24.- Aguas

Extensión y Validación de un Marco de Imputación de Datos Ambientales en la Cuenca del Río Santa Lucía, Uruguay para la Gestión Integrada de los Recursos Hídricos

Autor: Pou, Martina; mpou@fing.edu.uy

Co-autor(es): Pertusso, Pedro; Vilaseca, Federico

Orientador/a: Gorgoglione, Angela; agorgoglione@fing.edu.uy; Castro, Alberto; acastro@fing.edu.uy

Universidad de la República / Facultad de Ingeniería / Instituto de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA)

Resumen

Comprender los procesos naturales a escala de cuenca es crucial para la gestión eficaz de los recursos hídricos y el desarrollo sostenible. Sin embargo, el monitoreo ambiental, especialmente de las variables de calidad del agua, enfrenta desafíos debido a la variabilidad de los parámetros y la limitada disponibilidad de recursos, lo que dificulta una modelación integrada precisa. En nuestro trabajo anterior, desarrollamos un marco innovador para la imputación de datos faltantes meteorológicos, hidrométricos y de calidad del agua, integrando modelos basados en datos con restricciones físicas. Cada una de estas restricciones se implementó como una variable artificial que se añadió al *dataset* original, creando así una nueva variante. Este marco mostró resultados satisfactorios en la cuenca del río Santa Lucía Chico (Uruguay). El objetivo de este trabajo es extender este marco a toda la cuenca del río Santa Lucía, validando su robustez y efectividad a mayor escala. Los resultados demuestran un sólido desempeño del marco en diversos dominios ambientales, con más del 68% de los datos imputados con un NSE>0,45 (resultados satisfactorios). El *Huber Regressor* fue la técnica de imputación más exitosa, especialmente para las variables de precipitación y calidad del agua, seguido por el *Support Vector Regressor* para calidad del agua y *K-Nearest Neighbours Regressor* para variables hidrométricas. El 96,31% de las veces se identificó una variante o combinación de variantes como el mejor modelo-*dataset*, destacando la importancia de las restricciones físicas. Estos hallazgos subrayan la utilidad y la robustez del marco propuesto para la gestión integrada de recursos hídricos a escala de cuenca.

Palabras clave: Imputación de datos; Aprendizaje automático; Hidroinformática

Introducción

Comprender y estudiar los procesos naturales a escala de cuenca es fundamental para una gestión eficaz de los recursos hídricos y un desarrollo sostenible. Las cuencas son unidades fundamentales para el desarrollo de estudios hidrológicos y ecológicos, abarcando diversos procesos que se encuentran interconectados como el flujo de agua, el transporte de sedimentos y el ciclo de los nutrientes (Gorgoglione, y otros, 2020a). Estos procesos son vitales para mantener la salud de los ecosistemas, apoyar la biodiversidad y proporcionar recursos para actividades antrópicas. Conocerlos es esencial para predecir y mitigar los impactos de los cambios ambientales, tales como el cambio climático, las alteraciones en el uso de suelo y la contaminación (Gorgoglione, Castro, Chreties, & Etcheverry, 2020b). Sin embargo, el monitoreo ambiental, particularmente de la calidad del agua, enfrenta importantes desafíos y problemas. La gran variabilidad de los parámetros requiere una recopilación de datos exhaustiva que frecuentemente se dificulta debido a los recursos limitados (Sattari, Rezazadeh-Joudi, & Kusiak, 2017). Esto resulta en series temporales con altas cantidades de valores faltantes que dificulta el entendimiento y la modelación de los procesos.

Para atender a este problema, en la última década, ha habido un aumento notorio en el uso de modelos integrados para gestionar la calidad del agua a nivel de cuenca (Freni, Mannina, & Viviani, 2011). Estos modelos, capaces de simular interacciones entre sistemas físicos diversos como la atmósfera, suelo y cuerpos de agua, requieren datos sustanciales para su precisión. Diversas técnicas, desde imputaciones estadísticas hasta métodos de aprendizaje automático, se han explorado para abordar este problema (Chen, Xu, Jiang, & Yu, 2021). Las técnicas de aprendizaje automático supervisado pueden representar de manera efectiva las relaciones no lineales entre variables medidas en estaciones espacialmente distribuidas (Chivers, y otros, 2020). La imputación de datos ambientales de dominios interconectados, como meteorología, hidrología y calidad de agua, es esencial para mejorar la precisión de los modelos integrados y comprender la dinámica de la calidad del agua a nivel de cuenca.

En nuestro trabajo anterior (Pastorini, Rodríguez, Etcheverry, Castro, & Gorgoglione, 2024) , desarrollamos un marco (*framework*) innovador que aborda

eficazmente el desafío de imputar datos faltantes en diversos dominios ambientales (meteorología, hidrología y calidad del agua). Este marco integra modelos basados en datos con conocimientos físicos, lo que conduce a resultados satisfactorios en la imputación. En Pastorini et al. (2024), el marco se desarrolló para la cuenca del río Santa Lucía Chico, Uruguay.

Objetivos

El objetivo de este trabajo es extender el marco de imputación de datos ambientales a toda la cuenca del río Santa Lucía. Esta extensión es crucial para validar la robustez y efectividad del marco a una escala mayor, demostrando su capacidad de generalización.

Al aplicar el marco a un conjunto más amplio y diverso de condiciones ambientales dentro de toda la cuenca, buscamos confirmar su aplicabilidad y fiabilidad en diferentes subregiones y escenarios, mejorando así su utilidad para la gestión integrada de los recursos hídricos.

Materiales y Métodos

1- Área de estudio

El área de estudio es la cuenca del río Santa Lucía (Figura 1). Tiene una superficie de $13376\ km^2$ y se encuentra en la zona sur de Uruguay. El punto de cierre seleccionado se ubica sobre el cauce principal (río homónimo), en el punto más próximo a la estación de calidad de agua más cercana a la desembocadura en el río de la Plata (XCOL010). Es la cuenca con mayor importancia del país pues de ella se abastece de agua potable aproximadamente la mitad de la población. En la zona inferior se encuentran los Humedales del Santa Lucía, pertenecientes al Sistema Nacional de Áreas Protegidas desde febrero de 2015 (MVOTMA, 2015).

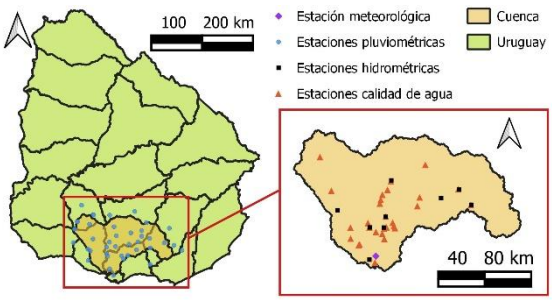

Figura 1: Zona de estudio y estaciones de medición

2- Datos disponibles

Los datos utilizados pueden agruparse en 3 grupos:

i) Hidrológicos: caudal (*Q*) $[m^3/s]$ y nivel (*h*) $[m]$ con datos diarios medidos en 9 estaciones desde 1980 hasta 2023.

ii) Meteorológicos: diarios de precipitación (*P*) $[mm]$ en 40 pluviómetros desde 1980 hasta 2023, temperatura del aire media (TA_{med}) [*°C (24h)*], temperatura del aire máxima (TA_{max}) [*°C*], temperatura del aire mínima (TA_{min})

[*ºC*], humedad relativa (*HR*) [*%*], radiación solar (*RS*) [W/m^2], heliofanía (*Hel*) [*hs*], evapotranspiración Penman (*ET*) [mm] y velocidad del viento (*VV*) [*2m/km/24h*] en una estación meteorológica ubicada cerca de la desembocadura con datos de 1980 a 2023 y muy pocos datos faltantes.

iii) Calidad del agua: se consideraron 25 estaciones desde 2011 hasta 2022 con datos mensuales, en las cuales se mide fósforo total (*FT*) [$\mu g/L\ P$], nitrógeno total (*NT*) [*mg/L N*], ion nitrato (NO_3^-) [*mg/L*], ion nitrito (NO_2^-) [*mg/L*], ion amonio (NH_4^+) [*mg/L*], ion fosfato (PO_4^{3-}) [$\mu g/L$], glifosato (*Glifosato*) [$\mu g/L$], solidos totales (*ST*) [*mg/L*], solidos suspendidos totales (SST) [mg/L], turbidez (*Turbidez*) [*NTU*], temperatura (*T*) [*ºC*], oxígeno disuelto (*OD*) [*mg/L*], demanda bioquímica de oxígeno (*DBO*) [*mg/L*], clorofila-a (*Clo-A*) [$\mu g/L$], potencial de hidrógeno (*pH*) y conductividad (*Cond*) [$\mu S/cm$].

Esto resulta en 429 pares variable – estación, de los cuales 19 no contaban con suficientes datos (<10) y se descartaron del proceso de imputación.

3- Metodología de imputación

Para la imputación de datos faltantes se utilizó una metodología que incluye cuatro fases (Figura 2).

En la *fase 1* se aplican restricciones físicas que consideran la variabilidad temporal y espacial, la correlación y los rangos de variación de las variables para evitar imputaciones irreales. Se utilizaron matrices de correlación de Pearson, Spearman y Kendall con límite $|0{,}5|$.

Hay dependencia positiva entre *T* y *RS*, *TA* y *Hel*; e inversa entre *T* y *OD*, mayores temperaturas implican menor *OD*. Además, Cond está muy influenciada por *T* y *Turbidez*. El aumento de la temperatura significa una mayor cantidad de iones presentes. *FT* y *T* están directamente correlacionadas, el aumento de temperatura favorece la actividad de microorganismos y la difusión del fósforo. El incremento de la temperatura ayuda al crecimiento de cianobacterias que contienen *Clo – A*.

Las dependencias espaciales (*SD*) consideran la posición de las estaciones de monitoreo. Por ejemplo, el caudal de una estación será influenciado por la precipitación de las estaciones aguas arriba. Además, las correlaciones espaciales (*SC*) representan el peso que se le otorga a las variables de ayuda de acuerdo a cuan alejadas están de la estación cuya variable se está imputando, según el método de Ponderación Inversa de la Distancia (*IDW*, por sus siglas en inglés):

$$Y_m = \frac{\sum_{i=1}^{n} Y_i d_{mi}^{-k}}{\sum_{i=1}^{n} d_{mi}^{-k}}$$

donde Y_m es la observación en la estación *m*, *n* es el número de estaciones, Y_i es la observación en la estación *i*, d_{mi} es la distancia entre *m* e *i*, *k* es un exponente que varía entre 1 y 6. Se asume $k = 2$.

Se utilizó la media móvil ponderada exponencialmente (*EWMA*) para contemplar la variabilidad temporal. Este método otorga

mayor peso a las observaciones más recientes:

$$EWMA(Y_n) = \frac{\sum_{i=0}^{t}(1-\alpha)^i Y_{n-1-i}}{\sum_{i=0}^{t}(1-\alpha)}$$

$$\alpha = \frac{2}{t+1}$$

donde *Y* es una serie temporal, *n* es el número de observaciones, *α* es el peso asignado, y *t* es la ventana temporal seleccionada basada en la variable a imputar: dos meses para las variables de calidad del agua, una semana para las variables hidrométricas y un día para las variables climáticas.

Cada una de estas restricciones se implementa como una variable artificial (columna) que se agrega al conjunto de datos original, creando, de esta manera, una nueva variante del conjunto de datos.

En la *fase 2*, se seleccionan las variables de ayuda para entrenar a los modelos. Son aquellas que tienen menos del 50% de los datos faltantes y que han sido imputadas con regresión lineal. Todas las variables se redujeron a escala mensual para coincidir a las variables de calidad de agua. Además, se utilizó una normalización *min-max* para igualar la importancia de las variables con diferentes escalas.

A continuación, en la *fase 3*, se utilizaron diversos modelos ya que era de gran importancia que fuera capaz de predecir con menos cantidad de datos. Se pueden agrupar en modelos univariados: *Inverse Distance Weighting (IDW)*; y multivariados: *Ridge Regressor (RR)*, *TheilSen Regressor (TR)*, *Huber Regressor (HR)*, *Bayesian Ridge Regressor (BRR)*, *Support Vector Regressor (SVR)*, *K-Nearest Neighbours Regressor (KNNR)*, *Random Forest Regressor (RFR)* y *Multivariate Imputation by Chained Equations (MICE)*.

Para el entrenamiento y validación de los modelos, se utiliza una validación cruzada de 10 pliegues. Para evaluar el desempeño de los modelos se calculan y se comparan los valores de la eficiencia de *Nash-Sutcliffe (NSE)*, sesgo porcentual *(PBIAS)* y eficiencia *Kling-Gupta (KGE)*. En la Tabla 1 se presentan los rangos de calificación.

Por último, en la *fase 4*, se sigue un proceso iterativo, comenzando con las variables que tienen menos datos faltantes. Una vez que se han imputado todas las variables, obtenemos el conjunto de datos completo y final. El resultado del marco es la mejor combinación entre el conjunto de datos y el modelo.

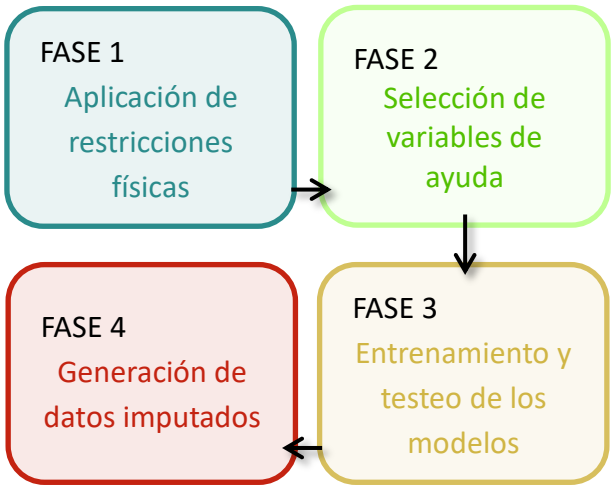

Figura 2: Conceptualización del marco de imputación

Tabla 1: Ratings de desempeño para los distintos tipos de variable (Chen, y otros, 2017); (Moriasi, Gitau, Pai, & Daggupati, 2015); (Rodríguez, y otros, 2021).

	Hidrométrica y climática	Calidad de agua físicas	Calidad de agua químicas
NSE			
Muy bueno	NSE > 0,80	NSE > 0,80	NSE > 0,65
Bueno	0,70 < NSE ≤ 0,80	0,70 < NSE ≤ 0,80	0,50 < NSE ≤ 0,65
Satisfactorio	0,50 < NSE ≤ 0,70	0,45 < NSE ≤ 0,70	0,35 < NSE ≤ 0,50
Insatisfactorio	NSE ≤ 0,50	NSE ≤ 0,45	NSE ≤ 0,35
PBIAS			
Muy bueno	|PBIAS| < 5	|PBIAS| < 10	|PBIAS| < 15
Bueno	5 ≤ |PBIAS| < 10	10 ≤ |PBIAS| < 15	15 ≤ |PBIAS| < 20
Satisfactorio	10 ≤ |PBIAS| < 15	15 ≤ |PBIAS| < 20	20 ≤ |PBIAS| < 30
Insatisfactorio	|PBIAS| ≥ 15	|PBIAS| ≥ 20	|PBIAS| ≥ 30
KGE			
Satisfactorio/Bueno	KGE ≥ -0,41	KGE ≥ -0,41	KGE ≥ -0,41
Insatisfactorio	KGE < -0,41	KGE < -0,41	KGE < -0,41

Resultados y Discusión

En la **Error! Reference source not found.** se presentan los resultados para cada tipo de variable cuya combinación variante y modelo maximizó el NSE.

Además, en Figura 4 la y la Figura 5 se presentan los resultados para PBIAS y KGE, respectivamente.

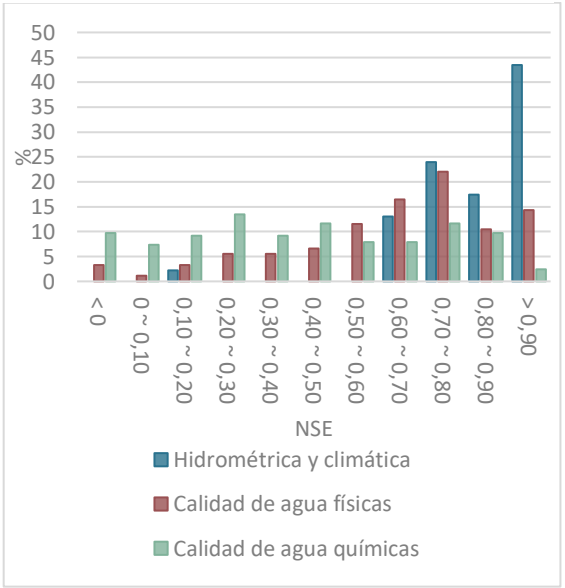

Figura 3: Resultados del framework de imputación en términos de NSE

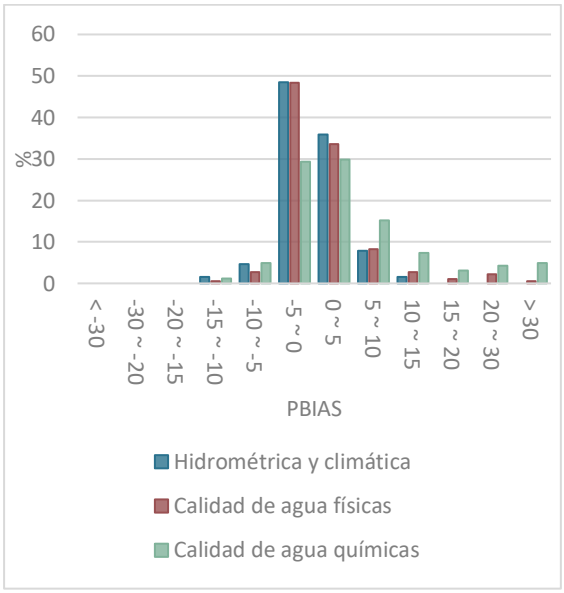

Figura 4: Resultados del framework de imputación en términos de PBIAS

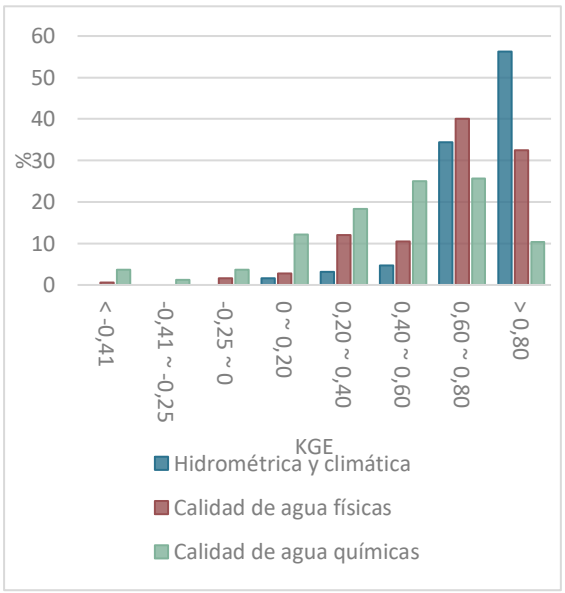

Figura 5: Resultados del framework de imputación en términos de KGE

En general, más del 68% de los datos imputados se caracterizan por un NSE>0,45 (resultados satisfactorios). En particular, el NSE mínimo calculado para las variables meteorológicas es 0,25, lo que significa que, de todas las imputaciones, el 40% puede considerarse buena y el 38% muy buenas. En relación a las estaciones pluviométricas, se realizó un análisis de calidad de datos que descarta la mitad de los pluviómetros por datos faltantes o anomalías sistemáticas. De esta manera, todos los pluviómetros considerados tienen un desempeño por lo menos satisfactorio, siendo el mínimo NSE de 0,67. Para las variables hidrométricas, el NSE es superior a 0,97 en todas las estaciones, excepto en dos, las cuales tienen un NSE de 0,64 (resultado satisfactorio) y 0,16 (resultado insatisfactorio). Esto muestra el buen rendimiento del *framework* propuesto. El desempeño tiende a disminuir cuando se trata de variables de calidad del agua. Más del 77% de las variables físicas de calidad del agua se caracterizan por un NSE>0,45 (resultados satisfactorios), y 49% de las variables químicas de calidad del agua alcanzan un NSE>0,35 (resultados satisfactorios). Para ambos dominios, el 64% de los datos imputados tiene un NSE>0,45 y el 71% tiene un NSE>0,35, lo que significa que, para casi todas las imputaciones de calidad del agua, el *framework* propuesto es

mejor que la función de media utilizada como imputador (NSE=0).

En las Figuras 6, 7 y 8, se presenta la variación espacial del NSE para algunas de las variables consideradas: *P* y *Q* se seleccionaron entre las variables hidrométricas y meteorológicas (Figura 6), *FT* y *NT*, como representativos de las descargas de nutrientes, se seleccionaron entre las variables químicas de calidad de agua (Figura 7), *Turbidez* y *OD* se eligieron entre las variables físicas de calidad de agua (Figura 8). El desempeño del *framework* para la variable *P* es bueno y muy bueno en todas las estaciones dentro de la cuenca, satisfactorio en las estaciones ubicadas en el borde de la cuenca e insatisfactorio en las estaciones fuera de la cuenca. Para la variable *Q*, el desempeño del *framework* es siempre muy bueno. Para las variables de calidad del agua, el *framework* presenta resultados menos favorables en las estaciones ubicadas en el noroeste de la cuenca (en la región de la sierra). En cuanto a las variables físicas de calidad del agua (en particular, *Cond* y *OD*), el *framework* muestra un mejor desempeño en las estaciones ubicadas más aguas arriba. Además, el desempeño del *framework* es siempre mejor en los embalses que en los ríos para todas las variables, siendo más notable en Paso Severino que en Canelón Grande. Sin embargo, esto no es así para las variables de calidad del agua que tienen pocos datos (*glifosato*, *Clo − A*, *SST* y *ST*).

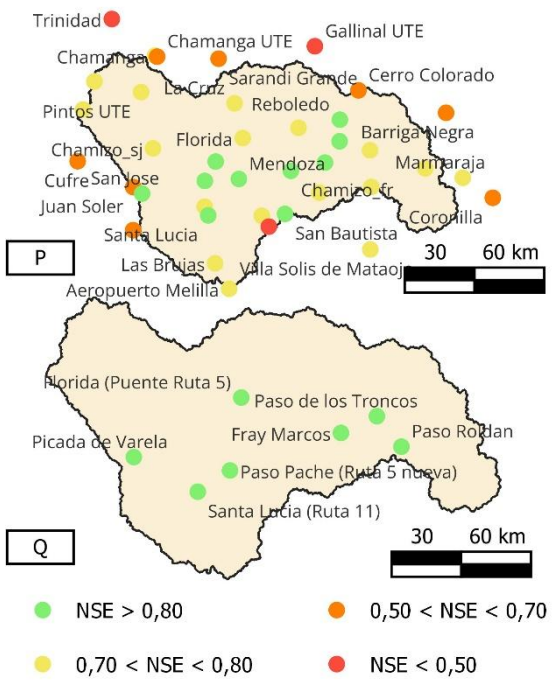

Figura 6: Variabilidad espacial del NSE para precipitación (P) y caudal (Q)

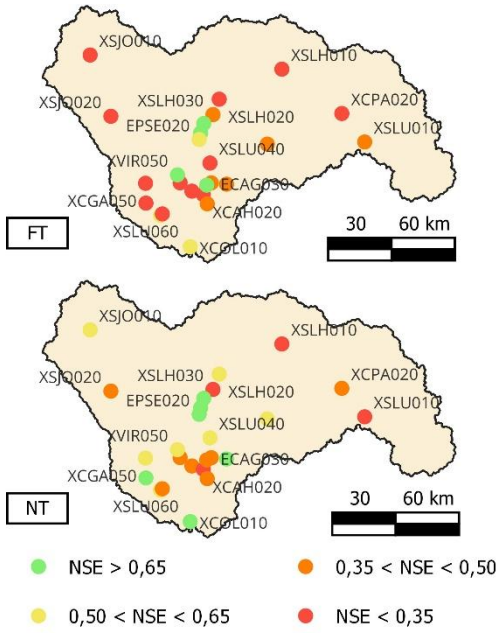

Figura 7: Variabilidad espacial del NSE para fósforo total (FT) y nitrógeno total (NT)

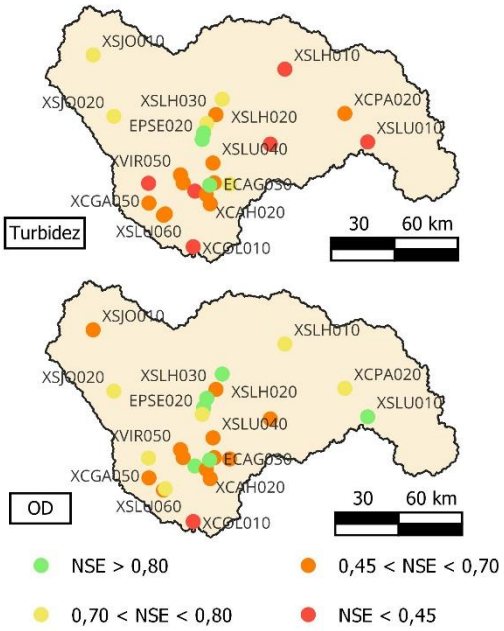

Figura 8: Variabilidad espacial del NSE para turbidez y oxígeno disuelto (OD)

En cuanto a los modelos, el *HR* es la técnica de imputación más exitosa, particularmente para las variables de calidad de agua y la precipitación (116 veces) (Tabla 2). Este modelo se eligió frecuentemente con las variantes temporales y espaciales, *SC+EWMA* fue seleccionado 26 veces, *SC+SD* 25 veces y *SC+SD+EWMA* 23 veces. El segundo modelo más seleccionado es el *SVR,* pero fue elegido un orden de magnitud menos (65 veces), que prevaleció en las variables de calidad de agua. En cambio, para las variables hidrométricas, el modelo *KNNR* es el más elegido y también se encuentra tercero en el *ranking* global (58 veces).

En cuanto a las variantes del *dataset*, el 96,31% de las veces se identificaron combinaciones de variantes del *dataset* original. Esto resalta el papel crucial de incluir restricciones físicas en el marco de aprendizaje automático para mejorar el rendimiento. Las variantes relacionadas a la variabilidad espacial fueron las preferidas (55,41%) y las de variabilidad temporal las

Tabla 2: Número de veces que cada variante fue escogida para imputar una variable, separado por modelo

	HR	SVR	KNNR	IDW	RFR	RR	BRR	TR	Total
SC + EWMA	26	9	9	4	9	4	6		67
SC + SD	25	1	3	4	2	3	2	1	41
SC + SD + EWMA	23	1	2	6	2	2	3	6	45
SC	17	4	22	4	4	2	5	1	59
MICE + SC + EWMA	5	3			1	3			12
SD + EWMA	5	5	5	10	7	2		1	35
MICE + SD	4	5				2			11
EWMA	3	10	2	6	7	3			31
MICE + EWMA	2	4			3	2	1		12
MICE + SC	2	1			2	1	3		9
Original	1	3	1	17	4	2			28
SD	1	4		7		1	1		14
MICE + SC + SD	1	3	3			2			9
MICE + SD + EWMA	1	4	4		2	1	1		13
MICE		5	7		3		1		16
MICE + SC + SD + EWMA		3					2		5
MICE + Ridge						3			3
Total	116	65	58	58	46	33	25	9	410

siguientes (29,02%). SC fue elegida 247 veces, SD 173 veces y EWMA 220 veces. El *dataset* original fue elegido solo 28 veces y 90 veces se utilizó MICE para mejorar el desempeño.

Tabla 3: Número de veces que cada variante fue escogida para imputar una variable

Variante	Número de variables imputadas
Original	28
SC	247
SD	173
EWMA	220
MICE	90

Conclusiones

Este estudio extiende a toda la cuenca del río Santa Lucía, Uruguay un marco de aprendizaje automático con restricciones físicas, previamente desarrollado para una cuenca más chica, para la imputación de datos ambientales. Los resultados demuestran un sólido desempeño del marco en diversos dominios ambientales, con más del 68% de los datos imputados tienen un *NSE>0,45* (resultados satisfactorios).

Para las variables meteorológicas, el NSE mínimo es 0,25, con el 40% de las imputaciones consideradas buenas y el 38% muy buenas. Las variables hidrométricas muestran un NSE superior a 0,97 en casi todas las estaciones, con solo dos excepciones. El rendimiento disminuye para las variables de calidad del agua: el 77% de las variables físicas tienen un *NSE>0,45* y el

49% de las variables químicas alcanzan un *NSE>0,35*.

El *Huber Regressor* fue la técnica de imputación más exitosa, especialmente para las variables de precipitación y calidad del agua, seguido por el *Support Vector Regressor* para calidad del agua y *K-Nearest Neighbours Regressor* para variables hidrométricas. El 96,31% de las veces se identificó una variante o combinación de variantes como el mejor modelo-*dataset*, destacando la importancia de las restricciones físicas. Las variantes espaciales fueron preferidas sobre las temporales, con la variante SC siendo la más frecuentemente seleccionada.

En resumen, el marco demuestra su robustez y efectividad en la imputación de datos faltantes en los diversos dominios ambientales en la cuenca del río Santa Lucía. Esto reafirma su utilidad para la gestión integrada de los recursos hídricos y destaca su potencial para una aplicación más amplia en contextos similares.

Referencias Bibliográficas

Chen, H., L. Y., Potter, C., Moran, P., Grieneisen, M., & Zhang, M. (2017). Modeling pesticide diuron loading from the San Joaquin watershed into the Sacramento-San Joaquin Delta using SWAT. *Water Research, 121*, 374-385. doi:10.1016/j.watres.2017.05.032

Chen, Z., Xu, H., Jiang, P., & Yu, S. L. (2021). A transfer learning-based LSTM strategy for imputing large-scale consecutive missing data and its application in a water quality prediction system. *Journal of Hydrology, 602*, 126573. doi:10.1016/j.jhydrol.2021.126573

Chivers, B., Wallbank, J., Cole, S., Sebek, O., Stanley, S., Fry, M., & Leontidis, G. (2020). Imputation of missing sub-hourly precipitation data in a large sensor network: A machine learning approach. *Journal of hydrology, 588*, 125126. doi:10.1016/j.jhydrol.2020.125126

Freni, G., Mannina, G., & Viviani, G. (2011). Assesment of the integrated urban water quality model complexity through identifiability analysis. *Water research, 45*, 37-50. doi:10.1016/j.watres.2010.08.004

Gorgoglione, A., Castro, A., Chreties, C., & Etcheverry, L. (2020b). Overcoming data scarcity in earth science. *Data, 5*(1), 5. doi:10.3390/data5010005

Gorgoglione, A., Gregorio, J., Ríos, A., Alonso, J., Chreties, C., & Fossati, M. (2020a). Influence of land use/land cover on surface-water quality of Santa Lucia river, Uruguay. *Sustainability, 12*, 4692. doi:10.3390/su12114692

Moriasi, D., Gitau, M., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE, 58*(6), 1763-1785. doi:10.13031/trans.58.10715

MVOTMA. (2015). *Presidencia de la República.* Obtenido de Decreto 55/015 Aprobación de la selección del área natural protegida denominada "humedales de Santa Lucia": https://www.gub.uy/presidencia/institucional/normativa/decreto-55015-aprobacion-seleccion-del-area-natural-protegida-denominada

Pastorini, M., Rodríguez, R., Etcheverry, L., Castro, A., & Gorgoglione, A. (2024). Enhancing environmental data imputation: A physically-constrained machine learning framework. *Science of The Total Envioronment, 926*, 171773. doi:10.1016/j.scitotenv.2024.171773

Rodríguez, R., Pastorini, M., Etcheverry, L., Chreties, C., Fossati, M., Castro, A., & Gorgoglione, A. (2021). Water-Quality Data Imputation with a High Percentage of Missing Values: A Machine Learning Approach. *Susteinability, 13*(11), 6318. doi:10.3390/su13116318

Sattari, M. T., Rezazadeh-Joudi, A., & Kusiak, A. (2017). Assessment of different methods for estimation of missing data in precipitation studies. *Hydrology Research, 48*(4), 1032-1044. doi:10.2166/nh.2016.364

Financiamiento

Este trabajo fue financiado por la Agencia Nacional de Investigación e Innovación (ANII) [numero: FMV_3_2022_1]