Fecha de recepción: 29 de septiembre de 2015
Fecha de aceptación y versión final: 2 de diciembre de 2015

RInCE

Revista de Investigaciones del Departamento de Ciencias Económicas de La Universidad Nacional de la Matanza

Artículo de investigación:

Aplicación de los Modelos Autorregresivos Integrados de Media Móvil (ARIMA) a las Series de Precipitaciones de Lluvia[1]

Autores[2]**:** (1) Guillermo Daniel Scheidereiter, (2) Omar Roberto Faure

Resumen:

Se presenta aquí un trabajo donde se buscan modelos matemáticos que expliquen el comportamiento de las series de precipitaciones de lluvia en Entre Ríos, Argentina. Para ello se seleccionaron veintidós estaciones de registro pluviométrico y se empleó la metodología de modelización ARIMA. Se realizaron pruebas de estacionalidad y de estacionariedad. Se identificaron los modelos candidatos a representar las series estudiadas, continuando una etapa de estimación de parámetros, análisis de significatividad de los mismos y pruebas de validación basadas en exámenes residuales y de correlación paramétrica. Se halló al menos un modelo para cada serie (veintisiete modelos en total), con período estacional de doce meses. Los resultados muestran modelos ARIMA con órdenes regulares y estacionales no mayores que 2 (salvo en tres casos) y catorce modelos con un orden de diferenciación (regular o estacional), que no dan margen para concluir que la estacionariedad sea una característica intrínseca de estas series.

Palabras Claves: Series de Tiempo, Modelos ARIMA, Precipitaciones en Entre Ríos, Estacionalidad, Raíces Unitarias.

[1] Basado en la Tesis de Licenciatura en Matemática Aplicada: "Aplicación de los Modelos Autorregresivos Integrados de Media Móvil (ARIMA) a las Series de Precipitaciones de Lluvia" de Guillermo Daniel Scheidereiter. Universidad Nacional de La Matanza. Diciembre de 2014. Director de tesis: Dr. Omar Roberto Faure.

[2] (1) Guillermo Daniel Scheidereiter, Facultad Regional Concordia, Universidad Tecnológica Nacional. Licenciatura en Matemática Aplicada, Universidad Nacional de La Matanza, correo de contacto: danielscheidereiter@gmail.com ; (2) Omar Roberto Faure, Facultad Regional Concepción del Uruguay, Universidad Tecnológica Nacional, correo de contacto: ofaure@frcu.utn.edu.ar

Title: Application of the Models Autoregressive Integrated Moving Average (ARIMA) to Series of Precipitations of Rainfall

Abstract

This article presents a research that seeks mathematical models to explain the behavior of the series of precipitations of rainfalls in Entre Rios, Argentina. For this research twenty two rainfall stations were selected and ARIMA modeling methodology was employed. Seasonality tests and stationarity tests were performed. The candidate models to represent the series under study were identified and this study was followed by a step of estimating parameters, significance analysis thereof and validation testing based on parametric tests and of residual correlation. It was found at least one model for each series (twenty seven models in total) with seasonal period of twelve months. The results show ARIMA models with seasonal and regular orders no larger than 2 (Except in three cases) and fourteen models with one order of differentiation (regular or seasonal), that does not leave room to conclude that the stationarity is an intrinsic characteristic of this series.

Key words: Time Series, ARIMA Models, Rainfall in Entre Rios, Seasonality, Unit Roots.

Título: Aplicação dos modelos autoregressivos integrados de média movil (ARIMA) ás Séries de Precipitações de Chuva

Resumo

Apresenta-se aqui um trabalho no qual busca-se modelos matemáticos que expliquem o comportamento das séries de precipitações de chuva em Entre Rios, Argentina. Para isso, selecionaram-se vinte e duas estações de registro pluviométrico e empregou-se a metodologia de modelado ARIMA. Realizaram-se testes de estacionalidade e de estacionariedade. Identificaram-se os modelos candidatos a representar as séries estudadas, continuando uma etapa de estimação de parâmetros, análise de significatividade dos mesmos e testes de validação baseados em exames residuais e de correlação paramétrica. Encontrou-se pelo menos um modelo para cada série. (vinte e sete modelos no total), com período estacional de doze meses. Os resultados mostram modelos ARIMA com ordens regulares e estacionais não maiores que dois (exceto trêscasos) e quatorze modelos com uma ordem de diferenciação (regular ou estacional), que não dão margem para concluir que a estacionaridade seja uma característica intrínseca destas séries.

Palavras chaves: Séries de Tempo, Modelos ARIMA, Precipitações em Entre Ríos, Estacionalidade, Raízes Unitárias.

Elección del Tema

El análisis estadístico de los cambios en la precipitación de lluvia a lo largo del tiempo constituye un aspecto fundamental en la comprensión y estudio de las características climáticas de una región. Es importante para el planeamiento y la implementación de políticas de producción agropecuaria, para el estudio de la propagación de plagas y/o parásitos que afectan la producción ganadera o citrícola de una región, para el análisis de eventos hidrometeorológicos que puedan causar inundaciones, para el diseño de estructuras hidráulicas, entre otros aspectos. Contar con modelos matemáticos que describan el comportamiento histórico de las precipitaciones ayuda a comprender este fenómeno y disponer de una herramienta imprescindible de análisis para la toma de decisiones.

Definición del problema

El presente trabajo se trata de una investigación estadística en el marco de la teoría de los procesos estocásticos, donde las unidades de análisis son series de registros de precipitaciones históricas de lluvia de la provincia de Entre Ríos. Estas series se estudiarán mediante la modelización ARIMA. A partir de esta metodología se trata de dar respuesta a la siguiente pregunta: ¿Cuáles son los modelos ARIMA que explican los procesos estocásticos generadores de las series de registros históricos de lluvia en la Provincia de Entre Ríos?

Antecedentes (marco teórico conceptual)

El estudio experimental de datos ordenados cronológicamente constituye uno de los problemas fundamentales del análisis estadístico y está presente en diversas áreas del conocimiento como economía, física, biología, epidemiología, entre otras (Shumway & Stoffer, 2011). El enfoque sistemático mediante el cual se estudia los datos espaciados regularmente en el tiempo se denomina en estadística análisis de series de tiempo. *Una serie de tiempo es una sucesión de observaciones tomadas secuencialmente en el tiempo* (Box, Jenkins, & Reinsel, 1994). En particular se trata de una realización muestral de un proceso estocástico. Supóngase que se denota por Ω al *espacio de* *sucesos elementales,* A un subconjunto de Ω y $\mathcal{A}$ una colección *sigma-álgebra* de tales conjuntos, tal que $P(A)$ es la probabilidad de ocurrencia de A. Así, $(\Omega, \mathcal{A}, P)$ denota un espacio de probabilidad. Un *proceso estocástico* de valor real es una función de valor real $y(t, \omega)$ definida en $T \text{x} \Omega$, donde T es un conjunto de índices, tal que para cada t fijo, $y(t, \omega)$ es una variable aleatoria en $(\Omega, \mathcal{A}, P)$. (Fuller, 1996)

En el estudio de una serie temporal es importante conocer el proceso estocástico que la genera y utilizar un modelo para caracterizar su comportamiento y realizar predicciones futuras. Para esto resulta fundamental que el proceso estocástico presente una estructura probabilística estable a lo largo del tiempo, ya que las predicciones se realizan en base a las regularidades manifiestas en la serie observada. Esta característica de ser estable en el tiempo, se conoce como *estacionariedad* del proceso y puede caracterizarse a partir de los momentos asociados a la serie (media, varianza y covarianzas). La teoría que se emplea para el estudio de los procesos estocásticos que generan las series de precipitaciones de lluvia es la metodología de modelización ARIMA (*Autoregressive Integrated Moving Average*). Esta fue propuesta en 1970 por George Box y Gwilym Jenkins en la obra *Time Series Analysis: Forecasting and Control* y constituye uno de los avances más notables en el estudio de las series temporales. La ecuación general de un modelo ARIMA tiene la siguiente forma:

$$\Phi_P(B^s)\varphi_p(B)\nabla_s^D\nabla^d y_t = \delta + \Theta_Q(B^s)\theta_q(B)A_t \qquad (1)$$

donde A_t es un proceso de ruido blanco[3] Gaussiano, y_t es la variable y δ un término constante . El modelo general se denota como $ARIMA(p,d,q)\text{x}(P,D,Q)_s$. Los polinomios $\varphi_p(B)$ y $\theta_q(B)$, donde B es el operador de retardos, recogen la estructura regular de la serie, mientras que $\Phi_P(B^s)$ y $\Theta_Q(B^s)$ explican la estructura estacional. $\nabla^d = (1-B)^d$ y $\nabla_s^D = (1-B^s)^D$ son el operador diferencia regular y el operador diferencia estacional, donde s es el período estacional. Tienen por finalidad corregir el comportamiento tendencial o no estacionario que ocasiona inestabilidad en la media de la serie.

Justificación del Estudio

Los modelos ARIMA son modelos paramétricos donde cada observación en un instante particular es modelada en función de los valores pasados. La ecuación general

[3] Un proceso de ruido blanco es una secuencia de variables aleatorias independientes, idénticamente distribuidas, con media constante, usualmente cero, varianza constante y covarianzas nulas.

permite explicar las componentes principales de una serie temporal como tendencia[4], estacionariedad, estacionalidad[5] y aleatoriedad que están presentes en las series de precipitaciones. En efecto, las lluvias, al igual que otros componentes del clima, se manifiestan de formas diferentes conforme a las estaciones del año. Por ejemplo, en primavera las precipitaciones son más frecuentes que en invierno. Además, estas series presentan una dinámica irregular producto de las condiciones aleatorias intrínsecas de estos fenómenos y puede darse, también, el caso de que se encuentre evidencia de un comportamiento tendencial. Los modelos ARIMA permiten recoger estas características lo que justifica su utilización en este estudio.

Limitaciones

La investigación se centra en la obtención de modelos matemáticos que representan las series de precipitaciones de lluvia en Entre Ríos. No se estudia la capacidad predictiva de estos modelos y tampoco se confronta su ajuste con el de otros modelos que puedan lograrse con metodologías alternativas.

Alcances del Trabajo

Este trabajo tiene, principalmente, interés técnico y analítico, mostrando una directriz procedimental específica para el análisis estadístico de las series de precipitaciones de lluvias. En este sentido, la investigación proporciona un aporte metodológico para otros estudios relacionados y no se descarta la posibilidad de que resulte de interés para la enseñanza de conceptos vinculados a las series de tiempo y a sus aplicaciones.

Objetivos

El objetivo principal es elaborar un modelo ARIMA para cada serie de registro de lluvia, de manera que las implicaciones teóricas de los modelos se ajusten y sean compatibles con los datos muestrales. Además, se espera que este estudio aporte información de interés metodológico en el análisis de investigaciones relacionadas.

[4] Movimiento o dirección general de la variable en períodos prolongados de tiempo.
[5] Característica según la cual una serie repite un comportamiento similar con cierta periodicidad en el tiempo.

Hipótesis

Hipótesis I: Los modelos que explican las series de precipitaciones acumuladas mensuales de lluvia de la Provincia de Entre Ríos son modelos ARIMA estacionales, con periodicidad estacional de 12 meses.

Hipótesis II: Las series de precipitaciones de lluvia son estacionarias.

Material y Métodos

Como se mencionó anteriormente, las unidades de análisis son series de precipitaciones. La variable de estudio es la cantidad acumulada mensual de lluvia registrada en milímetros. Se seleccionó un conjunto de estaciones de registro de la página oficial de la Dirección de Hidráulica de Entre Ríos (http://www.hidraulica.gob.ar/). En principio, se tomaron aquellas series con períodos completos de registros con continuidad mes a mes a través de los años. Si la serie presentó al menos un año sin observaciones, se tomó desde el primer año de registro a partir del cual la continuidad de datos mes a mes se prolonga hasta años recientes. Además, no se incluyeron estaciones con registros inferiores a los 15 años. Resultaron 22 series que se agruparon en intervalos de tiempo medidos en años (Tabla 1).

Tabla 1: Listado de las estaciones de registro pluviométrico seleccionadas y su agrupamiento de acuerdo a su longitud anual (entre paréntesis se indica el departamento).

Intervalo anual	Estación de Registro	
15 a 25 años	Los Charrúas (Concordia)	Lucas Sur (Viallaguay)
	Octavo Distrito (Gualeguay)	Puente de Hierro (Feliciano)
	San Víctor (Feliciano)	Santa Anita (Uruguay)
	Séptimo Distrito (Gualeguay)	Viale (Paraná)
26 a 35 años	Colón (Colón)	La Lila (Feliciano)
	Lucas González (Nogoyá)	Sta. María del Tatutí (Federación)
	Villa Elisa (Colón)	
36 a 45 años	Antelo (Victoria)	Febre (Nogoyá)
	Feliciano (Feliciano)	San Jaime (Federación)
	San Salvador (San Salvador)	
46 a 55 años	La Paz (La Paz)	San Gustavo (la Paz)
	Paraná (Paraná)	

Fuente: elaboración propia

Se trabajaron las siguientes etapas para la construcción del modelo:

1. Estudio de Estacionariedad

Se aplicó sobre cada serie la transformación de Box-Cox, $g(y_t) = \frac{(y_t+m)^{\lambda}-1}{\lambda}$, para lograr estabilidad en varianza, para lo cual se empleó el software Gretl obteniéndose valores de λ entre 0,37 y 0,5 y valores de m que no superaron 0,02.

Para el estudio de la estacionariedad en media se utilizaron dos test de raíces unitarias, el test HEGY (Hylleberg Engle, Granger y Yoo) y el contraste de Canova-Hansen[6]. Mediante el uso de Gretl se encontró evidencia de inestabilidad en ocho de las series estudiadas. La Tabla 2 consigna estos resultados:

Tabla 2: Series con evidencia de posibles raíces unitarias

Estación de Registro	**Test HEGY[7]**	**Test de Canova-Hansen (CH)**
Octavo Distrito	No hay evidencia de raíces unitarias estacionales[8]	$Valor-p<0{,}10$ en las frecuencias $\pm\frac{\pi}{2}$ y $\pm\frac{\pi}{2}$
San Víctor	$Valor-p>0{,}01$ en la frecuencia[9] $\pm\frac{\pi}{2}$	$Valor-p<0{,}05$ en la frecuencia $\pm\frac{\pi}{2}$
Santa Anita	La serie no presenta raíces unitarias estacionales[10]	$Valor-p<0{,}10$ en la frecuencia $\pm\frac{\pi}{2}$
Séptimo Distrito	$Valor-p>0{,}01$ en la frecuencia[11] $\pm\frac{\pi}{6}$	No hay evidencia de raíces unitarias estacionales
Viale	$Valor-p>0{,}01$ en la frecuencia cero[12].	$Valor-p<0{,}05$ en las frecuencias $\pm\frac{\pi}{2}$ y $\pm\frac{\pi}{2}$
Villa Paranacito	$Valor-p>0{,}01$ en la frecuencia cero[13].	No hay evidencia de raíces unitarias estacionales
Feliciano	No hay evidencia de raíces unitarias estacionales	$Valor-p<0{,}05$ en las frecuencias $\pm\frac{\pi}{6},\pm\frac{\pi}{2}\ y\ \frac{2\pi}{2}$
San Jaime	No hay evidencia de raíces unitarias estacionales	$Valor-p<0{,}05$ en las frecuencias $\pm\frac{\pi}{2}$ y $\pm\frac{2\pi}{2}$

Fuente: elaboración propia

[6] Sus detalles pueden consultarse en las publicaciones originales de estos tests citadas en la bibliografía.
[7] El software Gretl trabaja por defecto con el criterio Bayesiano de Schwarz (SBC).
[8] Con constante y criterio AIC el valor-p>0,01 en la frecuencia $\pm\frac{\pi}{6}$.
[9] Con constante y criterio AIC el valor-p>0,05 en las frecuencias $\pm\frac{\pi}{6}$ y $\pm\frac{\pi}{3}$ y el valor-p>0,01 en las frecuencias $\pm\frac{\pi}{2}\ y\ \pm\pi$. Con constante y criterio HQC, el valor-p>0,01 en la frecuencia $\pm\frac{\pi}{3}$.
[10] Con constante y criterio AIC el valor-p>0,05 en las frecuencias $0,\pm\frac{\pi}{6}\ y\ \pm\frac{\pi}{3}$.
[11] Con criterio AIC se confirma el resultado. Con criterio HQC se obtiene igual resultado.
[12] Con criterio AIC, el valor-p>0,05. Con criterio HQC, el valor-p>0,01.
[13] Igual resultado se obtiene con los criterios AIC y HQC.

ISSN 1852-3239 - http://rince.unlam.edu.ar

2. Identificación del modelo

La identificación de los modelos ARIMA candidatos a explicar el comportamiento de las series objeto de estudio se realiza mediante la observación y análisis de las funciones de autocorrelación simple y parcial junto con la información que aportan los contrastes de raíces unitarias. En esta etapa se procede a seleccionar los órdenes de los parámetros p, d, q, P, D y Q. Las características presentes en los correlogramas, no concordaron en todos los casos con los resultados obtenidos mediante los contrastes de raíces unitarias. Más aún, hay series para las que tampoco es unánime la conclusión de los test. Por lo tanto, los modelos se sugirieron teniendo en cuenta todas las alternativas.

3. Estimación de los parámetros

Una vez propuestos los modelos tentativos que explican el comportamiento de los procesos generadores de las series, se procede a estimar los parámetros del modelo. Se empleó el método de máxima verosimilitud y se estimaron los parámetros de los modelos. A saber, la media, μ, de la serie, varianza, σ_A^2, de A_t y el vector de parámetros $\left(\Phi_1, \dots, \Phi_P, \varphi_1, \dots, \varphi_p, \Theta_1, \dots, \Theta_Q, \theta_1, \dots, \theta_q\right)'$. Los valores $\hat{\mu}$, $\hat{\sigma}_A^2$ y $\left(\hat{\Phi}_1, \dots, \hat{\Phi}_P, \hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{\Theta}_1, \dots, \hat{\Theta}_Q, \hat{\theta}_1, \dots, \hat{\theta}_q\right)'$ que maximizan la función de verosimilitud asociada a cada modelo ARIMA son los estimadores máximo verosímiles de los parámetros del modelo.

4. Pruebas de significatividad paramétrica

La significación de los parámetros estimados en la fase anterior es fundamental para decidir su inclusión en los modelos propuestos. Si los parámetros que se estimaron no son significativamente diferentes de cero, entonces deben suprimirse del modelo para evitar la sobreparametrización. Si el vector a estimar es $h = \left(\delta, \Phi_1, \dots, \Phi_P, \varphi_1, \dots, \varphi_p, \Theta_1, \dots, \Theta_Q, \theta_1, \dots, \theta_q\right)'$, para determinar si el modelo está sobreestimado, se plantean los siguientes test de hipótesis:

$H_0: \delta = 0 \ versus \ H_1: \delta \neq 0 \quad (2)$ $\qquad$ $H_0: \Phi_i = 0 \ versus \ H_1: \Phi_i \neq 0 \quad (3)$

$H_0: \varphi_i = 0 \ versus \ H_1: \varphi_i \neq 0 \quad (4)$ $\qquad$ $H_0: \Theta_i = 0 \ versus \ H_1: \Theta_i \neq 0 \quad (5)$

$$H_0: \theta_i = 0 \ versus \ H_1: \theta_i \neq 0 \qquad (6)$$

Para muestras grandes, el estimador máximo verosímil $\hat{h}$ de h tiene una distribución aproximadamente normal con media h y matriz de covarianzas $Var[h]$. Además de estos test de hipótesis, se realizaron pruebas de colinealidad paramétrica mediante el cálculo de la matriz de correlación entre los parámetros. Valores altos para el coeficiente de correlación puede significar la existencia de factores comunes en el modelo. En otras palabras, los parámetros no son independientes y, por lo tanto, hay redundancia entre ellos. Una posible solución es eliminar del modelo los parámetros no significativos.

5. Validación del modelo

En la etapa de validación del modelo se tuvo en cuenta, en principio, el análisis residual. Las innovaciones de un modelo ARMA pueden escribirse mediante la expresión $A_t = \frac{\varphi(B)}{\theta(B)} w_t$ (7). Por lo tanto, es claro que si el modelo ARMA propuesto para el proceso generador de la serie es el correcto, (7) debe comportarse exactamente como un ruido blanco, con autocorrelaciones nulas, media cero y varianza constante. Para comprobar esto, se emplean estimaciones de A_t a partir de las cuales se calculan los *residuos* del modelo[14], $\hat{A}_t = \frac{\hat{\varphi}(B)}{\hat{\theta}(B)} w_t$. (8)

Primero se graficó los residuos contra el tiempo para comprobar que la serie oscile alrededor de cero con una variabilidad constante y se realizó el siguiente contraste a dos colas para probar que la media es cero:

$$H_0: E[A_t] = 0 \ versus \ H_1: E[A_t] \neq 0$$

El estadístico es un t habitual: $t = \frac{E[\hat{A}_t]}{\sqrt{\frac{Var[\hat{A}_t]}{n}}} \sim N(0,1)$. (9)

Para chequear si los residuos se distribuyen normalmente, se construyó un histograma de los residuos estandarizados, esto es $\frac{\hat{A}_t}{\hat{\sigma}_{\hat{A}}^2}$, y luego se realizó una comparación con la distribución normal empleando el test de Doornik-Hansen que sigue aproximadamente una distribución Chi-cuadrado con dos grados de libertad, $E_p \sim \chi^2_{(2)}$. Además, se calculó la ACF y la $PACF$ de $\hat{A}_t$, con el propósito de comprobar que los residuos no siguen ningún patrón que revele correlación, es decir que los coeficientes de la función de autocorrelación, ρ_k, no son estadísticamente significativos. Para probar

[14] La idea subyacente es que el modelo propuesto genera estimaciones de los valores reales observados de la serie. Así, la diferencia entre el valor observado y el estimado será un residuo.

esto, se contrasta la siguiente hipótesis: $H_0: \rho_1 = \rho_2 = \cdots = \rho_k = 0$ versus la hipótesis alternativa de que alguno de estos coeficientes es diferente de cero. Este contraste se realiza mediante el estadístico $Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$, (Ljung & Box, 1978) , que bajo la hipótesis nula tiene aproximadamente una distribución χ^2 con $h-(p+q)$ grados de libertad. (Wei, 2006)

Por último, se estudió la homocedasticidad residual mediante la observación de un gráfico rango-media para cada caso. Si un ajuste por regresión del gráfico señala que a mayor media aumenta la varianza, entonces existe heterocedasticidad y el diagrama suele presentar forma de “embudo”. Por el contrario, un conjunto de datos homocedástico tendrá una regresión con pendiente aproximadamente cero, puesto que las variables $\bar{x}_k$ y σ_k^2 no están correlacionadas linealmente y el diagrama no sigue ningún patrón que refleje alguna forma.[15]

Además de estos estudios, se comprobó que los modelos tengan raíces en módulo superior a la unidad para garantizar el cumplimiento de las condiciones de causalidad e invertibilidad. Por último, se utilizó el criterio de información de Akaike, AIC, el criterio bayesiano de Schwarz y el criterio de Hannan-Quinn, en los casos en que más de un modelo soportara las pruebas de validación.

Resultados

A continuación se muestra los modelos seleccionados para cada estación de registro. Los parámetros no significativos se suprimieron de los modelos:

Los Charrúas:

$$ARIMA(1,0,0)x(0,0,0)_{12}$$

$$(1-0{,}33805602B)g(y_t) = 8{,}311275676 + A_t$$

Lucas Sur:

$$ARIMA(0,0,2)x(1,0,0)_{12}$$

$$(1-0{,}174254.B^{12})g(y_t) = 9{,}9792 + (1+0{,}215065.B + 0{,}235373.B^2)A_t$$

Octavo Distrito:

Modelo 1:

$$ARIMA(0,0,2)x(1,1,1)_{12}$$

$$(1+0{,}140214B^{12})(1-B^{12})g(y_t) = (1+0{,}253185B^2)(1-0{,}701589B^{12})A_t$$

[15] Puede emplearse el test habitual que permite contrastar $H_0: \alpha = 0$ versus $H_1: \alpha \neq 0$ donde α es la pendiente de la recta de regresión $\sigma_k^2 = \alpha\bar{x}_k + constante$.

Modelo 2:

$$ARIMA(2,0,0)x(1,1,1)_{12}$$

$$(1-0{,}25624B^2)(1-B^{12})g(y_t)=(1-0{,}709892B^{12})A_t$$

Puente de Hierro:

$$ARIMA(0,0,5)x(1,0,1)_{12}$$

$$(1-0{,}535312B^{12})g(y_t) = 5{,}86861+(1+0{,}112918\mathrm{B}+0{,}208105B^2-0{,}136021B^4-0{,}191963B^5)A_t$$

San Víctor:

$$ARIMA(1,0,0)x(1,1,1)_{12}$$

$$(1-0{,}192611\mathrm{B})(1-B^{12})g(y_t)=(1-0{,}817495B^{12})A_t$$

Santa Anita:

Modelo 1:

$$ARIMA(1,1,1)x(1,1,1)_{12}$$

$$(1-B^{12})(1-B)g(y_t)=(1-0{,}899614\mathrm{B})(1-0{,}767096B^{12})A_t$$

Modelo 2:

$$ARIMA(1,0,1)x(1,1,1)_{12}$$

$$(1-0{,}0{,}600822\mathrm{B})(1-B^{12})g(y_t)=(1-0{,}782526B^{12})A_t$$

Séptimo Distrito:

$$ARIMA(1,0,2)x(1,1,1)_{12}$$

$$(1-B^{12})g(y_t)=(1+0{,}142392B^2)(1-0{,}752449B^{12})A_t$$

Viale:

$$ARIMA(1,0,0)x(0,1,1)_{12}$$

$$(1-0{,}222247\mathrm{B})(1-B^{12})g(y_t)=(1-0{,}687592B^{12})A_t$$

Villa Paranacito:

$$ARIMA(1,0,0)x(1,0,0)_{12}$$

$$(1-0{,}197151\mathrm{B})(1-0{,}185279B^{12})g(y_t)=9{,}1651+A_t$$

Colón:

$$ARIMA(1,0,0)x(1,1,1)_{12}$$

$$(1-0{,}161077\mathrm{B})(1-B^{12})g(y_t)=(1-0{,}836472B^{12})A_t$$

La Lila:

$$ARIMA(1,0,2)x(1,1,1)_{12}$$

$$(1-B^{12})g(y_t)=(1+0{,}110075B^2)(1-0{,}86073B^{12})A_t$$

Lucas González:

$$ARIMA(0,0,0)x(0,1,1)_{12}$$

$$(1-B^{12})g(y_t) = (1-0{,}854797B^{12})A_t$$

Santa María del Tatutí:

Modelo 1:

$$ARIMA(1,0,0)x(1,0,1)_{12}$$

$$(1-0{,}136604\text{B})(1-0{,}96121B^{12})g(y_t) = (1-0{,}88506B^{12})A_t$$

Modelo 2:

$$ARIMA(1,0,0)x(0,1,1)_{12}$$

$$(1-0{,}127983\text{B})(1-B^{12})g(y_t) = (1-0{,}927642B^{12})A_t$$

Villa Elisa:

$$ARIMA(1,0,0)x(0,0,0)_{12}$$

$$(1-0{,}296586\text{B})g(y_t) = 9{,}02309 + A_t$$

Antelo:

$$ARIMA(1,0,0)x(1,0,1)_{12}$$

$$(1-0{,}0800517\text{B})(1-0{,}900118B^{12})g(y_t) = 1{,}01489 + (1-0{,}758945B^{12})A_t$$

Febre:

Modelo 1:

$$ARIMA(1,0,1)x(1,0,1)_{12}$$

$$(1-0{,}558112\text{B})(1-0{,}938242B^{12})g(y_t) = 0{,}31656 + (1-0{,}550027\text{B})(1-0{,}827128B^{12})A_t$$

Modelo 2:

$$ARIMA(1,0,1)x(0,1,1)_{12}$$

$$(1+0{,}533195\text{B})(1-B)g(y_t) = (1-0{,}540941\text{B})(1-0{,}868914B^{12})A_t$$

Feliciano:

$$ARIMA(1,0,2)x(1,1,1)_{12}$$

$$(1-B^{12})g(y_t) = (1+0{,}458383\text{B}+0{,}173515B^2)(1-0{,}827634B^{12})A_t$$

San Jaime:

Modelo 1:

$$ARIMA(2,0,1)x(1,1,1)_{12}$$

$$(1-B^{12})g(y_t) = (1-0{,}840634B^{12})A_t$$

Modelo 2:

$$ARIMA(2,0,1)x(1,0,1)_{12}$$

$$(1-0{,}154501B^2)(1-0{,}291909B^{12})g(y_t) = 7{,}24863 + A_t$$

San Salvador:

$$ARIMA(1,0,0)x(2,0,0)_{12}$$

$$(1-0{,}206252\mathrm{B})(1-0{,}1483B^{12}-0{,}177606B^{24})g(y_t) = 5{,}97247 + A_t$$

La Paz:

$$ARIMA(0,0,5)x(2,0,0)_{12}$$

$$(1-0{,}144538\mathrm{B}^{12}-0{,}0810902\mathrm{B}^{24})g(y_t) = 8{,}45036 + (1+0{,}140548\mathrm{B}+0{,}15108B^2-0{,}109353B^5)A_t$$

San Gustavo:

$$ARIMA(1,0,1)x(2,0,0)_{12}$$

$$(1-0{,}435584\mathrm{B})(1-0{,}149665B^{12}-0{,}109613B^{24})g(y_t) = 4{,}74973 + A_t$$

Paraná:

$$ARIMA(0,0,6)x(1,0,1)_{12}$$

$$(1-0{,}974067B^{12})g(y_t) = 0{,}274851 + (1-0{,}0999834B^6)(1-0{,}887638B^{12})A_t$$

Discusión

Los instrumentos claves en la identificación de estos modelos son las funciones de autocorrelación y los contrastes de raíces unitarias. En ocho de las series estudiadas se detectó evidencia de raíz unitaria en alguna frecuencia. Pero en otros casos no es posible llegar a resultados concluyentes a partir de la información que aportan los test. En el caso del test HEGY el no rechazo de la hipótesis nula para algunas series se basa en un valor-p que supera a 0,01 pero no a 0,05, por lo que no puede concluirse presencia de raíz unitaria a un 5% de significación. Además, muchos de estos resultados dependieron del criterio de información utilizado (AIC, SBC y/o HQ). El contraste de Canova-Hansen parece menos ambiguo, aunque hubo casos en los que no detectó inestabilidad y el modelo con mejor ajuste sí incluyó un orden de diferenciación estacional sugerida por el test HEGY o por el comportamiento de la ACF (el modelo propuesto para la serie que proviene del registro de La Lila tiene un orden de diferenciación que no es detectada por el contraste de Canova y Hansen y apenas lo es con el test HEGY bajo ciertas restricciones). Aunque estos contrastes son sumamente valiosos para detectar la presencia de raíces unitarias no muestran, en este estudio, resultados unánimes. La decisión de incluir un orden de diferenciación estuvo en todos los casos sustentada en el conjunto de informaciones que revelaron los test, los correlogramas y el análisis de validación.

Un problema recurrente en muchos casos fue el de la alta correlación particularmente entre los parámetros estacionales. Esto representa una complicación cuando las estimaciones de los parámetros no pueden suprimirse del modelo, es decir, cuando son significativamente distintos de cero. Éste fue el caso de los modelos propuestos para las series procedentes de Febre y La Lila.

El análisis residual no mostró en todos los casos que los residuos se comporten perfectamente como un ruido blanco. Incrementar el orden del polinomio de media móvil regular permitió alcanzar residuos con covarianzas nulas, pero a riesgo de la pérdida de homocedasticidad o de normalidad.

Respecto de las hipótesis de investigación planteadas, en el caso de la Hipótesis I, es cierto que los modelos que explican los procesos generadores de las series son modelos ARIMA estacionales multiplicativos, pero hay registros tal que los modelos no incluyen órdenes estacionales. Concerniente a la Hipótesis II, existe evidencia de que ésta no se corroboró en todos los casos.

Los objetivos fijados en un inicio se lograron concretar. El trabajo muestra los aspectos básicos de la metodología de modelización ARIMA y cómo a partir de ella se logra obtener modelos que describen las características de las series de registros de lluvia.

Conclusiones

Los resultados muestran modelos que en su mayoría no presentan órdenes autorregresivos y de media móvil mayores que 2 (salvo en tres casos). Para algunas series (Octavo Distrito, Santa Anita, Santa María del Tatutí, Febre y San Jaime), se propuso dos modelos para explicar el proceso que las genera puesto que ambos muestran un buen ajuste residual y escasa diferencia en los criterios de información. Por lo que en total se consignaron veintisiete modelos, catorce de los cuales tienen un orden de diferenciación ya sea regular o estacional. Lo cual no deja margen para concluir que las series de precipitaciones de lluvia sean en su naturaleza estacionarias.

La primera línea de investigación que continúa inmediatamente es estudiar la capacidad predictiva de estos modelos. También, es importante analizar la presencia de outliers y explorar modelos que recojan esos valores extremos. Las causas de valores atípicos pueden encontrarse en cambios climáticos bruscos o algunos prolongados como por ejemplo, el fenómeno conocido como "*la corriente del niño*". Otra línea de investigación es considerar la existencia de otros modelos de series temporales, diferentes a los modelos ARIMA, que puedan presentar mejores ajustes y recojan con más precisión las características de las series de registros de lluvia. Por ejemplo, los modelos de *suavizado exponencial* o los modelos GARCH (*Generalized Autoregressive Conditional Heteroscedasticity*), que permiten estudiar series que tienen desviaciones grandes (pequeñas) con respecto a un nivel constante, seguidas inmediatamente de variaciones pequeñas (grandes). Se continuará estas líneas de investigación para obtener más y mejores resultados en la modelación estadística de las series de registros de lluvia, que puedan servir como aporte metodológico a otras investigaciones relacionadas.

Bibliografía

Beaulieu, J., & Miron, J. (1993). Seasonal Unit Roots in Aggregate U.S. Data. *Journal of Econometrics, 55*, 305-328.

Box, G., Jenkins, G., & Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control.* USA: Prentice-Hall International.

Canova, F., & Hansen, B. E. (Julio de 1995). Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability. *Journal of Business & Economic Statistics, 13*(3).

Doornik, J., & Hansen, H. (1994). An Omnibus Test for Univariate and Multivariate Normality. *Working Paper, Nuffield College, Oxord*.

Fuller, W. (1996). *Introduction to Statistical Time Series.* New York: Willey.

Hendry, D., & Mizon, G. (1978). Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England. *The Economic Journal, 88*(351), 549-563.

Hylleberg, S., Engle, R., Granger, C., & Yoo, B. (1990). Seasonal Integration and Cointegration. *Journal of Econometrics, 44*, 215 - 238.

Ljung, G. M., & Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*(2), 297-303.

Shumway, R. H., & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications.* New York: Springer.

Wei, W. W. (2006). *Time Series Analysis. Univariate and multivariate methods.* U.S.A.: Pearson Addison Wesley.