

*Modelamiento y pronósticos usando modelos TAR en
algunas series de tiempo hidrológicas y meteorológicas del
Tolima*

JUAN CAMILO GÓMEZ CEBALLOS
ESTUDIANTE DEL PROGRAMA DE MATEMÁTICAS CON ÉNFASIS EN ESTADÍSTICA
CÓDIGO: 070250142010

TRABAJO DE GRADO PRESENTADO PARA OPTAR AL TÍTULO DE
PROFESIONAL EN MATEMÁTICAS CON ÉNFASIS EN ESTADÍSTICA

DIRECTOR
JOAQUÍN GONZÁLEZ BORJA
PROFESOR DEL DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA



UNIVERSIDAD DEL TOLIMA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA
IBAGUÉ, TOLIMA
2018

*Modelamiento y pronósticos usando modelos TAR en
algunas series de tiempo hidrológicas y meteorológicas del
Tolima*

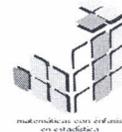
JUAN CAMILO GÓMEZ CEBALLOS

ESTUDIANTE DEL PROGRAMA DE MATEMÁTICAS CON ÉNFASIS EN ESTADÍSTICA

CÓDIGO: 070250142010



UNIVERSIDAD DEL TOLIMA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA
IBAGUÉ, TOLIMA
2018



UNIVERSIDAD DEL TOLIMA

FACULTAD DE CIENCIAS

PROGRAMA DE MATEMÁTICAS CON ÉNFASIS EN ESTADÍSTICA

ACTA DE SUSTENTACIÓN TRABAJO DE GRADO

TÍTULO: MODELAMIENTO Y PRONÓSTICOS USANDO MODELOS TAR EN ALGUNAS SERIES DE TIEMPO HIDROLÓGICAS Y METEOROLÓGICAS DEL TOLIMA

AUTORES: JUAN CAMILO GOMEZ CEBALLOS (código 070250142010)

DIRECTOR: JOAQUIN GONZÁLEZ BORJA

JURADOS: YURI MARCELA GARCIA SAAVEDRA
JAIRO ALFONSO CLAVIJO

CALIFICACIÓN: 4.9 (MERITO)

APROBÓ

REPROBÓ

OBSERVACIONES:

FIRMAS:

YURI MARCELA GARCÍA S.
Jurado 1

JAIRO ALFONSO CLAVIJO
Jurado 2

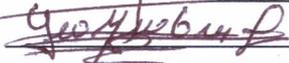
JOAQUIN GONZÁLEZ BORJA
Director del Trabajo

LEONARDO DUVAN RESTREPO
Director de programa (Ad Hoc)

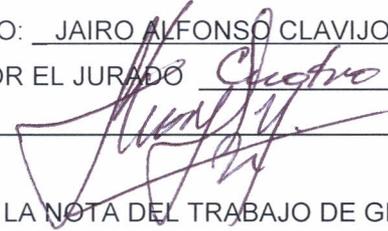
Ciudad y fecha: Ibagué, 30 de noviembre de 2018

6. Calificación

PRIMER JURADO:

NOMBRE DEL JURADO: YURI MARCELA GARCIA SAAVEDRANOTA OTORGADA POR EL JURADO 4.9FIRMA DEL JURADO 

SEGUNDO JURADO:

NOMBRE DEL JURADO: JAIRO ALFONSO CLAVIJO MENDEZNOTA OTORGADA POR EL JURADO Cuatro ocho (4.8)FIRMA DEL JURADO PROMEDIO FINAL DE LA NOTA DEL TRABAJO DE GRADO: 4.9 (CUATRO NUEVE)

7. RANGOS DE EQUIVALENCIA: (Acuerdo No. 030 de 2000 del Consejo de Facultad)

Calificación menor de tres cero (3.0)

REPROBADO

Calificación entre tres cero (3.0) y tres nueve (3.9)

APROBADO

Calificación entre cuatro cero (4.0) y cuatro cuatro (4.4.)

SOBRESALIENTE

Calificación entre cuatro cinco (4.5) y cuatro nueve (4.9)

MERITORIO

Calificación de cinco cero (5.0)

LAUREADO

FECHA DE SUSTENTACIÓN

30 - NOVIEMBRE - 2018

Título en español

Modelamiento y pronósticos usando modelos TAR en algunas series de tiempo hidrológicas y meteorológicas del Tolima.

Resumen: En este trabajo de grado se presenta el modelamiento y pronósticos de una serie de tiempo hidrológica y meteorológica a través de un modelo TAR con dos regímenes, en donde tenemos en cuenta épocas de mucha lluvia y épocas de poca lluvia. Se toman como variable de interés es el caudal del río Coello medido en una estación hidrológica del IDEAM en la vereda Coello Cocora de Ibagué, la precipitación se toma como variable de umbrales y es medida en el municipio de Cajamarca. Además la estimación de parámetros se calcula a través de métodos Bayesianos y técnicas MCMC. Se realiza un ejemplo simulado y otro con los datos empíricos mencionados anteriormente; dando resultados positivos en la estimación de parámetros y el cálculo de pronósticos, este último presenta desviaciones estándar grandes e intervalos de credibilidad anchos que pueden ser producto de trabajos futuros.

Palabras clave: Modelos TAR, técnicas MCMC, metodología Bayesiana, estimación de parámetros, cálculo de pronósticos.

Dedicado a

 Mi familia, a pesar de las adversidades que tuvimos en el camino siempre me apoyaron. A mi esposa y a mi hijo que fueron el motor para que hoy esté ante ustedes presentando mi trabajo.

Agradecimientos

El autor expresa sus agradecimientos a la Universidad del Tolima, Facultad de Ciencias, al programa de Matemáticas con Énfasis en Estadística y a todos aquellos que de una u otra forma han colaborado, contribuido o aportado en el desarrollo de este trabajo.

Índice general

Índice general	I
Índice de tablas	II
Índice de figuras	III
Objetivos	V
Introducción	VI
1. Consideraciones Teóricas	1
1.1. Especificación del modelo TAR	1
1.2. Función de verosimilitud condicional	2
1.3. Estimación Bayesiana de parámetros	2
1.4. Algoritmo MCMC para la estimación de parámetros	4
1.5. Validación del modelo TAR	5
1.6. Cálculo de Pronósticos	5
1.7. Metodología propuesta	6
2. Un Ejemplo Simulado	8
3. Una aplicación en la hidrología y meteorología	14
3.1. Análisis descriptivo de las series de tiempo hidrológicas y meteorológicas bajo estudio.	14
3.2. Proceso de un modelo TAR	20
4. Conclusiones	25
5. Bibliografía	26

Índice de tablas

2.1. Estimación de parámetros del modelo TAR simulado.	9
2.2. Pronósticos Variable X_t del modelo TAR simulado.	13
2.3. Pronósticos Variable Z_t del modelo TAR simulado.	13
3.1. Resumen estadístico de las series bajo estudio.	17
3.2. Prueba de no linealidad de umbrales sin covariable.	19
3.3. Parámetros estimados del modelo <i>TAR</i>	20
3.4. Pronósticos Variable X_t	23
3.5. Pronósticos Variable Z_t	24

Índice de figuras

2.1.	(a) Variable de interés $\{X_t\}$, (b) variable de umbrales $\{Z_t\}$	8
2.2.	Muestras de las cadenas MCMC de algunos parámetros estimados del modelo TAR simulado.	10
2.3.	Cumuplot para algunas cadenas MCMC del modelo TAR simulado.	11
2.4.	(a) Gráfica de la serie de residuales estandarizados, (b) histograma de los residuales estandarizados y (c) box plot de los residuales estandarizados, para el modelo TAR simulado.	11
2.5.	(a) función de autocorrelación muestral residuales estandarizados, (b) función de autocorrelación muestral de los residuales estandarizados al cuadrado, (c) función de autocorrelación parcial muestral residuales estandarizados y (d) función de autocorrelación parcial muestral de los residuales estandarizados al cuadrado.	12
2.6.	Gráfica CUSUM de los residuales del modelo TAR simulado.	12
3.1.	Estación hidrológica Puente Carretera del IDEAM	15
3.2.	(a) Caudal del Río Coello, (b) Precipitación de Cajamarca.	16
3.3.	(a) función de autocorrelación muestral caudal del río, (b) función de autocorrelación muestral precipitación,(c) función de autocorrelación parcial muestral caudal del río y (d) función de autocorrelación parcial muestral precipitación.	16
3.4.	(a) Caudal del río con ajuste estacional y (b) precipitación con ajuste estacional.	17
3.5.	(a) función de autocorrelación caudal del río con ajuste estacional, (b) función de autocorrelación precipitación con ajuste estacional,(c) función de autocorrelación parcial caudal del río con ajuste estacional y (d) función de autocorrelación parcial precipitación con ajuste estacional.	18
3.6.	Identificación de un modelo $ARMA(p, q)$ en la variable de umbrales.	18
3.7.	Identificación de un modelo $ARMA(p, q)$ en la variable de interés.	19
3.8.	Muestras de las cadenas MCMC de los parámetros estimados del modelo TAR.	21

3.9. Convergencia de las cadenas del modelo TAR.	21
3.10. (a) Gráfica de la serie de residuales estandarizados, (b) histograma de los residuales estandarizados y (c) box plot de los residuales estandarizados, para el modelo TAR.	22
3.11. (a) función de autocorrelación muestral residuales, (b) función de autocorrelación parcial muestral residuales, (c) función de autocorrelación muestral residuales al cuadrado y (d) función de autocorrelación parcial muestral residuales al cuadrado.	22
3.12. Gráfico CUSUM de los residuales estandarizados del modelo TAR.	23

Objetivos

Objetivo General

Modelar y pronosticar con un modelo TAR algunas series de tiempo hidrológicas y meteorológicas en la región de Cajamarca - Tolima.

Objetivos Específicos

- Utilizar una metodología Bayesiana y técnicas MCMC que permitan modelar y pronosticar con modelos TAR.
- Dar un ejemplo simulado, que ilustre el modelamiento y pronósticos con un modelo TAR.
- Aplicar la metodología propuesta con datos de series de tiempo hidrológicas y meteorológicas en la región de Cajamarca - Tolima.

Introducción

El uso sostenible del agua debe enmarcarse en una adecuada planificación, la cual requiere un profundo conocimiento del preciado líquido. Bajo esta perspectiva los pronóstico en series temporales son herramientas estadísticas importantes que ayudan en la toma de decisiones, es así como los modelos no lineales de tipo TAR (por sus siglas en inglés *Threshold Autoregressive*) implementados por Tong (1978, 1983) y Tong y Lim (1980) nos permiten describir series cronológicas y periódicas. Estos modelos tienen la capacidad de capturar el comportamiento dinámico de una serie de tiempo a través de sus regímenes, asimetría, bimodalidad, fenómenos de saltos y conglomerados de volatilidad que los modelos lineales no capturan, además, desempeñan un papel importante en la modelización de series de tiempo no lineales. Se asumen los valores de un proceso denominado umbrales $\{Z_t\}$, que determina la dinámica de un proceso estocástico $\{X_t\}$ (proceso de interés).

Tsay (1998) utilizó residuos predictivos para construir una estadística de prueba para detectar la no linealidad de umbrales en una serie temporal vectorial y proponer un procedimiento para construir un modelo de umbrales multivariado. Chen y Lee (1995) proponen una metodología Bayesiana y técnicas MCMC (Markov Chain Monte Carlo) para estimar los parámetros desconocidos en forma simultánea de un modelo TAR, Nieto (2005) desarrolló un procedimiento para procesos bivariados con datos faltantes, en modelos no lineales tipo TAR donde muestra que estos modelos son apropiados para analizar algunas series de tiempo hidrológicas y meteorológicas, Nieto (2008) desarrolla una fase de cálculo de pronósticos con modelos TAR con dos regímenes en variables hidrológicas y meteorológicas. Nieto, Zhang y Li (2013) desarrollan un nuevo procedimiento para identificar un modelo TAR y estimar sus parámetros siguiendo un procedimiento MCMC de salto reversible. Calderón y Nieto (2017) proponen una metodología para el cálculo de pronósticos en modelos TAR univariados.

El interés de este trabajo se centra en el modelamiento y pronósticos de un modelo TAR con $\{X_t\}$ como la variable de interés que será el caudal del río Coello medido en una estación hidrológica en la vereda Coello Cocora del municipio de Ibagué y $\{Z_t\}$ la variable de umbrales que es la precipitación del municipio de Cajamarca, Utiizando la metodología propuesta por Chen y Lee (1995) para la estimación de parámetros y la metodología de Calderón (2014) para el cálculo de pronósticos usando datos hidrológicos y meteorológicos. El trabajo se organiza de la siguiente manera:

En el capítulo 1 encontraremos las consideraciones teóricas en donde se encuentra una especificación del modelo TAR, la función de verosimilitud condicional, la estimación Bayesiana de parámetros, la validación del modelo y el cálculo de pronósticos, en el capítulo 2 se desarrolla la metodología propuesta para un modelo TAR de 2 regímenes con datos simulados y en el capítulo 3 se desarrolla la metodología propuesta para un modelo TAR de 2 regímenes con datos empíricos.

CAPÍTULO 1

Consideraciones Teóricas

1.1. Especificación del modelo TAR

Sean $\{X_t\}$ la variable de interés y $\{Z_t\}$ la variable de umbrales, procesos estocásticos relacionados por la ecuación modelo TAR, por sus siglas en inglés (*Threshold Autoregressive*)

$$X_t = a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)} X_{t-i} + h^{(j)} \varepsilon_t, \quad (1.1)$$

si Z_{t-d} pertenece al intervalo real $B_j = (r_{j-1}, r_j]$ para algún $j = 1, \dots, l$, donde $r_0 = -\infty$, $r_l = \infty$ y $d \geq 0$ entero positivo, siendo d el parámetro de rezago. Los números reales $a_i^{(j)}$ y los números enteros k_j denotan, respectivamente, los coeficientes y los órdenes autorregresivos de $\{X_t\}$ en el régimen $j = 1, \dots, l$. Los números reales r_j ; $j = 1, \dots, (l-1)$; son denominados valores umbrales del proceso $\{Z_t\}$ y son ellos quienes definen los l regímenes para el modelo. Los $h^{(j)}$ con $j = 1, \dots, l$, son las ponderaciones de las varianzas.

Además, $\{\varepsilon_t\}$ es un proceso de ruido blanco Gaussiano con media cero y varianza uno independiente de $\{Z_t\}$. Se usa el símbolo $TAR(l; k_1, \dots, k_l)$ para denotar un modelo TAR con l regímenes y k_1, \dots, k_l los órdenes autorregresivos del modelo en los respectivos regímenes. Los parámetros $l; r_1, \dots, r_{l-1}$ y k_1, \dots, k_l son denominados parámetros estructurales del modelo TAR y el resto de parámetros se llaman no estructurales.

Por otro lado, el proceso $\{Z_t\}$ es exógeno a $\{X_t\}$ con las propiedades de una cadena de Markov homogénea de orden p , con distribución estacionaria $f_p(\cdot)$ y kernel de transición $f(\cdot, \cdot)$.

El modelo TAR descrito en esta sección, es el desarrollado en Nieto (2005).

1.2. Función de verosimilitud condicional

Consideremos la siguiente función de verosimilitud condicional para los parámetros estructurales del modelo TAR

$$f(\mathbf{y}|\theta_x, \theta_z) = f(\mathbf{x}|\mathbf{z}, \theta_x, \theta_z)f(\mathbf{z}|\theta_x, \theta_z)$$

con $\mathbf{y} = (\mathbf{x}, \mathbf{z})$ con $\mathbf{x} = (x_1, \dots, x_T)$ y $\mathbf{z} = (z_1, \dots, z_T)$ los vectores de datos observados para el proceso $\{X_t\}$ y $\{Z_t\}$ respectivamente, en un periodo de $t = 1$ hasta $t = T$ y θ_x es el vector de los parámetros no estructurales. Ahora, podemos inicialmente estimar los parámetros de $\{Z_t\}$ para luego condicional sobre estos, se procede a estimar los parámetros del modelo TAR.

Principalmente nos enfocamos en calcular la función de verosimilitud condicional para el modelo TAR, además asumimos los primeros K valores de \mathbf{x} , como $\mathbf{x}_K = (x_1, \dots, x_K)$ con $K = \max\{k_1, k_2, \dots, k_l\}$. Si el proceso es un ruido blanco Gaussiano $\{\varepsilon_t\}$ entonces:

$$f(\mathbf{x}|\mathbf{z}, \theta_x, \theta_z) = (2\pi)^{-(T-K)/2} \left[\prod_{t=K+1}^T \{h^{j_t}\}^{-1} \right] \exp\left(-\frac{1}{2} \sum_{t=K+1}^T e_t^2\right), \quad (1.2)$$

donde

$$e_t = \frac{x_t - a_0^{(j_t)} - \sum_{j=1}^{p_{j_t}} a_j^{(j_t)} X_{t-j}}{h^{(j_t)}}$$

con $\{j_t\}$ la serie de tiempo observada en el proceso estocástico $\{J_t\}$ que es la sucesión de variables indicadoras definidas como $J_t = j$ si $Z_{t-d} \in (r_{j-1}, r_j]$ para algún $j = 1, \dots, l$. Asumimos que no hay relación entre los parámetros θ_x y θ_z y que la función de verosimilitud para x no depende de θ_z .

$\Theta_j = (a_0^{(j)}, a_1^{(j)}, \dots, a_p^{(j)})'$, $X_j^* = (X_{t_1, j}, \dots, X_{t_n, j})'$, $Z_j^* = (Z_{t_1, j}, \dots, Z_{t_n, j})'$ con $\{t_{1, j}, \dots, t_{n, j}\}$ como los índices en el tiempo.

1.3. Estimación Bayesiana de parámetros

Para adaptar la metodología propuesta por Chen y Lee (1995) es necesario encontrar las distribuciones a posteriori del vector θ_x . Para este fin, elegimos las a priori de la siguiente manera.

El vector Θ_j se distribuye como una normal tal que, $\Theta_j \sim \mathbf{N}(\theta_j, \mathbf{V}_j^{-1})$, para un $(\mathbf{h}^{(j)})^2 \sim \mathbf{IG}(\nu_j/2, \nu_j \lambda_j/2)$ donde "IG" denota la distribución gamma inversa y los hiperparámetros se asumen conocidos. Además, para el vector $\mathbf{r} = (r_1, \dots, r_{l-1})'$ decimos que su densidad a priori es constante $f(\mathbf{r}) = \frac{1}{C} \mathbf{I}(\mathbf{A})$ donde $C = \int \dots \int_{\mathbf{A}} dr_1 \dots dr_{l-1}$, \mathbf{A} es una

región que satisfice $a \leq r_1 < \dots < r_{l-1} \leq b$, con a y b cuantiles convenientes de Z_{t-d} y d sigue una distribución uniforme discreta $\mathbf{d} \sim \mathbf{U}_{\mathbf{d}}(0, 1, 2, \dots, d_0)$ siendo d_0 el rezago máximo considerado, Z_j^* son observaciones generadas en cada régimen, X_j^* son las observaciones ordenadas por los regímenes dados por Z_j^* y $\wp = \{X_t, Z_{j,t}, j = 1, \dots, K; t = 1, \dots, n\}$.

Nuestro interés radica en las distribuciones marginales a posteriori $\Theta_j, (h^{(j)})^2, r, d$. Usando técnicas Bayesianas estándar, obtenemos las siguientes distribuciones a posteriori condicionales:

1. La distribución condicional a posteriori para Θ_j está dada por:

$$p(\Theta_j | (h^{(j)})^2, r, d, \wp) \sim \mathbf{N}(\Theta_j^*, V_j^{-1}) \quad (1.3)$$

donde

$$\Theta_j^* = \left(\frac{X_j^{*'} X_j^*}{(h^{(j)})^2} + V_j \right)^{-1} \left(\frac{X_j^{*'} X_j^*}{(h^{(j)})^2} \hat{\Theta}_j + V_j \Theta_{0j} \right),$$

$$V_j^* = \left(\frac{X_j^{*'} X_j^*}{(h^{(j)})^2} + V_j \right)$$

$$\text{con } \hat{\Theta}_j = (X_j^{*'} X_j^*)^{-1} X_j^{*'} Z_j^*.$$

2. La distribución condicional a posteriori de $(h^{(j)})^2$ está dada por:

$$p((h^{(j)})^2 | \Theta_j, r, d, \wp) \sim \mathbf{IG} \left(\frac{v_j + n_j}{2}, \frac{v_j \lambda_j + n_j s_j^2}{2} \right) \quad (1.4)$$

$$\text{es decir, } \frac{v_j \lambda_j + n_j s_j^2}{(h^{(j)})^2} \sim \chi_{v_j + n_j}^2, \quad j = 1, \dots, l,$$

donde n_j es el número de observaciones de cada régimen y

$$n_j s_j^2 = (Z_j^* - X_j^* \Theta_j)' (Z_j^* - X_j^* \Theta_j).$$

3. La función de probabilidad condicional a posteriori de r está dada por:

$$p(r | \Theta_j, (h^{(j)})^2, d, \wp) \propto \exp \left\{ \frac{1}{2} \sum_{j=1}^l \frac{(Z_j^* - X_j^* \Theta_j)' (Z_j^* - X_j^* \Theta_j)}{(h^{(j)})^2} \right\} \quad (1.5)$$

la anterior función no posee una distribución estándar conocida.

4. La función de probabilidad condicional a posteriori para d tiene una distribución multinomial con probabilidad:

$$p(d | \Theta_j, (h^{(j)})^2, r, \wp) = \frac{\mathbf{L}(\Theta_j, (h^{(j)})^2, r, d | \wp)}{\sum_{d=0}^{d_0} \mathbf{L}(\Theta_j, (h^{(j)})^2, r, d | \wp)}, \quad (1.6)$$

donde $d = 0, 1, 2, \dots, d_0$ y

$$L(\Theta_j, (h^{(j)})^2, r, d|\varphi) = \frac{\exp \left\{ - \sum_{j=1}^l \frac{1}{2(h^{(j)})^2} (Z_j^* - X_j^* \Theta_j)' (Z_j^* - X_j^* \Theta_j) \right\}}{(h^{(j)})^{n_j}}$$

observe que n_j son funciones de r .

1.4. Algoritmo MCMC para la estimación de parámetros

Como lo mencionamos anteriormente la función de probabilidad condicional de r no es conocida, para solucionar este problema utilizaremos técnicas MCMC con un tamaño de muestra lo suficientemente grande, descartando un periodo de calentamiento inicial y las iteraciones restantes se usan para el análisis. Se usa el siguiente algoritmo para construir las muestras a posteriori:

Paso 1. Se genera Θ_j , dada en 1.3

Paso 2. Se genera $(h^{(j)})^2$, dada en 1.4

Paso 3. Se genera r empleando el algoritmo de Metrópolis Hastings de caminata aleatoria donde $f(\cdot)$ está dada en 1.5

Paso 4. Se genera d , dada en 1.6

Podemos resumir el algoritmo Metrópolis Hastings usado en el paso (iii), de la siguiente forma:

i) Definir un valor inicial $r^{(0)}$, el indicador de cada etapa $j = 0$

ii) Se genera $r^* = r^{(j)} + N(0, \sigma^2)$

iii) Definir $\alpha = \min \left\{ 1, \frac{f(r^*)}{f(r^{(j)})} \right\}$, como la probabilidad de aceptación

iv) Tomar:

$$r^{(j+1)} = \begin{cases} r^*, & \text{con probabilidad } \alpha \\ r^{(j)}, & \text{con probabilidad } 1 - \alpha \end{cases}$$

v) Vuelva a ii) aumentando el indicador de etapa hasta que el proceso llegue a una distribución estacionaria.

En la práctica el tamaño de paso del algoritmo Metrópolis Hastings se puede adecuar para producir una rápida convergencia.

1.5. Validación del modelo TAR

Para poder validar el modelo *TAR* usaremos los residuales estandarizados

$$\hat{\epsilon}_t = \frac{X_t - X_{t|t-1}}{h^{(j)}}$$

si $r_{j-1} < Z_{t-d} \leq r_j$ para $j = 1, \dots, l$ donde

$$X_{t|t-1} = a_0^{(j)} + \sum_{i=1}^{p_j} a_i^{(j)} X_{t-i} \quad (1.7)$$

es el predictor un paso adelante para X_t , $t = (K + 1), \dots, T$ con $K = \max\{k_1, \dots, k_l\}$ y T es la longitud del periodo muestral de $\{X_t\}$. $\{\epsilon_t\}$ debe tener un comportamiento muy parecido al de un ruido blanco Gaussiano. Esto lo podemos ver en los gráficos CUSUM para verificar que hay evidencia de una especificación correcta del modelo. Además, es necesario garantizar a través de pruebas los supuestos de normalidad, homocedasticidad e independencia.

1.6. Cálculo de Pronósticos

En esta sección trabajaremos la fase de pronóstico para modelos TAR. Adaptaremos la metodología propuesta por Calderón y Nieto (2017).

Para este fin, necesitamos encontrar $X_{T+h} = E(\hat{X}_{T+h} | \mathbf{x}, \mathbf{z})$, que es la mejor predicción en el sentido de MECM (mínimo error cuadrático medio) para un modelo con l regímenes y $h \geq 1$ con h como los pasos adelante a pronosticar. Por lo tanto, utilizando el análisis Bayesiano y la función de pérdida cuadrática como criterio de optimalidad, procedemos hallar las distribuciones predictivas $p(x_{T+h} | x_{1:T}, z_{1:T})$ para $h \geq 1$ con el que podemos obtener las expectativas condicionales objetivo. Sin embargo, nos centramos en la distribución predictiva conjunta $p(x_{T+1:T+h}, z_{T+1:T+h} | x_{1:T}, z_{1:T})$ a partir de la cual se pueden obtener las distribuciones marginales de interés, siendo θ_x el vector de parámetros de $\{X_t\}$ y θ_z el vector de parámetros de $\{Z_t\}$. La distribución predictiva conjunta se puede obtener de la siguiente manera:

$$p(x_{T+1:T+h}, z_{T+1:T+h} | x_{1:T}, z_{1:T}) = \int p(x_{T+1:T+h}, z_{T+1:T+h} | x_{1:T}, z_{1:T}, \theta_x) p(\theta_x | x_{1:T}, z_{1:T}) d\theta_x \quad (1.8)$$

Para especificar la distribución marginal, todas las suposiciones sobre el modelo TAR son aceptadas.

$$p(x_{T+1}, \dots, x_{T+h}, z_{T+1}, \dots, z_{T+h} | x_{1:T}, z_{1:T}, \theta_x) = \prod_{i=1}^h p(z_{T+i} | z_{1:T+i-1}) p(x_{T+i} | z_{T+i}, x_{1:T+i-1}, z_{1:T+i-1}, \theta_x) \quad (1.9)$$

- i) $p(\theta_x | x_{1:T}, z_{1:T})$ es la densidad a posteriori de los parámetros del modelo TAR.
- ii) $p(z_{T+i} | z_{1:T+i-1})$ es la densidad del kernel de transición en la cadena de Markov homogénea $\{Z_t\}$.
- iii) $p(x_{T+i} | z_{T+i}, x_{1:T+i-1}, z_{1:T+i-1}, \theta_x)$ tiene una distribución normal con media $a_0^{(j)} + \sum_{m=1}^{p_j} a_m^{(j)} X_{T+i-m}$ y varianza $(h^{(j)})^2$ si $r_{j-1} < z_{T+i-1} \leq r_j$.

Además $x_{1:T+i-1} = (x_1, \dots, x_{T+i-1})$ y $z_{1:T+i-1} = (z_1, \dots, z_{T+i-1})$. Así, mediante técnicas de simulación, se pueden generar muestras de $z_{T+1}, x_{T+1}, z_{T+2}, x_{T+2}, \dots, z_{T+h}, x_{T+h}$ en forma secuencial de la siguiente manera:

Algoritmo:

- Paso 1. Se genera una muestra MCMC de $\theta_x^{(k)}$ de las densidades a posteriori de $p(\theta_x | x_{1:T}, z_{1:T})$ aplicando el esquema muestral de la sección 1.4.
- Paso 2. Se genera $z_{T+1}^{(k)}$ con una densidad kernel $p(z_{T+1} | z_{1:T})$
- Paso 3. Se genera $x_{T+1}^{(k)}$ con una densidad $p(x_{T+1} | z_{T+1}^{(k)}, x_{1:T}, z_{1:T}, \theta_x^{(k)})$
- Paso 4. Se genera $z_{T+i}^{(k)}$ con una densidad kernel $p(z_{T+i} | z_{T+i-1}^{(k)}, z_{1:T})$
- Paso 5. Se genera $x_{T+i}^{(k)}$ con una densidad $p(x_{T+i} | z_{T+i}^{(k)}, x_{1:T}, z_{T+i-1}^{(k)}, x_{1:T}, z_{1:T}, \theta_x^{(k)})$
- Paso 6. Repetir pasos 4 y 5 para $i = 2, 3, \dots, h$

Con el conjunto de muestras $\{z_{T+h}^{(k)}, x_{T+h}^{(k)}\}$, es posible calcular: 1) la media de la distribución predictiva, 2) la matriz de covarianza de la distribución predictiva y 3) los intervalos creíbles para el pronóstico del punto en la variable umbral como en la variable de interés.

1.7. Metodología propuesta

Se propone la siguiente metodología para trabajar con series de tiempo hidrológicas y meteorológicas.

- Paso 1. Identificar el modelo TAR apropiado.
- Paso 2. Se asumen conocidos o estimados en forma preliminar el número de regímenes y los órdenes autorregresivos del modelo.

-
- Paso 3. Se adapta la metodología Bayesiana y técnicas MCMC propuestas para estimar los parámetros desconocidos de un modelo TAR.
- Paso 4. Se adapta la metodología propuesta para el cálculo de pronósticos en modelos TAR univariados.

Un Ejemplo Simulado

La metodología propuesta es ilustrada con un ejemplo simulado. Vamos a considerar el modelo $TAR(2; 3, 5)$ relacionados en la ecuación

$$X_t = \begin{cases} 1.32 + 0.20X_{t-1} + 0.50X_{t-2} - 0.10X_{t-3} + 2\varepsilon_t & \text{si } Z_{t-1} < 4.7 \\ 1.92 + 0.20X_{t-1} + 0.30X_{t-2} + 0.50X_{t-3} - 0.10X_{t-4} - 0.15X_{t-5} + 3\varepsilon_t & \text{si } Z_{t-1} \geq 4.7 \end{cases} \quad (2.1)$$

donde $Z_t = 1.8 + 0.6Z_{t-1} + \varepsilon_t$ con $\varepsilon_t \sim RBG(0, 1)$. La simulación se realizó con un periodo muestral de 600 observaciones.

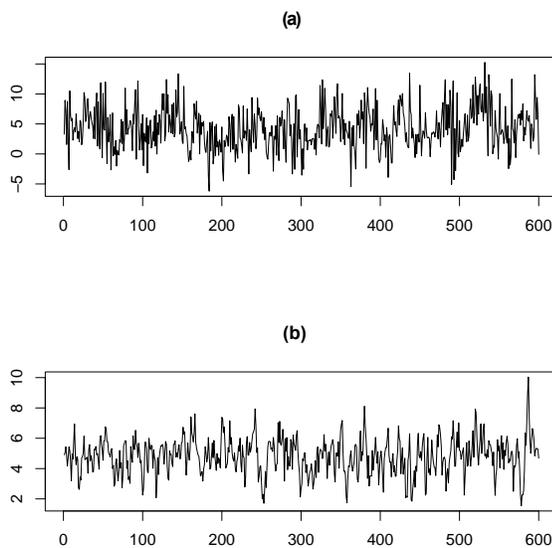


FIGURA 2.1. (a) Variable de interés $\{X_t\}$, (b) variable de umbrales $\{Z_t\}$.

Podemos ver en la figura 2.1 que ambas series tienen un comportamiento estable a través del tiempo. La variable de interés está construida bajo un modelo TAR con dos regímenes y la variable de umbrales se construyó con un modelo $AR(1)$.

Se estiman de forma preliminar los valores de $l = 2, p_1 = 3, p_2 = 5$ y se estiman el resto de los parámetros del modelo TAR usando el algoritmo descrito en la sección 1.4

Las condiciones iniciales para el vector $\Theta_j = 0.05$, para $(h^{(j)})^2 = 0.20$, para $d = 3$ y para r la media de la variable de umbrales rezagada d períodos, d una uniforme discreta de parámetros $\{0, 1, 2, \dots, d_0\}$ con $d_0 = 5$ como el rezago máximo y se fija un escalar positivo $a^2 = 0.005$ que requiere el algoritmo Metropolis-Hastings de caminata aleatoria. Se realizan 100 repeticiones, 12.000 iteraciones, eliminando las primeras 6.000 como periodo de calentamiento. Por último garantizamos por lo menos un 10 % de las observaciones en cada régimen, con el propósito de obtener buenas estimaciones sobre el valor de r .

TABLA 2.1. Estimación de parámetros del modelo TAR simulado.

Parámetro	Valor parámetro	Estimación parámetro	d.e.	I.C. 95 %	Cobertura	Sesgo
$a_0^{(1)}$	1.32	1.34	0.23	(0.89, 1.78)	92	0.02
$a_1^{(1)}$	0.20	0.20	0.04	(0.12, 0.29)	92	0.00
$a_2^{(1)}$	0.50	0.49	0.04	(0.41, 0.57)	95	-0.01
$a_3^{(1)}$	-0.10	-0.10	0.04	(-0.18, -0.02)	95	0.00
$a_0^{(2)}$	1.92	1.91	0.41	(1.11, 2.71)	95	-0.01
$a_1^{(2)}$	0.20	0.20	0.07	(0.07, 0.32)	94	0.00
$a_2^{(2)}$	0.30	0.30	0.07	(0.17, 0.43)	92	0.00
$a_3^{(2)}$	0.50	0.50	0.07	(0.35, 0.64)	95	0.00
$a_4^{(2)}$	-0.10	-0.10	0.07	(-0.24, 0.04)	94	0.00
$a_5^{(2)}$	-0.15	-0.15	0.07	(-0.28, -0.01)	96	0.00
$(h^{(1)})^2$	4	4.03	0.32	(3.45, 4.70)	98	0.03
$(h^{(2)})^2$	9	9.11	0.81	(7.66, 10.84)	98	0.11
r	4.7	4.69	0.03	(4.64, 4.73)	86	-0.01

En la tabla 2.1 encontramos el valor verdadero de los parámetros, la estimación de cada parámetro, la desviación estándar (d.e), el intervalo de confianza al 95 % (I.C. 95 %), la cobertura que es el porcentaje de veces que el valor verdadero cae en el intervalo de confianza y el sesgo que es la diferencia que hay entre el valor verdadero y el estimado de cada parámetro. Podemos observar que el sesgo y la desviación estándar son muy cercana a cero, lo que es un buen indicador. Además, todos los intervalos de confianza al 95 % contiene los parámetros verdaderos.

Necesitamos garantizar que las cadenas MCMC de la muestras a posteriori convergen correctamente, revisaremos las gráficas de las series de la cadena MCMC y realizaremos algunos gráficos de autocorrelación muestral de las últimas 6.000 iteraciones de algunos

parámetros.

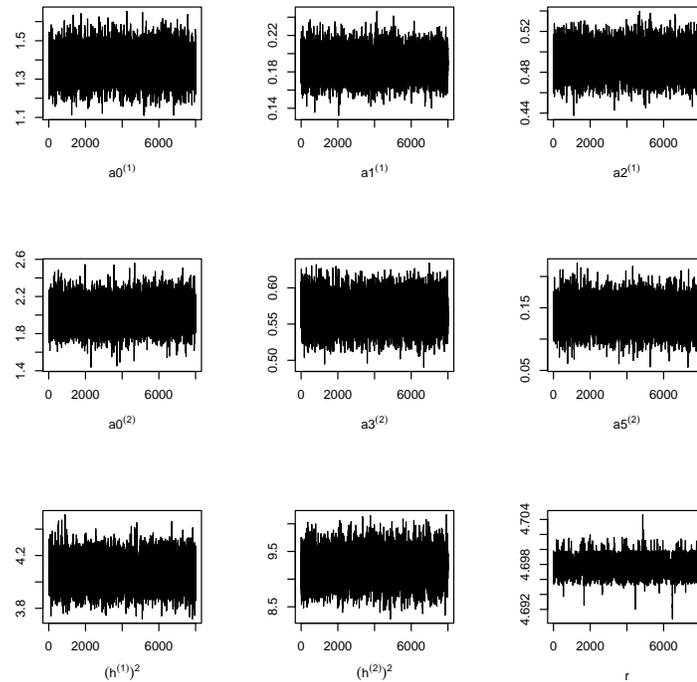


FIGURA 2.2. Muestras de las cadenas MCMC de algunos parámetros estimados del modelo TAR simulado.

Podemos observar en la figura 2.2 que las cadenas MCMC graficadas presentan un comportamiento estacionario, pero el parámetro r presenta un comportamiento diferente pero natural dado a que su función a posteriori no posee una distribución estándar conocida.

La gráfica *cumuplot*, de la librería *coda* del programa estadístico R, realiza una traza de la evolución de los cuantiles de muestra, en función del número de iteraciones como se muestra en la figura 2.3 donde las cadenas MCMC convergen rápidamente, sin embargo, el *cumuplot* de r presenta un comportamiento diferente debido a su función de densidad. Con esto podemos asegurar que las cadenas MCMC tomadas son suficientes para obtener la convergencia.

Continuando, se valida que el modelo cumpla con los supuestos trabajados en la sección 1.5. Así construimos los gráficos de la figura 2.4 con los residuales estandarizados del modelo TAR simulado que presentan un comportamiento estable y una distribución simétrica.

Realizamos los gráficos de autocorrelación muestral y autocorrelación parcial muestral de los residuales estandarizados y los residuales estandarizados al cuadrado en la figura 2.5, el cual nos indica no correlación serial en los residuales estandarizados y no presenta heterocedasticidad en los residuales estandarizados al cuadrado.

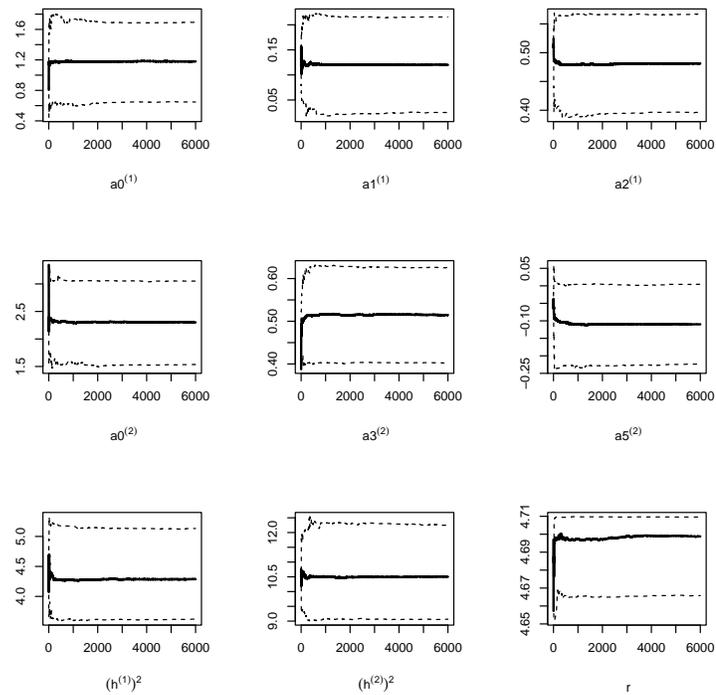


FIGURA 2.3. Cumulplot para algunas cadenas MCMC del modelo TAR simulado.

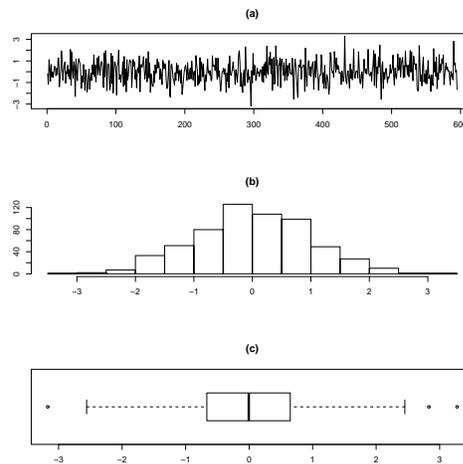


FIGURA 2.4. (a) Gráfica de la serie de residuales estandarizados, (b) histograma de los residuales estandarizados y (c) box plot de los residuales estandarizados, para el modelo TAR simulado.

En las pruebas estadísticas para probar normalidad y no autocorrelación, realizamos la prueba de Jarque y Bera con un valor p de 0.48 y la prueba de Ljung-Box un valor p de 0.93 indicando que los datos son normales y no están autocorrelacionados. También se realizó un gráfico CUSUM el cual nos evidencia que hay una especificación correcta del modelo.

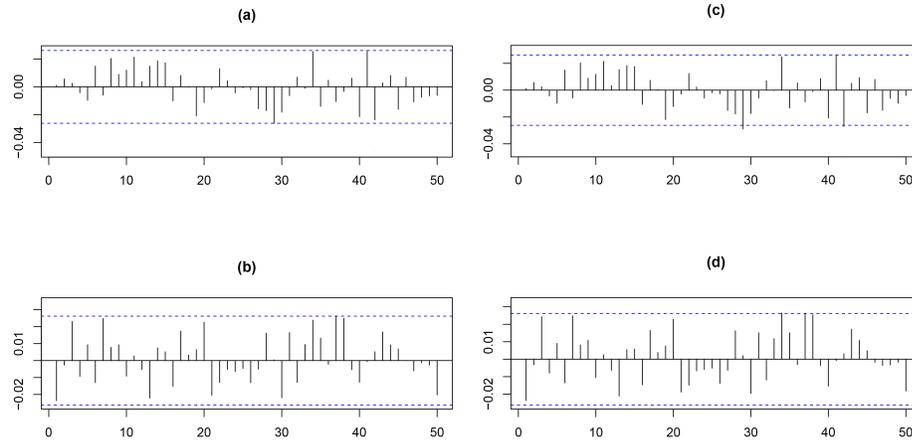


FIGURA 2.5. (a) función de autocorrelación muestral residuales estandarizados, (b) función de autocorrelación muestral de los residuales estandarizados al cuadrado, (c) función de autocorrelación parcial muestral residuales estandarizados y (d) función de autocorrelación parcial muestral de los residuales estandarizados al cuadrado.

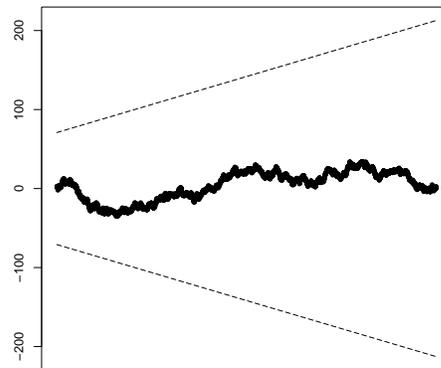


FIGURA 2.6. Gráfica CUSUM de los residuales del modelo TAR simulado.

Seguimos el estudio del modelo con la fase de pronósticos, para verificar el comportamiento de las predicciones se realizan 100 réplicas. También tuvimos en cuenta la incertidumbre de la estimación de parámetros para el cálculo de las densidades predictivas. Se fijan las últimas 12 observaciones de las series $\{X_t\}$ y $\{Z_t\}$ como horizonte de predicción, las cuales compararemos con predicciones obtenidas.

Podemos encontrar los pronósticos de las series de tiempo $\{X_t\}$ y $\{Z_t\}$ en las tablas 2.2 y 2.3 respectivamente, también encontramos (*d.e.*) que es la desviación estándar y los intervalos creíbles al 90% y 95%. Se observa que los datos pronosticados son muy cercanos a los datos observados en cada una de las tablas, esto evidencia que la metodología propuesta nos puede presentar unas buenas estimaciones, sin embargo, las *d.e.* son muy grandes y los intervalos creíbles son muy anchos.

TABLA 2.2. Pronósticos Variable X_t del modelo TAR simulado.

Horizonte (h)	X_{T+h}	\widehat{X}_{T+h}	d.e.	I.C. 90 %	I.C. 95 %
1	4.59	4.89	2.45	(0.84, 8.92)	(0.07, 9.68)
2	5.26	4.92	2.78	(0.60, 9.69)	(-0.22, 10.73)
3	4.97	4.82	2.98	(0.06, 9.83)	(-0.84, 10.88)
4	4.91	4.79	3.18	(-0.18, 10.26)	(-1.14, 11.49)
5	5.11	4.80	3.30	(-0.30, 10.53)	(-1.27, 11.84)
6	5.27	4.76	3.34	(-0.41, 10.54)	(-1.40, 11.85)
7	5.06	4.75	3.39	(-0.46, 10.62)	(-1.44, 11.99)
8	5.04	4.76	3.41	(-0.48, 10.71)	(-1.48, 12.11)
9	4.59	4.74	3.42	(-0.51, 10.68)	(-1.52, 12.10)
10	4.86	4.73	3.43	(-0.54, 10.69)	(-1.56, 12.11)
11	4.86	4.73	3.44	(-0.56, 10.71)	(-1.55, 12.14)
12	4.75	4.73	3.45	(-0.56, 10.73)	(-1.56, 12.13)

TABLA 2.3. Pronósticos Variable Z_t del modelo TAR simulado.

Horizonte (h)	Z_{T+h}	\widehat{Z}_{T+h}	d.e.	I.C. 90 %	I.C. 95 %
1	4.62	4.52	1.00	(2.88, 6.16)	(2.57, 6.48)
2	4.68	4.51	1.16	(2.60, 6.42)	(2.23, 6.78)
3	4.59	4.51	1.22	(2.51, 6.51)	(2.13, 6.89)
4	4.68	4.51	1.24	(2.48, 6.54)	(2.09, 6.93)
5	4.49	4.50	1.25	(2.45, 6.55)	(2.06, 6.94)
6	4.51	4.50	1.25	(2.44, 6.56)	(2.05, 6.95)
7	4.47	4.50	1.25	(2.45, 6.56)	(2.06, 6.95)
8	4.32	4.50	1.25	(2.44, 6.55)	(2.06, 6.94)
9	4.53	4.50	1.25	(2.45, 6.55)	(2.06, 6.95)
10	4.57	4.50	1.25	(2.45, 6.56)	(2.06, 6.95)
11	4.50	4.50	1.25	(2.45, 6.55)	(2.07, 6.95)
12	4.29	4.50	1.25	(2.45, 6.55)	(2.06, 6.95)

Una aplicación en la hidrología y meteorología

3.1. Análisis descriptivo de las series de tiempo hidrológicas y meteorológicas bajo estudio.

El desarrollo industrial, agrícola, pecuario y el surgimiento de grandes concentraciones humanas, pusieron de presente la necesidad de aprovechar de manera más eficiente el recurso agua. El incremento del uso del agua obliga a tener un conocimiento más profundo del comportamiento hidrológico de las cuencas hidrográficas, para poderla aprovechar de manera integral en los diferentes usos.

El Río Coello presenta vital importancia en el desarrollo del departamento del Tolima. Dentro de su cuenca se encuentran localizadas las fuentes abastecedoras de los cascos urbanos de los municipios de Cajamarca (Quebradas Chorros Blancos y Dos Quebradas), Ibagué (Río Combeima y quebrada Cay), El Espinal (Río Coello), Flandes y de otros centros habitacionales de menor densidad poblacional.

Igualmente, dentro del área de su Cuenca Mayor se ubica el Distrito de Riego del Río Coello, localizado en la zona central del Departamento del Tolima y abarca una extensión de 71900 hectáreas, de las cuales las aguas del proyecto irrigan 48000 hectáreas, que pertenecen a los municipios de El Espinal, Coello y Flandes, y además son utilizadas para generar energía eléctrica en la planta de La Ventana ,con capacidad instalada de 80 MWh.

El Río Coello nace en la Cordillera Central (Páramo de Don Simón), a una altura de 3850 m.s.n.m.. En su parte inicial recibe el nombre de Río Toche, al que tributa sus aguas el río Tohecito, recibiendo desde allí el nombre de Río Coello, más adelante los ríos Bermellón, Cocora, Andes, Gallego y Combeima le vierten sus aguas, constituyéndose en sus principales afluentes. La longitud del cauce principal es de aproximadamente 141 kilómetros, presentando una pendiente media del 3.20 % y desemboca en la margen izquierda en el sentido aguas abajo del Río Magdalena a una cota aproximada de 300 m.s.n.m.

Las series de tiempo han sido suministradas por el IDEAM que es una institución pública de apoyo técnico y científico al Sistema Nacional Ambiental, que genera conocimiento, produce información confiable, consistente y oportuna, sobre el estado y las dinámicas de los recursos naturales y del medio ambiente.



FIGURA 3.1. Estación hidrológica Puente Carretera del IDEAM

La variable de interés es el caudal del río Coello desde enero de 1969 hasta diciembre de 2007, medido en la estación hidrológica puente carretera ubicado en el municipio de Ibagué, vereda Coello Cocora, como se muestra en la figura 3.1. La variable de umbrales es la precipitación del municipio de Cajamarca medida desde enero de 1969 hasta diciembre de 2007. Ambas series de tiempo fueron recolectadas y suministradas por el IDEAM, tienen 456 observaciones, con una frecuencia mensual.

En la figura 3.2 encontramos que hay conglomerados de valores altos y bajos de volatilidad, parece haber una relación dinámica entre las dos variables y las series presentan un comportamiento estable en media.

En la figura 3.3 hay presencia de estacionalidad en ambas series de tiempo. Se realizan las pruebas de Dickey-Fuller y Osborn-Chui-Smith-Birchenhall (1988) para ver si hay raíces unitarias ordinarias y/o estacionales, respectivamente, dando como resultado la estadística de prueba 0 en ambas pruebas, lo cual, nos indica que las series no tienen ningún tipo de raíces unitarias ni estacionales, luego se ajustan estacionalmente las dos series para eliminar la estacionalidad presentada.

Para realizar el ajuste estacional utilizamos el software estadístico R y su paquete *seasonal* que trabaja con el algoritmo $X - 13ARIMA - SEATS$ creado por la US Census Bureau en alianza con el Banco de España. Su principal función es el ajuste estacional a través de componentes estacionales basados en modelos *ARIMA*, haciendo un procedimiento automatizado que permite producir rápidamente buenos ajustes

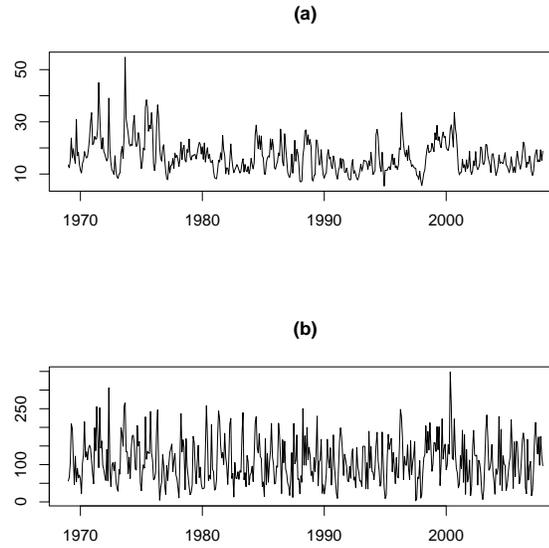


FIGURA 3.2. (a) Caudal del Río Coello, (b) Precipitación de Cajamarca.

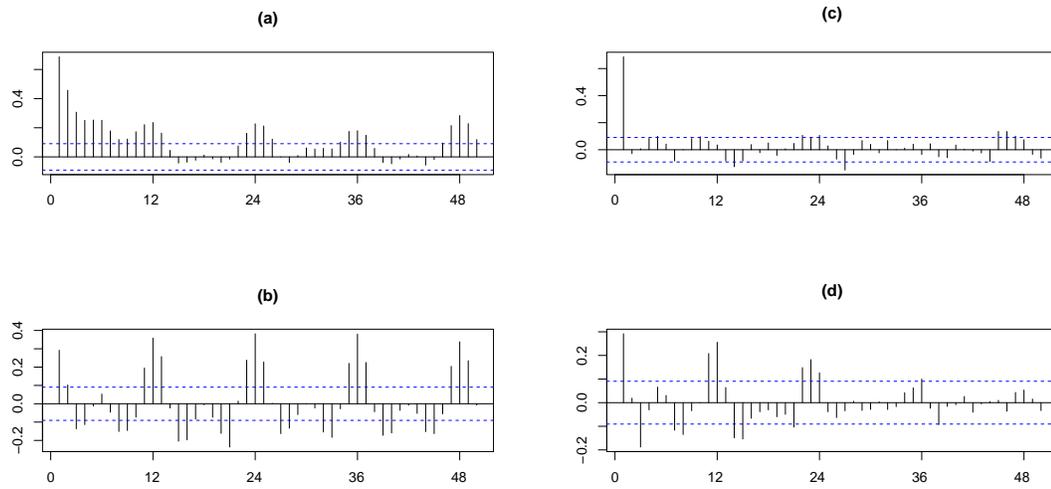


FIGURA 3.3. (a) función de autocorrelación muestral caudal del río, (b) función de autocorrelación muestral precipitación, (c) función de autocorrelación parcial muestral caudal del río y (d) función de autocorrelación parcial muestral precipitación.

estacionales de series de tiempo.

De aquí en adelante vamos a estudiar las series de tiempo obtenidas después del ajuste estacional, la variable de interés será el caudal del río Coello con ajuste estacional $\{X_t\}$ y la variable de umbrales será la precipitación de Cajamarca con ajuste estacional $\{Z_t\}$. Hasta predecir 12 pasos adelante de estas series.

TABLA 3.1. Resumen estadístico de las series bajo estudio.

	Media	Mediana	s.d.	Mínimo	Máximo	Sesgo	Exceso de Kurtosis	Prueba de normalidad
$\{X_t\}$	2.752	2.719	0.318	1.691	4.008	0.264	0.274	0.029
$\{Z_t\}$	10.032	10.184	2.297	2.692	17.031	-0.207	0.157	0.14

En la tabla 3.1 se muestran medidas de tendencia central como lo son la media y la mediana que ayudan a dar una idea de la distribución de los datos, en este caso observamos que la media y la mediana no difieren por una cantidad numérica grande lo cual nos lleva a pensar que la distribución de estas, para ambos casos, es aproximadamente simétrica. Por otro lado, obtenemos el coeficiente de exceso de kurtosis que nos muestra que tan aproximada en forma es nuestra distribución con respecto a una distribución normal; aquí apreciamos que en los dos casos tiene una forma leptocúrtica siendo para $\{X_t\}$ un poco más que para $\{Z_t\}$ este es un síntoma de simetría en cada serie, que se ve también en el sesgo debido a que sus valores son muy cercanos al 0. La desviación estándar de la precipitación $\{Z_t\}$ es mayor que la del caudal del río $\{X_t\}$; esto concuerda con el hecho de que la precipitación es más volátil que el caudal del río como se observa en la figura 3.4; estas características también se muestra en las estadísticas de pruebas de normalidad de Jarque-Bera.

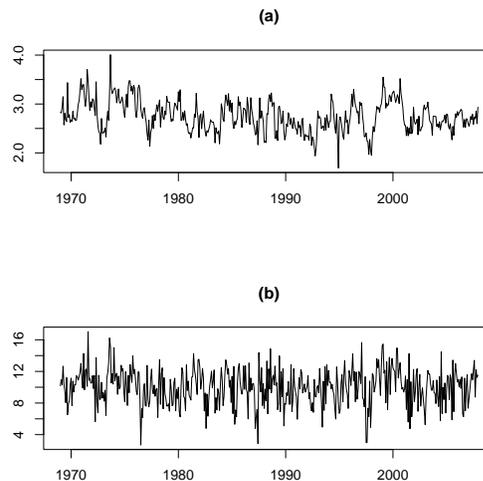


FIGURA 3.4. (a) Caudal del río con ajuste estacional y (b) precipitación con ajuste estacional.

En la figura 3.4 se observa como las series se estabilizan mucho más en media, además, en la figura 3.5 ya no hay presencia evidente de comportamiento estacional en las dos series de tiempo.

Usando el comando `armasubsets(.)` del paquete TSA de R ajustamos nuestra nueva variable de umbrales $\{Z_t\}$ a un proceso autorregresivo de media móvil $ARMA(p, q)$.

Se encontró en la figura 3.6 de un conjunto de modelos candidatos, considerando $p_{max} = 3$ y $q_{max} = 3$, un $ARMA(2, 0)$ bajo el mínimo Criterio de Información Bayesiano (BIC) y

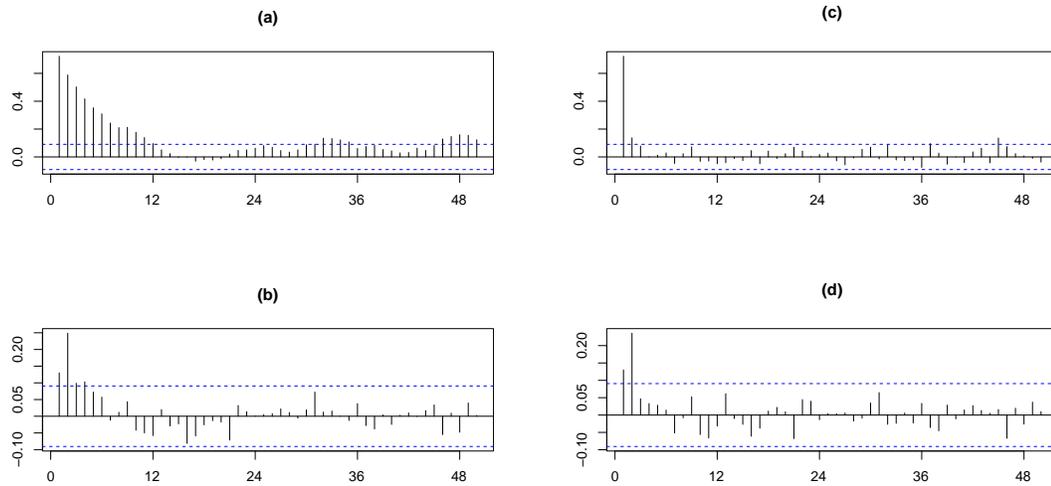


FIGURA 3.5. (a) función de autocorrelación caudal del río con ajuste estacional, (b) función de autocorrelación precipitación con ajuste estacional, (c) función de autocorrelación parcial caudal del río con ajuste estacional y (d) función de autocorrelación parcial precipitación con ajuste estacional.

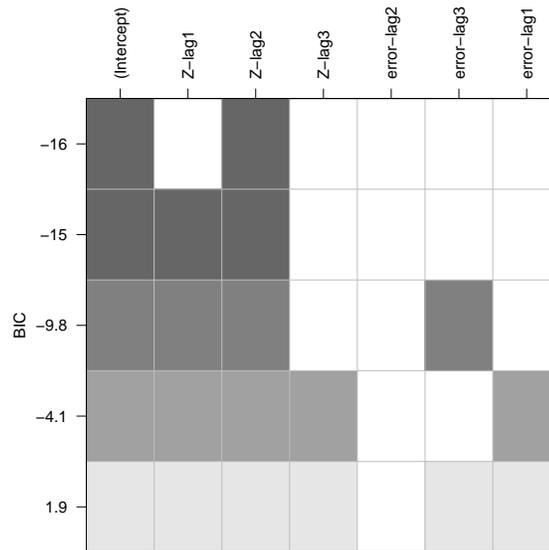


FIGURA 3.6. Identificación de un modelo $ARMA(p, q)$ en la variable de umbrales.

el mínimo Criterio de Información Akaike (AIC). Así, ajustamos la variable de umbrales a un modelo $AR(2)$, donde $Z_t = 10.0167 + 0.2479Z_{t-2}$ con $\sigma^2 = 4.985$, un mínimo $BIC = -16$ y un mínimo $AIC = 2030.7$.

Se realiza la prueba de no linealidad de umbrales entre la variable de interés y la variable de umbrales, para esto emplearemos la estadística de prueba propuesta por Tsay(1998), en la cual se contrasta una hipótesis nula de linealidad $AR(p)$ versus la alternativa de no

linealidad tipo *TAR*.

Esta prueba se realiza para diferentes rezagos de la variable de umbrales Z_{t-d} con $d = 0, 1, \dots, 7$. Para determinar los máximos órdenes autorregresivos de p , se ajusta a un modelo autorregresivo lineal $AR(p)$ la serie de tiempo $\{X_t\}$ usando nuevamente el comando *armasubsets(.)* de un conjunto de modelos candidatos, considerando $p_{max} = 3$ y $q_{max} = 3$, el mejor es un $ARMA(3,0)$ según el mínimo Criterio de Información Bayesiano ($BIC = -350$) y de acuerdo con el criterio de información de la Devianza ($DIC = -141.616$) los valores estimados para $p = 3$.

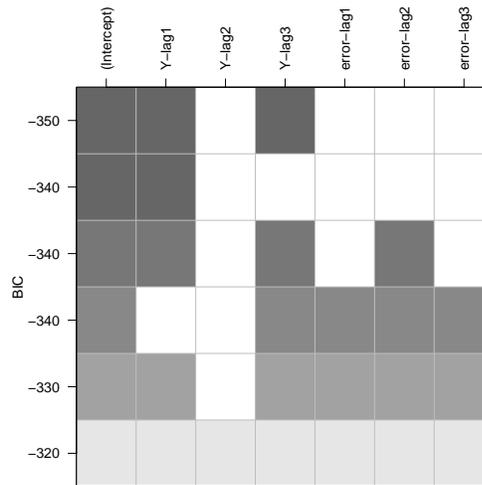


FIGURA 3.7. Identificación de un modelo $ARMA(p, q)$ en la variable de interés.

TABLA 3.2. Prueba de no linealidad de umbrales sin covariable.

Rezagos de la variable umbral								
$p = 3$								
<i>Test</i>	0	1	2	3	4	5	6	7
<i>F</i>	17.134	1.783	0.745	1.060	2.099	0.152	1.004	1.071
<i>P - valor</i>	0.000	0.131	0.562	0.376	0.080	0.962	0.405	0.371

La hipótesis nula de linealidad de umbrales se rechaza con un rezago de $d = 0$ con un nivel de significancia del 10 %. Este hecho es lógico debido a que los datos tomados tienen una frecuencia mensual, lo cual no permite establecer un rezago mayor a 0. Con esto podemos establecer que un modelo *TAR* es apropiado para las series de tiempo consideradas, fijando un orden autorregresivo máximo $p = 3$ como órdenes autorregresivos iniciales para los regímenes del modelo *TAR*.

3.2. Proceso de un modelo TAR

Para poder seleccionar los regímenes del modelo *TAR* tendremos en cuenta sólo dos ciclos meteorológicos que se presentan en esta zona geográfica, uno es periodos de mucha lluvia y el otro periodos de poca lluvia, fijando el valor de $l = 2$. Para estimar el resto de los parámetros que son desconocidos en el modelo se estimará de forma simultánea siguiendo la metodología descrita en la sección 1.7. Se realizan 20.000 iteraciones con un periodo de calentamiento de 12.000, las condiciones iniciales de arranque del modelo son las mismas que se tomaron en el ejemplo simulado.

TABLA 3.3. Parámetros estimados del modelo *TAR*.

	regímen 1		regímen 1	
	coeficientes	I.C. 95 %	coeficientes	I.C. 95 %
$a_0^{(j)}$	0.831 (0.24)	(0.356, 1.294)	0.932 (0.11)	(0.714, 1.150)
$a_1^{(j)}$	0.609 (0.11)	(0.394, 0.822)	0.650 (0.05)	(0.556, 0.745)
$a_2^{(j)}$	0.018 (0.07)	(-0.025, 0.262)	-0.017 (0.07)	(-0.162, 0.123)
$a_3^{(j)}$	0.004 (0.06)	(-0.006, 0.255)	0.061 (0.05)	(-0.055, 0.178)
$(h^{(j)})^2$	0.113 (0.02)	(0.088, 0.0147)	0.045 (0.00)	(0.038, 0.053)
r	12.510 (0.19)	(12.206, 12.820)	Tasa de aceptación	93.1 %

El valor estimado del rezago es $d = 0$ que coincide con la prueba de no linealidad Tsay (1998). En los resultados de la tabla 3.3 se encuentra la estimación de los coeficientes en cada régimen, entre paréntesis el error estándar, también se encuentran los intervalos de credibilidad al 95 %. Además se estimó $r = 12.510$ con una tasa de aceptación del 93.1 %.

La dinámica de $\{X_t\}$ esta determinada por los valores de precipitación tal que $Z_t < 12.510$. Es así, como los regímenes están de tal manera que para los índices de precipitación menores a 12.510 que corresponde a épocas de poca lluvia estarán asociados al régimen 1 y para los índices mayores o iguales a 12.510 que corresponden a épocas de mucha lluvia estarán asociados al régimen 2.

A continuación, presentamos los coeficientes que resultaron significativos para el modelo, dando como resultado un modelo *TAR*(2; 1, 1) que se ajusta a los datos en estudio:

$$X_t = \begin{cases} 0.831 + 0.609X_{t-1} + 0.336\varepsilon_t & \text{si } Z_t < 12.510 \\ 0.932 + 0.650X_{t-1} + 0.211\varepsilon_t & \text{si } Z_t \geq 12.510 \end{cases} \quad (3.1)$$

donde $Z_t = 10.0167 + 0.2479Z_{t-2}$ con $\hat{\sigma}^2 = 4.985$.

Ahora, se presentan algunas gráficas que permiten ver la convergencia de la cadenas de MCMC del modelo TAR.

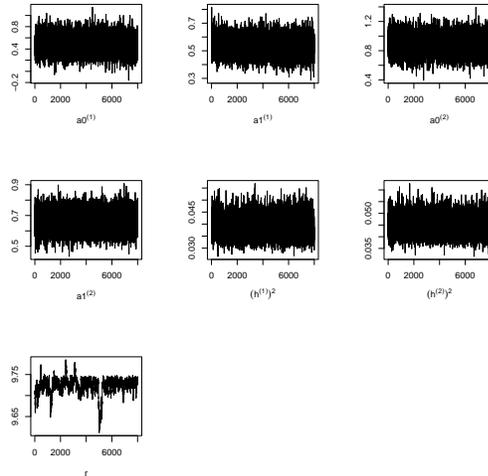


FIGURA 3.8. Muestras de las cadenas MCMC de los parámetros estimados del modelo TAR.

La figura 3.8 nos muestra que hay un comportamiento estacionario en las cadenas MCMC estimadas del modelo TAR.

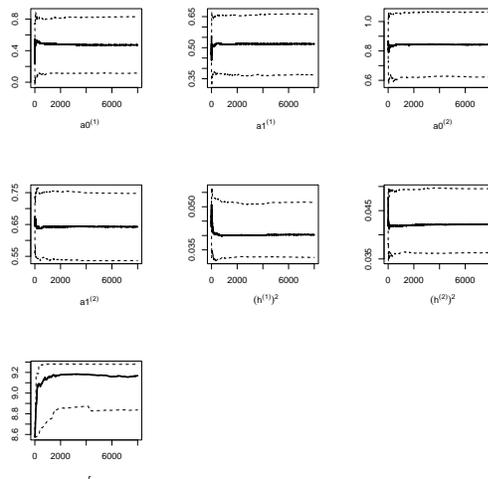


FIGURA 3.9. Convergencia de las cadenas del modelo TAR.

La figura 3.9 muestra las gráficas *cumuplot*, del modelo en donde se garantiza la convergencia de las cadenas MCMC.

Para verificar que los residuales estandarizados cumplen con los supuestos de normalidad y no autocorrelación, usamos algunas pruebas estadísticas que nos permitirán comprobarlo, la función de autocorrelación muestral y la función de autocorrelación parcial muestral de los residuales estandarizados y los residuales estandarizados al cuadrado, también la

gráfica CUSUM. La prueba de normalidad de Jarque y Bera da un $\text{valor} - p = 0$ La prueba de autocorrelación de Ljung-Box da un $\text{valor} - p = 0.956$ no existe autocorrelación en los residuales estandarizados y adicionalmente en la figura 3.10 se observa un comportamiento estable y una distribución simétrica.

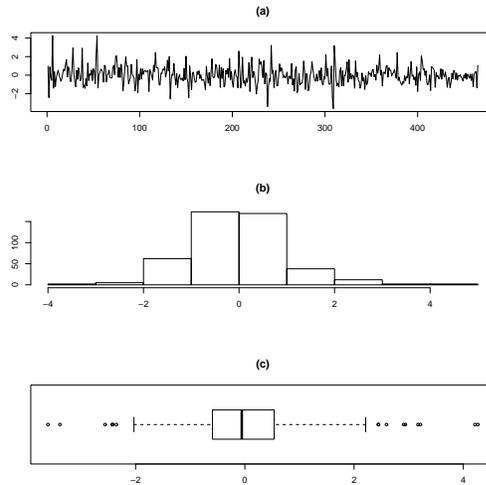


FIGURA 3.10. (a) Gráfica de la serie de residuales estandarizados, (b) histograma de los residuales estandarizados y (c) box plot de los residuales estandarizados, para el modelo TAR.

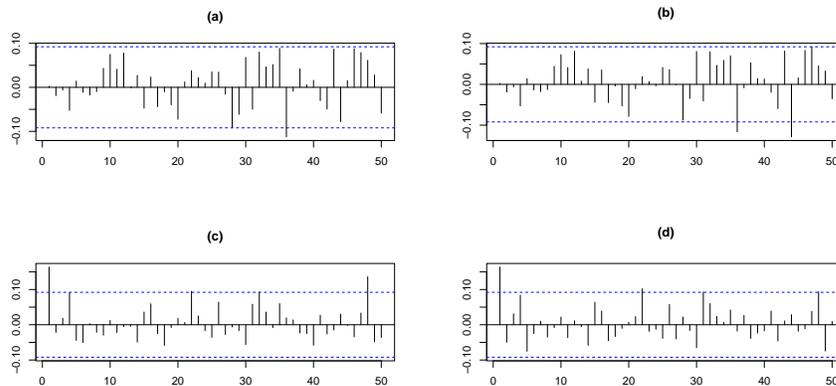


FIGURA 3.11. (a) función de autocorrelación muestral residuales, (b) función de autocorrelación parcial muestral residuales, (c) función de autocorrelación muestral residuales al cuadrado y (d) función de autocorrelación parcial muestral residuales al cuadrado.

La figura 3.11 nos indica que no hay correlación serial, ni presencia de heterocedasticidad en los residuales estandarizados.

En la figura 3.12 se realizó un gráfico CUSUM el cual nos evidencia que hay una especificación correcta del modelo.

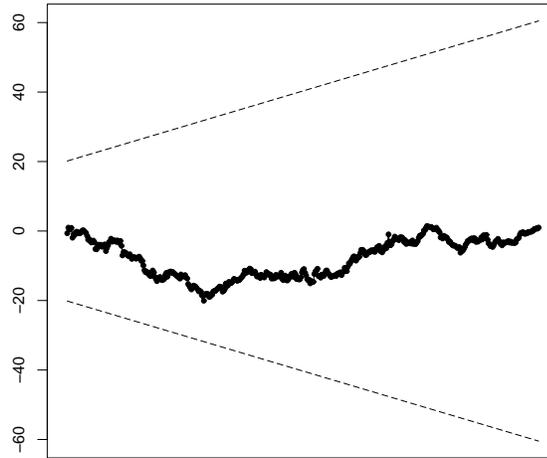


FIGURA 3.12. Gráfico CUSUM de los residuos estandarizados del modelo TAR.

Para la fase del cálculo de pronósticos en el modelo TAR, reservamos las últimas 12 observaciones, recordemos que estos datos se encuentran con un ajuste estacional.

En las tablas 3.4 y 3.5 podemos observar los pronósticos para las variables caudal del río y precipitación con ajuste estacional junto con sus intervalos creíbles al 90 % y 95 %.

TABLA 3.4. Pronósticos Variable X_t .

Horizonte (h)	X_{T+h}	\hat{X}_{T+h}	d.e.	I.C. 90 %	I.C. 95 %
1	2.61	2.48	0.28	(1.97, 2.90)	(1.85, 2.98)
2	2.55	2.46	0.33	(1.89, 2.99)	(1.76, 3.08)
3	2.61	2.47	0.36	(1.85, 3.02)	(1.71, 3.12)
4	2.75	2.46	0.36	(1.83, 3.03)	(1.69, 3.13)
5	2.68	2.46	0.37	(1.82, 3.04)	(1.68, 3.14)
6	2.79	2.46	0.37	(1.82, 3.04)	(1.68, 3.14)
7	2.58	2.46	0.37	(1.81, 3.04)	(1.67, 3.14)
8	2.73	2.45	0.37	(1.81, 3.04)	(1.67, 3.14)
9	2.70	2.45	0.37	(1.81, 3.04)	(1.67, 3.14)
10	2.85	2.45	0.37	(1.81, 3.04)	(1.67, 3.14)
11	2.60	2.45	0.37	(1.81, 3.04)	(1.67, 3.14)
12	2.93	2.45	0.37	(1.81, 3.04)	(1.67, 3.14)

Las predicciones de las variables presentan un acercamiento importante a los datos observados en los primeros 4 pasos adelante, de ahí en adelante pareciera presentar problemas la predicción, también observamos que las *d.e.* son muy grandes y los intervalos creíbles son anchos.

TABLA 3.5. Pronósticos Variable Z_t .

Horizonte (h)	Z_{T+h}	\widehat{Z}_{T+h}	d.e.	I.C. 90 %	I.C. 95 %
1	12.34	13.32	2.23	(9.64, 16.98)	(8.94, 17.70)
2	12.48	13.38	2.23	(9.70, 17.05)	(9.01, 17.75)
3	12.65	13.32	2.30	(9.53, 17.10)	(8.80, 17.82)
4	12.93	13.34	2.30	(9.56, 17.11)	(8.84, 17.84)
5	13.70	13.32	2.31	(9.53, 17.11)	(8.81, 17.85)
6	12.94	13.32	2.30	(9.53, 17.11)	(8.81, 17.84)
7	12.18	13.32	2.31	(9.52, 17.11)	(8.79, 17.84)
8	13.40	13.32	2.31	(9.53, 17.11)	(8.80, 17.84)
9	8.75	13.32	2.30	(9.52, 17.11)	(8.80, 17.83)
10	12.20	13.32	2.30	(9.53, 17.11)	(8.80, 17.83)
11	11.23	13.32	2.30	(9.53, 17.11)	(8.81, 17.84)
12	11.52	13.32	2.30	(9.53, 17.11)	(8.81, 17.83)

El modelo TAR ajustado a los datos empíricos son herramientas importantes que nos permiten pronosticar y así poder tomar decisiones que nos permitan estar preparados a futuros sucesos. También, cabe resaltar que los residuales estandarizados del modelo TAR con datos empíricos no presentan normalidad, pero sí simetría, para lo cual un posible mejor modelo sería un TAR con ruido t-Student.

En este trabajo, también se consideró ajustar un modelo TAR con covariable, siendo la covariable la misma variable de umbrales (modelo TARX - *Threshold Autoregressive Model with eXogenous*) pero debido a que estas series tienen una frecuencia mensual es difícil determinar un rezago diferente de cero ($d \neq 0$). Mientras se realizaban las estimaciones de los parámetros desconocidos el modelo siempre estimó valores d mayores a 0.

Conclusiones

En este trabajo de grado se estudió un modelo TAR de dos regímenes con series de tiempo hidrológicas (Caudal del río Coello $\{X_t\}$) y meteorológicas (Precipitación de Cajamarca $\{Z_t\}$), ambas series de tiempo con un ajuste estacional. Se adapta una metodología Bayesiana y técnicas MCMC dadas por Chen y Lee (1995) para la estimación simultánea de parámetros, además, se adapta la metodología propuesta por Calderón y Nieto (2017) para el cálculo de pronósticos. Se obtuvo un buen resultado en la estimación de parámetros y en el cálculo de pronóstico a corto plazo, aunque para estos últimos se observa que las desviaciones estándar son grandes y los intervalos creíbles anchos.

En trabajos futuros se espera poder reducir el ancho de los intervalos creíbles y las desviaciones estándar. Comparar los pronósticos con otros tipos de modelos no lineales (SETAR, TARX, STAR, etc.) como con un modelo TAR con estacionalidad.

Bibliografía

- CALDERÓN, S. (2014), *Análisis Bayesiano de Modelos Multivariados Autoregresivos de Umbrales con datos Faltantes*, Tesis de Doctorado, Departamento de Estadística, Universidad Nacional de Colombia, Bogotá.
- CASELLA, G. (1985), An introduction to empirical bayes data analysis, *The American Statistician*, **39**, 83-87.
- CHEN, W.S. CATHY Y LEE, J.C. (1995), Bayesian inference of threshold autoregressive models, *Journal of Time Series Analysis*, **16**, 483-492.
- CHIB, S. (1995), Marginal likelihood from the the Gibbs output, *Journal of the American Statistical Association*, **90**, 1313-1321.
- GEMAN, S. Y GEMAN, D. (1984), Stochastic relaxation, Gibbs distributions and the bayesian restoration of images, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **6**, 721-741.
- HASTINGS, W. K. (1970), Monte carlo sampling methods using markov chains and their applications, *Biometrika*, **57**, 97-109.
- NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AUGUSTA H. TELLER Y EDWARD TELLER (1953) Equation of State Calculations by Fast Computing Machines, *The Journal of Chemical Physics*, **21**, 1087-1093
- NIETO, F.H. (2005), Modeling Bivariate Threshold Autoregressive Processes in the Presence of Missing Data, *Communications in Statistics-Theory and Methods*, **34**, 905-930.
- NIETO, F.H. (2008), Forecasting with Univariate TAR Models, *Statistical Methodology*, **5**, 263-276.
- NIETO, F.H., ZHANG, H. Y LI, W. (2013), Using the reversible jump MCMC procedure for identifying and estimating univariate TAR models, *Communications in Statistics-Simulation and Computation* **42**, 814-840.
- TONG, H. (1978), *On a Threshold Model in Pattern Recognition and Signal Processing*, Chen, C.H.,ed, Sijhoff and Noordhoff, Amsterdam.
- TONG, H. Y LIM, K.S. (1980), Threshold autoregression, limit cycles and cyclical data, *Journal of the Royal Statistical Society, Series B* **42**, 245-292.
- TONG, H. (1990), *Nonlinear Times Series: Dynamical System Approach*, Oxford University Press, Oxford.

TSAY, R.S. (1998), Testing and modeling multivariate threshold models, *Journal of the American Statistical Association*, **93**, 1188-1202.

 Universidad del Tolima	PROCEDIMIENTO DE FORMACIÓN DE USUARIOS AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL	Página 1 de 3
		Código: GB-P04-F03
		Versión: 03
		Fecha Aprobación: 15 de Febrero de 2017

Los suscritos:

<u>Juan Camilo Gómez Ceballos</u>	con C.C N°	<u>1.110.503.436</u>
_____	con C.C N°	_____
_____	con C.C N°	_____
_____	con C.C N°	_____
_____	con C.C N°	_____

Manifiesto (an) la voluntad de:

Autorizar

No Autorizar

Motivo: _____

La consulta en físico y la virtualización de **mi OBRA**, con el fin de incluirlo en el repositorio institucional de la Universidad del Tolima. Esta autorización se hace sin ánimo de lucro, con fines académicos y no implica una cesión de derechos patrimoniales de autor.

Manifestamos que se trata de una OBRA original y como de la autoría de LA OBRA y en relación a la misma, declara que la UNIVERSIDAD DEL TOLIMA, se encuentra, en todo caso, libre de todo tipo de responsabilidad, sea civil, administrativa o penal (incluido el reclamo por plagio).

Por su parte la UNIVERSIDAD DEL TOLIMA se compromete a imponer las medidas necesarias que garanticen la conservación y custodia de la obra tanto en espacios físico como virtual, ajustándose para dicho fin a las normas fijadas en el Reglamento de Propiedad Intelectual de la Universidad, en la Ley 23 de 1982 y demás normas concordantes.

La publicación de:

Trabajo de grado	<input checked="" type="checkbox"/>	Artículo	<input type="checkbox"/>	Proyecto de Investigación	<input type="checkbox"/>
Libro	<input type="checkbox"/>	Parte de libro	<input type="checkbox"/>	Documento de conferencia	<input type="checkbox"/>
Patente	<input type="checkbox"/>	Informe técnico	<input type="checkbox"/>		
Otro: (fotografía, mapa, radiografía, película, video, entre otros)					<input type="checkbox"/>

 Universidad del Tolima	PROCEDIMIENTO DE FORMACIÓN DE USUARIOS AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL	Página 2 de 3
		Código: GB-P04-F03
		Versión: 03
		Fecha Aprobación: 15 de Febrero de 2017

Producto de la actividad académica/científica/cultural en la Universidad del Tolima, para que con fines académicos e investigativos, muestre al mundo la producción intelectual de la Universidad del Tolima. Con todo, en mi condición de autor me reservo los derechos morales de la obra antes citada con arreglo al artículo 30 de la Ley 23 de 1982. En concordancia suscribo este documento en el momento mismo que hago entrega del trabajo final a la Biblioteca Rafael Parga Cortes de la Universidad del Tolima.

De conformidad con lo establecido en la Ley 23 de 1982 en los artículos 30 “...**Derechos Morales. El autor tendrá sobre su obra un derecho perpetuo, inalienable e irrenunciable**” y 37 “...**Es lícita la reproducción por cualquier medio, de una obra literaria o científica, ordenada u obtenida por el interesado en un solo ejemplar para su uso privado y sin fines de lucro**”. El artículo 11 de la Decisión Andina 351 de 1993, “**los derechos morales sobre el trabajo son propiedad de los autores**” y en su artículo 61 de la Constitución Política de Colombia.

- Identificación del documento:

Título completo: **Modelamiento y pronósticos usando modelos TAR en algunas series de tiempo hidrológicas y meteorológicas del Tolima**

- Trabajo de grado presentado para optar al título de:

Profesional en Matemáticas con Énfasis en Estadística

- Proyecto de Investigación correspondiente al Programa (No diligenciar si es opción de grado “Trabajo de Grado”):

- Informe Técnico correspondiente al Programa (No diligenciar si es opción de grado “Trabajo de Grado”):

- Artículo publicado en revista:

- Capítulo publicado en libro:

- Conferencia a la que se presentó:

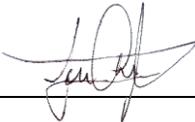
 Universidad del Tolima	PROCEDIMIENTO DE FORMACIÓN DE USUARIOS AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL	Página 3 de 3
		Código: GB-P04-F03
		Versión: 03
		Fecha Aprobación: 15 de Febrero de 2017

Quienes a continuación autentican con su firma la autorización para la digitalización e inclusión en el repositorio digital de la Universidad del Tolima, el:

Día: **03** Mes: **diciembre** Año: **2018**

Autores:

Firma

Nombre:	Juan Camilo Gómez Ceballos		C.C.	1.110.503.436
Nombre:	_____	_____	C.C.	_____
Nombre:	_____	_____	C.C.	_____
Nombre:	_____	_____	C.C.	_____

El autor y/o autores certifican que conocen las derivadas jurídicas que se generan en aplicación de los principios del derecho de autor.