

Tarea 2
Modelamiento Estocástico y Simulación
Universidad Técnica Federico Santa María
Departamento de Informática

Daniel San Martín Reyes
<daniel.sanmartinr@sansano.usm.cl>

22 de agosto de 2018

Pregunta 1

En las siguientes preguntas debe analizar los datos mensuales de flujos (cms) de un río. Los datos pueden ser encontrados en el siguiente link (en formato csv). Primero proponga uno (o más) modelos $ARIMA(p, d, q)$ para la serie. Siga la metodología de Box & Jenkins (Identificación del Modelo, Estimación de Parámetros y Diagnóstico). Recuerde explicar cada paso claramente y con gráficos. ¿Que puede concluir?.

El primer paso realizado fue una exploración general de los datos. Primero, se muestra un gráfico de la serie de tiempo (Figura 1) con sus principales estadísticas (Cuadro 1). Luego, obtenemos los gráficos de ACF y $PACF$ de la serie para analizar su comportamiento.

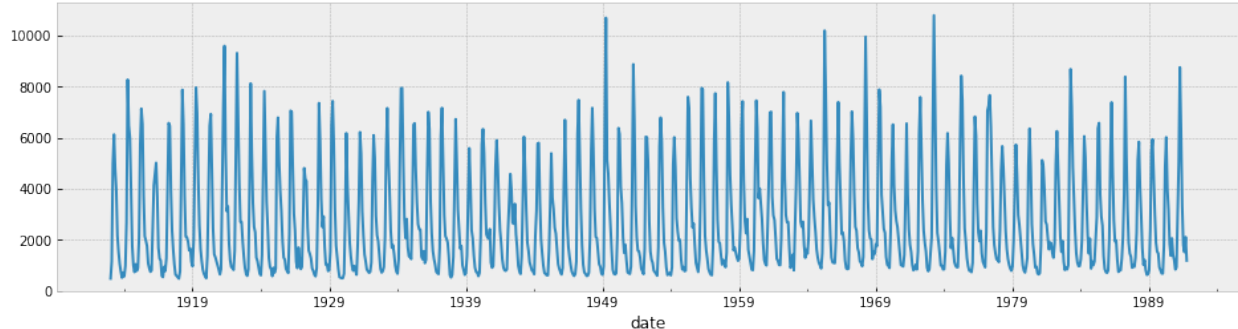


Figura 1: Serie de tiempo.

Muestras	946
Media	2708,57
Desv.	2123,23
Min	482
25 %	1010
50 %	1915
75 %	3947
Max	10800

Cuadro 1: Estadísticas Principales

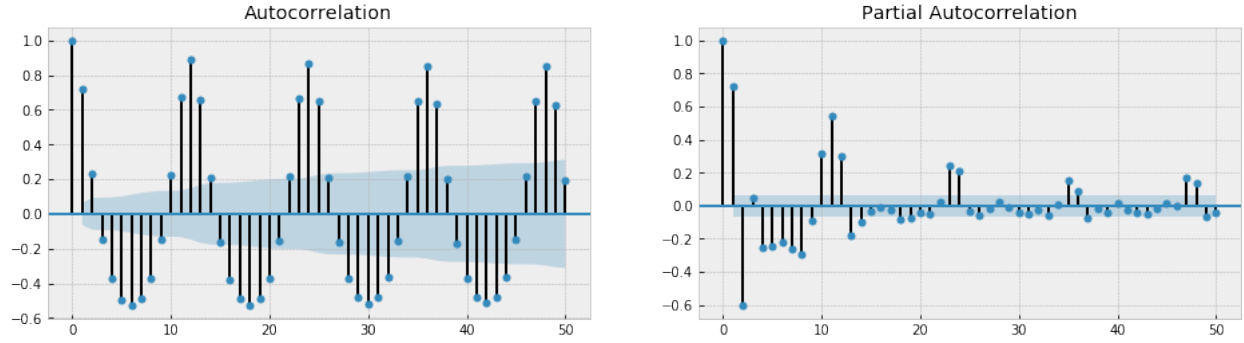


Figura 2: Gráficos de ACF y $PACF$.

De los gráficos de la Figura 2 podemos notar el comportamiento estacional de la serie. Esto no nos permite asegurar la condición de *estacionariedad* para el uso de un modelo $ARIMA$ según la teoría. Para intentar superar este problema realizaremos una descomposición de la serie asumiendo un modelo ingenuo aditivo subyacente. El resultado de la descomposición se presenta en la Figura 3.

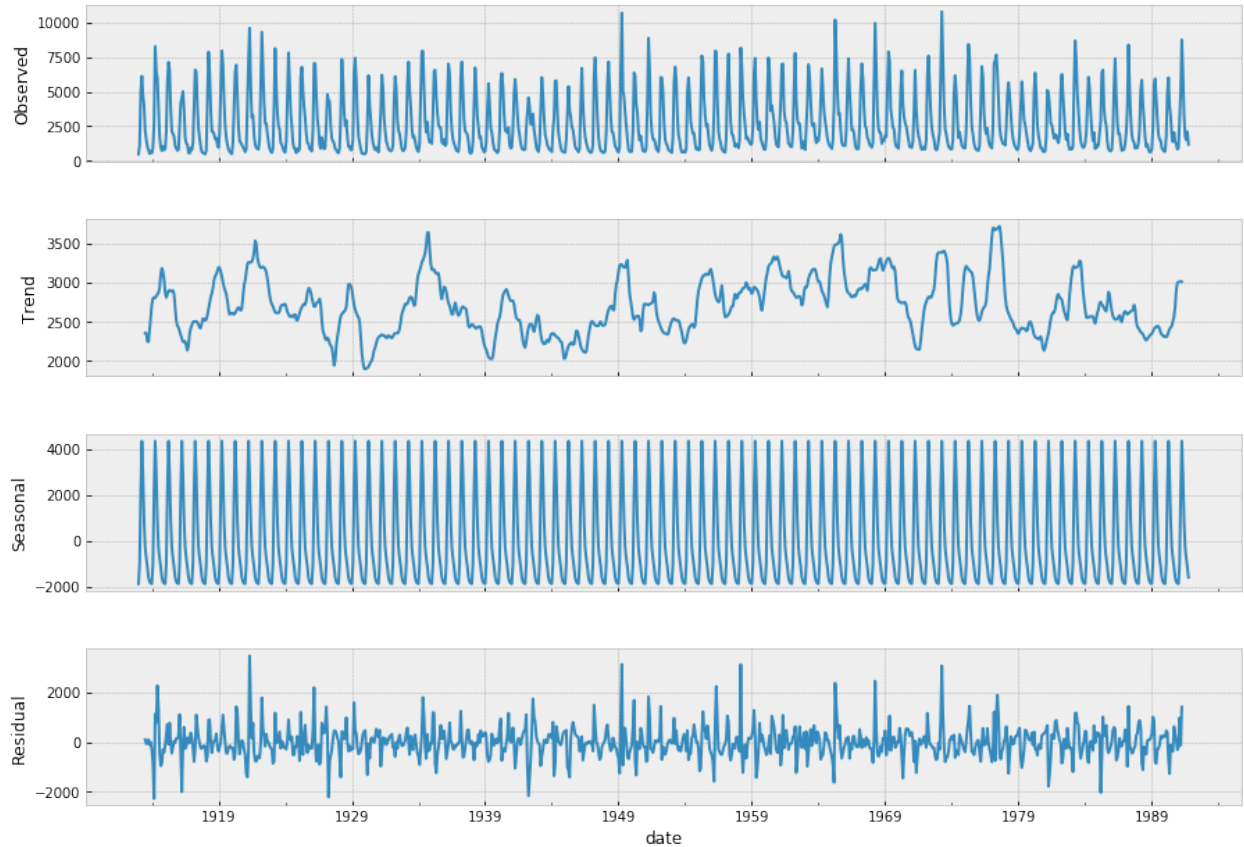


Figura 3: Descomposición de la serie.

Ahora, analizaremos una nueva serie definida por $y_t = x_t - \text{componente estacional}$. Los nuevos gráficos de ACF y $PACF$ se presentan en la Figura 4.

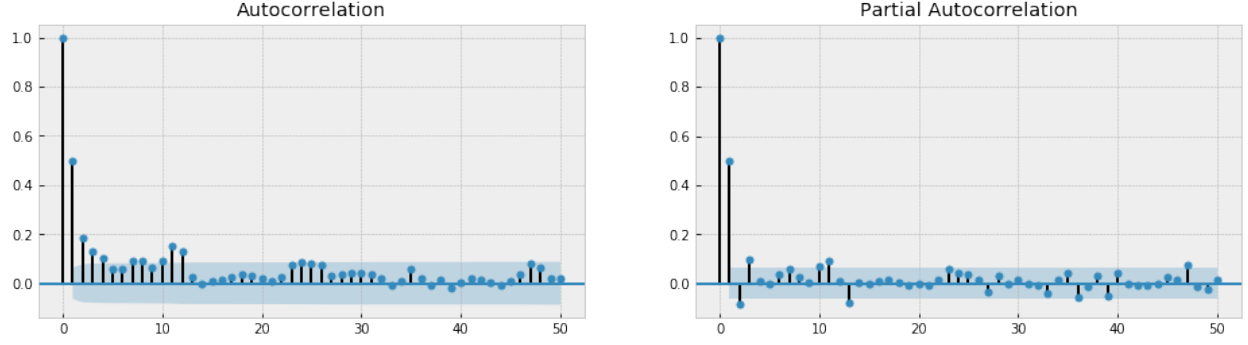


Figura 4: Gráficos de ACF y $PACF$ de la nueva serie y_t .

De la Figura 4, notamos ahora que el ACF no decae lentamente lo que nos permite suponer que la nueva serie es *estacionaria*. En particular, pareciera que el ACF decae de forma exponencial y que el $PACF$, según la interpretación, corta después del *lag* 1 o *lag* 3, por ende los dos primeros modelos propuestos son un $ARIMA(1,0,0)$ y $ARIMA(3,0,0)$. Otra forma de abordar el problema, asumiendo que un modelo $ARIMA$ no se ajustará bien a datos que presentan una fuerte componente estacional, es realizar una búsqueda exhaustiva de parámetros sin aplicar ninguna modificación a la serie. Bajo este enfoque, el siguiente modelo propuesto corresponde a un $ARIMA(3,0,3)$. El resumen de las métricas reportadas en los modelos propuestos se presentan en el Cuadro 2.

p, d, q	AIC	BIC
1, 0, 0	16483,259	16497,816
3, 0, 0	16054,938	16079,200
3, 0, 3	15703,402	15742,220

Cuadro 2: Métricas reportadas para modelos $ARIMA$

Finalmente, luego de la estimación de parámetros es necesaria la validación de los modelos propuestos. De acuerdo a los test de Ljung-Box, de normalidad y análisis gráfico de los residuos se puede concluir que ninguno de los modelos $ARIMA$ propuestos son capaces de describir el comportamiento de los datos. En la Figura 5 se presenta el mejor residuo obtenido. Se aprecia en el correlograma que no hay evidencia suficiente para suponer que el residuo es un ruido, por esto no podemos validar directamente el modelo.

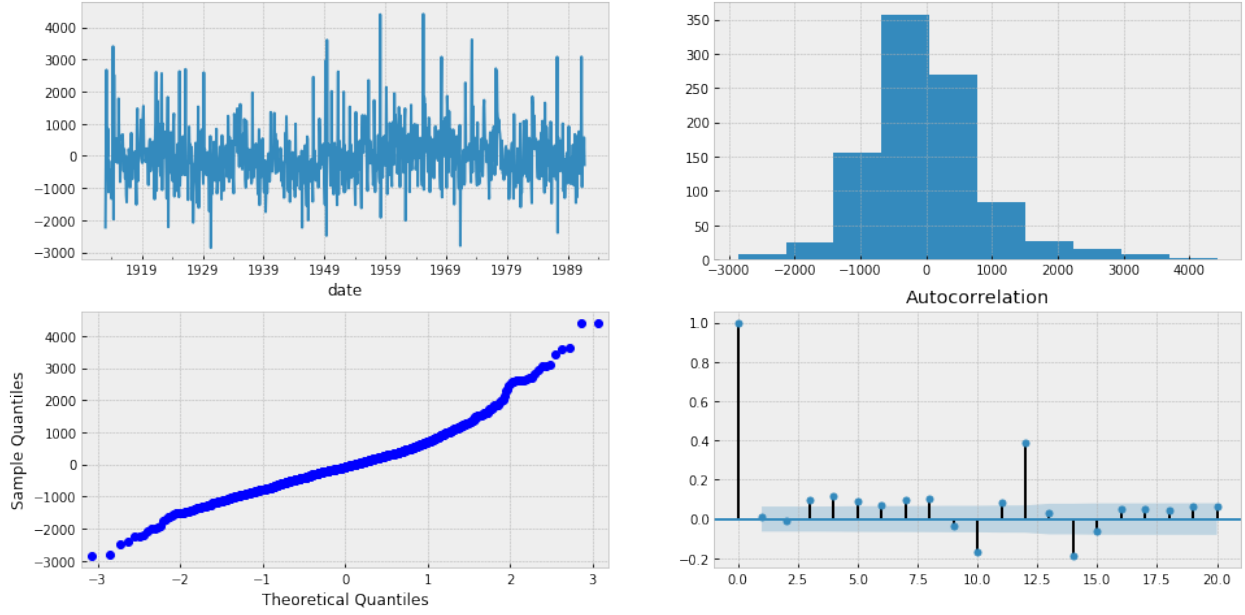


Figura 5: Diagnóstico del mejor resultado de residuo del modelo $ARIMA$.

Pregunta 2

Ahora proponga un modelo $ARIMA(p, d, q) \times (P, D, Q)_s$. Nuevamente explique cada uno de sus pasos cuidadosamente. ¿Que puede concluir?.

Para mejorar los resultados del modelo anterior, se propone utilizar un modelo $SARIMA$ dada la componente estacional presentada en los datos. Lo primero es notar que esta componente se repite de forma anual, y con datos mensuales definimos el valor de $s = 12$. Tendiendo en cuenta esto, debemos diferenciar al menos 1 vez para construir la serie $y_s = \{x_{13} - x_1, x_{25} - x_2, \dots\}$ esto significa que el parámetro $D = 1$.

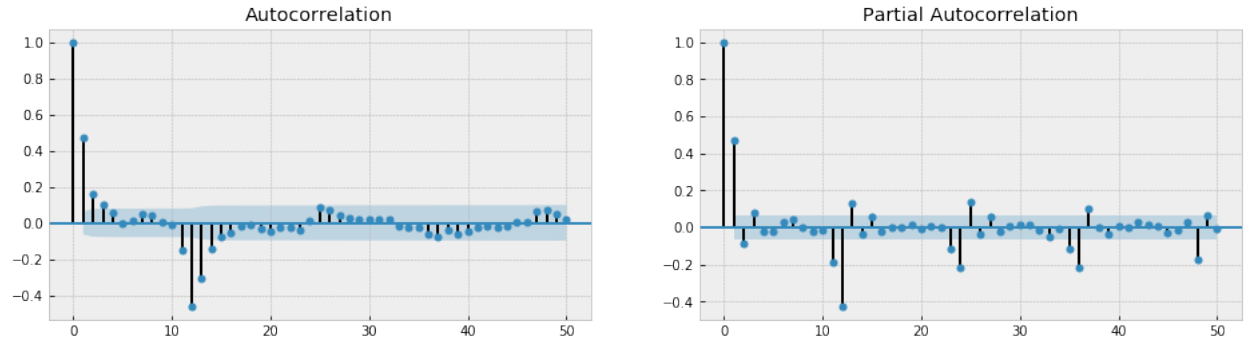


Figura 6: Gráficos de ACF y $PACF$ de la serie y_s .

De la Figura 6 del ACF notamos que la parte estacional se corta en el lag 12, esto es $Q = 2$ y el $PACF$ va disminuyendo cada 12 $lags$ por lo que se puede asumir un MA . Para la parte no estacional vemos un decaimiento exponencial en el ACF mientras que en el $PACF$

se corta del *lag* 3. De esta forma el modelo propuesto es $ARIMA(3, 0, 0) \times (0, 1, 2)_{12}$.

p, d, q, P, D, Q, s	AIC	BIC
3, 0, 0, 0, 1, 2, 12	14315,21	14344,084

Cuadro 3: Métricas reportadas para el modelo *SARIMA*

De acuerdo a las métricas de evaluación notamos una mejora en el modelo en comparación al modelo *ARIMA*. Sobre la validación del modelo, la evaluación del residuo en la Figura 7 nos permite suponer que el residuo se comporta como un ruido blanco gaussiano, por lo que el modelo es válido para explicar el proceso asociado al fenómeno.

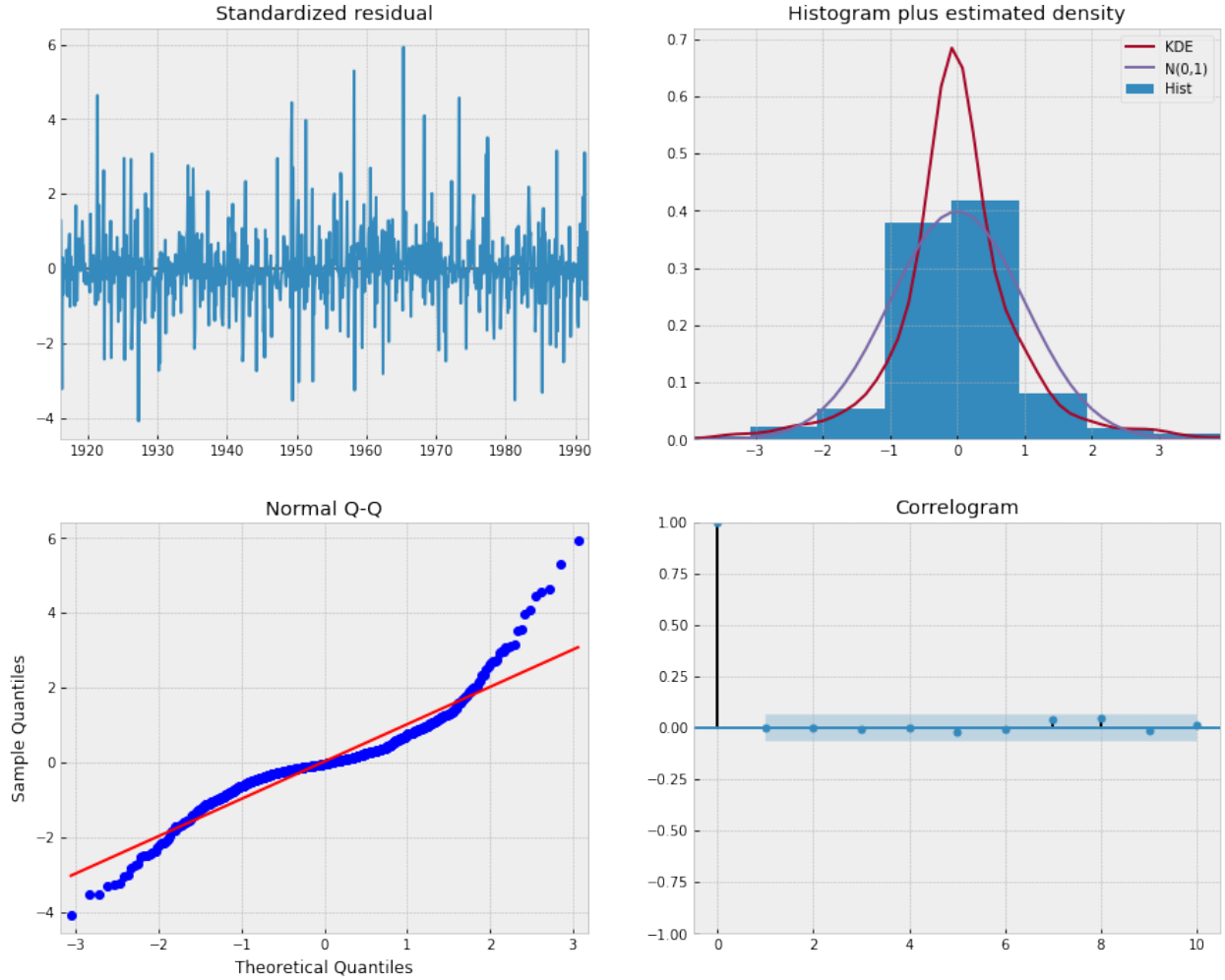


Figura 7: Diagnóstico del residuo del modelo *SARIMA*.

Pregunta 3

Utilizando los modelos entrenados realice predicciones para cada uno de los próximos 6 meses, incluyendo los intervalos de confianza, ¿Que puede concluir?.

En la Figura 8 y Figura 9 se muestra las predicciones de los próximos 6 meses junto al intervalo de confianza respectivo de cada modelo.

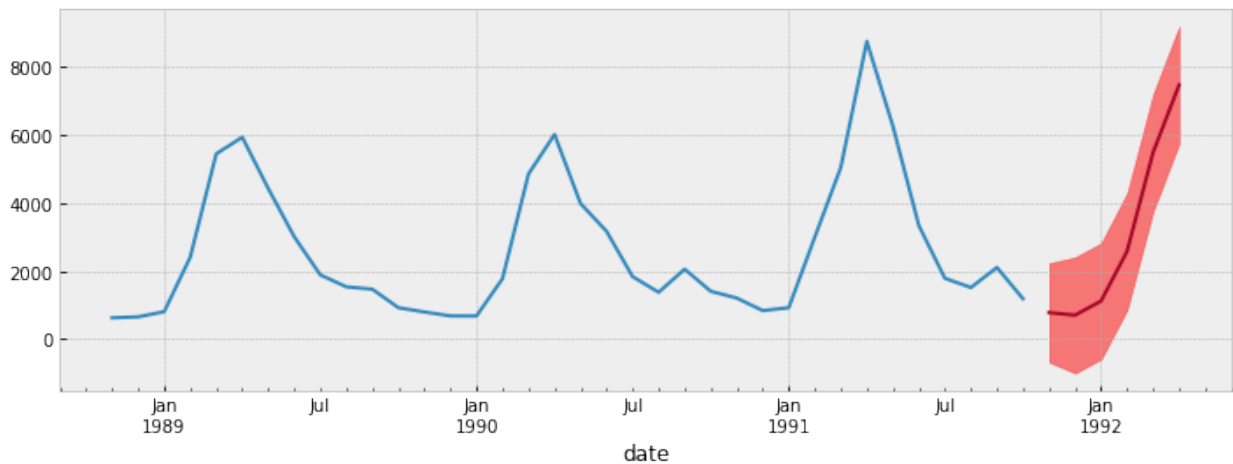


Figura 8: Pronóstico realizado con el mejor modelo *ARIMA* obtenido.

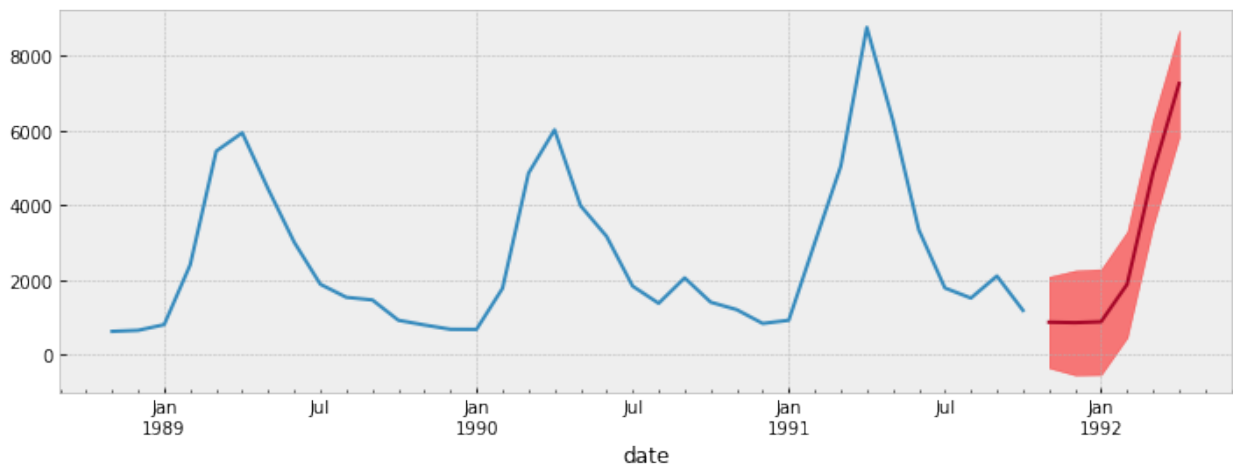


Figura 9: Pronóstico realizado con el modelo *SARIMA*.

Podemos notar que cualitativamente son muy parecidas las predicciones, pero el intervalo de confianza asociado al modelo *SARIMA* nos permite estimar con mayor certeza. Además, no es posible confiar en su totalidad de los resultados del modelo *ARIMA* dado que en la pregunta 1 no pudo ser validado con el análisis de los residuos asociados a este.

Pregunta 4

En esta pregunta utilice uno de los modelos propuestos en las preguntas anteriores. Simule 500 realizaciones de 2 años utilizando el modelo, luego conteste ¿Cuan probable es obtener un flujo mayor que 8000 en alguno de los meses?.

De acuerdo a las simulaciones utilizando el modelo *SARIMA* (Figura 10) no es posible obtener un flujo mayor a 8000 en algunos de los meses dentro de los años simulados.

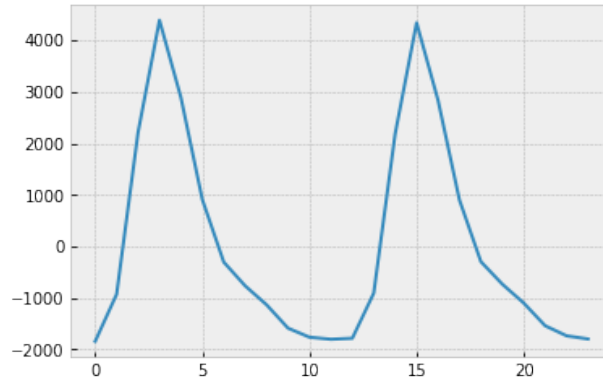


Figura 10: Simulación realizada con el modelo *SARIMA*.

Referencias

- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Shumway, R. H. & Stoffer, D. S. (2005). *Time Series Analysis and Its Applications (Springer Texts in Statistics)*. Berlin, Heidelberg: Springer-Verlag.