ANÁLISIS DE SERIE DE TIEMPO DE CAUDALES DEL RÍO EL TALA PERIODO 1937-1960

Verón, Juan Antonio* ; Herrera, Carlos Gabriel*; Rodríguez, Norma Leonor**
* Facultad de Tecnología y Ciencias Aplicada de la UNCa.
** Facultad de Ciencias Exactas y Naturales de la UNCa.
javeron@tecno.unca.edu.ar Maximio Victoria 55, CP 4700 - Catamarca

PLANTEO DEL PROBLEMA

El río El Tala es una de las principales fuentes de provisión de agua para el consumo humano de la ciudad de Catamarca, tanto por su aporte superficial como por las reservas subterráneas.
Sin embargo, es escasa la información disponible de las variables hidrológicas de su cuenca. No hubo instalado un pluviómetro en ella y por lo tanto no existen registros hidrométricos de precipitaciones.
La única serie disponible es la de caudales históricos que registró la antigua Obra Sanitaria de la Nación (OSN) responsable del servicio antes de que se transfiriera al ámbito provincial, todos obtenidos en una estación de registro ubicada en el paraje de La Brea, cerca del lugar donde el caudal es tomado para derivarlo a la presa de El Jumeal.

OBJETIVO

Analizar la serie temporal de los caudales del Río El Tala entre 1936 y 1960 con el objeto de entender y describir los procesos aleatorios que generan las observaciones, predecir valores futuros de la variable y el control óptimo de esta clase de procesos.

MARCO TEÓRICO

El análisis de las series temporales se enmarca en los procesos estocásticos. Una sucesión de variables aleatorias X_t (t= 1, 2,…,n), donde cada valor observado puede considerarse una muestra aleatoria de tamaño uno de la variable. Además X_a y X_b están separadas por *k* retardos si $|a-b|$ = *k.*

Se denomina función de autocorrelación (ACF) de un proceso a la que describe las correlaciones en dos variables cualesquiera del proceso:

$$\rho(t,k) = Corr(X_t, X_{t+k}) \qquad \forall t,k$$

Para poder aplicar la metodología se debe convertir este proceso estocástico en un proceso estacionario. Un proceso estocástico es estacionario si:

- La media de X_t es constante
- La varianza de X_t es constante
- La correlación entre X_t y X_{t+k} depende únicamente del número de retardos k.

Para la estimación de los parámetros del modelo se utilizan las funciones:

Autocorrelación simple de orden k (ρ_k): . Es la representación de los coeficientes de autocorrelación en función de los retardos, es decir, es la correlación entre variables separadas por k retardos
Autocorrelación parcial de orden k (α_k): Es la correlación parcial entre variables separadas por k retardos eliminando el efecto de las k – 1 variables intermedias.
Para ajustar el modelo se analizan los residuos que se obtienen en un proceso que se denomina de ruido blanco (a_t). Un proceso estocástico es ruido blanco si:

- La esperanza de a_t es igual a cero
- La varianza de a_t es constante
- La Correlación de los a_k es igual a cero
- a_t sigue una distribución normal

Las alternativas del modelo se identifican según los órdenes de los parámetros que se estiman, distinguiéndose :

- Autoregresivos AR(p)
- De media móvil MA(q)
- Autoregresivo de media móvil ARMA(p,q)
- Autoregresivo integrado de media móvil ARIMA(p,d,q)
- Autoregresivo integrado de media móvil estacional ARIMA(p,d,q)(P,D,Q) para un proceso estacional con período conocido.

METODOLOGÍA

Se trata de una matriz de promedios mensuales de caudales del río El Tala de 24 años, entre los años 1937 y 1960, obtenida en el paraje La Brea, cerca del lugar donde su curso es capturado, para ser derivado a la presa de El Jumeal.

No fue posible disponer de la información primaria, es decir la relevada por OSN, en períodos menores de tiempo a fin de verificar los valores que se dispone y con los que se realizará el presente análisis.

Con esta información disponible, se realiza el análisis de la serie temporal. La base se incluye en Anexo I tal como fue provista por la Dirección Provincial de Recursos Hídricos.

El análisis de la serie temporal se realiza con el Modelo Autoregresivo Integrado de Media Móvil conocido como ARIMA (Auto Regresive Integrated Moving Average) propuesto por Box – Jenkins. El nombre genérico deriva de sus tres componentes: autoregresivo (AR), integrado (I) y de medias móviles (MA).

El modelo permite describir un valor como una función lineal de datos anteriores y errores debidos al azar, pudiendo incluir componentes cíclica y estacional, además de la tendencia.

Para obtener el modelo de series temporales se siguen los siguientes pasos:

1. Ingreso de Datos: Se incorporan los datos disponibles (en este caso mensuales) como un vector, es decir en una única columna con el nombre X="caudal" (m^3/seg.).
2. Gráfica de la serie: Con el graficador de SPSS se representa la serie histórica a fin de decidir la estacionariedad de la serie.

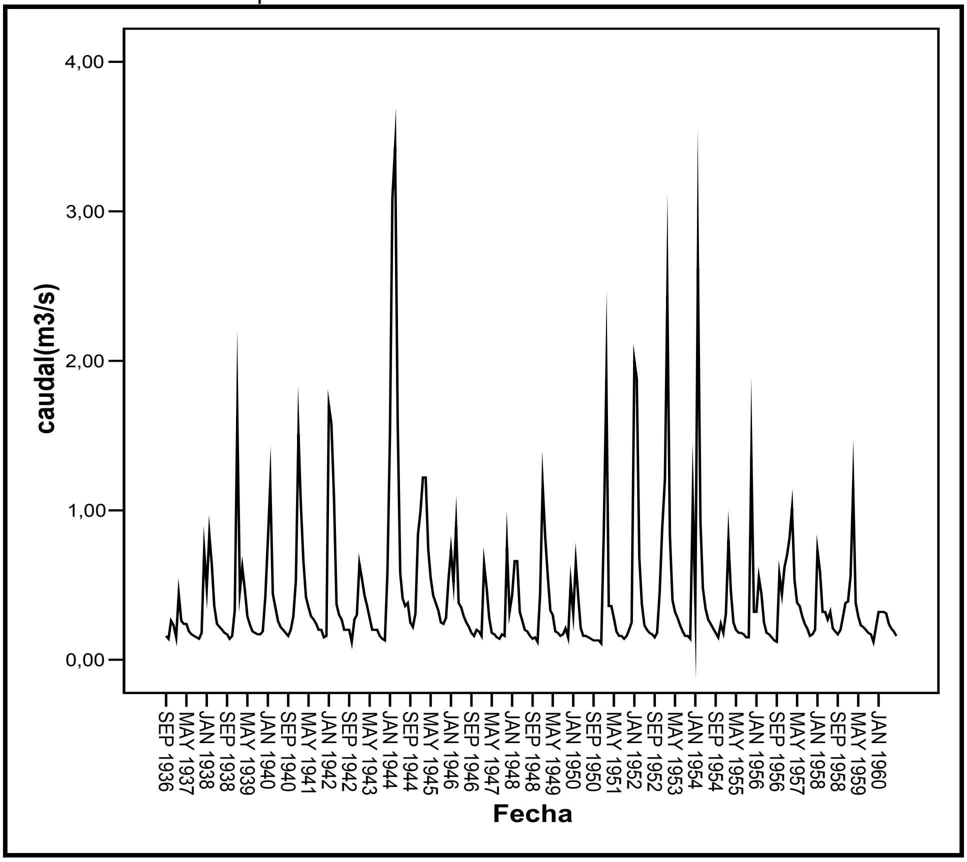
3. Transformaciones: En muchos casos conviene, antes de aplicar el método, transformar la serie original X, según una serie de alternativas que ofrece el programa. Esta transformación es imprescindible en caso de que la varianza no sea constante.
4. Eliminación de la tendencia: La observación del gráfico indicará si existe o no tendencia.
5. Identificación del modelo: Determinar el tipo de modelo más adecuado para la serie. La decisión se tomara en base de los correlogramas de las funciones de autocorrelación (ACF) y de autocorrelación parcial (PACF) que corresponden a los procesos autoregresivos y de medias móviles de las componentes regular y estacional. En caso de duda puede considerarse varias alternativas que, luego de contrastarse su validez se define el modelo adecuado.
6. Estimación de los coeficientes del modelo como se trata de un proceso iterativo de cálculo, pueden indicarse valores iniciales.
7. Contraste de validez del modelo: Se usan procedimientos para evaluar el o los modelos preseleccionados, contraste de significación de parámetros, covarianza entre estimadores, coeficiente de correlación, suma de cuadrado de errores, etc.
8. Análisis de los errores: Las diferencias entre los valores observados y los estimados por el modelo es una fuente de interés para valorarlo.
9. Selección del modelo: con los resultados de las etapas anteriores se decide sobre el modelo definitivo.
10. Predicción: el modelo seleccionado se usará como fórmula de predicción.-

Para el procesamiento de la información se utilizó el paquete SPSS (Statistics Package Social Science), versión 10.0.

RESULTADOS

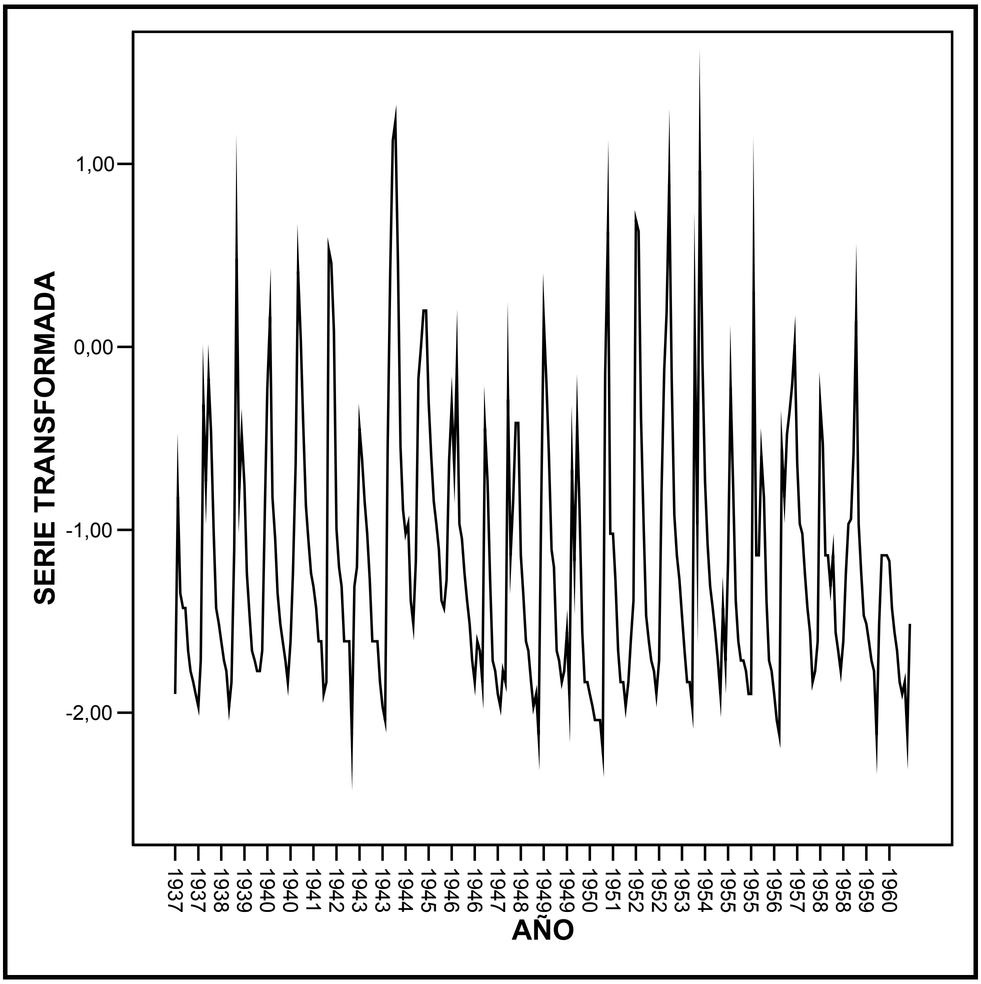
Una primera aproximación al análisis de la serie se realiza graficando la serie original CAUDAL (Gráfico Nº 1). En ella se observa gran variabilidad, con valores muy altos en algunos meses. También se visualiza el fuerte efecto estacional en cada uno de los 24 años de la serie, con valores modales en febrero-marzo de cada año y depresiones en los meses de invierno.

Gráfico Nº 1: Serie temporal de caudales de río El Tala. Período 1937 -1960



Fuente: Dirección Provincial de Recursos Hídricos de la provincia de Catamarca.

Gráfico Nº 2: Serie de Caudales del Río El Tala transformada por Logaritmo Natural. Período 1937 -1960



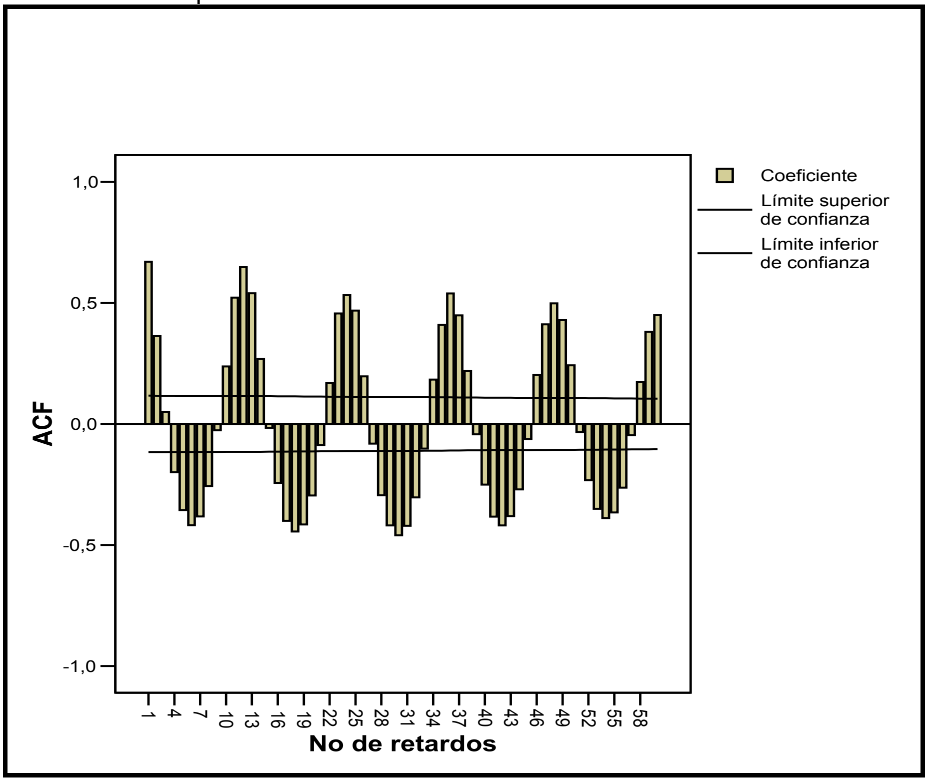
Fuente: Dirección Provincial de Recursos Hídricos y elaboración propia.

Para lograr que la media sea constante, se debe diferenciar regularmente la serie tantas veces como sea necesario para estabilizarla, identificando de esta forma el parámetro “d”. Como además la serie es estacional, se tomarán diferencias de ciclo para eliminarla, identificando este parámetro como “D”. En la serie objeto de estudio se consideró d= 0 y D = 1.

Para analizar la estabilidad de la varianza en la serie transformada se aplica la prueba de Levenne, obteniéndose una significación de p=0.067, es decir que no hay evidencia para rechazar la hipótesis nula de homogeneidad.

Para determinar si la serie presenta tendencia se recurre a la función de autocorrelación simple. (Gráfico Nº 3)

Grafico Nº 3: Función de autocorrelación simple de la serie de caudales sobre los 60 primeros retardos. Rio El Tala 1937-1960



Las líneas paralelas al eje que pasa por el valor cero, corresponden a los límites del intervalo de confianza al 95 % para el cero del coeficiente de correlación. Si la serie presentara tendencia creciente o decreciente, el gráfico mostraría un decrecimiento lento hacia 0. Sin embargo en este caso, se observa que los retardos de múltiplos de 12 (12, 24, 36, etc) presentan estructura positiva con decrecimiento hacia 0; lo que corrobora la estacionalidad de periodo 12 que mostraba el gráfico de la serie original.

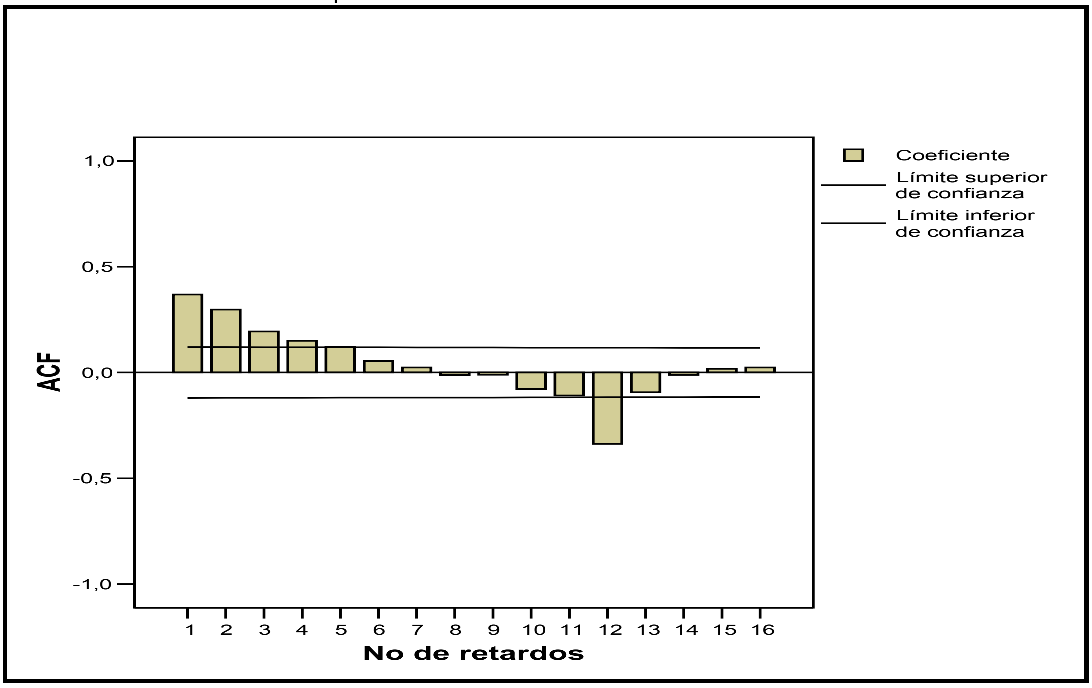
Dado que la serie no presenta tendencia, no será necesario diferenciarla y el parámetro d será igual a 0.

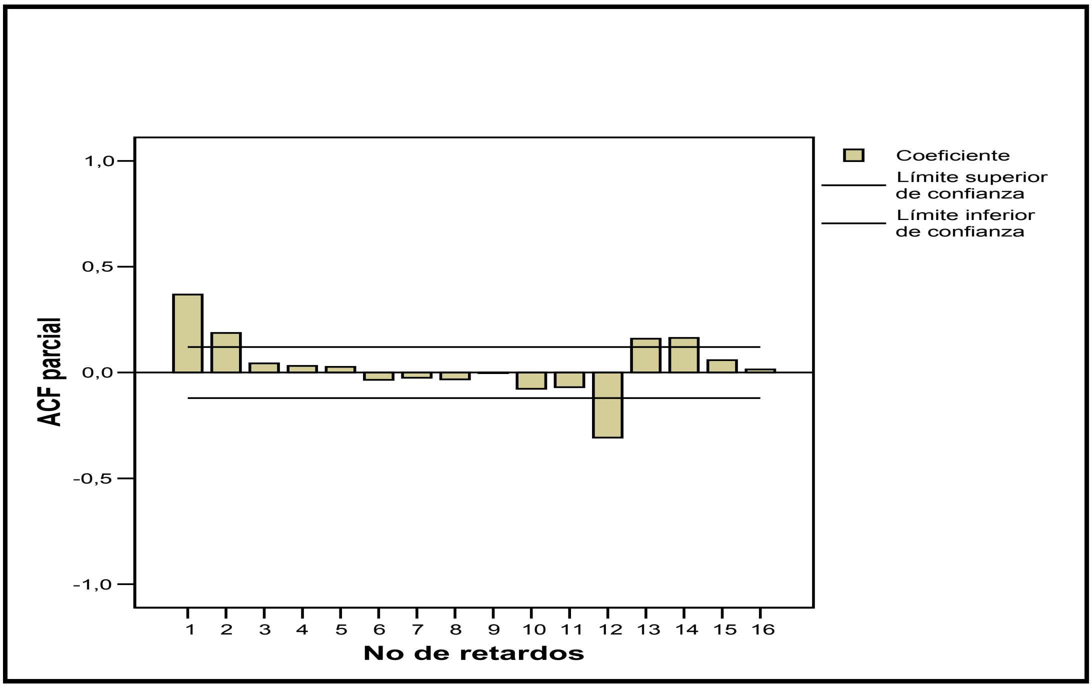
Se considera ahora las funciones de autocorrelación simple y parcial sobre los 16 primeros retardos después de tomar una diferenciación estacional (D=1) que se presenta en el Gráfico Nº 4. Los parámetros “p” y “q” se estiman a partir de estas funciones.

La serie observada tiene estacionalidad de período 12, y por lo tanto el modelo se ajustaría a un ARIMA (p, d, q) $(P, D, Q)_{12}$

Como los dos primeros coeficientes de la PACF son no nulos y el resto tienden a cero, y en tanto los coeficientes en la ACF decrecen con el retardo en forma exponencial o sinusoidal, el modelo será autoregresivo AR(p), de orden p= 2 y q=0.

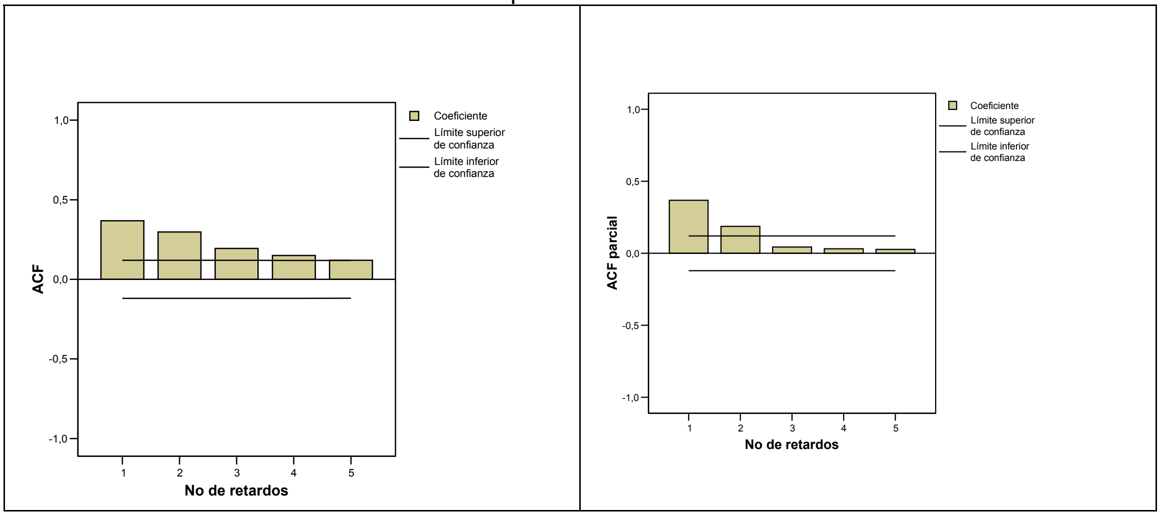
Grafico Nº 4: Funciónes de autocorrelación simple y parcial de la serie caudales, diferenciada estacionalmente sobre los 16 primeros retardos**.**





Para obtener los valores de P y Q, se realiza similar análisis que el determinado para p y q, pero con cinco retardos solamente, que se presenta en el gráfico N° 5. Se observa en el ACF dos valores no significativos, lo que determina que P = 2 y Q = 0.

Gráfico Nº 5: Función de aurtocorrelación y de autocorrelación parcial de la serie de caudales diferenciada estacionalmente sobre los 5 primeros retardos estacionales.



El modelo obtenido será entonces ARIMA (2, 0, 0)(2, 1, 0)$_{12}$.
Corresponde ahora calcular los parámetros que respondan a las ecuaciones del modelo. En el Cuadro Nº 1 se observa la estimación de los parámetros del modelo seleccionado, indicándose en la columna B los valores para cada uno de los coeficientes del modelo y en la última columna la significación de las estimaciones.

Cuadro Nº1: Calculo de las variables del modelo ARIMA y su significación

	B	SEB	T-RATIO	APPROX. PROB.
AR1	,36946581	,05946850	6,2127988	,00000000
AR2	,17487823	,05969728	2,9294172	,00368496
SAR1	-,53644843	,05691450	-9,4255136	,00000000
SAR2	-,35994331	,05795765	-6,2104544	,00000000
CONSTANT	,00270233	,03279103	,0824105	,93438110

La constante calculada en el modelo tiene una significación cercana a uno, por lo que puede eliminarse y realizar el ajuste del modelo sin considerarla, obteniéndose los valores que se presentan en el Cuadro Nº 2, en el cual pueden observarse leves modificaciones en los valores de la variable.

Cuadro Nº 2: Ajuste del modelo ARIMA (2, 0, 0)(2, 1, 0)$_{12}$.

	B	SEB	T-RATIO	APPROX. PROB.
AR1	,36942607	,05935895	6,2235954	,00000000
AR2	,17483035	,05959510	2,9336364	,00363570
SAR1	-,53643553	,05680856	-9,4428652	,00000000
SAR2	-,35994655	,05784884	-6,2221910	,00000000

Se efectúo luego la validación del modelo utilizando los errores que se obtienen como diferencia entre los valores de la serie Transformada y la resultante del modelo.(ANEXO). Se verificó que la media sea igual a cero y la varianza estable. La primera es consecuencia de los métodos de estimación de los coeficientes del modelo, en tanto que la estabilidad de las varianzas fue comprobada al inicio de este análisis.
También se verificó la normalidad de los errores a través de la prueba de Kolmogorov– Smirnov a un nivel de significación 5 % (p= 0.1512).

CONCLUSIONES Y DISCUSIÓN

El modelo desarrollado permite realizar predicciones para períodos de tiempo anteriores y posteriores al considerado en este análisis (1937-1960). Del Gráfico Nº 1 puede observarse la excesiva variabilidad de la serie, lo que dificulta la obtención de un modelo que contemple tales variaciones. Si bien el modelo es perfectible en el sentido de que se pueden introducir variables que no responden a ningún patrón sistemático de comportamiento (ARIMA con intervención), los resultados que se presentan en el ANEXO, muestran que los valores obtenidos a través del modelo no difieren de los valores de la serie original transformada.
Al respecto conviene aclarar que para volver a la serie de los datos del caudal en (m3/s) debe aplicarse la función inversa del Logaritmo Natural, con la cual se ajustó el modelo.

BIBLIOGRAFÍA

- Ferrán Aranaz, Magdalena. (1996). SPSS para Windows: Programación y análisis estadístico. McGraw-Hill, Madrid, España,. Capítulo 20 Series temporales, pág 451 a 512 y Capítulo 6 Pruebas Estadística para una muestra, pág. 135 a 140.

- Box, G. E. P.,Jenkins, G. M. (1976). Time series analysis Forecasting and control (rev.ed.) Holder-Day, San Francisco.

- Norusis, M. J. (1994). SPSS para Windows, Versión 6.0. Chicago.

- Leiva, R (2002). Introducción al análisis a la Serie de Tiempo. Universidad Nacional de Cuyo. Cátedra de Estadística II. http://fce.uncu.edu.ar

- Pérez, César (2001). Técnicas Estadísticas con SPSS. Prentice Hall. Madrid Capítulo 18 Análisis de series temporales, pág. 535 a 557.

ANEXO I

CUADRO A1

DATOS DE AFOROS DEL RIO : EL TALA

- EVARSA -

AÑO	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC	MED.ANUAL
1937	0,150	0,400	0,260	0,240	0,240	0,190	0,170	0,160	0,160	0,140	0,260	0,230	**0,226**
1938	0,480	0,840	0,630	0,360	0,240	0,220	0,200	0,180	0,150	0,140	0,180	0,730	**0,394**
1939	1,620	0,460	0,620	0,470	0,290	0,230	0,190	0,180	0,170	0,140	0,160	0,330	**0,508**
1940	0,800	1,180	0,440	0,350	0,260	0,220	0,200	0,180	0,170	0,170	0,190	0,420	**0,454**
1941	1,510	1,040	0,660	0,420	0,350	0,290	0,270	0,240	0,160	0,200	0,290	0,530	**0,598**
1942	1,700	1,580	1,090	0,370	0,300	0,270	0,200	0,200	0,200	0,200	0,150	0,160	**0,714**
1943	0,640	0,540	0,430	0,360	0,280	0,200	0,200	0,200	0,200	0,120	0,270	0,300	**0,356**
1944	1,520	3,090	3,430	1,600	0,580	0,410	0,360	0,380	0,160	0,140	0,130	0,580	**1,421**
1945	1,000	1,220	1,220	0,740	0,550	0,430	0,380	0,330	0,250	0,220	0,310	0,840	**0,734**
1946	0,730	0,520	0,890	0,380	0,350	0,290	0,250	0,220	0,250	0,240	0,280	0,540	**0,454**
1947	0,160	0,640	0,480	0,280	0,180	0,170	0,150	0,140	0,180	0,160	0,200	0,190	**0,275**
1948	0,430	0,660	0,660	0,320	0,260	0,200	0,190	0,160	0,170	0,160	0,750	0,320	**0,360**
1949	1,170	0,820	0,560	0,330	0,300	0,190	0,180	0,160	0,140	0,150	0,120	0,440	**0,464**
1950	0,310	0,650	0,410	0,210	0,160	0,160	0,150	0,140	0,170	0,210	0,150	0,510	**0,274**
1951	0,860	1,870	0,360	0,360	0,280	0,190	0,160	0,160	0,130	0,130	0,130	0,110	**0,530**
1952	2,000	1,880	0,680	0,370	0,230	0,200	0,180	0,170	0,140	0,160	0,200	0,250	**0,714**
1953	1,210	2,430	0,830	0,400	0,320	0,280	0,230	0,190	0,150	0,187	0,450	0,880	**0,736**
1954	0,380	2,620	0,920	0,480	0,340	0,270	0,240	0,210	0,160	0,160	0,140	1,070	**0,683**
1955	0,310	0,800	0,460	0,250	0,200	0,180	0,180	0,170	0,180	0,150	0,240	0,180	**0,319**
1956	0,320	0,550	0,440	0,250	0,180	0,170	0,150	0,130	0,150	0,150	1,350	0,320	**0,274**
1957	0,700	0,810	1,010	0,530	0,380	0,360	0,290	0,240	0,120	0,570	0,440	0,620	**0,540**
1958	0,730	0,590	0,320	0,320	0,270	0,320	0,210	0,190	0,210	0,160	0,170	0,200	**0,369**
1959	0,390	0,560	1,150	0,380	0,290	0,230	0,220	0,200	0,170	0,200	0,290	0,380	**0,428**
1960	0,320	0,320	0,320	0,310	0,240	0,210	0,190	0,160	0,180	0,170	0,120	0,220	**0,259**

Cuadro A2: Comparación de la serie de caudales del Río El Tala transformada por Logaritmo Natural y los valores obtenidos por el modelo ARIMA para los 100 primeros casos.

CASO	MES	SERIE TRANSFORMADA	SERIE ARIMA	CASO	MES	SERIE TRANSFORMADA	SERIE ARIMA
1	1	-1,897119985	.	51	3	-0,415515444	-0,449858614
2	2	-0,820980552	.	52	4	-0,867500568	-0,879578867
3	3	-1,347073648	.	53	5	-1,049822124	-1,274507177
4	4	-1,427116356	.	54	6	-1,237874356	-1,371939874
5	5	-1,427116356	.	55	7	-1,30933332	-1,465697426
6	6	-1,660731207	.	56	8	-1,427116356	-1,553661778
7	7	-1,771956842	.	57	9	-1,609437912	-1,639732929
8	8	-1,832581464	.	58	10	-1,609437912	-1,645788207
9	9	-1,897119985	.	59	11	-1,897119985	-1,435215593
10	10	-1,966112856	.	60	12	-1,832581464	-0,955902601
11	11	-1,714798428	.	61	1	0,530628251	-0,103774383
12	12	-0,314710745	.	62	2	0,457424847	-0,328666265
13	1	-0,733969175	-1,897119985	63	3	0,086177696	-0,218948374
14	2	-0,174353387	-0,302555985	64	4	-0,994252273	-0,518551319
15	3	-0,46203546	-0,907895166	65	5	-1,203972804	-1,115710742
16	4	-1,021651248	-0,99187736	66	6	-1,30933332	-1,406239416
17	5	-1,427116356	-1,130026237	67	7	-1,609437912	-1,472290899
18	6	-1,514127733	-1,601381986	68	8	-1,609437912	-1,615394382
19	7	-1,609437912	-1,7358481	69	9	-1,609437912	-1,73875601
20	8	-1,714798428	-1,774891254	70	10	-2,120263536	-1,636671447
21	9	-1,771956842	-1,869610888	71	11	-1,30933332	-1,845672153
22	10	-1,966112856	-1,965989946	72	12	-1,203972804	-1,210175409
23	11	-1,832581464	-1,807045828	73	1	-0,446287103	0,331743425
24	12	-1,108662625	-0,498226819	74	2	-0,616186139	0,037817277
25	1	0,482426149	-1,375589128	75	3	-0,84397007	-0,779133779
26	2	-0,776528789	0,120686042	76	4	-1,021651248	-1,33859269
27	3	-0,478035801	-0,650131185	77	5	-1,272965676	-1,329323207
28	4	-0,755022584	-1,124877961	78	6	-1,609437912	-1,392150524
29	5	-1,237874356	-1,220133878	79	7	-1,609437912	-1,652579245
30	6	-1,46967597	-1,441213483	80	8	-1,609437912	-1,676535704
31	7	-1,660731207	-1,624178765	81	9	-1,832581464	-1,696894707
32	8	-1,714798428	-1,771790543	82	10	-1,966112856	-1,897996708
33	9	-1,771956842	-1,854888268	83	11	-2,040220829	-1,456604424
34	10	-1,771956842	-2,017821308	84	12	-0,544727175	-1,37221695
35	11	-1,660731207	-1,845907157	85	1	0,418710335	0,129815978
36	12	-0,867500568	-0,872819706	86	2	1,128171091	0,049759128
37	1	-0,223143551	-0,481534714	87	3	1,232560261	0,028737602
38	2	0,165514438	-0,521264022	88	4	0,470003629	-0,081228495
39	3	-0,820980552	-0,40943298	89	5	-0,544727175	-0,344319875
40	4	-1,049822124	-0,90725969	90	6	-0,891598119	-0,937629483
41	5	-1,347073648	-1,347304943	91	7	-1,021651248	-1,194055394
42	6	-1,514127733	-1,55014758	92	8	-0,967584026	-1,273716256
43	7	-1,609437912	-1,68117513	93	9	-1,386294361	-1,416128247
44	8	-1,714798428	-1,721176076	94	10	-1,514127733	-1,643543718
45	9	-1,832581464	-1,786962525	95	11	-1,171182982	-1,673024326
46	10	-1,609437912	-1,874449907	96	12	-0,174353387	-0,808940748
47	11	-1,237874356	-1,614729492	97	1	0	0,777768175
48	12	-0,634878272	-0,489857589	98	2	0,198850859	0,631849473
49	1	0,412109651	-0,171701063	99	3	0,198850859	0,259490802
50	2	0,039220713	0,146846296	100	4	-0,301105093	-0,480802768

NOTA: los datos de las series están transformados por el logaritmo natural, por lo tanto para volver a los datos originales que muestra el Cuadro A1, se debe aplicarse a los datos la inversa de esa función.