Ingeniare. Revista chilena de ingeniería, vol. 15 Nº 3, 2007, pp. 302-312

CARACTERIZACIÓN DEL SOI USANDO ANFIS CON RESIDUALES HETEROCEDÁSTICOS

SOI CHARACTERIZATION USING ANFIS WITH HETEROCEDASTIC RESIDUALS

Elizabeth C. Zapata[1] **Juan D. Velásquez**[2] **Ricardo Smith Q.**[3]

Recibido 30 de noviembre de 2006, aceptado 22 de octubre de 2007
Received: November 30, 2006 Accepted: Octoberr 22, 2007

RESUMEN

En este trabajo se propone la representación de la dinámica del Índice de Oscilación del Sur usando una nueva clase de modelo híbrido no lineal. En este nuevo modelo, la no-linealidad en la media es representada usando un sistema adaptativo neurodifuso de inferencia (ANFIS), mientras que la varianza es representada usando una componente autoregresiva heterocedástica condicional. En adición, una estrategia de especificación es desarrollada para el modelo propuesto, la cual está basada en una batería de pruebas estadísticas. Los resultados indican que el modelo propuesto captura de mejor forma las principales características de la serie de tiempo estudiada, en comparación con otros modelos competitivos.

Palabras clave: ANFIS, ARCH, SOI, heterocedasticidad, series temporales, modelos no lineales.

ABSTRACT

In this paper, we propose the representation of the Southern Oscillation Index (SOI) dynamics using a new class of nonlinear hybrid model. In this new model, the non-linearity in the mean is represented by an adaptive neurofuzzy inference system (ANFIS), while the variance is represented using an autoregressive conditional heterocedastic (ARCH) component. In addition, a specification strategy for the proposed model, which is based on a battery of statistical tests is developed. The results indicate that the proposed model captures in a better way the main features of the studied time series, when compared to other competitive models.

Keywords: ANFIS, ARHC, SOI, heterocedasticity, nonlinear time series modelling.

INTRODUCCIÓN

El fenómeno climático de El Niño – Oscilación del Sur (ENSO) se presenta como el resultado de la interacción océano-atmósfera en el Océano Pacífico ecuatorial, afectando de forma importante las condiciones climatológicas de las regiones comprometidas durante sus fases cálida y fría, conocidas como periodos de El Niño y La Niña, respectivamente [11]. Diversos estudios han determinado que El Niño es un evento recurrente y aperiódico que se presenta con un periodo comprendido entre 2 y 7 años, y que tiene una duración aproximada de 12 meses [6]; su intensidad depende de la magnitud de las anomalías y del área cubierta por las mismas.

Ambas fases del ENSO producen eventos extremos hidrometeorológicos, que tienen unos importantes impactos ambientales, sociales y económicos, que han sido documentados en diversas investigaciones [6]. Dichos impactos han confirmado la necesidad de entender y predecir este fenómeno con el fin de emprender tareas de planificación para la mitigación de sus efectos [8].

Debido a la naturaleza caótica del clima, la interacción de fenómenos que actúan en distintos lugares y momentos, y su aperiodicidad, el pronóstico del ENSO representa un problema complejo [15] que ha sido abordado por dos enfoques fundamentales. Por un lado se encuentran los modelos físicos que consideran los sistemas océano-

[1] Escuela de Sistemas. Universidad Nacional de Colombia. Medellín, Colombia. E-mail: eczapata@unal.edu.co
[2] Escuela de Sistemas. Universidad Nacional de Colombia. Medellín, Colombia. E-mail: jdvelasq@unal.edu.co. Autor para correspondencia.
[3] Escuela de Geociencias y Medio Ambiente. Universidad Nacional de Colombia. Medellín, Colombia. E-mail: rasmith@unalmed.edu.co

Libro INGENIERIA.indb 302 8/1/08 16:44:45

atmosféricos, mientras que por el otro se encuentran los modelos estadísticos [7] en que se pretende encontrar representaciones matemáticas de la dinámica que siguen las variables que pueden ser usadas para medir el fenómeno.

El SOI (Southern Oscillation Index) es un índice para medir la fuerza y la fase de la Oscilación del Sur, el cual permite determinar la presencia de los eventos de El Niño y La Niña; los valores extremos negativos corresponden al fenómeno de El Niño, mientras que los valores extremos positivos corresponden a La Niña. Las características ya mencionadas del ENSO se ven reflejadas precisamente en la dinámica que sigue el SOI, lo que ha llevado a abordar su modelado a partir de técnicas estadísticas no lineales. Particularmente, en [7] se ha realizado el modelado de esta serie usando un modelo STAR (Smooth Transition AutoRegressive Model) [13], quienes construyen un modelo LSTAR (Logistic Smooth Transition AutoRegressive Model) univariado para representar el SOI; los resultados del diagnóstico del modelo obtenido en [7] muestran que los residuales no presentan correlaciones residuales remanentes, pero su volatilidad es cambiante en el tiempo, por lo que el modelo podría ser rechazado, y consecuentemente se requeriría considerar otras especificaciones alternativas u otras clases de modelos. En esta investigación se avanza en dicha dirección, al considerar un modelo alternativo para representar la dinámica del SOI.

El modelo utilizado en este trabajo se basa en una generalización del trabajo de Velásquez, Dyner y Souza [14], quienes demuestran que un modelo STR (Smooth Transition Regressio) es equivalente a un sistema adaptativo de inferencia neurodifusa (ANFIS) [9], cuando a este último se le imponen ciertas restricciones. La generalización presentada se basa en permitir que los residuales del modelo definido en [14] sigan un proceso ARCH.

Consecuentemente con las ideas planteadas, el objetivo de este trabajo es caracterizar la serie del SOI con un modelo que logre aproximar tanto la no linealidad de la serie como su volatilidad cambiante en el tiempo. Se propone entonces un modelo ANFIS-ARCH como un híbrido entre el modelo ANFIS tradicional [9], bajo algunas restricciones, y un modelo ARCH [5] que capture la volatilidad de dicha serie.

A continuación se presenta la definición del modelo ANFIS-ARCH y una estrategia para su especificación. Asímismo, se hace una breve descripción del SOI y se presentan los resultados obtenidos a modelar esta serie con el modelo propuesto.

EL MODELO ANFIS-ARCH Y SU ESPECIFICACIÓN

Sistemas adaptativos de inferencia neurodifusa

Un sistema adaptativo de inferencia neurodifusa [9] o ANFIS (adaptive neuro-fuzzy inference system) puede definirse como un tipo de red neuronal artificial, el cual emula el proceso de inferencia que se realiza en un sistema difuso. Puede ser considerado como una técnica no paramétrica de regresión no lineal, que puede aproximar cualquier función definida en un dominio compacto, de tal forma que se le considera como un aproximador universal de funciones. Ha sido ampliamente utilizado en el contexto de la regresión, especialmente para aplicaciones en el área de la Ingeniería de Control [1].

La arquitectura típica de ANFIS es presentada en la figura 1. Donde x e y son las variables tanto del antecedente como del consecuente, y z la salida del sistema. El proceso de inferencia para las entradas (x, y) implica [14]:

1. Calcular las funciones de pertenencia $\mu A_1(x)$, $\mu A_2(x)$, $\mu B_1(y)$ y $\mu B_2(y)$.
2. Estimar para cada regla el producto $w_j = \mu A_j(x) \cdot \mu B_j(y)$.
3. Establecer el porcentaje en que cada regla aporta a la solución final $w_j = w_j \Big/ \sum_{i=1}^{N} w_i$.
4. Y finalmente calcular el resultado del sistema como: $\sum_i w_i \cdot f_i$. Donde f_i es usualmente una combinación lineal de las variables del consecuente.

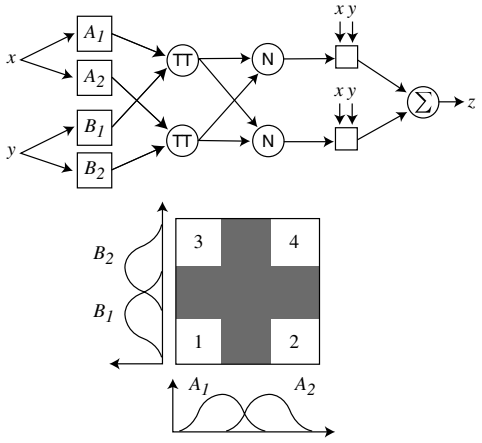

Figura 1. Arquitectura típica de ANFIS.

Libro INGENIERIA.indb 303 8/1/08 16:44:48

El proceso de especificación tradicionalmente usado [10] se basa en el uso de CART para encontrar la especificación inicial del sistema neurodifuso, que consiste en definir las variables de entrada, la variable de salida y la cantidad de conjuntos difusos que se usará para especificar cada variable.

Velásquez [14] argumenta que el modelo anterior puede ser generalizado para el problema específico del modelado y la predicción de series temporales. Considera el caso en que el antecedente tiene una sola variable de entrada *z* cuyo dominio está cubierto por dos conjuntos difusos, y los consecuentes pueden considerar un conjunto de regresores *x*. Así, el sistema de inferencia descrito está conformado por dos reglas:

$$\begin{aligned} Si\ z\ &\in A_1 \rightarrow y_t = \hat{a}_0^{(1)} + \sum_{k=1}^{U} \hat{a}_k^{(1)} \cdot x_t^{(k)} \\ Si\ z\ &\in A_2 \rightarrow y_t = \hat{a}_0^{(2)} + \sum_{k=1}^{U} \hat{a}_k^{(2)} \cdot x_t^{(k)} \end{aligned} \quad (1)$$

donde los regresores $x_t^{(k)}$ (para k=1,…,U) y la variable de transición z pueden corresponder a rezagos de la serie o variables exógenas que puedan explicar el comportamiento de la serie. Los dos conjuntos difusos que cubren el dominio de z pueden especificarse a través de las funciones $S()$ y $Z()$:

$$\begin{aligned} S &= \frac{1}{1+\exp(-\gamma_1(x-c_1))} \\ Z &= 1 - \frac{1}{1+\exp(-\gamma_2(x-c_2))} \end{aligned} \quad (2)$$

donde c_i es un parámetro que define el punto de inflexión de la función de transición, y γ_i define su pendiente. Tal como es mostrado en [14], la formulación propuesta en (2) es equivalente a un modelo LSTR cuando se hace $c_1 = c_2$ y $\gamma_1 = \gamma_2$, ya que para la ponderación asignada a cada regla se obtiene que w_1=1-w_2, de acuerdo con el algoritmo de cálculo previamente descrito.

La función de transición puede especificarse de igual modo por medio de una función exponencial dada por

$$\begin{aligned} S &= \exp(-\gamma_1(x-c_1)^2) \\ Z &= 1 - \exp(-\gamma_2(x-c_2)^2) \end{aligned} \quad (3)$$

Un problema de esta función es que converge a un valor constante cuando *ã tiende a cero. Igualmente, cuando se imponen las mismas restricciones anteriores, el modelo es equivalente a un ESTAR (Exponential STAR).*

Modelo ARCH

Los modelos ARCH [5] (autoregressive conditional heteroscedatic model) y sus generalizaciones como GARCH [2] han sido una forma efectiva de modelar la no linealidad en varianza o heterocedasticidad o varianza cambiante en el tiempo, que poseen algunas series temporales. Su uso tradicional ha estado relacionado con los modelos ARIMA de Box y Jenkins, pero recientemente diversos modelos híbridos han sido presentados, donde los modelos no lineales intentan aproximar la media de la serie, y los modelos ARCH modelan su varianza [4, 3].

Engle en [5] propone representar una serie y_t como

$$y_t = f\left(I_{t-1}\right) + h\left(I_{t-1}\right) \cdot \varepsilon_t \quad (4)$$

donde I_t es la información disponible hasta el instante *t*. Así el valor esperado de y_t es *f(.),* mientras que *h(.)* representa su varianza. Si *g(.)* es lineal se dice que la serie es lineal en la media. Si *h(.)* es constante se dice que es homocedástica, o de varianza constante; de lo contrario, se dice que es heterocedástica. De esta forma, la heterocedasticidad puede definirse como la presencia de una estructura causal en la varianza que la hace dependiente del tiempo.

Una forma de modelar la heterocedasticidad es representar la varianza de los residuales como una combinación lineal de los shocks anteriores, así:

$$\sigma_t^2 = \omega_0 + \sum_{i=1}^{I} \omega_i \, e_{t-i}^2 \quad (5)$$

La cual define un modelo ARCH de parámetros w_i, i=1:I.

Modelo ANFIS-ARCH

Un modelo ANFIS - ARCH consiste en una extensión del modelo desarrollado en [14] que se obtiene al considerar que los residuales siguen un proceso ARCH. De esta forma, un modelo ANFIS-ARCH se define como:

$$\hat{y}_t = f(\Psi, x) + e_t, \quad e_t = \sigma_t^2 \varepsilon_t \quad (6)$$

Libro INGENIERIA.indb 304 8/1/08 16:44:51

donde $f()$ es definida en (4), Ψ es un vector de parámetros; y σ_t^2 es la varianza de los residuales, la cual sigue un proceso ARCH definido en (5).

Los parámetros del modelo, Ψ, son obtenidos maximizando el logaritmo de la función de verosimilitud de los residuales:

$$L = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log\sigma_t^{\,2} - \frac{1}{2}\sum_{t=1}^{T}\frac{e_t^2}{\sigma_t^{\,2}} \qquad (7)$$

El uso de (7) es la base del procedimiento estadístico de inferencia para la especificación del modelo, el cual se describe más adelante.

Estrategia para la especificación de Modelos ANFIS-ARCH

El procedimiento de especificación del modelo es descrito en el siguiente algoritmo, el cual se basa principalmente en [12]:

1. Se selecciona un modelo autorregresivo de orden *P* que aproxime la serie, usando un criterio de información tal como el de Akaike, Hannan-Quinn o Schwartz.

2. Se realiza una prueba de contraste para determinar si un modelo STR [12] puede representar mejor la dinámica de la serie versus un modelo lineal. Si se acepta la hipótesis nula favoreciendo la alternativa lineal se detiene el proceso; de lo contrario, se sigue con el paso 3.

3. Se determina el tipo de función de transición más adecuada (logística o exponencial), así como la variable de transición, siguiendo el proceso de especificación propio de los modelos STR; la prueba aparece descrita en [13].

4. Se considera el modelo ANFIS con dos reglas presentado en (1). Cada regla es un modelo autorregresivo cuyo orden *P* fue determinado en el paso 1. Igualmente se considera que la varianza de los errores es constante.

5. Se realiza la estimación del vector de parámetros óptimos del modelo, Ψ, maximizando (7).

6. Se realizan pruebas de diagnóstico para demostrar si los residuales e_t para $t=1,\ldots,T$, son normales, incorrelacionados y homocedásticos. Cuando se cumplen estas tres condiciones se termina el proceso de especificación. La homocedasticidad es determinada a partir del contraste de Engle [5].

7. A partir de los resultados del contraste de Engle, se postula un modelo ARCH para representar la varianza.

8. Se estima el modelo como en el paso 5.

9. Se realizan las pruebas de diagnóstico descritas en el paso 6 a los residuales normalizados. Si se rechaza el contraste de Engle, se retorna al paso 7; en caso contrario, el modelo ya está especificado.

DESCRIPCIÓN DE LOS DATOS

El Niño consiste en un calentamiento anómalo de las aguas superficiales del centro y este del Pacífico tropical, lo que produce una profundización de la termoclima; está asociado a un debilitamiento de los vientos alisios del este y al desplazamiento del centro de convección del oeste al centro del Pacífico.

La Oscilación del Sur es una onda estacionaria de masa atmosférica que produce un diferencial de presiones entre el oeste y el este del Pacífico ecuatorial. El centro de alta presión se encuentra en la isla de Tahití, mientras que el de baja presión se halla en Darwin, Australia. El SOI (Índice de Oscilación del Sur) determina el gradiente de presiones y se establece como la diferencia entre las presiones de los centros de presión dividida por la desviación estándar. De esta manera, valores positivos del SOI denotan el evento La Niña y valores negativos del SOI sugieren el evento El Niño.

La figura 2(a) presenta la serie del SOI, en ella se visualizan de El Niño y La Niña. Así, por ejemplo, los fenómenos El Niño de 1982- 1983 y 1997-1998 se consideran muy fuertes en intensidad, los de 1957-1958, 1965-1966, 1972-1973 y 1991-1992, fuertes y los de 1976-1978 y 1986-1987, moderados. No obstante, el efecto climático y el impacto socioeconómico de estos fenómenos dependen de otros factores como la vulnerabilidad de las regiones [8].

La información usada en este estudio corresponde a las mediciones mensuales del SOI, desde enero de 1876 hasta marzo de 1997 para un total de 1.455 observaciones, las cuales pueden obtenerse en la página web Bureau of Meteorology.

Libro INGENIERIA.indb 305 8/1/08 16:44:52

CARACTERIZACIÓN DE LA SERIE E INTERPRETACIÓN DE RESULTADOS

En esta sección se presentan e interpretan los resultados obtenidos al modelar la serie del SOI con un modelo ANFIS-ARCH, siguiendo la estrategia de especificación planteada. Distintos modelos ANFIS-ARCH pueden ser obtenidos al considerar: diferentes órdenes *P* para el modelo AR, el cual especifica los consecuentes de las reglas; diferentes rezagos para especificar la variable de transición, y diferentes especificaciones para el modelo ARCH.

En el modelado de series temporales se busca encontrar un modelo que reproduzca las principales características de la serie estudiada, cumpliendo al mismo tiempo con las suposiciones usadas para la derivación del modelo mismo. Así, por ejemplo, la derivación del modelo ANFIS en su forma tradicional asume que los residuales son normales, homocedásticos e incorrelacionados; ello implica que las pruebas de diagnóstico del modelo a la serie estudiada deben confirmar estas hipótesis, y la violación de alguno de estos supuestos permite rechazar el modelo postulado. En este caso de estudio, se pretende encontrar un modelo del SOI que pueda representar simultáneamente su no-linealidad en la media y sus cambios en la varianza. El estudio realizado en [7] confirma la existencia de estas dos características.

La figura 3 resume las propiedades estadísticas del SOI. La función de autocorrelación simple o FAC (figura 3(a)) indica que se podría aproximar con un modelo AR(*P*); mientras que la función de autocorrelación parcial o FACP muestra que posiblemente *P* sea 3. Sin embargo, hay presencia de rezagos importantes alrededor de *P*=14. En el espectro de potencias (figura 3(c)) se muestra una concentración de energía de frecuencias muy bajas, debidas posiblemente a que el periodo de ocurrencia de El Niño es amplio y de largo plazo. El histograma, además, muestra que los datos de la serie siguen una distribución aproximadamente normal.

Paso 1: En la tabla 1 se reportan los valores de los diferentes criterios de información estimados para modelos autorregresivos de orden *P*=1, …, 18. El orden óptimo *P* del modelo AR(*p*) para las reglas de los consecuentes es 16. En la tabla 1, CAK es el criterio de Akaike, CHQ el de Hannan-Quinn y CSC el de Schwartz.

Paso 2: Los resultados de las pruebas para especificar el tipo de modelo son reportados en la tabla 2. H0 es un contraste cuya hipótesis nula es que la serie puede ser representada linealmente, versus la hipótesis alternativa en que la serie puede ser representada por un modelo STR. El valor reportado en esta columna es el estadístico de dicha prueba, mientras que la cantidad entre paréntesis corresponde a su valor crítico. Cuando se rechaza la hipótesis nula, se calculan los estadísticos H1, H2 y H3, los cuales permiten determinar el tipo de función de transición entre los regímenes así: si H2 < H1 y H2 < H3 entonces la función es Exponencial; de lo contrario, es Logística. Para las diferentes variables de transición consideradas, se selecciona aquella que tenga el menor valor crítico. De acuerdo con los resultados obtenidos, se selecciona a y_{t-3} como variable de decisión, y a la función logística para realizar la transición.

Tabla 1. Valor de los criterios de información para el modelo *AR* (*p*).

p	CAK	CHQ	CSC
1	7.019	7.023	7.030
2	6.979	6.984	6.993
3	6.971	6.978	6.989
4	6.966	6.974	6.988
5	6.965	6.975	6.991
6	6.966	6.977	6.995
7	6.965	6.978	**6.998**
8	6.965	6.979	7.002
9	6.967	6.982	7.007
10	6.959	6.976	7.003
11	6.960	6.978	7.008
12	6.961	6.980	7.011
13	6.959	6.979	7.013
14	6.951	**6.972**	7.009
15	6.951	6.974	7.012
16	**6.950**	6.974	7.015
17	6.951	6.977	7.020
18	6.952	6.979	7.025

Tabla 2. Resultados de la Prueba STR.

Var	H0	H1	H2	H3	Tipo
Lag 1	1,37 (0,050)	1,62 (0,058)	1,15 (0,300)	1,31 (0,181)	L
Lag 2	1,42 (0,033)	1,52 (0,084)	1,21 (0,252)	1,50 (0,090)	L
Lag 3	**1,54 (0,011)**	**1,36 (0,156)**	**1,56 (0,071)**	**1,69 (0,043)**	**L**
Lag 4	0,93 (0,616)				
Lag 5	1,38 (0,046)	1,62 (0,056)	1,20 (0,256)	1,28 (0,199)	L
Lag 6	1,27 (0,107)				
Lag 7	0,95 (0,571)				
Lag 8	1,17 (0,200)				
Lag 9	1,22 (0,149)				
Lag 10	0,88 (0,711)				
Lag 11	1,11 (0,288)				
Lag 12	0,90 (0,669)				
Lag 13	1,25 (0,116)				
Lag 14	1,27 (0,103)				
Lag 15	0,98 (0,510)				
Lag 16	0,94 (0,592)				

Libro INGENIERIA.indb 306 8/1/08 16:44:54

Pasos 3, 4 y 5: La tabla 3 presenta los resultados obtenidos al estimar un modelo ANFIS homocedástico; LLF es el logaritmo de la función de verosimilitud de los residuales; Box-Pierce y Ljung-Box son los estadísticos para determinar correlaciones residuales remanentes; McLeod-Lee es un estadístico para determinar si existen relaciones no lineales remanentes en los residuales, y ARCH es la prueba de Engle para determinar si los residuales son heterocedásticos. Las pruebas de normalidad realizadas rechazan la hipótesis nula de normalidad de los residuales, lo cual es consistente con el resultado de la prueba ARCH, el cual indica que ellos son heterocedásticos. No existen autocorrelaciones remanentes en los residuales. Este análisis indica que el modelo ANFIS propuesto no captura completamente la dinámica de la serie, y se debe proceder a incluir una componente heterocedástica a través de un modelo ARCH de orden 1. La presencia de una estructura ARCH en los residuales indica una persistencia en el cambio mes a mes de la serie; esto es, que una variación mensual grande en el SOI tiende a ser seguida por otra variación grande y que variaciones pequeñas tienden a ser seguidas por nuevas variaciones pequeñas.

El modelo final elegido corresponde a un modelo ANFIS-ARCH con un orden en los consecuentes de 16 rezagos y un orden ARCH de 1 rezago. Los parámetros para este modelo son presentados en la tabla 4.

La figura 4 resume las principales propiedades estadísticas para los residuales normalizados; la FAC y el espectro indican que los residuales parecen ser incorrelacionados; además el histograma indica que son aproximadamente normales. Las pruebas estadísticas para este modelo se encuentran reportadas en la tabla 5. Aunque los residuales no siguen una distribución normal, los contrastes aplicados a ellos indican que son homocedásticos y que no existen correlaciones remanentes.

Las figuras 2(b) y 2(c) presentan las innovaciones y varianzas del modelo propuesto. Aunque la serie presenta una componente importante en su volatilidad, es importante resaltar la presencia de un número importante de cambios extremos en el SOI de un mes a otro. Estos se evidencian en la figura 2(c) como varianzas puntuales extremadamente altas en relación al resto de la serie.

La figura 5 muestra la función de transición del modelo en el tiempo y como función de la variable de transición. En dicha figura se muestra cómo la función de transición fluctúa rápidamente de un régimen al otro; ello claramente confirma que la dinámica de la serie es no lineal.

Tabla 3. Pruebas estadísticas para el modelo ANFIS homocedástico.

Tamaño de la Muestra				**1455**
Sumatoria				0.172
Media				1,18E-04
Mediana				0,005333
Desviación Estándar				1,006
Asimetría				-0,08791
Curtosis				3 813
Exceso de curtosis				0,8936
Mínimo				-4,846
Máximo				4,1240
Test de Normalidad de Jarque-Bera				49,7 (0,000)
Test de Normalidad Lin-Mudholkar				1,148 (0,126)
LLF				**-5,0221E+03**
Lag	**Box-Pierce**	**Ljung-Box**	**McLeod-Lee**	**ARCH**
1	0,02 (0,889)	0,02 (0,889)	6,80 (0,009)	**6,84 (0,009)**
2	0,10 (0,954)	0,10 (0,953)	7,40 (0,025)	3,89 (0,143)
3	0,10 (0,992)	0,10 (0,992)	13,19 (0,004)	4,74 (0,192)

Tabla 4. Parámetros del modelo ANFIS-ARCH.

Parámetro	**Valor**	**Error Estándar**	**Estadístico T**
Center Z()	1,35950	34,800	**0,0391**
Gamma Z()	0,04374	579,760	**0,0001**
Center S()	0,98385	34,800	**0,0283**
Gamma S()	0,60587	812,640	**0,0007**
Constant1	-1,13030	255,500	-0,0044
AR1 Lag[1]	0,43481	13,642	0,0319
AR1 Lag[2]	0,19183	6,743	0,0284
AR1 Lag[3]	-0,03865	23,988	-0,0016
AR1 Lag[4]	0,05474	72,545	0,0008
AR1 Lag[5]	0,04668	7,313	0,0064
AR1 Lag[6]	0,00425	10,062	0,0004
AR1 Lag[7]	0,01603	17,191	0,0009
AR1 Lag[8]	-0,01492	51,192	-0,0003
AR1 Lag[9]	0,12360	20,397	0,0061
AR1 Lag[10]	-0,11776	90,805	-0,0013
AR1 Lag[11]	0,06870	33,941	0,0020
AR1 Lag[12]	-0,08133	27,530	-0,0030
AR1 Lag[13]	-0,04987	23,847	-0,0021
AR1 Lag[14]	-0,08370	19,654	-0,0043
AR1 Lag[15]	0,01880	9,620	0,0020
AR1 Lag[16]	0,03620	25,201	0,0014
Constant2	1,84260	507,630	0,0036
AR2 Lag[1]	0,49224	6,269	**0,0785**
AR2 Lag[2]	0,05242	74,794	0,0007
AR2 Lag[3]	0,00070	340,960	0,0000
AR2 Lag[4]	0,06602	92,441	0,0007
AR2 Lag[5]	0,03447	3,153	0,0109
AR2 Lag[6]	0,17630	112,160	0,0016
AR2 Lag[7]	-0,08950	47,412	-0,0019
AR2 Lag[8]	-0,00138	63,357	0,0000
AR2 Lag[9]	-0,05001	116,150	-0,0004
AR2 Lag[10]	0,00772	196,180	0,0000
AR2 Lag[11]	-0,09185	122,820	-0,0007
AR2 Lag[12]	0,12749	186,150	0,0007
AR2 Lag[13]	0,07510	24,742	0,0030
AR2 Lag[14]	-0,08866	47,883	-0,0019
AR2 Lag[15]	-0,18384	164,780	-0,0011
AR2 Lag[16]	0,06245	31,920	0,0020
Var Constant	57,61900	1640,900	**0,0351**
ARCH Lag[1]	0,04272	6,369	**0,0067**

Libro INGENIERIA.indb 307 8/1/08 16:44:56

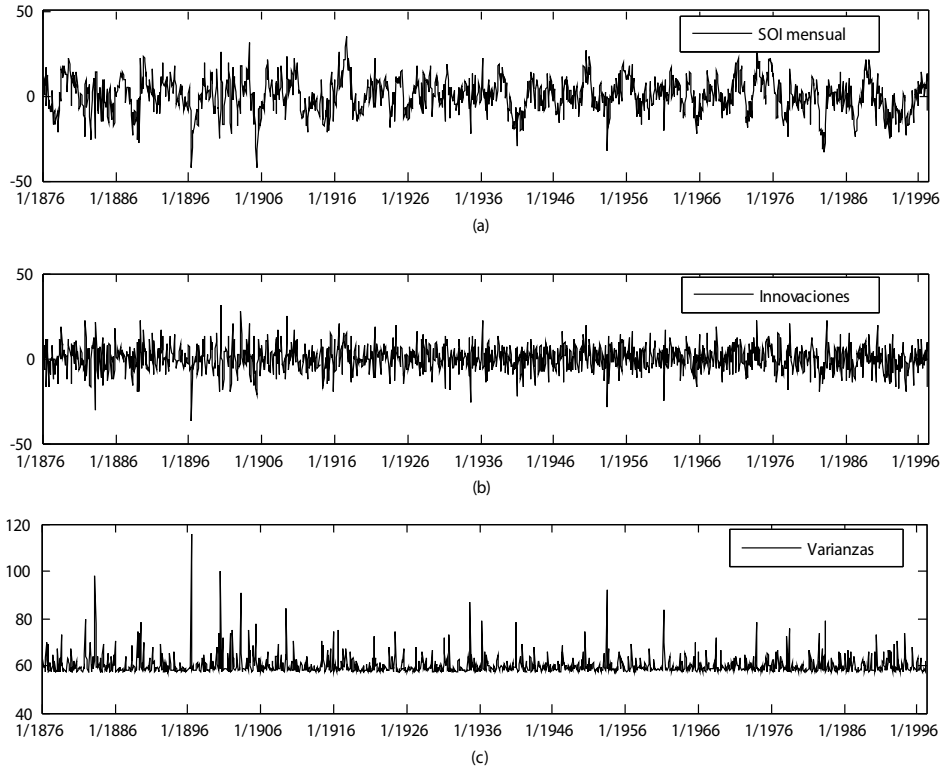

Figura 2. (a) SOI, (b) Innovaciones del modelo ANFIS-ARCH y (c) Varianzas del modelo ANFIS-ARCH.

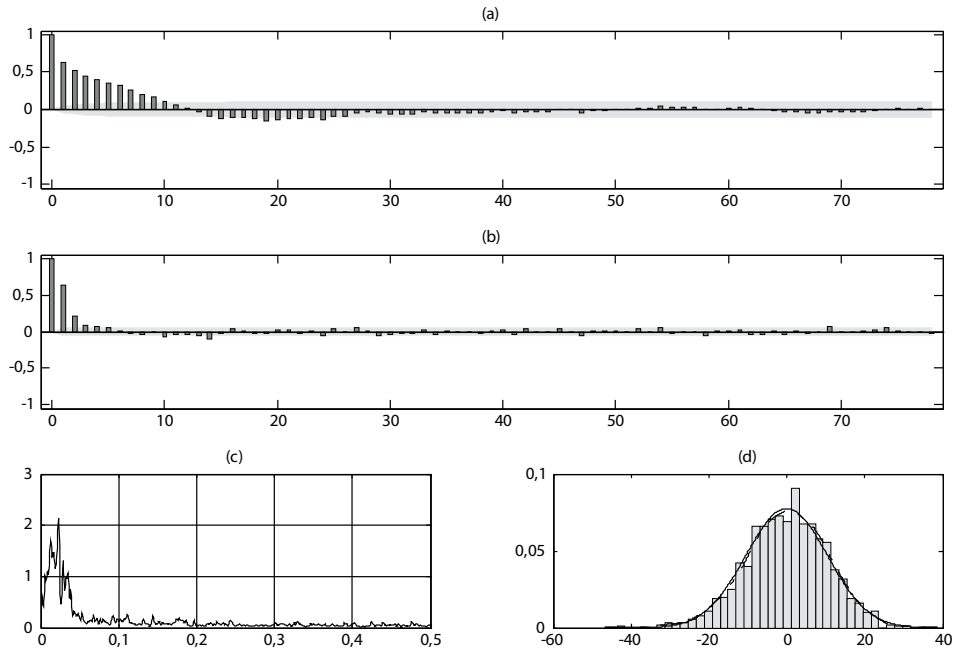

Figura 3. Propiedades estadísticas de la serie del SOI. (a) FAC; (b) FACP; (c) Espectro; (d) Histograma.

Libro INGENIERIA.indb 308 8/1/08 16:44:59

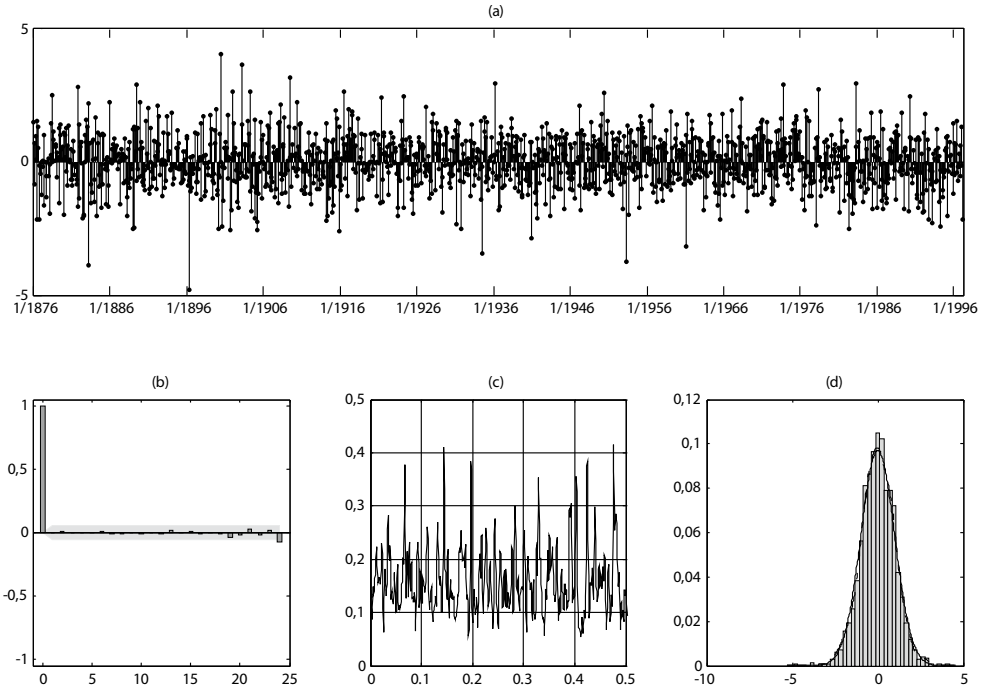

Figura 4. Propiedades estadísticas de los residuales del modelo ANFIS-ARCH. (a) Residuales; (b) FAC; (c) Espectro; (d) Histograma.

La figura 6 presenta el pronóstico determinístico para la serie; y_{T+1} se calculó utilizando los datos históricos. Para calcular y_{T+2} se utilizaron los datos reales hasta y_T, y el pronóstico para y_{T+1}; los restantes valores se calcularon de forma similar; en todos los casos se asume que los residuales son cero. Se observa cómo la serie pronosticada converge rápidamente a un punto de equilibrio donde las entradas y la salida del modelo son iguales numéricamente.

La figura 7 muestra la convergencia a largo plazo del modelo; ella se obtiene al realizar el pronóstico determinístico utilizando diferentes puntos de arranque obtenidos de los datos históricos. En general, los modelos no lineales pueden presentar procesos de convergencia a uno o más puntos estables; procesos cíclicos y procesos divergentes; es común que, dependiendo del punto de inicio, un mismo modelo presente distintos comportamientos. Para el modelo obtenido se presenta un proceso de convergencia a un único punto de equilibrio, y la velocidad con que el modelo converge a él es aproximadamente constante e independiente del tipo de fenómeno extremo.

Otra forma de analizar qué tan bien son reproducidas las propiedades estadísticas de los datos es generar una serie sintética utilizando el modelo estimado y comparar las propiedades estadísticas de los datos y de la serie sintética. En la tabla 6 se presentan las propiedades estadísticas de la serie original y de una serie sintética de 10.000 datos de longitud; puede concluirse que el modelo reproduce bastante bien las propiedades de la serie original.

Tabla 5. Pruebas estadísticas para el modelo ANFIS-ARCH.

Tamaño de la Muestra				1455
Sumatoria				-0,5652
Media				-0,0003885
Mediana				-0,002709
Desviación Estándar				0,984
Asimetría				-0,08914
Curtosis				3,927
Exceso de Curtosis				0,9272
Mínimo				-4,805
Máximo				4,061
Test de Normalidad de Jarque-Bera				53,43 (0,000)
Test de Normalidad Lin-Mudholkar				1,17 (0,121)
LLF				**-5,0193E+03**
Lag	**Box-Pierce**	**Ljung-Box**	**McLeod-Lee**	**ARCH**
1	0,00 (0,998)	0,00 (0,998)	0,68 (0,409)	0,68 (0,409)
2	0,12 (0,943)	0,12 (0,942)	1,59 (0,452)	0,82 (0,665)
3	0,14 (0,986)	0,14 (0,986)	7,78 (0,051)	2,68 (0,443)

Libro INGENIERIA.indb 309 8/1/08 16:45:01

Hall, Skalin y Teräsvirta [7] proponen un modelo LSTAR de orden 13 para caracterizar la serie del SOI; los estadísticos de sus residuales son presentados en la tabla 7, paralelamente con los estadísticos de los residuales normalizados del presente trabajo. El criterio de Akaike, a pesar de ser independiente de si están o no normalizados los residuales, no puede ser usado para comparar los modelos entre sí, ya que no poseen la misma estructura. Sin embargo, el R^2 del modelo ANFIS-ARCH es menor al igual. La diferencia en las desviaciones estándar es debida a la normalización de los residuales de este trabajo. La prueba ARCH, en el caso LSTAR, indica que efectivamente estos modelos requerían un modelado adicional de la varianza de los residuales, que se resuelve con los modelos ANFIS-ARCH.

Tabla 6. Momentos de la serie real y el modelo.

Modelo	Media	Varianza	Kurtosis	Asimetría
Real	0,0559	108,5393	3,3799	-0,2116
ANFIS-ARCH	0,1061	114,6941	2,9895	-0,0107

Tabla 7. Estadísticos del los modelos LSTAR y ANFIS.

Estadístico	LSTAR		ANFIS-ARCH	
T	1455		1455	
R2	0,467		0,463	
AIC	4,12		6,954	
S	7,769		0,984	
S Lineal	7,824		7,711	
Asimetría	-0,071		-0,089	
Exceso kurtosis	0,974		0,927	
LJB	58,788	(p 0,000)	53,43	(p 0,000)
LMARCH	2,544	(p 0,038)	2,02	(p 0,733)

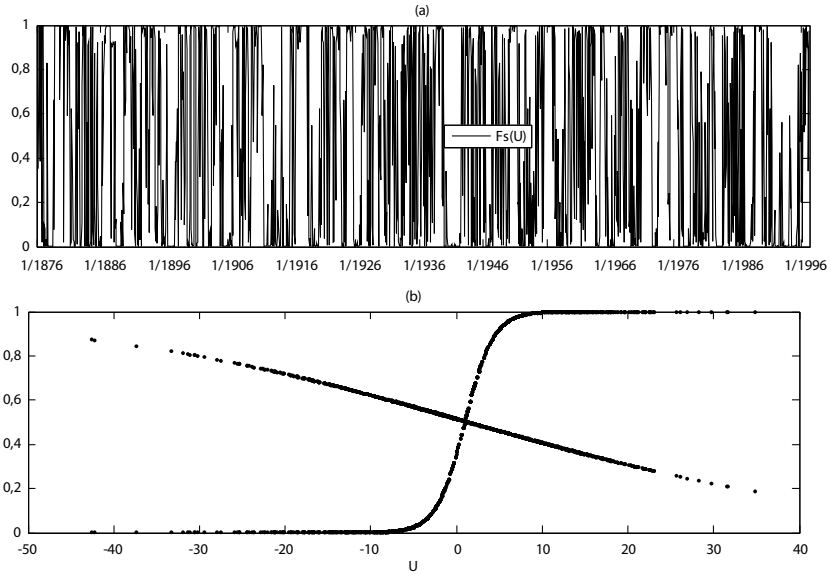

Figura 5. Función de transición para el modelo ANFIS-ARCH. (a) En el tiempo, (b) Como función de la variable de transición.

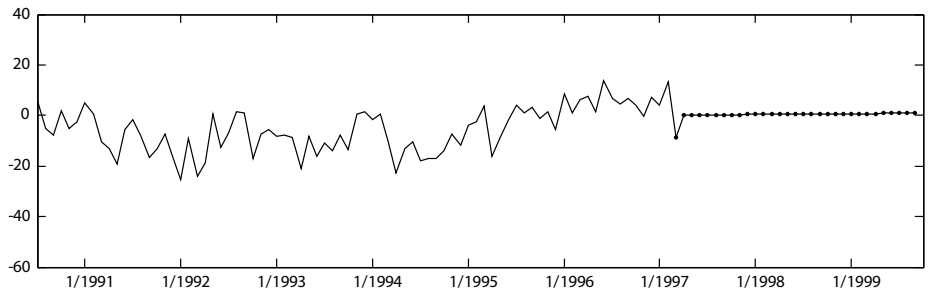

Figura 6. Pronóstico Determinístico.

Libro INGENIERIA.indb 310 8/1/08 16:45:05

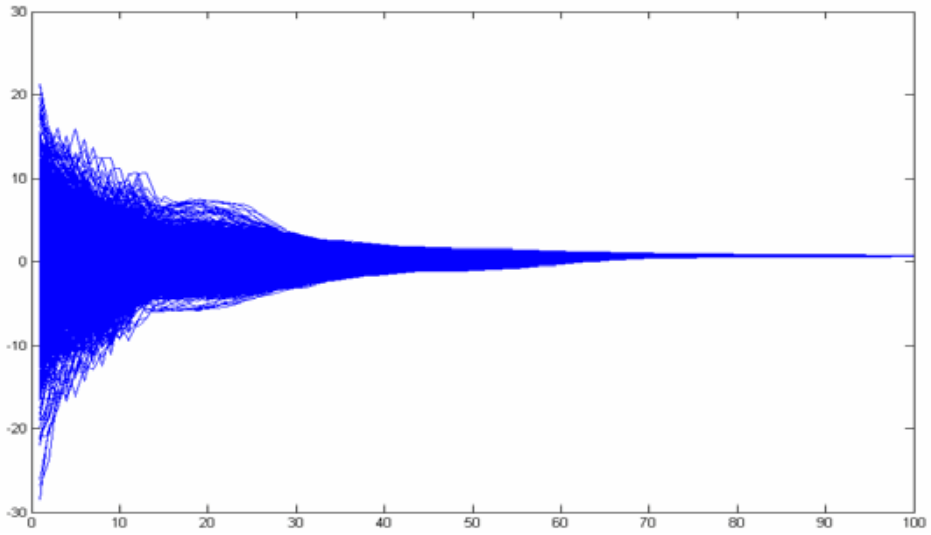

Figura 7. Convergencia de Largo Plazo.

CONCLUSIONES

En este artículo se ha desarrollado un modelo del SOI que permite representar simultáneamente la estructura no lineal en la media y la varianza que posee la serie. El proceso de especificación realizado está compuesto por una batería de pruebas de especificación que permiten probar, desde principios estadísticos, las bondades del ajuste del modelo a la serie estudiada. Como ejercicio pedagógico, este trabajo ha permitido una mayor comprensión del modelado de series temporales no lineales, del uso de las técnicas de Inteligencia Computacional y del fenómeno mismo. Es difícilmente justificable que un modelo univariado, como el propuesto, pueda alcanzar mejores resultados que los alcanzados en este trabajo o en otros similares, por lo que la siguiente fase de estudio corresponde al modelado de esta misma serie, incluyendo variables exógenas que podrían explicar su comportamiento.

REFERENCIAS

[1] A. Abraham. "Neuro-fuzzy systems: State-of-the-art modeling techniques, connectionist models of neurons, learning processes and artificial intelligence". Technical Report. Springer-Verlag. Germany. 2001.

[2] T. Bollerslev. "Generalized autoregressive conditional heteroskedasticity". Journal of Econometrics. Vol. 31, pp. 307-327. 1986.

[3] F. Chan, D. Marinova and M. McAleer. "Modelling thresholds and volatility in US ecological patents". Environmental Modelling & Software. Volume 20 Issue 11, pp. 1369-1378. November, 2005.

[4] B.R. Chang. "Applying nonlinear generalized autoregressive conditional heteroscedasticity to compensate ANFIS outputs tuned by adaptive support vector regression". Fuzzy Sets and Systems. Vol. 157 Issue 13, pp. 1832-1850. 2006.

[5] R. Engle. "Autoregressive conditional heterocedasticity with estimates of the variance of United Kingdom inflations". Econometrica. Vol. 50, pp. 987-1007. 1982.

[6] M.H. Glantz. "Currents of Change El Niño`s impacts on climate and society". Cambridge University Press. 1996.

[7] A.D. Hall, J. Skalin J. and T. Teräsvirta. "A nonlinear time series model of El Niño". Stockholm School of Economics Working Paper Series in Economics and Finance. Nº 263. 1998.

[8] IDEAM. "Efectos naturales and socioeconómicos del Fenómeno El Niño en Colombia". Ministerio del Medio Ambiente, Instituto de Hidrología Meteorología y Estudios Ambientales. República de Colombia. Bogotá D.C. 2002.

Libro INGENIERIA.indb 311 8/1/08 16:45:06

[9] J. Jang. "Anfis: Adaptive-network-based fuzzy inference system". IEEE Transactions on Systems, Man, and Cybernetics. Vol. 23, pp. 665-684. 1993.

[10] R. Jang. "Structure determination in fuzzy modeling: a fuzzy cart approach". IEEE international conference on fuzzy systems. 1994.

[11] S.G. Philander. "El Niño, La Niña and The Southern Oscillation". Academic Press Inc. San Diego, California. 1990.

[12] T. Teräsvirta, C.F. Lin and C.W.J. Granger. "Power of the neural network linearity test". Journal of Time Series Analysis. Nº 14, pp. 209-220. 1993.

[13] T. Teräsvirta. "Specification, estimation, and evaluation of smooth transition autoregressive models". Journal of the American Statistical Association. Vol. 89, pp. 208-218. 1994.

[14] J.D. Velásquez, I. Dyner y R.C. Souza. "Modelación de Series Temporales usando ANFIS". Revista Iberoamericana de Inteligencia Artificial. Vol. 23 Nº 17. 2004.

[15] A.T. Wittenberg. "ENSO Response to Altered Climates". Tesis para optar al grado de doctor. Princeton University. 2002.

Libro INGENIERIA.indb 312 8/1/08 16:45:07