

Análisis temporal del caudal del río San Juan

[Ignacio A. Benemérito](#)¹ , [Myriam B. Herrera](#)² , [María Romagnano](#)³ 

¹ Universidad Nacional de San Juan, Argentina, ignacio.benemerito@gmail.com

² Universidad Nacional de San Juan, Argentina, myriamhrrr@gmail.com

³ Universidad Nacional de San Juan, Argentina, maritaroma@iinfo.unsj.edu.ar

Resumen. La predicción del caudal de ríos es fundamental para la gestión de los recursos hídricos en regiones áridas. En este trabajo se analiza la serie temporal del caudal medio mensual del río San Juan mediante modelos SARIMA. La metodología comprende la transformación Box-Cox, el análisis de estacionariedad mediante las funciones ACF y PACF y las pruebas ADF y KPSS, la selección de modelos utilizando AIC y validación cruzada, y el análisis de residuos. Los resultados muestran que el modelo SARIMA $(2,0,1)(0,1,1)_{12}$ obtuvo el mejor desempeño predictivo, superando tanto al modelo seleccionado mediante AIC como al modelo de referencia Seasonal Naïve. Estos resultados respaldan el uso de modelos SARIMA para el pronóstico del caudal del río San Juan.

Palabras clave: Series temporales, SARIMA, Predicción de caudales, Recursos hídricos.

Time analysis of the San Juan River streamflow

Abstract. River streamflow forecasting is essential for water resource management in arid regions. This work analyzes the monthly streamflow series of the San Juan River using SARIMA models. The methodology includes a Box-Cox transformation, stationarity assessment through ACF, PACF, ADF, and KPSS, model selection using AIC and cross-validation, and residual analysis. Results show that the SARIMA $(2,0,1)(0,1,1)_{12}$ model achieved the best predictive performance, outperforming both the AIC-selected model and a Seasonal Naïve benchmark. These findings support the use of SARIMA models for streamflow forecasting in the San Juan River.

Keywords: Time series, SARIMA, Streamflow forecasting, Water resources.

1 Introducción

En la provincia de San Juan, el río homónimo constituye la principal fuente hídrica para el riego agrícola, el abastecimiento poblacional y diversas actividades productivas. Debido a la elevada variabilidad temporal del caudal, su predicción resulta fundamental para la gestión del recurso. En este trabajo se propone un modelo basado en series temporales para generar predicciones del caudal del río.

2 Antecedentes

La predicción de caudales de ríos ha sido ampliamente abordada mediante modelos de series temporales, siendo ARIMA y su extensión estacional SARIMA algunas de las técnicas más utilizadas debido a su capacidad para modelar dependencia temporal y patrones periódicos.

Diversos estudios han demostrado la eficacia de los modelos SARIMA en series hidrológicas. Por ejemplo, en Tadesse et al. (2017) aplicaron un modelo SARIMA para predecir el caudal mensual del río Waterval, obteniendo un buen ajuste y predicciones consistentes a partir de datos históricos. En la misma línea, en Asare (2025) se emplea este tipo de modelos para el análisis de series de caudal del río Missouri, destacando su capacidad para capturar la estacionalidad y mejorar la precisión predictiva en contextos hidrológicos.

Estos trabajos muestran que los modelos SARIMA permiten representar adecuadamente la estacionalidad y la autocorrelación de las series de caudal, con un número reducido de parámetros y sin requerir variables exógenas. Por ello, continúan siendo una herramienta relevante en aplicaciones hidrológicas, especialmente cuando se busca un equilibrio entre precisión e interpretabilidad.

3 Experimentación

3.1 Datos

Se utilizó una serie temporal del caudal medio mensual del río San Juan, medido en metros cúbicos por segundo (m^3/s), desde julio de 1971 hasta julio de 2024.

Esta serie presenta un comportamiento altamente variable, con la presencia de patrones estacionales y una tendencia general, factores a tener en cuenta para su modelado y predicción, como se muestra en la Fig. 1.

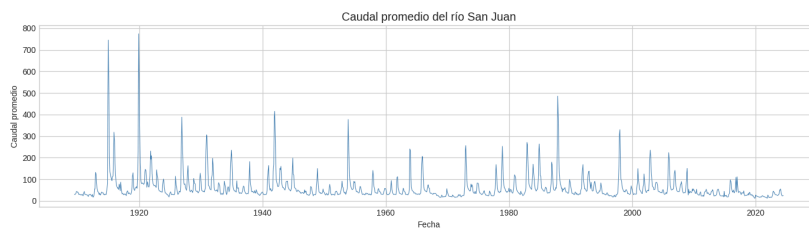


Fig. 1. Serie temporal del caudal medio mensual del río San Juan.

3.2 Modelos

Como se mencionó en los antecedentes, para la predicción del caudal del río se decidió trabajar con modelos clásicos de series temporales, en particular con la familia SARIMA (Seasonal Autoregressive Integrated Moving Average), debido a su capacidad para representar series que presentan tendencia y estacionalidad.

3.3 Metodología

Los modelos fueron implementados en Python utilizando la biblioteca pmdarima, considerando tanto enfoques automáticos como configuraciones definidas manualmente.

En primer lugar, se estabilizó la varianza mediante una transformación de Box-Cox ($\lambda=-0,54$). Posteriormente se analizó la serie mediante las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), las cuales se muestran en la Fig. 2, y se realizaron pruebas de ADF y KPSS a fin de determinar la estacionariedad de la serie.

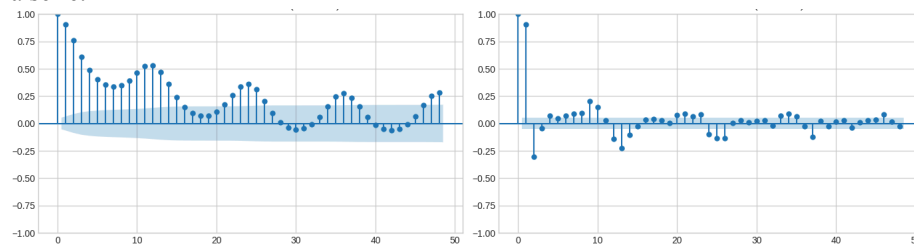


Fig. 2. ACF y PACF de la serie transformada.

La ACF presenta una oscilación de período 12, mientras que las pruebas de ADF y KPSS rechazaron ambas sus hipótesis nulas con un nivel de significancia $\alpha=0,05$; siendo resultados contradictorios con respecto a la estacionariedad de la serie, por lo que se decidió aplicar el operador diferencia de orden 12 y evaluar el comportamiento de la serie nuevamente con la ACF y la PACF (Fig. 3), y las pruebas de ADF y KPSS.

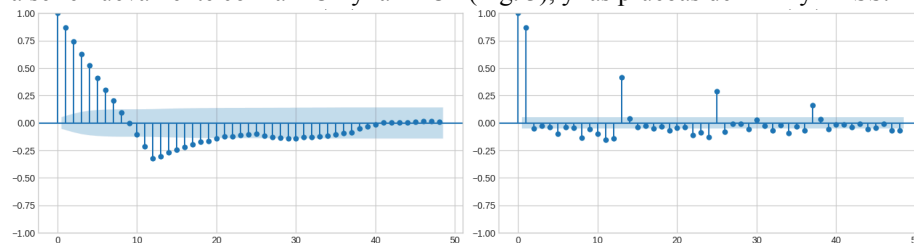


Fig. 3. ACF y PACF de la serie diferenciada.

Tras la diferenciación de la serie, la prueba de ADF rechazó su hipótesis nula, mientras que la prueba de KPSS no la rechazó, concluyendo que la serie es estacionaria y no se necesitan más diferenciaciones.

Tras esto, se realizó la búsqueda automática de un modelo utilizando el Criterio de Información de Akaike (AIC) y, debido al interés en pronosticar el caudal de un año siguiente, la búsqueda de los modelos con el menor MAE y RMSE al predecir los 12 meses posteriores en validación cruzada (reservando el 20% final de los datos para realizar las diferentes evaluaciones). En ambos casos se acotó la búsqueda a modelos de la forma SARIMA(p, 0, q) (P, 1, Q)₁₂, de acuerdo a lo observado en las funciones de autocorrelación y las pruebas ADF y KPSS.

Además, se utilizó un modelo de referencia (Seasonal Naïve) para comparar el rendimiento de los modelos SARIMA.

Por último, se realizó un análisis de los residuos a fin de determinar si el modelo capturó adecuadamente la dependencia temporal de los datos.

4 Resultados

A partir de la búsqueda automática con AIC se obtuvo un modelo SARIMA (1, 0, 0) (3, 1, 0)₁₂, mientras que con validación cruzada el modelo SARIMA (2, 0, 1) (0, 1, 1)₁₂ tuvo el menor MAE y RMSE.

En la Tabla 1 se presenta el AIC, el RMSE y el MAE de ambos modelos, junto con el RMSE y MAE del modelo de referencia.

Tabla 1. Comparación de modelos

Modelo	MAE en m ³ /s	RMSE en m ³ /s	AIC
Seasonal Naïve	19,73	33,46	-
SARIMA(1, 0, 0) (3, 1, 0) ₁₂	19,03	35,63	-6064,86
SARIMA (2, 0, 1) (0, 1, 1) ₁₂	14,45	24,72	-6056,87

Debido a que el modelo SARIMA (1, 0, 0) (3, 1, 0)₁₂ no presenta una mejora significativa con respecto al modelo de referencia, se decidió continuar con el modelo SARIMA (2, 0, 1) (0, 1, 1)₁₂, cuyo análisis de residuos se muestra en la Fig. 4.

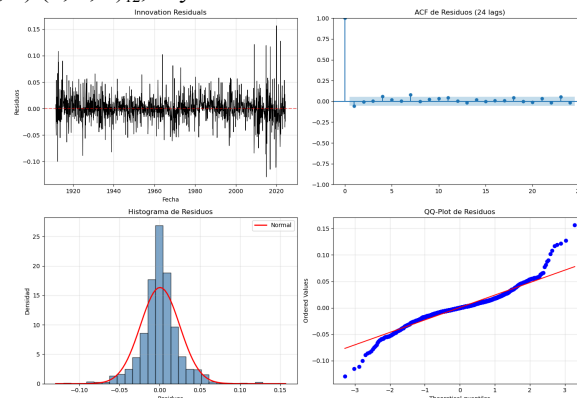


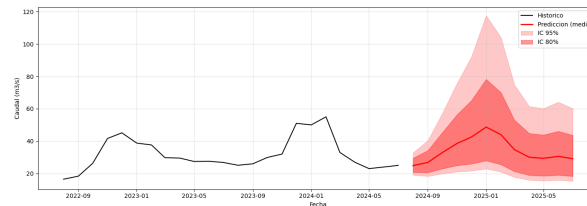
Fig. 4. Análisis de residuos de SARIMA (2, 0, 1) (0, 1, 1)₁₂.

Como muestra la ACF, los residuos no presentan evidencia de autocorrelación. Asimismo, la prueba de Ljung-Box, aplicada con un retardo máximo de 24 siguiendo la recomendación de

Hyndman et al. (2025), obtuvo un p-valor de 0,098. Por lo tanto, con un nivel de significancia de 0,05, no se rechazó la hipótesis nula de ausencia de autocorrelación.

Sin embargo, al observar el histograma de residuos y el gráfico Q-Q no se observa que dichos residuos estén normalmente distribuidos, por lo que no se puede garantizar que los intervalos de predicción obtenidos al revertir la transformación de Box-Cox sean correctos.

Por último, a modo de ilustración, en la Fig. 5 se muestran las predicciones de este modelo para el año siguiente a la totalidad de los datos junto a sus intervalos de predicción.



5 Conclusiones

Los resultados muestran que los modelos SARIMA constituyen un enfoque adecuado para la predicción del caudal del río San Juan. En particular, el modelo seleccionado mediante validación cruzada superó tanto al obtenido mediante AIC como al modelo de referencia Seasonal Naïve, evidenciando la importancia de evaluar los modelos con métricas acordes al objetivo de predicción. Estas predicciones resultan de utilidad para la planificación y gestión del recurso hídrico, especialmente en el sector agrícola.

Como líneas de trabajo futuro, se propone extender el modelo hacia enfoques SARIMAX que incorporen variables exógenas potencialmente relevantes para la predicción del caudal, tales como indicadores de acumulación nival y el Índice Oceánico de El Niño (ONI). Asimismo, se plantea el desarrollo de una aplicación web que permita la visualización y consulta de las predicciones de manera sencilla y accesible, con el objetivo de facilitar su utilización por parte de productores agrícolas, organismos de gestión del recurso hídrico y demás actores involucrados en la planificación y toma de decisiones.

6 Referencias

- Asare, B. (2025). Forecasting Missouri river flow with SARIMA models: A data-driven framework for adaptive water resource management. *International Journal of Statistics and Applied Mathematics*, 10(7), 147-159. <https://doi.org/10.22271/math.2025.v10.i8b.2138>.
- Hyndman, R.J., Athanasopoulos, G., Garza, A., Challu, C., Mergenthaler, M. y Olivares, K.G. (2025). *Forecasting: Principles and Practice, the Pythonic Way*. OTexts. <https://otexts.com/fpppy>.
- Tadesse, K. B., Dinka, M. O. (2017). Application of SARIMA model to forecasting monthly flows in Waterval River, South Africa. *Journal of Water and Land Development*, (35), 229-236. <https://doi.org/10.1515/jwld-2017-0088>.