Revista Colombiana de Estadística
Volumen 28 N^{o} 2. pp. 113 a 124. Diciembre 2005

Ajuste de un modelo no lineal a la variable precipitación en una estación hidro-meteorológica de Colombia

Fitting a nonlinear model to the precipitation variable in a Colombian hydrological/meteorological station

Ángela P. Briñez[*]
Fabio H. Nieto[**]

Resumen

En la literatura sobre análisis de series temporales, se ha establecido en años recientes que las series hidrológicas y meteorológicas se describen apropiadamente por modelos no lineales, en particular por los modelos SETAR (Self-exciting threshold autoregressive models). Combinamos dos métodos propuestos en la literatura para ajustar un modelo SETAR a datos de precipitación, medida en una estación hidro-meteorológica de Colombia.

Palabras Claves: Modelos SETAR, series de tiempo no lineales, precipitación

Abstract

In the literature on time series analysis it has been established in recent years that the hydrological/meteorological time series are well described by nonlinear models, in particular by the SETAR (Self- exciting threshold autoregressive) models. In this paper, a nonlinear SETAR model is fitted to the precipitation variable that is observed in a certain Colombian hydrological/meteorological station. In the fitting process, two alternative methodologies, which have been proposed in the literature about nonlinear time series models, are combined.

Keywords: SETAR models, Nonlinear Time Series, Precipitation

[*]Estadística de la Universidad Nacional de Colombia. E-mail: apbrinezr@unal.edu.co
[**]Profesor Titular, Departamento de Estadística de la Universidad Nacional de Colombia. E-mail: fhnietos@unal.edu.co

1. Introducción

Se ha encontrado constantemente que muchas series de tiempo son de tipo no lineal y su presencia en la vida cotidiana es más frecuente de lo que se piensa. Particularmente, los procesos de tipo hidrológico y meteorológico quedan bien descritos por modelos no lineales como los SETAR (Self-Exciting Threshold Autoregressive Models) (Tsay 1989, Tong 1990)). En nuestro medio se han venido aplicando de manera inapropiada modelos lineales a este tipo de procesos, lo que hace que se presenten problemas en el área de interpolación y de predicción.

Tong (1990) y Tsay (1989) propusieron alternativas para ajustar modelos SETAR a procesos no lineales. La característica de estos procesos es que su espacio muestral está particionado en regímenes delimitados por umbrales. De esta manera, la metodología propuesta por Tong y Tsay se fundamenta en localizar e identificar los umbrales para luego ajustar modelos autoregresivos en cada régimen.

En particular, Tsay (1989) propone la aplicación de autoregresiones sobre los datos ordenados por magnitud para localizar umbrales. Formula también una prueba para verificar la linealidad del proceso y estimar el parámetro de rezago que determina la variable de umbrales. Tong propone otro método consistente en tomar determinados cuantiles de la serie como candidatos a umbrales, alternando combinaciones de posibles umbrales y de órdenes autoregresivos en cada uno de los regímenes, de tal forma que se haga mínimo el criterio de información de Akaike normalizado (NAIC) (Tong 1990).

La propuesta de Tong (1990) y Tsay (1989) constituye una solución al problema representado por la modelación de series temporales no lineales, pues resulta más eficiente que la aplicación de modelos lineales. En este trabajo se ajusta un modelo no lineal SETAR a la serie diaria de precipitación registrada en la estación Laguna de San Rafael, municipio de Puracé, departamento del Cauca, utilizando las metodologías descritas anteriormente. La idea es aplicar un procedimiento que no ha sido utilizado antes en nuestro medio.

En la Sección 2 se presenta el modelo SETAR, la construcción de la estadística F de la prueba de linealidad y los pasos para ajustar el modelo a un conjunto de datos. En la Sección 3 se muestran los resultados del ajuste del modelo SETAR a la serie de precipitación y finalmente, en la Sección 4 se encuentran las conclusiones.

2. El modelo SETAR

Un proceso estocástico $\{Y_t\}$ obedece un modelo SETAR (Self-Exciting Threshold Autoregressive) si satisface la ecuación:

$$Y_t = \Phi_0^{(j)} + \sum_{i=1}^{p_j} \Phi_i^{(j)} Y_{t-i} + h^{(j)} a_t \quad \text{si} \quad r_{j-1} < Y_{t-d} \leq r_j \tag{1}$$

donde $1 \leq j \leq k$ para un cierto entero positivo k, los números reales $\Phi_i^{(j)}$, $i = 1, \ldots, p_j$; $j = 1, \ldots, k$; son los coeficientes autoregresivos, $\{a_t\}$ es un proceso ruido

blanco de media cero y varianza uno y $h^{(j)}$, $j = 1, 2, \ldots, k$, es un número real positivo. Los números reales $r_1 < r_2 < \cdots < r_{k-1}$ son los umbrales del modelo. Por convención $r_0 = -\infty$ y $r_k = +\infty$. A la variable Y_{t-d} se le llama la variable de umbrales y d es el parámetro de retraso. El modelo (1) es denotado SETAR $(k; p_1, \ldots, p_k; d)$. El intervalo $(r_{j-1}, r_j]$ es el j-ésimo régimen de $\{Y_t\}$. Inicialmente supondremos que $p_1 = p_2 = \cdots = p_j = p$, el orden autoregresivo identificado para toda la serie.

Tsay (1989) propone una prueba estadística para la hipótesis nula de linealidad del proceso contrastada con la alternativa que el proceso obedece a un modelo SETAR, la cual encuentra su fundamento en el ajuste de autoregresiones ordenadas, de tal forma que en el recorrido de $\{Y_t\}$ se encuentren k regímenes delimitados por $k-1$ umbrales. Para probar la linealidad de la serie, se debe proceder de la siguiente manera:

Paso 1: Para el modelo (1), el proceso $\{Y_{t-d}\}$ se reduce a $\{Y_h, \ldots, Y_{n-d}\}$ donde $h = \text{máx}\{1,\ p + 1 - d\}$. Se ordena $\{Y_{t-d}\}$ por magnitud, de esta manera, si los índices de tiempo de la variable ordenada se denotan por π_i, Y_{π_1} es la observación más pequeña, Y_{π_2} es la segunda observación más pequeña, etc.

Paso 2: Se realiza una autoregresión tomando como variable respuesta a Y_{π_i+d} (la variable de umbrales ordenada adelantada d períodos) y como regresores a las variables $\{Y_{\pi_i+d-v}\}$, con $v = 1, \ldots, p$ (los p rezagos de la variable respuesta). Las estimaciones obtenidas mediante el método de mínimos cuadrados recurrentes se calculan de la siguiente forma:

Con las primeras $m = \left[\frac{n}{10} + p\right]$ observaciones de la variable respuesta y de los regresores[1], calcule la matriz $P_m = [X'X]^{-1}$, donde X es la matriz de diseño. El vector de parámetros estimados mediante el método de mínimos cuadrados ordinarios es $\hat{\beta}_m$ y el vector x_{m+1} es tal que contiene los valores de los regresores correspondientes a la observación $m + 1$.

Determine las estimaciones de los parámetros según las recursiones:

$$\hat{\beta}_{m+1} = \hat{\beta}_m + K_{m+1}[Y_{d+\pi_{m+1}} - x'_{m+1}\hat{\beta}_m]$$

$$D_{m+1} = 1 + x'_{m+1}P_m x_{m+1}$$

$$K_{m+1} = \frac{P_m x_{m+1}}{D_{m+1}}$$

$$P_{m+1} = \left(I - P_m \frac{x_{m+1}x'_{m+1}}{D_{m+1}}\right)P_m$$

donde $m = \left[\frac{n}{10} + p\right] + 1, \ldots, n - d - h + 1$ con d y h definidos anteriormente. Los residuales están dados por:

$$\hat{a}_{d+\pi_{m+1}} = a_{d+\pi_{m+1}} - x'_{m+1}\hat{\beta}_m$$

[1] [x] denota la parte entera de x

y los residuales estandarizados por:

$$\hat{e}_{d+\pi_{m+1}} = \frac{\hat{a}_{d+\pi_{m+1}}}{\sqrt{D_{m+1}}}$$

Realice la regresión, por mínimos cuadrados ordinarios, entre los residuales estandarizados $\hat{e}_{\pi_i+d}$ y las variables $\{Y_{\pi_i+d-v}\}$, con $v = 1, \ldots, p$:

$$\hat{e}_{d+\pi_{m+1}} = \omega_0 + \sum_{v=1}^{p} \omega_v Y_{d+\pi_i-v} + \varepsilon_{\pi_i+d}$$

Construya la siguiente estadística, denotada F, con los residuales estandarizados y con los obtenidos en la regresión del paso anterior $\hat{\varepsilon}_i$:

$$\hat{F}(p,d) = \frac{\sum(\hat{e}_i^2 - \hat{\varepsilon}_i^2)/(p+1)}{\sum \hat{\varepsilon}_i^2/(n-d-m-p-h)}$$

Bajo la hipótesis nula de linealidad, la estadística $\hat{F}(p,d)$ tiene distribución asintótica F con $(p+1)$ grados de libertad en el numerador y $(n-d-m-p-h)$ en el denominador. Además, $(p+1)\hat{F}(p,d)$ se distribuye asintóticamente Ji-cuadrado con $(p+1)$ grados de libertad (Tsay 1989).

Si se detecta no linealidad en la serie, se procede a identificar el número de regímenes k y los valores de los umbrales $r_1, \ldots, r_{k-1}$. Para esto, Tsay (1989) propone realizar diagramas de dispersión entre los residuales estandarizados y la variable de umbrales. La idea es que si la serie es no lineal, los residuales deben presentar quiebres en su trayectoria alrededor de los valores de los umbrales, los cuales pueden encontrarse mediante los diagramas de dispersión. Las razones t de las estimaciones recurrentes pueden sustituir a los residuales estandarizados.

Para identificar los umbrales de la serie, Tong (1990) propone un esquema consistente en tomar como candidatos ciertos cuantiles de la serie y considerar todas las posibles combinaciones ordenadas con ellos. Se estudian también diferentes órdenes autoregresivos en cada régimen posible. La idea es seleccionar los umbrales y órdenes autoregresivos que minimicen el criterio NAIC:

$$NAIC = \frac{\sum_{j=1}^{k} AIC_j}{\sum_{j=1}^{k} n_j}$$

donde AIC_j y n_j denotan el criterio de información de Akaike (AIC) y el número de observaciones en el j-ésimo régimen, respectivamente, y k es el número de regímenes.

Posterior a la localización de los umbrales, se procede a ajustar el modelo SETAR mediante los siguientes pasos:

Paso 1: Seleccione un orden autoregresivo p para $\{Y_t\}$ y el conjunto de posibles valores del parámetro de retraso d, llamado S. Defina entonces $S = 1, \ldots, p$.

Paso 2: Ajuste autoregresiones ordenadas para el valor de p dado y todo elemento d de S. Calcule la estadística F y seleccione el valor de d_p tal que

$$\hat{F}(p, d_p) = \max_{v \in S}\{\hat{F}(p, v)\}$$

Paso 3: Para p y d_p dados, localice los valores de los umbrales usando los diagramas de dispersión descritos anteriormente.

Paso 4: Reestime los órdenes autoregresivos p y los valores de los umbrales, si es necesario, utilizando en cada régimen técnicas de autoregresión lineal y el criterio de información de Akaike (AIC), como lo describe Tsay (1989).

Paso 5: Estime los parámetros desconocidos del modelo usando los datos ordenados y el procedimiento de mínimos cuadrados ordinarios en cada régimen (Véase justificación en Tsay 1989).

Paso 6: Use técnicas estándar de verificación y validación del modelo estimado, utilizando los residuales estandarizados y sus cuadrados (Véase Tsay 1989, Tong 1990).

Teóricamente, los modelos SETAR pueden ser utilizados para pronosticar la serie modelada h pasos adelante, para $h = 1$ entero. Para $h = d$, los pronósticos pueden ser calculados de forma analítica, y en el caso donde $h > d$, es necesario recurrir a técnicas de simulación (Tong 1990). Para $h = 1$, obtenemos que:

$$Y_{N+1} = \Phi_0^{(j)} + \sum_{i=1}^{p_j} \Phi_i^{(j)} Y_{N+1-i} + h^{(j)} a_{N+1}$$

si $r_{j-1} < Y_{N+1-d} \leq r_j$, y

$$E(Y_{N+1}|Y_N, \ldots, Y_1) = Y_{N+1|N} = \Phi_0^{(j)} + \sum_{i=1}^{p_j} \Phi_i^{(j)} Y_{N+1-i}$$

si $r_{j-1} < Y_{N+1-d} \leq r_j$.

Como el proceso de ruido $\{a_t\}$ es de media cero y varianza uno, entonces:

$$ECM(Y_{N+1|n}) = E(h^{(j)} a_{N+1})^2 = (h^{(j)})^2$$

3. Ajuste del modelo

La serie de precipitación fue registrada diariamente en la estación Laguna de San Rafael, ubicada en la cuenca del río Bedón, municipio de Puracé, departamento del Valle del Cauca. Contiene un total de 11127 observaciones tomadas desde el 5 de febrero de 1970 hasta el 30 de noviembre de 2000, además de 130 datos faltantes.

La serie toma valores mayores o iguales que cero, con una alta frecuencia de este último valor (20.52 %), lo cual indica que $Pr[Z = 0] > 0$, con Z la variable precipitación. Debido a esta característica se encuentra que su distribución no tiene función de densidad (Nieto 2002), lo cual impide suponer que el proceso de ruido del modelo sea Gaussiano. Sin embargo, esto no restringe la aplicabilidad del método de Tsay (1989).

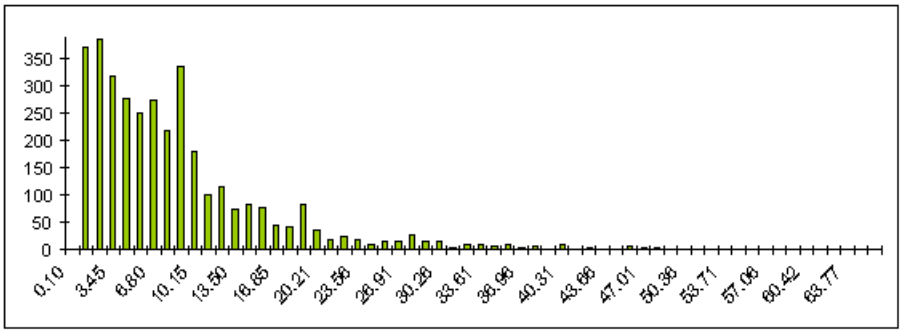

Figura 1: Histograma de la serie de precipitación excluyendo los valores iguales a cero

En cuanto a sus cuantiles, el cuantil 0.25 es igual a 3, el cuantil 0.5 es 6, el cuantil 0.75 es 10.47 y el valor máximo en la serie es 66. La mayor parte de los datos son menores que 10.5, lo cual indica que es poco usual el registro de valores de precipitación altos en la estación donde fueron tomados.

Se encuentra que es necesario estimar los datos faltantes de la serie de precipitación debido a los requerimientos de la metodología de Tsay (1989), la cual se basa en datos igualmente espaciados. La estimación aproximada de los datos faltantes se hizo mediante la metodología de Nieto & Ruiz (2002) diseñada para implementarse con el programa TRAMO (Gomez & Maravall 1996). El modelo identificado es un $AR(1)$ y con él se realizan las estimaciones de los datos faltantes.

De aquí en adelante se analizan los últimos 12 años de la serie completada, la cual contiene 4384 datos correspondientes al período comprendido entre el 30 de noviembre de 1988 y 30 de noviembre de 2000, en virtud de la magnitud de los cálculos computacionales. Este período incluye el fenómeno del Niño.

Prueba de linealidad: Para implementar la prueba de linealidad es necesario reemplazar los valores iguales a cero por un valor pequeño como 10^{-7}, para evitar problemas en el cómputo de matrices inversas en el método de mínimos cuadrados recurrentes sobre el cual está basado el cálculo de la estadística de prueba. Al observar la función de autocorrelación parcial, se encuentra que un orden autoregresivo razonable para la serie es 7. Por lo tanto, el conjunto S de posibles valores del rezago de la variable de umbrales d está dado por $S = \{1, 2, 3, 4, 5, 6, 7\}$.

A excepción del caso $d = 2$, se rechaza la hipótesis de linealidad en todas las ocasiones. Los resultados de los cálculos se presentan en la tabla 1 y se observa que el valor de d que maximiza la estadística F de la prueba de no linealidad es 1, para el cual $F = 111.77218$ con un p-valor < 0.001. En síntesis, la serie de precipitación es no lineal porque en la mayoría de los casos se rechaza la hipótesis de linealidad y el valor de d identificado para el modelo es igual a 1.

Identificación del modelo SETAR: Para encontrar el número de umbrales y sus localizaciones, se obtienen los residuales, su estandarización y las razones t de los parámetros estimados de forma recurrente, tomando $d = 1$.

Tabla 1: p-valores de las estadísticas F calculadas para cada valor de d

d	$F_{8,\ 3924}$	p-valor
1	111.77218	< 0.001
2	0.7106268	0.682441602
3	2.6827729	0.006111454
4	6.5375984	< 0.001
5	5.5992641	< 0.001
6	3.8611298	0.000153021
7	2.0225969	0.040148565

El diagrama de dispersión de los residuales estandarizados contra la variable de umbrales Y_{t-1} proporciona poca información en cuanto a la ubicación de los umbrales pues si la serie no es lineal se espera encontrar un comportamiento heterogéneo con algún tipo de partición. Se opta por la recomendación de graficar las razones t de los parámetros estimados de forma recurrente contra la variable de umbrales Y_{t-1} (Tsay 1989). Los coeficientes autoregresivos significativos según los valores de las razones t fueron $\hat{\phi}_3$, $\hat{\phi}_4$, $\hat{\phi}_5$ y $\hat{\phi}_6$.

Los diagramas de dispersión en las figuras 2 a 5 no proporcionan información contundente en cuanto al número y localización de los umbrales. En el caso de $\hat{\phi}_3$, los posibles umbrales se sitúan alrededor de 4, 9 y 17, donde se presentan quiebres; en el diagrama correspondiente a $\hat{\phi}_4$ se observa una partición cerca de 10; para el caso de $\hat{\phi}_5$ los candidatos a umbrales son aproximadamente 1, 3 y 8 debido a que en estos puntos se presentan quiebres y en el diagrama de $\hat{\phi}_6$ hay particiones en 7 y 9. En virtud de lo anterior, se hace necesario utilizar otro procedimiento para localizar e identificar los umbrales de la serie.

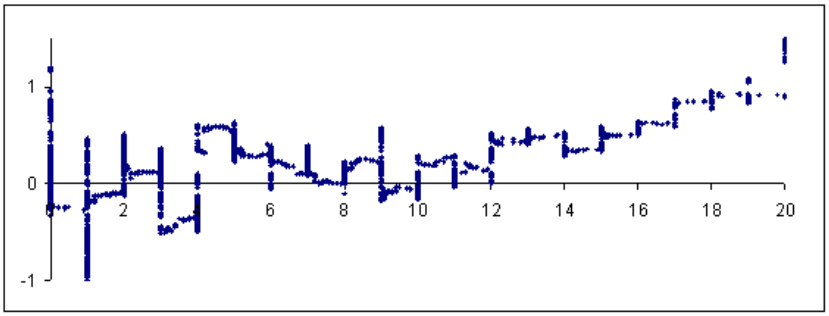

Figura 2: Diagrama de dispersión de $\hat{\phi}_3$ contra la variable de umbrales ordenada Y_{t-1}

Se considera la alternativa propuesta por Tong (1990). Como candidatos a umbrales se escogen los deciles de la serie: 1, 2, 3, 5, 6, 8, 9.3, 12 y 17.5 y se supone como máximo la existencia de cuatro posibles regímenes a la luz de las figuras 2 a 5. Se probaron en total 129 combinaciones donde se contempla la existencia de

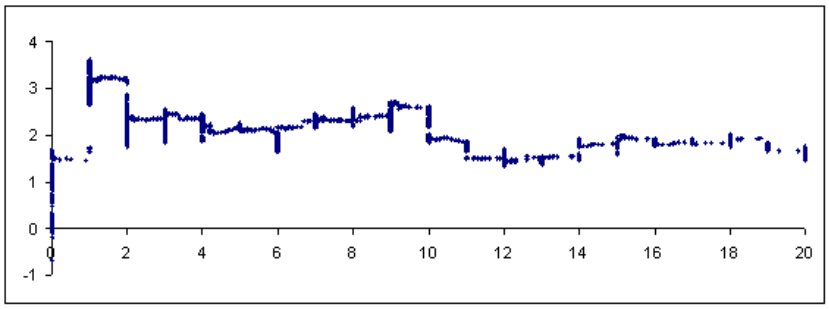

Figura 3: Diagrama de dispersión de $\hat{\phi}_4$ contra la variable de umbrales ordenada Y_{t-1}

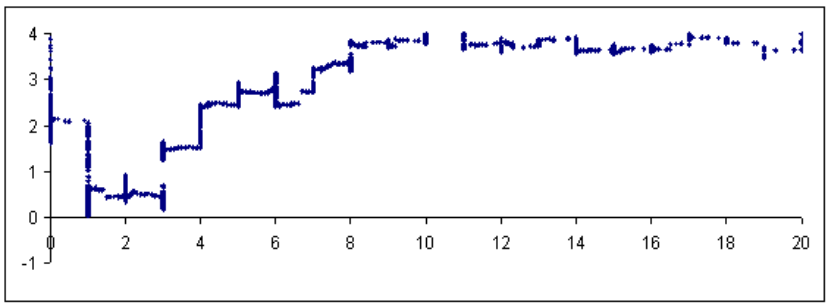

Figura 4: Diagrama de dispersión de $\hat{\phi}_5$ contra la variable de umbrales ordenada Y_{t-1}

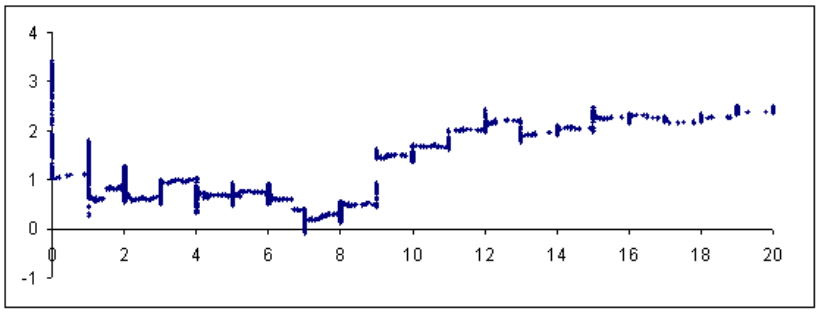

Figura 5: Diagrama de dispersión de $\hat{\phi}_6$ contra la variable de umbrales ordenada Y_{t-1}

un umbral (r_1), de dos umbrales ($r_1 < r_2$) y de tres umbrales ($r_1 < r_2 < r_3$). Para escoger el orden autoregresivo en cada uno de los posibles regímenes se observa la función de autocorrelación parcial. El mínimo NAIC ocurrió para tres umbrales: 8, 9.3 y 17.5. Los órdenes autoregresivos identificados son 6 en el régimen I, 7 en el régimen II, 1 en el régimen III y 4 en el régimen IV. El valor del criterio NAIC es 10.4349. Se encuentran modelos con valores NAIC menores, pero éste fue el único que superó todas las pruebas de la etapa de verificación.

Tabla 2: Algunos casos donde se considera la existencia de tres umbrales

r_1	r_2	r_3	n_1	n_2	n_3	n_4	p_1	p_2	p_3	p_4	NAIC	6	12	18	24
5	8.0	17.5	2452	679	896	350	11	3	1	4	10.1719	0.2959	0.0939	0.0104	0.0002
5	9.3	12.0	2452	883	361	681	11	1	1	3	10.1785	0.0625	0.0304	0.0020	< 0.0001
5	9.3	17.5	2452	883	692	350	11	1	1	4	10.1674	0.2320	0.0621	0.0048	< 0.0001
5	12.0	17.5	2452	1244	331	350	11	1	1	4	10.2347	0.2251	0.0672	0.0041	< 0.0001
6	8.0	9.3	2725	406	204	1042	11	6	7	3	10.3312	0.5801	0.0799	0.0074	0.0009
6	8.0	12.0	2725	406	565	681	11	6	1	3	10.2562	0.4348	0.0251	0.0008	< 0.0001
6	8.0	17.5	2725	406	896	350	11	6	1	4	10.2716	0.6815	0.0336	0.0014	< 0.0001
6	9.3	12.0	2725	610	361	681	11	1	1	3	10.2601	0.1097	0.0020	< 0.0001	< 0.0001
6	9.3	17.5	2725	610	692	350	11	1	1	4	10.2490	0.2683	0.0034	< 0.0001	< 0.0001
6	12.0	17.5	2725	971	331	350	11	1	1	4	10.2900	0.2671	0.0035	< 0.0001	< 0.0001
8	9.3	12.0	3131	204	361	681	6	7	1	3	10.4460	0.8281	0.0371	0.0015	< 0.0001
8	**9.3**	**17.5**	**3131**	**204**	**692**	**350**	**6**	**7**	**1**	**4**	**10.4349**	**0.9141**	**0.0398**	**0.0018**	< 0.0001
8.0	12	17.5	3131	565	331	350	6	1	1	4	10.4134	0.8454	0.0074	< 0.0001	< 0.0001
9.3	12	17.5	3335	361	331	350	6	1	1	4	10.5168	0.9466	0.0305	0.0009	< 0.0001

A manera de ilustración se incluyen en la tabla 2 los resultados obtenidos para 14 de las 129 combinaciones consideradas. Las columnas encabezadas con r_i contienen el valor del i-ésimo umbral considerado; las encabezadas con p_i muestran los órdenes autoregresivos identificados en el i-ésimo régimen por medio del estudio de la función de autocorrelación parcial; seguidamente se encuentran n_i que indica el número de observaciones en el i-ésimo régimen y la columna NAIC que indica el valor del criterio de información NAIC correspondiente a cada caso. En las últimas cuatro columnas se encuentran los p-valores de la prueba Q aplicada a los residuales, correspondientes a los rezagos 6, 12, 18 y 24. Se resaltan los resultados obtenidos para la combinación de umbrales del modelo seleccionado.

Estimación del modelo: En los regímenes determinados por los umbrales 8, 9.3 y 17.5, se estiman modelos autoregresivos por el método de mínimos cuadrados ordinarios y se obtiene que el modelo SETAR para la serie de precipitación está dado por (desviaciones estándar en paréntesis):

$$Y_t = \begin{cases} \underset{(0.1757)}{2.4436} + \underset{(0.0388)}{0.5600}Y_{t-1} + \underset{(0.0191)}{0.0207}Y_{t-2} - \underset{(0.0175)}{0.0017}Y_{t-3} + \underset{(0.0171)}{0.0343}Y_{t-4} + \\ \qquad \underset{(0.0169)}{0.0542}Y_{t-5} + \underset{(0.0157)}{0.0037}Y_{t-6} + 5.3812a_t^{(1)}, & \text{si } Y_{t-1} \leq 8 \\ \underset{(18.2454)}{10.6705} - \underset{(2.0367)}{0.5410}Y_{t-1} + \underset{(0.0696)}{0.0399}Y_{t-2} - \underset{(0.0626)}{0.0557}Y_{t-3} + \underset{(0.0627)}{0.0877}Y_{t-4} + \\ \underset{(0.0633)}{0.0158}Y_{t-5} + \underset{(0.0660)}{0.1568}Y_{t-6} + \underset{(0.0653)}{0.0655}Y_{t-7} + 5.9330a_t^{(2)}, & \text{si } 8 < Y_{t-1} \leq 9.3 \\ \underset{(1.6956)}{3.7159} + \underset{(0.1323)}{0.48199}Y_{t-1} + 7.7290a_t^{(3)}, & \text{si } 9.3 < Y_{t-1} \leq 17.5 \\ \underset{(2.0677)}{1.8206} + \underset{(0.0762)}{0.4239}Y_{t-1} - \underset{(0.0592)}{0.0958}Y_{t-2} + \underset{(0.0762)}{0.1535}Y_{t-3} + \underset{(0.0692)}{0.2488}Y_{t-4} \\ \qquad + 11.5051a_t^{(4)}, & \text{si } 17.5 < Y_{t-1} \end{cases}$$

El número de observaciones por régimen es 3131 en el primero, 204 en el segundo, 692 en el tercero y 350 en el cuarto. Las varianzas de los residuales en los regímenes I, II, III y IV son 28.95, 35.20, 59.73 y 132.36, respectivamente. Se considera que el régimen I corresponde a precipitación nula, el régimen II a poca precipitación, el régimen III a precipitación normal y el régimen IV a precipitación alta.

En el primer régimen hay una gran cantidad de observaciones en comparación con los demás; esto se debe a la alta frecuencia del valor cero y a que la estación meteorológica donde fueron registrados los valores de la serie se encuentra en un lugar de clima seco. En particular, la varianza de los residuales en el régimen IV es alta si se le compara con la de los otros regímenes; esto se explica por la magnitud del rango de variación de los datos correspondientes a éste régimen, ya que la serie de precipitación toma valores entre 17.5 y 66.

Verificación del modelo: En cada régimen, se dividen los residuales por su correspondiente desviación estándar, luego se agrupan todos en un solo conjunto

y se reordenan con respecto al índice de tiempo original. Estos valores constituyen la serie temporal para el proceso. Se encuentra que no se rechaza la hipótesis de ruido blanco en los residuales usando la estadística de Ljung-Box con rezago 6 con un p-valor de 0.9141. De igual manera, la prueba Q de los residuales al cuadrado no rechaza la hipótesis nula hasta el rezago 6 con un p-valor de 0.1492.

En las figuras 6 y 7 se muestran las cartas CUSUM y CUSUMSQ construidas con una confianza del 99 %. Se aprecia que el modelo está razonablemente especificado, en vista que las funciones calculadas no sobrepasan las bandas de confianza.

Por lo tanto, a la luz de los criterios anteriores, el modelo está ajustado razonablemente.

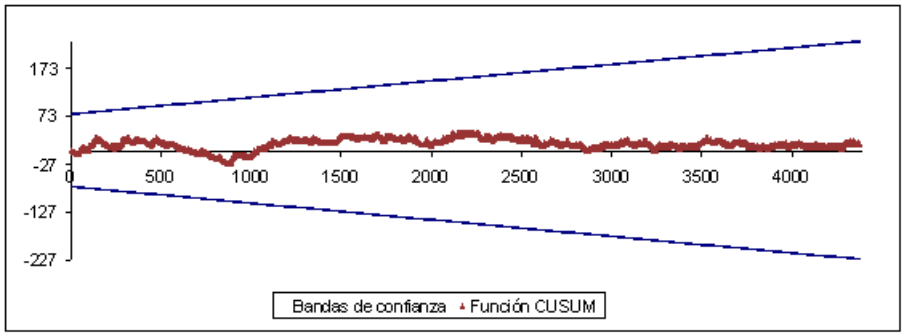

Figura 6: Carta CUSUM para los residuales del modelo con banda de confianza del 99 %

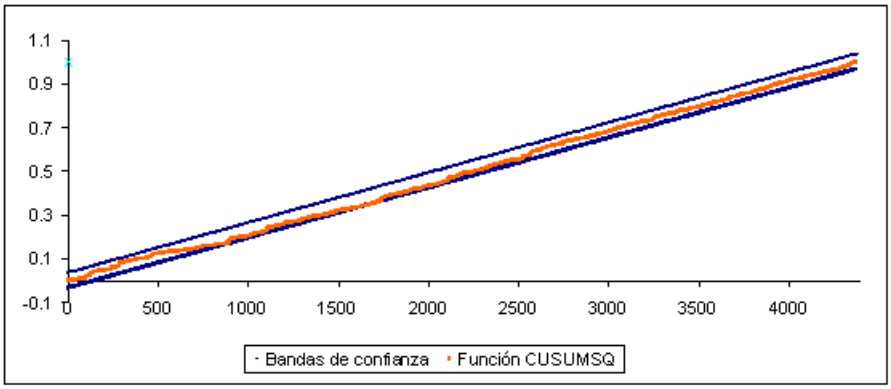

Figura 7: Carta CUSUMSQ para los residuales del modelo con banda de confianza del 99 %

Cálculo de pronósticos: La predicción para la serie de precipitación un paso adelante es igual a 7.5537 y corresponde al día 1 de diciembre de 2000, con error cuadrático medio igual a 28.95837. No se conoce la función de densidad de proba-

bilidades de la variable precipitación, debido a que es una variable de tipo mixto (Nieto 2002). Por lo tanto, no es posible calcular un intervalo de confianza para el pronóstico calculado.

4. Conclusiones

El modelo ajustado a la serie de precipitación registrada en la estación Laguna de San Rafael, ubicada en la cuenca del río Bedón, municipio de Puracé, departamento del Valle del Cauca es un $SETAR(4; 6, 7, 1, 4; 1)$, el cual se obtiene mediante la combinación de las metodologías de (Tsay 1989) y (Tong 1990). La primera metodología no es directamente aplicable en virtud de la alta subjetividad que se debía utilizar en la identificación de los umbrales.

En el modelo se identifican cuatro regímenes, interpretados como niveles de lluvia nula, poca, normal y alta. El valor del parámetro de rezago de la variable de umbrales indica que la precipitación contemporánea se ve determinada de cierta manera por la registrada el día anterior.

El modelo se ajustó a una variable aleatoria de tipo mixto, es decir, una variable con partes discreta y continua. Esta característica impide suponer Gaussianidad del ruido blanco y en consecuencia, impide igualmente examinar esta hipótesis nula con los residuales.

Recibido: 23 de Marzo de 2005
Aceptado: 3 de Junio de 2005

Referencias

Gomez, V. & Maravall, A. (1996), *Programs TRAMO and SEATS: Instructions for the user*, Banco de España-Servicio de estudios, Madrid.

Nieto, F. (2002), *Interpolation of nonlinear TAR models.*, Unidad de Investigación, Departamento de Estadística, Universidad Nacional de Colombia, Bogotá. Reporte Interno No. 1.

Nieto, F. & Ruiz, F. (2002), 'About a prompt strategy for estimating missing data in long time series', *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales* **100**(26), 411–418.

Tong, H. (1990), *Nonlinear Time Series, A Dynamical System Approach*, Oxford University Press, Oxford.

Tsay, R. S. (1989), 'Testing and modeling threshold autoregressive processes', *Journal of the American Statistical Association* **84**, 231–240.