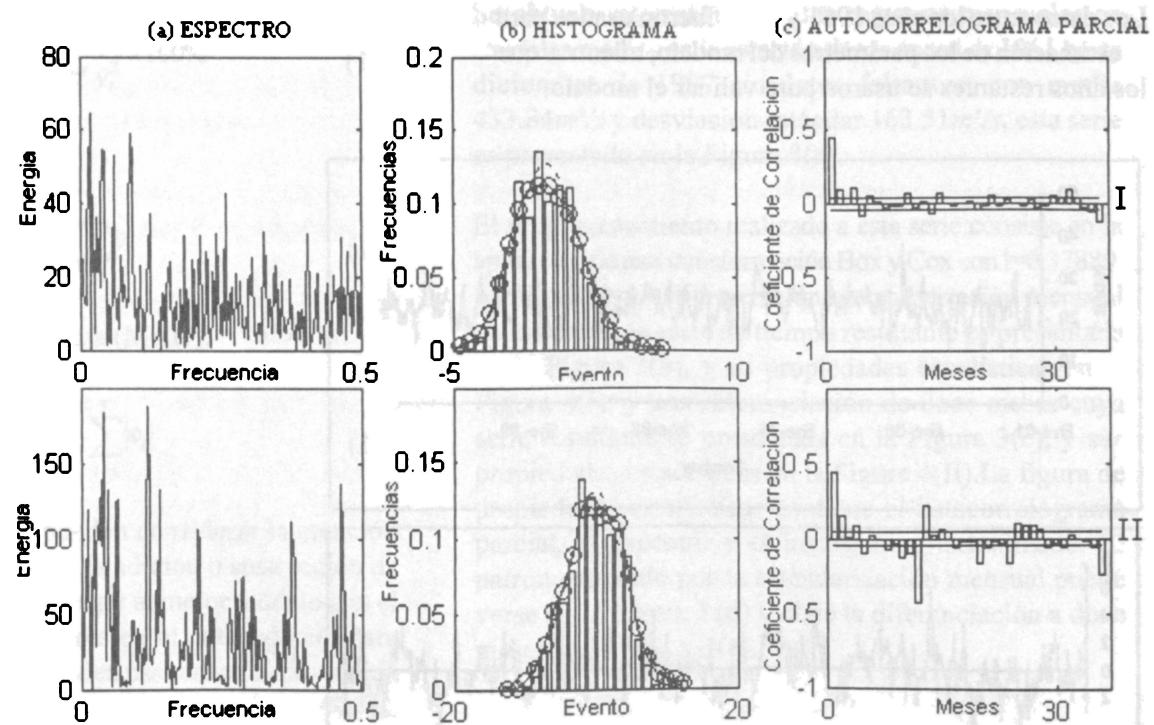


FIGURA 4. Propiedades Estadísticas para la Serie (I) con Estandarización Mensual y (II) con Diferenciación a Doce Meses

5. RESULTADOS

La aplicación de ANFIS en series de caudal, involucra no solo la construcción de modelos, sino también la especificación del procesamiento adecuado que debe hacerse a los datos antes de desarrollar dicho modelo. Con este fin, se usaron dos métodos convencionales: el primero usado en el procesamiento de series de caudal llamado estandarización mensual de los datos o normalización, y el segundo utilizado en series de tiempo llamado diferenciación a doce meses de la serie de tiempo. En la aplicación específica se construyeron modelos autorregresivos tanto para la serie con estandarización mensual, como para la serie con diferenciación a doce meses. De ellos se partió para la contrucción de modelos ANFIS para ambos casos de acuerdo a la metodología propuesta.

5.1. Modelo para la Serie con Estandarización Mensual

Para la formulación del modelo, para la serie con estandarización mensual, se construyó un modelo autorregresivo con el número de regresores óptimo por el criterio de Akaike. Este procedimiento arrojó como resultado la selección del siguiente modelo:

$$z = 0.34913R_1 + 0.16948R_2 + 0.09406R_4 + 0.14551 + \vartheta_t \quad (17)$$

donde R_1, R_2, R_4 son los rezagos de la serie en uno, dos y cuatro meses respectivamente, con una media de $436.86m^3/s$ y una desviación estándar de $123.15m^3/s$. Los residuales tienen un coeficiente de asimetría de 0.8929 y una curtosis de 2.5798.

Los mismos regresores fueron utilizados en el modelo ANFIS, construyendo modelos tomando una sola variable en la premisa y con estos rezagos como variables del consecuente, variando la variable de la premisa entre los rezagos antes mencionados y combinaciones lineales de estos. El mejor modelo encontrado tiene como variable en la premisa una combinación lineal de los rezagos a uno y cuatro meses de la serie de caudal, con el siguiente conjunto de reglas:

$$\begin{aligned} \text{If } x \in A_1 \rightarrow f_1 &= 0.2632R_1 + 0.1994R_2 + 0.1105R_4 + 0.12696 \\ \text{If } x \in A_2 \rightarrow f_2 &= 0.2632R_1 + 0.1994R_2 + 0.1105R_4 + 0.12696 \end{aligned} \quad (18)$$

donde x es una combinación lineal de la siguiente forma:

$$x = 0.2657R_1 + 0.7343R_4 \quad (19)$$

En la Figura 5 se pueden observar la función de transición de los conjuntos difusos y la pertenencia de los conjuntos uno y dos.

La relación entre las varianzas del modelo autorregresivo y el modelo ANFIS, durante la estimación de parámetros, es igual a 0.97, y la comparación entre sus errores se presenta en la Tabla 1.

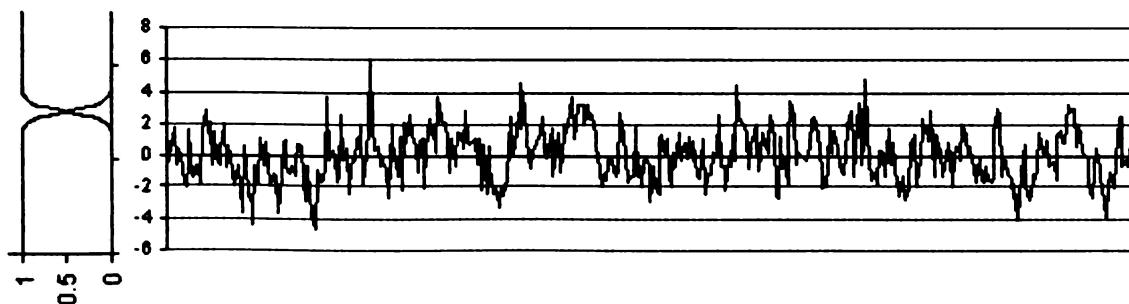


FIGURA 5. Función de Transición para el Modelo ANFIS de la Serie con Estandarización Mensual

TABLA 1. Tabla Comparativa entre los Modelos AR y ANFIS, para el periodo de estimación de parámetros, para el Modelo de la Serie con Estandarización Mensual.

	MSE	MAPE
AR	111.70	21.74%
ANFIS	108.91	21.62%

La varianza del predictor es de $12200.5(m^3/s)^2$ y una desviación media absoluta del $86.2 m^3/s$. Las propiedades estadísticas del modelo se presentan en la Figura 7. en ellas se puede observar como el modelo se ajusta a la serie de tal manera que en el espectro solo se ve ruido blanco, la distribución que se aprecia en el histograma es aproximadamente normal, no se observan correlaciones significativamente diferentes de cero y el comportamiento de la gráfica de residuales acumulados sugiere que estos tienen dependencia entre sí, en la medida en que la gráfica es creciente. La curtosis de los residuales es 2.7883 y el coeficiente de asimetría 0.2454, de acuerdo a la prueba de Snedecor-Cochran no hay suficiente evidencia para afirmar que los residuales son normales.

5.2. Modelo para la Serie con Diferenciación Mensual

Para la formulación del modelo, para la serie con diferenciación mensual, se procedió de manera similar al modelo formulado anteriormente. Este procedimiento arrojó como resultado el siguiente modelo

$$z = 0.30183R_1 + 0.04836R_2 + 0.11627R_4 + 0.04664R_5 + 0.29772R_{12} - 0.0703R_{13} + 0.02166 + \nu_t \quad (20)$$

donde $R_1, R_2, R_4, R_5, R_{12}, R_{13}$ son los rezagos de la serie en uno, dos, cuatro, cinco, doce y trece meses respectivamente.

Para el modelo ANFIS el procedimiento también fue similar y el mejor modelo encontrado tiene como variable en la premisa una combinación lineal de los rezagos a uno y dos meses de la serie de caudal, con el siguiente conjunto de reglas:

$$\begin{aligned} \text{If } x \in A_1 \rightarrow z = & 0.48734R_1 - 0.03230R_2 + 0.15773R_4 + 0.18638R_5 \\ & + 0.15713R_{12} + 0.10982R_{13} + 0.12949 \\ \text{If } x \in A_2 \rightarrow z = & 0.21807R_1 + 0.19054R_2 - 0.00950R_4 + 0.05014R_5 \\ & + 0.13621R_{12} - 0.12377R_{13} + 0.01155 \end{aligned} \quad (21)$$

siendo x una combinación lineal de la siguiente forma:

$$x = 0.4869R_1 + 0.5131R_2 \quad (22)$$

En la Figura 6 se pueden observar la función de transición de los conjuntos difusos y la pertenencia de los conjuntos uno y dos.

La relación entre las varianzas del modelo autorregresivo y el modelo ANFIS, durante el periodo de estimación de parámetros es igual a 0.945, y la comparación entre sus errores se presenta en la Tabla 2.

TABLA 2. Tabla Comparativa entre los Modelos AR y ANFIS, para el periodo de estimación de parámetros, para el Modelo de la Serie Diferenciada.

	MSE	MAPE
AR	130.98	24.75%
ANFIS	123.81	22.70%

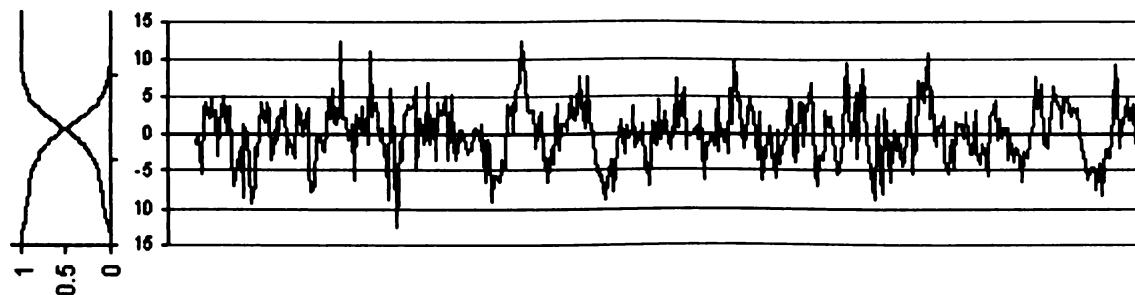


FIGURA 6. Función de Transición para el Modelo ANFIS de la Serie con Diferenciación 12 Meses

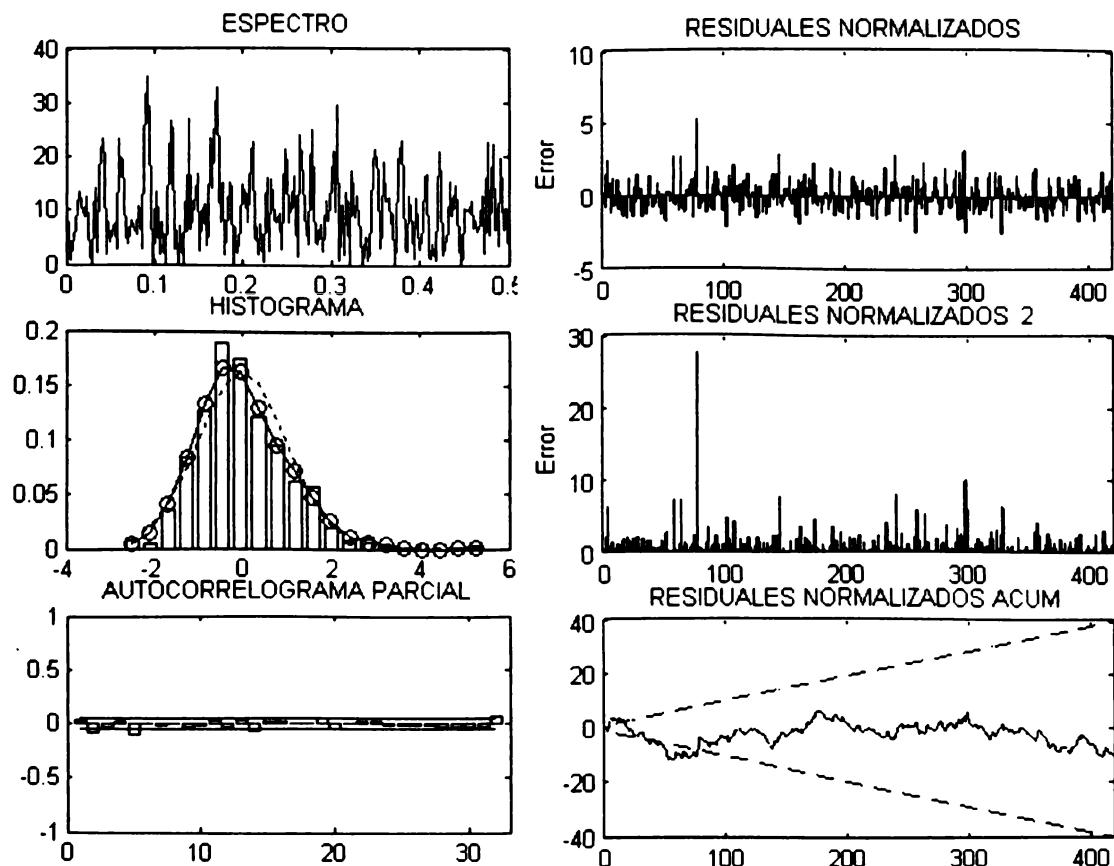


FIGURA 7. Propiedades Estadísticas de los Residuales Normalizados para el Modelo con Estandarización Mensual

La varianza del predictor es de $20708.2(m^3/s)^2$ y una desviación media absoluta de $153.5m^3/s$. Las propiedades estadísticas del modelo se presentan en las Figura 8, en el espectro se observa que aún hay picos en frecuencias altas, es decir que hay ciclos de corto plazo que no se lograron capturar con el modelo. La distribución que se aprecia en el histograma no es simétrica. En el autocorrelograma, se pueden identificar varias correlaciones significativamente diferentes de cero, especialmente a largo plazo, lo que podría explicarse como los ciclos de largo plazo de la serie que están influenciados por el clima. Por último el comportamiento de la gráfica de residuales sugiere que estos no tienen dependencia entre sí, ya que no se observan tendencias en la gráfica. La curtosis de los residuales es 3.0321 y el coeficiente de asimetría 0.0625, de acuerdo a la prueba de Snedecor-Cochran se acepta la hipótesis de que los residuales son normales.

6. CONCLUSIONES

Se ha presentado ANFIS como una generalización de modelos autorregresivos, que lo introduce como un método de modelamiento de series de tiempo. En este contexto se define una metodología para la formulación del modelo, que permite incorporar criterios estadísticos usados comúnmente en las metodologías tradicionales de series de tiempo.

Los resultados permiten concluir que los modelos ANFIS mejoran los modelos tradicionales autorregresivos, siendo capaces de capturar mejorar la estructura de la serie, confirmando que existe cierto comportamiento no-lineal en ella, que puede ser explicado por la influencia que tienen las variables climáticas sobre las series de caudal y la probada no linealidad de estas últimas. Además, se observó que en el modelamiento de series de caudal con ANFIS se obtienen mejores resultados cuando se preprocesa la información con la técnica usada tradicionalmente para esto, es decir, la estandarización mensual de los datos, la cual elimina la componente estacional anual de los datos.

Los trabajos futuros involucran la incorporación de variables macro-climáticas que estén en capacidad de capturar parte de la estructura de la serie que escapa a los modelos univariados, y la evaluación de modelo híbridos que incluyan el modelamiento de la componente estacional de las series de caudal.

7. AGRADECIMIENTOS

Al posgrado en Recursos Hídricos de la Facultad de Minas, a los profesores Ricardo Smith y Jaime Ignacio Vélez, y a los demás miembros del proyecto de investigación “Modelos de Predicción de Caudales en Ríos Colombianos”.

8. BIBLIOGRAFIA

- Box, G.E.P. y Jenkins, G.M., 1970. Time Series Anaysis: Forecasting and Control, HoldenDay Inc., ESTados Unidos.
- Carvajal, L., 1994. Acerca de la Predicción no Lineal en Hidrología, Tesis Maestría, Universidad Nacional de Colombia, Medellín.
- Jang, R., 1993. Anfis: Adaptive-network-based fuzzy inference system, IEEE Transactions (3).
- Mesa, O. y Poveda G., 1993. Análisis Espectral Singular y Redes Neuronales Aplicadas a Hidrología, Avances en Recursos Hídricos (2).
- Mesa, O., Poveda, G. y Carvajal, L., 1997. Introducción al Clima de Colombia, Universidad Nacional de Colombia, Bogotá.
- Poveda, G. y Mesa, O., 1993. Metodologías de Predicción de la Hidrología Colombiana Considerando el fenómeno El Niño/Oscilación del Sur, Revista Atmósfera (20).
- Poveda, G. y Mesa, O., 1996. Caudales Medios Mensuales de Ríos Colombianos durante El Niño y La Niña, IV Congreso Colombiano de Meteorología, IDEAM, Bogotá, Colombia.
- Poveda, G., Mesa, O., Carvajal, L., Hoyos, C., Mejía, J., Cuartas, A. y Pulgarín, A., 2002. Predicción de Caudales Medios Mensuales en Ríos Colombianos usando Métodos no Lineales, Meteorología Colombiana (6).
- Prada, L. F. y Obregón, N., 2004. Consideraciones Prácticas para la Calibración de RNA Aplicadas a la Predicción de Series de Tiempo de Caudales, III Congreso y I Conferencia Andina Internacional de Investigación de Operaciones, Cartagena, Colombia.
- Rendón, L., 1997. Beneficios de la Predicción Hidrológica en el Sector Eléctrico Colombiano,

Considerando la Variabilidad Climática.

- Salas, J. D., Delleur, J. W., Yevjevich, V. y Lane, W. L., 1980. *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Colorado, Estados Unidos.
- Salazar, J., Mesa, O., Poveda, G. y Carvajal, L. F., 1994. Aplicación de un Modelo Continuo No Lineal al Estudio de Series Hidrológicas, XXI Congreso Latinoamericano de Hidráulica e Hidrología, IAHR, Número 4.
- Salazar, J., Mesa, O., Poveda, G. y Carvajal, L. F., 1994. Ecuación Diferencial Estocástica No-Lineal para Modelar la Hidrología, X Seminario Nacional de Hidráulica e Hidrología.
- Salazar, J., Mesa O. y Poveda, G., 1994. Aplicación de dos Modelos No Lineales a la Simulación de Series Hidrológicas, Avances en Recursos Hidráulicos (2).
- Svanidze, G., 1980. *Mathematical Modeling of Hydrologic Series*, Water Resources Publications, Colorado, Estados Unidos.
- Takagi, T. y Sugeno, M., 1982. Derivation of Fuzzy Control Rules from Human Operator's Control Actions, In Proceedings of the IFAC Symposium of Fuzzy Information, Knowledge Representation and Decision Analysis.
- Tong, H., 1990. *Non-Linear Time Series, a Dynamical System Approach*, Oxford Statistical Science Series, U.K.
- Velásquez, J., Dyner, I. y Souza, R., 2004. Statistical Modelling of the Brazilian Electricity Spot Price Using Anfis.