

# REDES NEURONALES ARTIFICIALES PARA ESTIMACIÓN DE DATOS FALTANTES EN SERIES DE TIEMPO HIDROCLIMÁTICAS

David De León Perez<sup>1\*</sup>, Nicolás Cortés-Torres<sup>1\*</sup>, Efraín Domínguez Calle<sup>2\*</sup> y Félix Francés<sup>1x</sup>

<sup>1</sup>Universitat Politècnica de València - España, <sup>2</sup>Pontificia Universidad Javeriana – Colombia, \*Colombia, <sup>x</sup>España.

[ddeleo1@doctor.upv.es](mailto:ddeleo1@doctor.upv.es), [ncortor@doctor.upv.es](mailto:ncortor@doctor.upv.es), [ffrances@upv.es](mailto:ffrances@upv.es)

## Introducción

Las series de tiempo hidroclimáticas (ST) son fundamentales para diversos análisis, sin embargo, frecuentemente presentan datos faltantes debido a factores como fallas en los equipos de medición, errores de transcripción o falta de registro oportuno (WMO, 1994). Para cumplir con los criterios de continuidad, homogeneidad y longitud establecidos por la Organización Meteorológica Mundial (WMO), es necesario estimar estos valores ausentes. La complejidad de esta tarea varía según la variable hidroclimática analizada, por ejemplo, la precipitación exhibe una distribución espacial irregular debido a los núcleos de lluvia, lo que dificulta la estimación de datos faltantes en estas series (De León & Domínguez, 2021). Por lo tanto, se requiere emplear técnicas robustas que permitan inferir de manera precisa los valores faltantes en ST hidroclimáticas para garantizar la integridad y calidad de los análisis posteriores (Zhang & Post, 2018).

## Materiales y métodos

Se recopilaron ST mensuales de 8 variables hidroclimáticas (Brillo Solar, Evaporación, Humedad Relativa, Punto de Rocío, Precipitación, Caudal, Temperatura y Tensión de Vapor) de la base de datos del IDEAM (Colombia), totalizando 9.268 ST. La complementación de las ST se desarrolló mediante una RNA del tipo Perceptrón Multicapa. Los predictores se seleccionaron realizando una correlación cruzada con estaciones dentro de un búfer de 250 km (Excepto para precipitación que debido a su gran densidad se usó un buffer de 150 km), con 25 rezagos (-12 hasta 12 meses). Para el entrenamiento, ajuste y evaluación de precisión de la RNA, se empleó el módulo MLPRegressor de la librería SciKit-Learn de Python (Rossum, 1995). Una de las métricas clave de esta librería es el Coeficiente de Determinación, el cual provee una estimación cuantitativa del grado de ajuste alcanzado por la RNA a los datos de entrenamiento, permitiendo evaluar su desempeño y precisión.

Se aplicó adicional un Análisis de Componentes Principales (ACP) al 85% de la varianza para reducir la dimensionalidad. Con los nuevos predictores del ACP, se construyó el conjunto de entrenamiento para la RNA y posteriormente se predijeron los

valores faltantes. El flujo de trabajo se describe en la Figura 1.

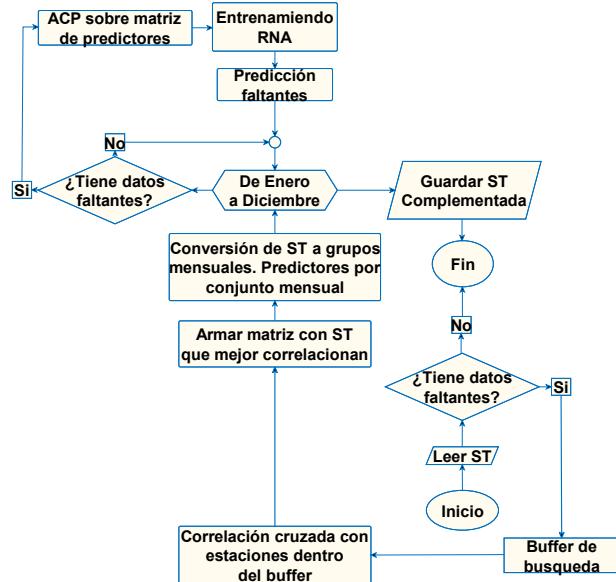


Figura 1. Diagrama de flujo de la metodología planteada

## Resultados y discusión

Los resultados obtenidos del entrenamiento de la RNA para la estimación de datos faltantes en ST hidroclimáticas son prometedores. El desempeño del modelo se evaluó mediante el margen de error medio alcanzado para cada variable, observándose valores aceptables en general.

Para la variable Brillo Solar, se logró un margen de error del 9.87%, indicando una capacidad notable de la RNA para capturar los patrones subyacentes en esta serie temporal. Un resultado similar se obtuvo para la Evaporación, con un 10.89% de error, demostrando la habilidad del enfoque para modelar adecuadamente las complejas interacciones que gobiernan esta variable.

Las variables Humedad Relativa, Punto de Rocío y Tensión de Vapor presentaron márgenes de error más elevados, (13.65%, 13.08% y 15.17% respectivamente). Aunque aceptables, sugieren una mayor dificultad de la RNA para capturar las sutilezas y no linealidades intrínsecas a estas variables atmosféricas relacionadas con la humedad.

En las Figuras 2 y 3 se presenta el límite donde se ubica el 70% de las estaciones evaluadas, para la precipitación fue del 16% mientras que para el caudal fue del 20% el error límite.

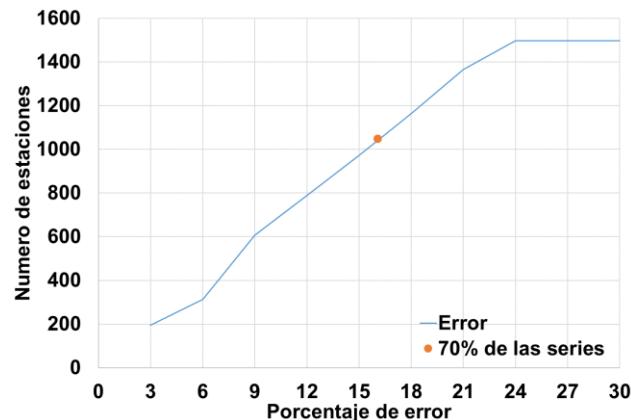


Figura 2. Evaluación del error en las series de precipitación evaluadas

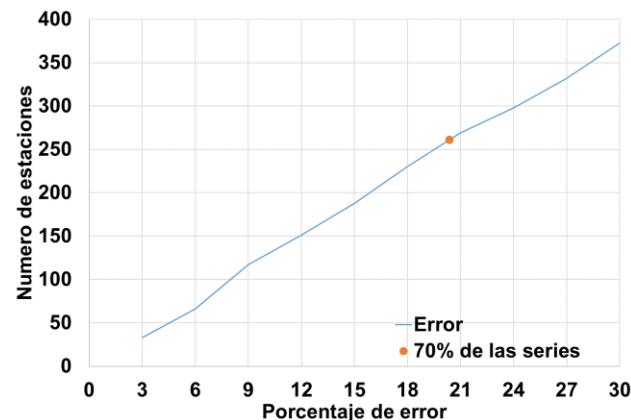


Figura 3. Evaluación del error en las series de caudal evaluadas

En cuanto a la Precipitación, variable de gran importancia hidrológica, el modelo exhibió un prometedor 12% de error. Resultado destacable, considerando los patrones altamente variables y aleatorios que caracterizan a los eventos de precipitación conociendo que trabajos previos (Fagandini et. al. 2024) han reportado la dificultad en la estimación precisa de datos faltantes para esta variable utilizando técnicas tradicionales.

Para el Caudal, variable clave en aplicaciones hidráulicas, el error fue 15.4%, ligeramente más elevado que para otras variables. Este hallazgo sugiere posibles complicaciones al representar las intrincadas relaciones lluvia-escorrentía y otras variables que influyen en el flujo de los cauces (Dembélé et. Al. 2019). No obstante, el resultado es alentador y deja margen para mejoras futuras. Finalmente, la Temperatura presentó el menor error (3.66%), indicando gran precisión de la RNA.

## Conclusiones

En conjunto, estos resultados demuestran la

capacidad general de las Redes Neuronales Artificiales para estimar datos faltantes en series hidroclimáticas complejas con niveles razonables de precisión, a excepción de las variables dependientes de la humedad atmosférica, que parece representar un desafío particular para este enfoque. Aunque hay margen para mejorar aún más el desempeño, el enfoque propuesto supera a muchos métodos convencionales, particularmente para variables difíciles de modelar como la precipitación y el caudal.

Futuras investigaciones podrían explorar diferentes arquitecturas y algoritmos de entrenamiento de RNA para optimizar los resultados, con especial énfasis en las variables problemáticas. Además, la incorporación de información adicional, como datos espaciales o variables exógenas relevantes (correlación con diferentes variables), podría mejorar la representación de los procesos físicos subyacentes y conducir a estimaciones aún más precisas.

## Agradecimientos

Investigación financiada por la Generalitat Valenciana con el proyecto Water4Cast (PROMETEO/2021/074), el Ministerio de Ciencia e Innovación de España con el proyecto TETISPREDIT (PID2022-141631OB-I00), el Ministerio de Ciencia, Tecnología e Innovación de Colombia (MINCIENCIAS) a través de la Convocatoria de Doctorados en el Exterior 885-2 y COLCIENCIAS con la convocatoria 672 de 2014. En póstumo a Efraín Domínguez Calle (2021), director de la tesis de Maestría (donde inició de este estudio) de David De León.

## Referencias

- De León Pérez, D.; Domínguez, E. (2021). *Determinación de áreas hidroclimáticamente homogéneas. Una propuesta técnica*. Ingeniería del agua. 25(2):97-114. <https://doi.org/10.4995/ia.2021.14659>
- Dembélé M., Oriani F., Tumbulto J., Mariéthoz G., Schaeffli D. 2019. *Gap-filling of daily streamflow time series using Direct Sampling in various hydroclimatic settings*. Journal of Hydrology, 569-2019, p.p. 573-586, ISSN 0022-1694, <https://doi.org/10.1016/j.jhydrol.2018.11.076>.
- Fagandini, C., Todaro, V., Tanda, M.G. Pereira, J., Azevedo, L., Zanini, A. (2024). Missing Rainfall Daily Data: A Comparison Among Gap-Filling Approaches. Math Geosci 56, 191–217. <https://doi.org/10.1007/s11004-023-10078-6>
- Rossum, G. (1995). *Python reference manual. Technical report*, CWI (Centre for Mathematics and Computer Science), Amsterdam, The Netherlands, The Netherlands. 30, 52, 53.
- WMO (1994). Guía de prácticas hidrológicas: adquisición y proceso de datos, análisis, predicción y otras aplicaciones. Organización Meteorológica Mundial. 20, 29
- Zhang, Y. & Post, D. (2018). *How good are hydrological models for gap-filling streamflow data?*, Hydrol. Earth Syst. Sci., 22, 4593–4604, <https://doi.org/10.5194/hess-22-4593-2018>.