

DOI: 10.24850/j-tyca-18-02-02

Artículos

Comparativa de modelos tradicionales vs machine learning para el relleno de datos faltantes en series de caudales diarios en cuencas representativas del Perú
Comparative analysis of traditional models vs machine learning for daily streamflow infilling in representative Peruvian basins

Becky Berrocal¹, ORCID: <https://orcid.org/0009-0001-3364-3307>

Jamil Ramirez², ORCID: <https://orcid.org/0009-0009-3386-8571>

¹Universidad Tecnológica del Perú, Lima, Perú, c26212@utp.edu.pe

²Universidad Nacional Mayor de San Marcos, Lima, Perú, jamil_ramirez.m@hotmail.com

Autora para correspondencia: Becky Berrocal, c26212@utp.edu.pe

Resumen

Los registros confiables de caudales son fundamentales para la gestión hídrica, aunque los datos faltantes limitan frecuentemente el análisis hidrológico. Este estudio evalúa comparativamente métodos tradicionales (persistencia, climatología, regresión) y de machine learning (algoritmos de árboles, ensambles y stacking) para el relleno de datos faltante de



caudal diario en cuencas representativas de las principales regiones naturales del Perú: Costa, Sierra y Selva. A partir de datos observados, se generaron escenarios sintéticos con pérdidas de información del 20 % al 50 %. El desempeño se cuantificó mediante la eficiencia de Nash-Sutcliffe (NSE), el error cuadrático medio (RMSE) y el sesgo porcentual (PBIAS). Los resultados demuestran que no existe un método universalmente superior. En escenarios de baja pérdida, ambos enfoques presentan resultados comparables; sin embargo, ante mayores vacíos de información, el machine learning muestra una alta sensibilidad en cuencas de la selva debido a su variabilidad hidrológica. En contraste, los métodos tradicionales evidencian mayor robustez en las regiones de la costa y sierra. Finalmente, la limitada precisión obtenida en sistemas intercuenca subraya las restricciones actuales de estas técnicas para modelar transferencias complejas.

Palabras clave: medición hidrológica, inteligencia artificial, análisis estadístico, procesamiento de datos, modelo matemático, cuenca, Perú.

Abstract

Reliable streamflow records are essential for hydrological analysis and water resource management; however, missing data remains a frequent limitation in hydrometric networks. This study comparatively evaluates traditional methods (persistence, climatology, and regression) and machine learning approaches (tree-based algorithms, ensemble models, and stacking) for infilling missing daily flow data across representative basins in Peru's main natural regions: Coast, Highlands (Sierra), and Rainforest (Selva). Based on observed records, synthetic information loss scenarios were generated ranging from 20% to 50%. Performance was



assessed using efficiency and error metrics, such as the Nash–Sutcliffe efficiency (NSE), Root Mean Square Error (RMSE), and Percent Bias (PBIAS). The results indicate that no single method is universally superior. In scenarios with low data loss, both approaches show comparable performance. However, as the percentage of missing data increases, several machine learning models exhibit higher sensitivity, particularly in rainforest basins with high hydrological variability. In contrast, traditional methods prove more robust in coastal and highland basins. Furthermore, the limited performance of all methods in inter-basin systems highlights the current constraints of these techniques for modeling complex hydrological transfers.

Keywords: hydrological measurement, artificial intelligence, statistical analysis, data processing, mathematical models, basins, Peru.

Recibido: 20/01/2026

Aceptado: 17/05/2026

Publicado ahead of print: 27/05/2026

1. Introducción

La disponibilidad de series hidrológicas continuas y confiables es fundamental para el análisis y la gestión de los recursos hídricos, particularmente a partir de registros de caudales diarios utilizados en estudios de balance hídrico, sequías e inundaciones, diseño de infraestructura hidráulica y evaluación de impactos del cambio climático. Sin embargo, en muchos países en desarrollo, como el Perú, las redes hidrométricas presentan discontinuidades asociadas a fallas de medición,



limitaciones operativas y condiciones climáticas extremas, lo que genera series con datos faltantes y afecta la confiabilidad de los análisis hidrológicos (Little & Rubin, 2002; Moriasi et al., 2007).

El relleno de datos faltantes en series de caudales se ha abordado tradicionalmente mediante métodos estadísticos simples, como interpolaciones temporales y regresiones lineales. Aunque estos enfoques son de fácil implementación e interpretación, su desempeño se ve limitado en regímenes hidrológicos complejos, caracterizados por no linealidades, alta variabilidad estacional y eventos extremos, debido a supuestos de linealidad y estacionariedad que no siempre representan adecuadamente los procesos hidrológicos (Hirsch, 1982; Krause et al., 2005).

En este contexto, los métodos basados en machine learning han surgido como alternativas para el relleno de datos hidrológicos faltantes, dada su capacidad para modelar relaciones no lineales y capturar patrones complejos. No obstante, la evidencia disponible sobre su desempeño en series de caudales diarios es heterogénea y no muestra una superioridad consistente frente a los métodos tradicionales, especialmente cuando se consideran métricas sensibles a los extremos y a la variabilidad hidrológica (Coulibaly & Evora, 2007; Arriagada et al., 2021; Zhou et al., 2023).

Un aspecto menos explorado es la dependencia del desempeño de los métodos de relleno respecto a las condiciones hidrológicas regionales y al porcentaje de datos faltantes. En países con alta heterogeneidad hidroclimática, como Perú, las cuencas de la costa, sierra y selva presentan regímenes de caudal contrastantes, con patrones áridos e intermitentes en la Costa, estacionales en la Sierra y de alta variabilidad en la Selva, por lo que extrapolar conclusiones generales sin considerar estas diferencias puede conducir a aplicaciones poco robustas. En este

sentido, la literatura reciente coincide en que no existe un método universalmente superior y que el desempeño de las técnicas para completar datos faltantes depende tanto de las características hidrológicas del sistema analizado como del nivel de información ausente (Addi et al., 2022; Hamzah et al., 2021).

Asimismo, muchos estudios han evaluado los métodos bajo condiciones ideales o con pérdidas reducidas de información, lo que no siempre refleja la realidad operativa de las redes hidrométricas. La generación de escenarios controlados de pérdida de datos permite analizar de forma sistemática la sensibilidad y estabilidad de los métodos bajo condiciones más exigentes. En este marco, el presente estudio evalúa comparativamente métodos tradicionales y de machine learning para el relleno de datos faltantes en series de caudales diarios, considerando escenarios de pérdida de información entre 20 % y 50 % y se aplica en cuencas representativas de las principales regiones naturales de Perú, con el fin de aportar criterios prácticos para la selección de técnicas para completar datos faltantes en estudios hidrológicos.

2. Materiales y métodos

2.1. Datos Hidrometeorológicos

El estudio se desarrolló a partir de series de caudales diarios observados en nueve estaciones hidrométricas correspondientes a nueve cuencas representativas de las principales regiones naturales de Perú: costa, sierra y selva. Se consideraron tres cuencas por región y una estación representativa por cuenca, con el fin de mantener una distribución balanceada entre regiones hidrológicas contrastantes. La selección

priorizó estaciones con suficiente continuidad temporal, disponibilidad de registros diarios y una baja proporción de datos faltantes en su condición original, de modo que las series fueran adecuadas para la generación controlada de escenarios sintéticos de pérdida de información y la evaluación comparativa de los métodos de relleno de datos.

Si bien las estaciones presentan diferencias en longitud temporal y porcentaje original de datos faltantes, todas las series seleccionadas contaban con suficiente continuidad y disponibilidad de registros para permitir la generación de escenarios sintéticos comparables de pérdida de información. En este sentido, el estudio no buscó aislar completamente el efecto de las características propias de cada serie, sino evaluar el comportamiento relativo de los métodos de relleno bajo distintos contextos hidrológicos y niveles controlados de información ausente.

Previamente al análisis, las series fueron sometidas a un control básico de calidad que incluyó la revisión de valores atípicos, la consistencia temporal y la verificación de unidades. A partir de las series depuradas, se construyó un conjunto de datos de referencia (ground truth), utilizado para la generación de escenarios sintéticos de pérdida de información y la evaluación del desempeño de los métodos aplicados.

Las nueve estaciones hidrométricas analizadas, junto con sus principales características (cuenca, región natural, período de análisis y porcentaje de datos faltantes en su condición original), se presentan en la Tabla 1.

Tabla 1. Estaciones hidrométricas y características generales

Estación	Cuenca/Río	Región natural	Período de análisis	% datos faltantes
Puente Sánchez Cerro	Cuenca Piura	Costa	2014–2025	0.42 %
Racarumi	Cuenca Chancay - Lambayeque	Costa	1983–2025	19.08 %
Salinar	Cuenca Chicama	Costa	1983–2025	18.09 %
Pongor	Cuenca Mantaro	Sierra	2017–2025	1.02 %
Condorcerro	Cuenca Santa	Sierra	1983–2025	3.21 %
Puente Carretera Ilave	Cuenca Ilave	Sierra	1983–2025	3.47 %
Bellavista	Cuenca Napo	Selva	1998–2024	4.55 %
Bocatoma San Pablo	Intercuenca Medio Alto Huallaga	Selva	2018–2025	1.42 %
Los Naranjos	Cuenca Utcubamba	Selva	2001–2020	9.14 %

El porcentaje de datos faltantes corresponde a la condición original de las series de caudales diarios antes de la generación de los escenarios sintéticos de pérdida de información.

Fuente: Elaboración propia a partir de datos hidrométricos oficiales.

La distribución espacial de las cuencas analizadas se muestra en la Figura 1, donde se presenta su localización en el territorio peruano y su agrupación por región natural (costa, sierra y selva), lo que permite contextualizar las diferencias en los regímenes hidrológicos y facilitar la interpretación comparativa de los resultados.

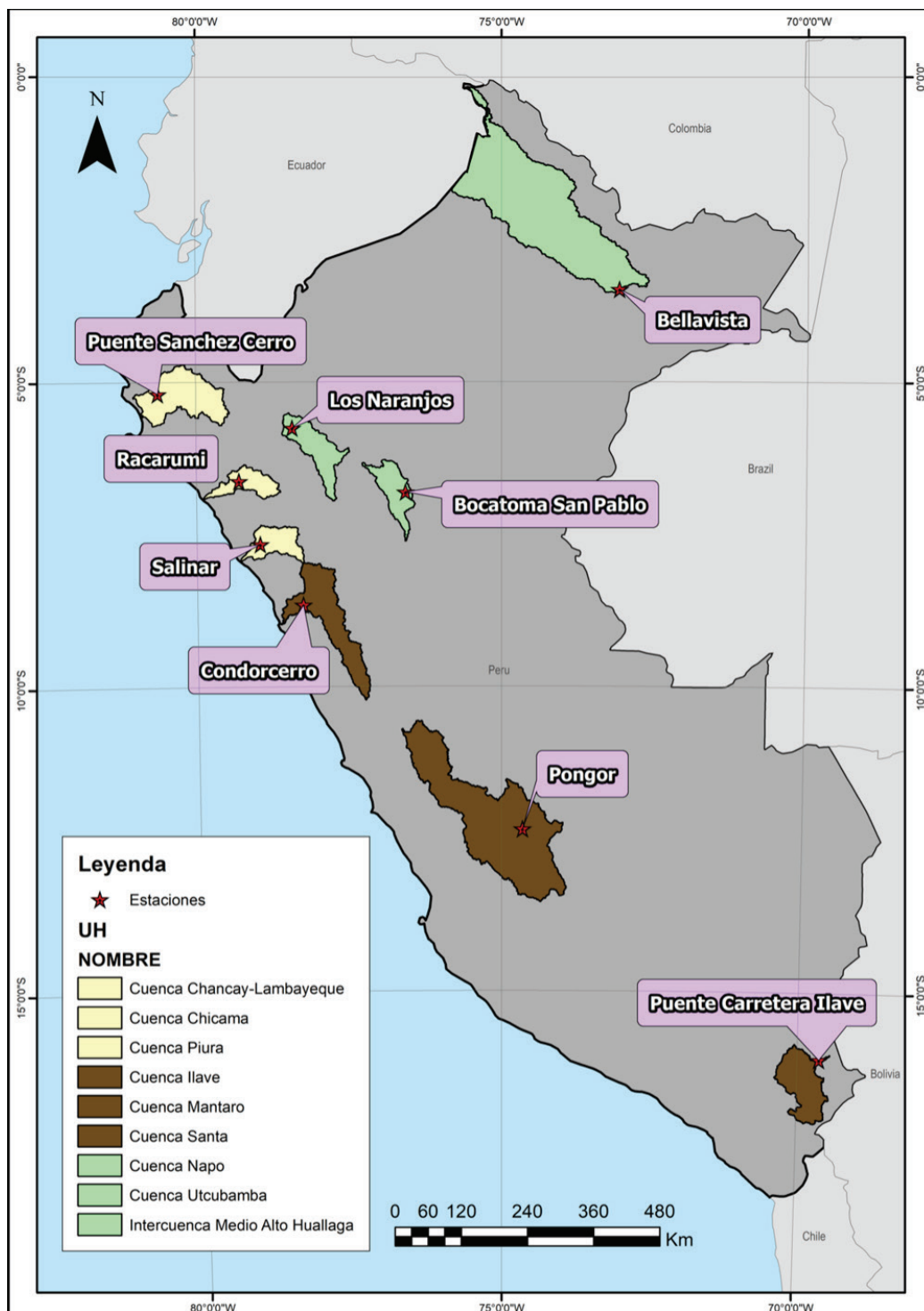


Figura 1. Localización de las estaciones hidrométricas y cuencas analizadas en el Perú, agrupadas según región natural (costa en "amarillo", sierra en "marrón" y selva en "verde").

2.2. Generación de escenarios de datos faltantes

Con el fin de analizar de manera sistemática el comportamiento de los distintos métodos ante diferentes niveles de pérdida de información, se generaron escenarios controlados de datos faltantes. A partir de las series originales completas, se eliminaron de forma aleatoria porcentajes crecientes de datos diarios, correspondientes a 20 %, 30 %, 40 % y 50 %, los cuales se denominaron esc20, esc30, esc40 y esc50, respectivamente.

La eliminación de datos se realizó de forma aleatoria y distribuida a lo largo de todo el período de análisis, con el fin de preservar la representatividad temporal de los distintos regímenes hidrológicos presentes en cada serie. Este enfoque permitió generar escenarios controlados y comparables de pérdida progresiva de información para la evaluación relativa de los métodos de relleno. No se simuló vacíos continuos extensos debido a que estos representan un problema metodológico distinto, más asociado a la reconstrucción de períodos completos y a posibles sesgos derivados de la eliminación de años húmedos, años secos o eventos hidrológicos particulares.

2.3. Métodos tradicionales de relleno de datos

Como referencia metodológica, se aplicaron métodos tradicionales que se han utilizado en estudios hidrológicos para el relleno de datos faltantes en series de caudales diarios. Estos enfoques se caracterizan por su simplicidad conceptual, bajo requerimiento computacional y por constituir

prácticas habituales en el procesamiento y manejo operativo de registros hidrométricos.

Los métodos evaluados incluyen técnicas basadas en la persistencia temporal, estadísticas descriptivas de carácter estacional e interpolaciones en el dominio del tiempo, así como modelos de regresión lineal apoyados en rezagos del propio caudal. En conjunto, estos enfoques aprovechan la información temporal disponible de cada serie para estimar los valores ausentes y, en general, asumen comportamientos relativamente estables y relaciones funcionales simples. La Tabla 2 presenta un resumen de los métodos tradicionales implementados y sus principales supuestos.

La selección de estos métodos responde a su uso extendido en estudios hidrológicos y a su reconocimiento como referencia metodológica en la evaluación de series de caudales diarios (Hirsch, 1982; Moriasi et al., 2007).

Tabla 2. Métodos tradicionales evaluados y configuración

Nombre del método	Tipo	Idea / supuesto principal	Variables de entrada	Aplicación del relleno (regla)
Persistencia de 1 día (Lag-1)	Persistencia	El caudal faltante es igual al caudal del día anterior	Q_{t-1}	Si falta Q_t y existe Q_{t-1} , entonces $Q_t = Q_{t-1}$
Persistencia iterativa (LOCF)	Persistencia	Mantiene el último valor observado disponible	Serie Q	Propaga el último valor observado hacia adelante en todo el gap
Climatología diaria (mediana por día del año)	Climatología	El caudal típico depende del día del año (estacionalidad)	DOY, Q	Si falta Q_t , se reemplaza por la mediana histórica del DOY correspondiente

Nombre del método	Tipo	Idea / supuesto principal	Variables de entrada	Aplicación del relleno (regla)
Climatología mensual (mediana por mes)	Climatología	El caudal típico depende del mes (estacionalidad más agregada)	Mes, Q	Si falta Q_t , se reemplaza por la mediana del mes correspondiente
Interpolación lineal temporal	Interpolación	El caudal evoluciona de forma aproximadamente lineal dentro del gap	Serie Q, tiempo	Interpola linealmente entre puntos vecinos; incluye extrapolación en extremos
Interpolación por splines	Interpolación	Suaviza la trayectoria temporal mediante una curva continua	Serie Q, tiempo	Interpolación spline y ajuste para evitar valores negativos
Media móvil (ventana 7 días)	Suavizado	El caudal faltante se aproxima por el promedio local	Serie Q	Calcula media móvil centrada de 7 días y la usa para completar
Regresión lineal multirrezago + índices ENSO	Regresión	Relación lineal entre Q_t y rezagos del propio caudal, con forzantes climáticas	$Q_{t-1}, Q_{t-2}, Q_{t-7}+$ ONI/SOI/Niño1+2 (rezagos)	Ajuste con datos disponibles y predicción en días faltantes (si hay datos suficientes)
Regresión lineal simple (Lag-1 + ENSO)	Regresión	Versión parsimoniosa cuando el modelo completo no converge	$Q_{t-1} +$ ONI/Niño1+2 (rezagos)	Se aplica como "fallback" cuando no se ajusta el modelo completo
Método híbrido: interpolación (gaps cortos) + climatología DOY (gaps largos)	Híbrido	Estrategia por longitud del gap	Interpolación + DOY	Si $\text{gap} \leq 7$ días: interpolación; si $\text{gap} > 7$ días: climatología DOY

DOY = día del año

La interpolación lineal incluye extrapolación en los extremos. En el método híbrido se aplica interpolación para brechas ≤ 7 días y climatología diaria para brechas > 7 días. En el caso de splines se impuso $Q \geq 0$. Los modelos de regresión se ajustaron solo cuando existió información suficiente.

2.4. Métodos basados en machine learning (ML)

De manera adicional a los enfoques tradicionales, se evaluaron métodos de relleno de datos basados en algoritmos de machine learning. En este estudio se utilizaron modelos de tipo árbol y enfoques de ensamble, se seleccionaron por su capacidad para manejar interacciones complejas entre variables, por trabajar con predictores correlacionados y por su uso extendido en aplicaciones hidrológicas.

La aplicación de algoritmos basados en árboles y ensambles se reporta en la literatura como una alternativa adecuada para el análisis de series hidrológicas, aunque con un desempeño fuertemente condicionado por la estructura temporal de los datos y la disponibilidad de información (Papacharalampous & Tyrallis, 2020; Tyrallis & Papacharalampous, 2017).

Los modelos se entrenaron utilizando exclusivamente la información disponible en cada escenario sintético de datos faltantes, incorporando como predictores retardos temporales del caudal, estadísticas móviles y variables de calendario. La evaluación se realizó mediante un esquema de enmascaramiento sintético, donde los valores removidos artificialmente fueron comparados posteriormente con la serie observada original. En este sentido, no se utilizó una división fija convencional entrenamiento-validación, sino que el conjunto de evaluación estuvo definido por el porcentaje de datos eliminados en cada escenario (20 %, 30 %, 40 % y 50 %).

Adicionalmente, se exploraron enfoques de ensamble mediante la combinación de distintos modelos base, con el objetivo de evaluar si la integración de múltiples predictores permitía mejorar el desempeño global del relleno de datos. En el caso del modelo de stacking, se aplicó

validación cruzada de cinco particiones para generar predicciones out-of-fold durante el entrenamiento del meta-modelo y reducir el riesgo de fuga de información (data leakage). Todos los modelos se implementaron utilizando software de código abierto. La Tabla 3 resume los métodos de machine learning que se utilizaron con sus principales características.

Tabla 3. Métodos ML (machine learning) evaluados y configuración

Nombre del método	Tipo	Idea / supuesto principal	Variables de entrada	Aplicación del relleno (regla)
Árbol de decisión (CART)	ML (árbol simple)	Calendario (DOY, mes, año), rezagos Q_{t-1} , Q_{t-2} , Q_{t-7} , promedios móviles (7 y 30 días), índices ENSO (ONI, Niño1+2, SOI y rezagos)	rpart (parámetros por defecto)	Entrenamiento con datos no faltantes; predicción directa sobre días faltantes
Random Forest	ML (ensamble)	Idem anterior	ranger, 500 árboles, $mtry \approx \sqrt{p}$	Idem anterior
Extra Trees (o proxy)	ML (ensamble)	Idem anterior	ranger, 500 árboles; splitrule="extratrees" si está disponible; si no, proxy con $mtry = 1$	Idem anterior
Gradient Boosting (XGBoost)	ML (boosting)	Idem anterior	xgboost: 600 iteraciones, max_depth=5, eta=0.05, subsample=0.8, colsample_bytree=0.8, objetivo MSE	Idem anterior
Ensamble por stacking (meta-modelo lineal)	Ensamble (meta-learning)	Predicciones de modelos base (dtr, rf, etr, xgb)	Meta-modelo lm entrenado con predicciones out-of-fold (OOF)	Stacking con validación cruzada 5-fold para generar OOF (sin fuga de información) y luego predicción en faltantes

Los predictores incluyeron componentes temporales (DOY, mes), rezagos de caudal, promedios móviles e índices ENSO (ONI, Niño1+2 y SOI). El *stacking* se implementó mediante predicciones *out-of-fold* usando validación cruzada 5-fold para evitar fuga de información. En Extra Trees se utilizó splitrule = "extratrees" cuando estuvo disponible.

2.5. Estrategia de evaluación

El desempeño de los métodos tradicionales y de machine learning se evaluaron comparando los valores estimados con los caudales observados originales eliminados en cada escenario de pérdida de información. Para ello, se emplearon métricas comúnmente utilizadas en hidrología: el coeficiente de eficiencia de Nash-Sutcliffe (NSE), utilizado para evaluar la capacidad del modelo para reproducir la dinámica general de la serie; la raíz del error cuadrático medio (RMSE), empleada para cuantificar la magnitud del error absoluto; y el porcentaje de sesgo (PBIAS), utilizado para identificar tendencias sistemáticas de sobreestimación o subestimación. Estas métricas permitieron evaluar de manera complementaria la eficiencia global, el error y el sesgo de los métodos analizados.

La evaluación se realizó de manera independiente para cada estación, escenario y región natural. Dado que métricas agregadas como el NSE pueden ocultar deficiencias específicas del ajuste, su interpretación se realizó de forma conjunta con indicadores de error y sesgo (Gupta et al., 2009; Gupta et al., 2011).

2.6. Análisis comparativo por región natural

Los resultados se analizaron de manera diferenciada para las regiones costa, sierra y selva, con el fin de identificar patrones regionales en el desempeño de los métodos y evaluar la influencia del régimen hidrológico y de la variabilidad climática sobre la efectividad de las técnicas de relleno

de datos faltantes, así como la transferibilidad de los resultados entre regiones con características contrastantes.

3. Resultados

3.1. Desempeño general de los métodos de relleno de datos faltantes

En términos generales, el desempeño de los métodos de relleno de datos varía en función del porcentaje de información ausente y del enfoque empleado. En escenarios con baja pérdida de datos (esc20), tanto los métodos tradicionales como los basados en machine learning reproducen adecuadamente el comportamiento general de las series de caudales diarios, con diferencias relativamente reducidas entre enfoques. No obstante, a medida que aumenta el porcentaje de datos faltantes (esc30–esc50), se observa una mayor dispersión en el desempeño, evidenciándose una pérdida de estabilidad en algunas estimaciones, particularmente en ciertos modelos de machine learning, que muestran incrementos notables del error relativo a pesar de mantener valores elevados en métricas de eficiencia global. Este comportamiento es consistente con resultados reportados en la literatura, donde bajo escenarios de baja pérdida de información los métodos tradicionales presentan un desempeño comparable al de los modelos de machine learning (Baddoo et al., 2021; Hamzah et al., 2021).

3.1.1. Comparación entre métodos tradicionales vs machine learning – Región Costa

En las estaciones de la región Costa (Puente Sánchez Cerro, Racarumi y Salinar), los métodos tradicionales igualan o superan de manera consistente a los enfoques de machine learning en la mayoría de los escenarios de pérdida de datos. En esc20, los métodos tradicionales alcanzan las mayores eficiencias hidrológicas ($NSE > 0.93$), destacando la interpolación temporal y la persistencia de un día, mientras que los modelos de machine learning presentan valores aceptables, pero sistemáticamente inferiores. Al incrementarse el porcentaje de datos faltantes (esc30–esc40), los métodos tradicionales mantienen una mayor estabilidad, con errores relativamente controlados, evidenciando la ventaja de enfoques híbridos y de regresión con rezagos e información climática frente a modelos puramente data-driven. En el escenario más exigente (esc50), todos los métodos muestran una degradación del desempeño; sin embargo, los enfoques tradicionales conservan una ventaja relativa en términos de Nash–Sutcliffe Efficiency (NSE) y Root Mean Square Error (RMSE), mientras que los modelos de machine learning exhiben mayor variabilidad y sensibilidad a la pérdida de información. La Tabla 4 resume, para cada estación y escenario, el mejor método tradicional y de machine learning, junto con métricas complementarias de error y sesgo.

Tabla 4. Comparación del mejor método tradicional y de machine learning (ML) por estación y escenario en la región Costa

Estación	Escenario	Mejor Método Tradicional	NSE	RMSE	PBIAS	Mejor Método ML	NSE	RMSE	PBIAS
Puente Sánchez Cerro	esc20	Interpolación por splines	0.94	49.24	-3.56	Random Forest	0.88	67.01	-0.63
	esc30	Interpolación por splines	0.93	42.41	4.81	Random Forest	0.88	54.88	1.32
	esc40	Regresión lineal multirrezago + índices ENSO	0.89	44.76	-5.83	Random Forest	0.79	80.80	-5.95
	esc50	Persistencia de 1 día (Lag-1)	0.85	73.48	1.53	Random Forest	0.80	78.79	4.46
Racarumi	esc20	Interpolación lineal temporal	0.95	9.22	0.70	Random Forest	0.91	11.61	0.55
	esc30	Método híbrido: interpolación (gaps cortos) + climatología DOY (gaps largos)	0.93	10.71	-0.18	Random Forest	0.90	12.74	-0.35
	esc40	Interpolación lineal temporal	0.91	11.87	-0.54	Ensamble por stacking (meta-modelo lineal)	0.87	14.21	-0.37
	esc50	Regresión lineal multirrezago + índices ENSO	0.89	12.47	-1.35	Ensamble por stacking (meta-modelo lineal)	0.81	17.61	-1.06
Salinar	esc20	Persistencia de 1 día (Lag-1)	0.94	13.15	-0.24	Gradient Boosting (XGBoost)	0.90	16.93	1.79
	esc30	Persistencia de 1 día (Lag-1)	0.94	12.95	1.02	Ensamble por stacking (meta-modelo lineal)	0.95	11.77	0.56
	esc40	Método híbrido: interpolación (gaps cortos) + climatología DOY (gaps largos)	0.91	20.13	-1.53	Random Forest	0.89	22.84	-0.70
	esc50	Interpolación lineal temporal	0.94	14.63	-0.32	Gradient Boosting (XGBoost)	0.92	17.84	0.02

NSE = Nash-Sutcliffe Efficiency

RMSE = Root Mean Square Error

PBIAS = Percent Bias

El mejor método se seleccionó en función del mayor valor de NSE. En caso de valores similares, se priorizó el método con menor RMSE y menor valor absoluto de PBIAS.

3.1.2. Comparación entre métodos tradicionales vs machine learning – Región Sierra

Para las estaciones hidrométricas de la región Sierra, los métodos tradicionales muestran un desempeño consistentemente alto en todos los escenarios de pérdida de datos, con valores de NSE generalmente superiores a 0.95 incluso bajo condiciones severas (esc40–esc50). La interpolación temporal (lineal y por splines) presenta recurrentemente algunos de los mejores desempeños relativos, lo que sugiere una fuerte continuidad temporal y una estacionalidad bien definida en los regímenes de caudal. En el escenario esc20, algunos modelos de machine learning, principalmente Gradient Boosting, alcanzan eficiencias cercanas a las de los métodos tradicionales, especialmente en la estación Pongor; sin embargo, estos últimos tienden a mantener métricas ligeramente más favorables en términos de NSE y RMSE. A medida que aumenta el porcentaje de datos faltantes (esc30–esc50), los métodos tradicionales muestran un comportamiento relativamente más estable, mientras que los modelos de machine learning presentan incrementos del RMSE y, en ciertos casos, sesgos más pronunciados, particularmente en Puente Carretera Ilave. Aunque los enfoques de ensamble mejoran parcialmente el desempeño en escenarios intermedios, sus resultados no evidencian una superioridad clara y consistente respecto a los métodos tradicionales. La Tabla 5 presenta la comparación del método tradicional y de machine learning con mejor desempeño relativo por estación y escenario, junto con métricas complementarias de error y sesgo.

Tabla 5. Comparación del mejor método tradicional y de machine learning (ML) por estación y escenario en la región Sierra

Estación	Escenario	Mejor Método Tradicional	NSE	RMSE	PBIAS	Mejor Método ML	NSE	RMSE	PBIAS
Pongor	esc20	Interpolación por splines	0.98	35.22	-0.70	Gradient Boosting (XGBoost)	0.97	42.20	-1.63
	esc30	Interpolación lineal temporal	0.98	37.55	-0.45	Gradient Boosting (XGBoost)	0.96	47.11	-1.18
	esc40	Interpolación por splines	0.97	46.31	-0.87	Gradient Boosting (XGBoost)	0.95	57.14	-3.01
	esc50	Regresión lineal multirrezago + índices ENSO	0.96	52.20	-1.33	Ensamble por stacking (meta-modelo lineal)	0.94	60.39	0.69
Condorcero	esc20	Interpolación lineal temporal	0.97	22.60	0.19	Ensamble por stacking (meta-modelo lineal)	0.95	28.94	-0.16
	esc30	Interpolación lineal temporal	0.97	22.80	0.26	Random Forest	0.95	28.92	0.03
	esc40	Interpolación lineal temporal	0.96	26.88	-0.13	Random Forest	0.94	33.39	-0.46
	esc50	Persistencia de 1 día (Lag-1)	0.96	27.76	0.08	Random Forest	0.92	36.50	0.34
Puente Carretera Ilave	esc20	Interpolación lineal temporal	0.95	17.03	0.65	Gradient Boosting (XGBoost)	0.92	20.74	-1.64
	esc30	Interpolación lineal temporal	0.92	18.70	0.27	Gradient Boosting (XGBoost)	0.88	22.80	-0.68
	esc40	Interpolación lineal temporal	0.92	18.38	0.97	Ensamble por stacking (meta-modelo lineal)	0.89	22.28	-0.84
	esc50	Regresión lineal multirrezago + índices ENSO	0.91	18.37	-1.00	Random Forest	0.85	25.46	-2.27

NSE = Nash–Sutcliffe Efficiency

RMSE = Root Mean Square Error

PBIAS = Percent Bias

El mejor método se seleccionó en función del mayor valor de NSE. En caso de valores similares, se priorizó el método con menor RMSE y menor valor absoluto de PBIAS.

3.1.3. Comparación entre métodos tradicionales vs machine learning – Región Selva

En la región Selva, el desempeño de los métodos para completar los datos faltantes, presenta un comportamiento contrastante entre estaciones, fuertemente condicionado por la naturaleza del sistema hidrológico. En la estación Bellavista (cuenca del Napo), tanto los métodos tradicionales como los basados en *machine learning* alcanzan eficiencias muy elevadas ($NSE > 0.98$) en todos los escenarios; sin embargo, los enfoques tradicionales —particularmente la interpolación por *splines* y la regresión con rezagos e índices ENSO— muestran de manera sistemática menores errores, especialmente bajo mayores porcentajes de datos faltantes, lo que evidencia la alta efectividad de métodos basados en la persistencia temporal en cuencas naturales. En contraste, la estación Bocatoma San Pablo, representativa de un sistema intercuenca influenciado por derivaciones y operaciones hidráulicas externas, presenta bajos valores de NSE para todos los métodos, incluso en escenarios de baja pérdida de información, lo que sugiere limitaciones en la capacidad de las técnicas basadas únicamente en la dinámica temporal local para representar este tipo de sistemas regulados. Por su parte, la estación Los Naranjos muestra un desempeño intermedio, con diferencias reducidas entre enfoques, aunque con una ligera ventaja de los métodos tradicionales y semihíbridos en términos de estabilidad y control del sesgo. La Tabla 6 presenta la comparación del método tradicional y de machine learning con mejor desempeño relativo por estación y escenario, junto con métricas complementarias de error y sesgo.

Tabla 6. Comparación del mejor método tradicional y de machine learning (ML) por estación y escenario en la región Selva

Estación	Escenario	Mejor Método Tradicional	NSE	RMSE	PBIAS	Mejor Método ML	NSE	RMSE	PBIAS
Bellavista	esc20	Interpolación por splines	1.00	134.04	-0.11	Random Forest	0.99	220.29	-0.05
	esc30	Interpolación por splines	0.99	183.27	-0.03	Random Forest	0.99	269.14	-0.10
	esc40	Interpolación por splines	1.00	175.51	-0.07	Random Forest	0.99	284.87	-0.04
	esc50	Regresión lineal multirrezago + índices ENSO	0.99	197.00	-0.23	Random Forest	0.98	368.21	-0.29
Bocatoma San Pablo	esc20	Método híbrido: interpolación (gaps cortos) + climatología DOY (gaps largos)	0.21	48.98	-6.21	Extra Trees (o proxy)	0.17	50.26	-13.82
	esc30	Regresión lineal multirrezago + índices ENSO	0.09	54.50	-12.49	Extra Trees (o proxy)	0.19	43.83	-8.39
	esc40	Regresión lineal multirrezago + índices ENSO	0.14	38.46	-1.47	Extra Trees (o proxy)	0.19	41.44	-0.92
	esc50	Interpolación lineal temporal	0.27	40.62	-3.54	Random Forest	0.25	41.29	-5.30
Los Naranjos	esc20	Interpolación lineal temporal	0.74	54.20	0.01	Random Forest	0.75	52.88	-1.10
	esc30	Regresión lineal multirrezago + índices ENSO	0.76	42.30	0.59	Random Forest	0.77	43.62	1.09
	esc40	Método híbrido: interpolación (gaps cortos) + climatología DOY (gaps largos)	0.77	48.37	-1.66	Random Forest	0.72	52.64	-1.21
	esc50	Regresión lineal multirrezago + índices ENSO	0.78	43.58	-3.36	Extra Trees (o proxy)	0.72	51.58	-1.13

NSE = Nash–Sutcliffe Efficiency

RMSE = Root Mean Square Error

PBIAS = Percent Bias

El mejor método se seleccionó en función del mayor valor de NSE. En caso de valores similares, se priorizó el método con menor RMSE y menor valor absoluto de PBIAS.

3.1.4. Resultados por región natural

El análisis regional evidencia diferencias en el desempeño de los métodos de relleno asociadas al régimen hidrológico predominante y al tipo de sistema analizado. En la Sierra, los métodos tradicionales mantuvieron desempeños elevados incluso bajo escenarios severos de pérdida de información, mientras que los modelos de machine learning mostraron una mayor sensibilidad al incremento de datos faltantes. En la Costa, los enfoques tradicionales y semihíbridos presentaron un comportamiento relativamente más estable frente al aumento de información ausente, sin que los modelos de machine learning evidenciaran ventajas consistentes. En la Selva, el desempeño estuvo condicionado por el tipo de sistema hidrológico: en cuencas naturales se obtuvieron eficiencias elevadas para ambos enfoques, mientras que en sistemas intercuenca todos los métodos mostraron limitaciones importantes.

La Tabla 7 sintetiza los métodos con mejor desempeño relativo para cada región natural y escenario de datos faltantes, considerando métricas promedio regionales.

Tabla 7. Método con mejor desempeño relativo de relleno por región natural y escenario de porcentaje de datos faltantes

Región	Escenario	Método con mejor desempeño relativo	NSE promedio	RMSE promedio	PBIAS promedio
Costa	esc20	Método híbrido: interpolación (gaps cortos) + climatología DOY (gaps largos)	0.91	29.64	0.99
	esc30	Interpolación lineal temporal	0.93	22.08	1.24
	esc40	Interpolación lineal temporal	0.89	34.30	-1.24
	esc50	Interpolación lineal temporal	0.87	35.64	2.55
Sierra	esc20	Interpolación lineal temporal	0.97	25.09	-0.09
	esc30	Interpolación lineal temporal	0.96	26.35	0.03
	esc40	Interpolación lineal temporal	0.95	30.77	-0.32
	esc50	Regresión lineal multirezago + índices ENSO	0.94	33.55	-0.94
Selva	esc20	Interpolación lineal temporal	0.65	81.32	-2.09
	esc30	Regresión lineal multirezago + índices ENSO	0.61	99.00	0.05
	esc40	Interpolación lineal temporal	0.63	93.76	-0.99
	esc50	Regresión lineal multirezago + índices ENSO	0.68	93.30	0.30

NSE = Nash-Sutcliffe Efficiency

RMSE = Root Mean Square Error

PBIAS = Percent Bias

La selección del método con mejor desempeño relativo se realizó considerando principalmente el NSE promedio regional y utilizando RMSE y PBIAS como métricas complementarias para la interpretación del error y sesgo asociado.

3.2. Análisis visual del ajuste temporal

El análisis visual de las series reconstruidas mediante diagramas de dispersión y comparación temporal con los caudales observados permite complementar la interpretación de las métricas cuantitativas. En escenarios con baja pérdida de información, la mayoría de los métodos reproduce adecuadamente la tendencia general y la estacionalidad del caudal. Sin embargo, a medida que aumenta el porcentaje de datos faltantes, se observan mayores dispersiones y dificultades para representar ciertos comportamientos hidrológicos, particularmente en rangos de caudal con alta variabilidad.

Como caso ilustrativo, se analizó el ajuste obtenido en la estación Bellavista (cuenca del Napo) bajo el escenario esc50 utilizando el modelo de stacking. Las Figuras 2 y 3 presentan, respectivamente, la relación entre caudales observados y estimados y la distribución del error según el régimen hidrológico, permitiendo identificar patrones de error que no siempre son evidentes mediante métricas agregadas como el NSE.

Bellavista – esc50 – stacking

RMSE=363.66 | NSE=0.980 | R=0.990 | n=4623 (de 4641 S/D)

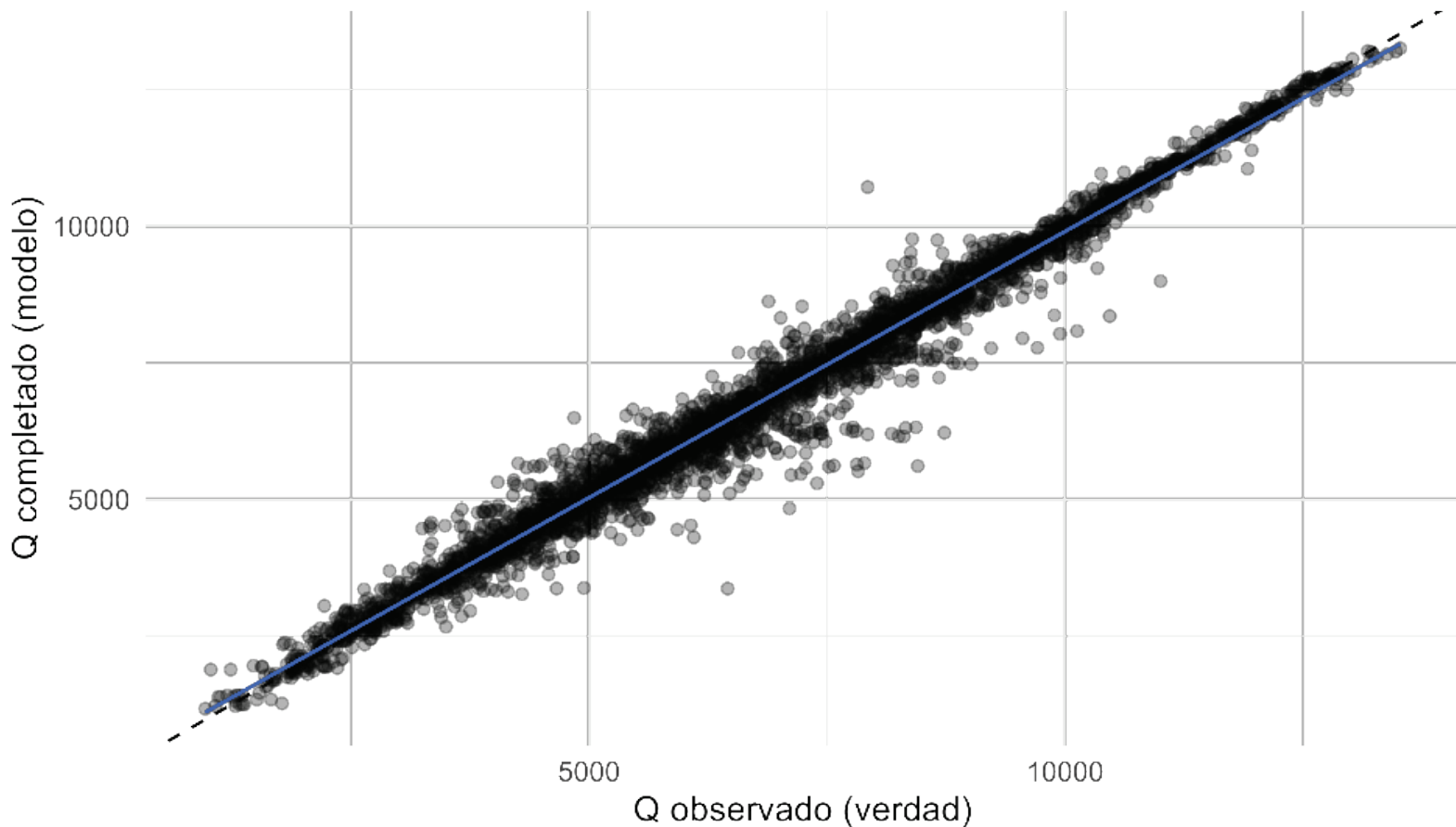


Figura 2. Diagrama de dispersión entre caudales observados y estimados mediante machine learning (modelo de stacking) para la estación Bellavista bajo el escenario esc50.

La Figura 2 muestra una relación lineal general adecuada entre los caudales observados y estimados; sin embargo, también se evidencia una dispersión creciente en determinados rangos de caudal y una tendencia a la concentración de predicciones en algunos intervalos específicos.

Error absoluto por régimen – Bellavista

Método: stacking | Escenario: esc50

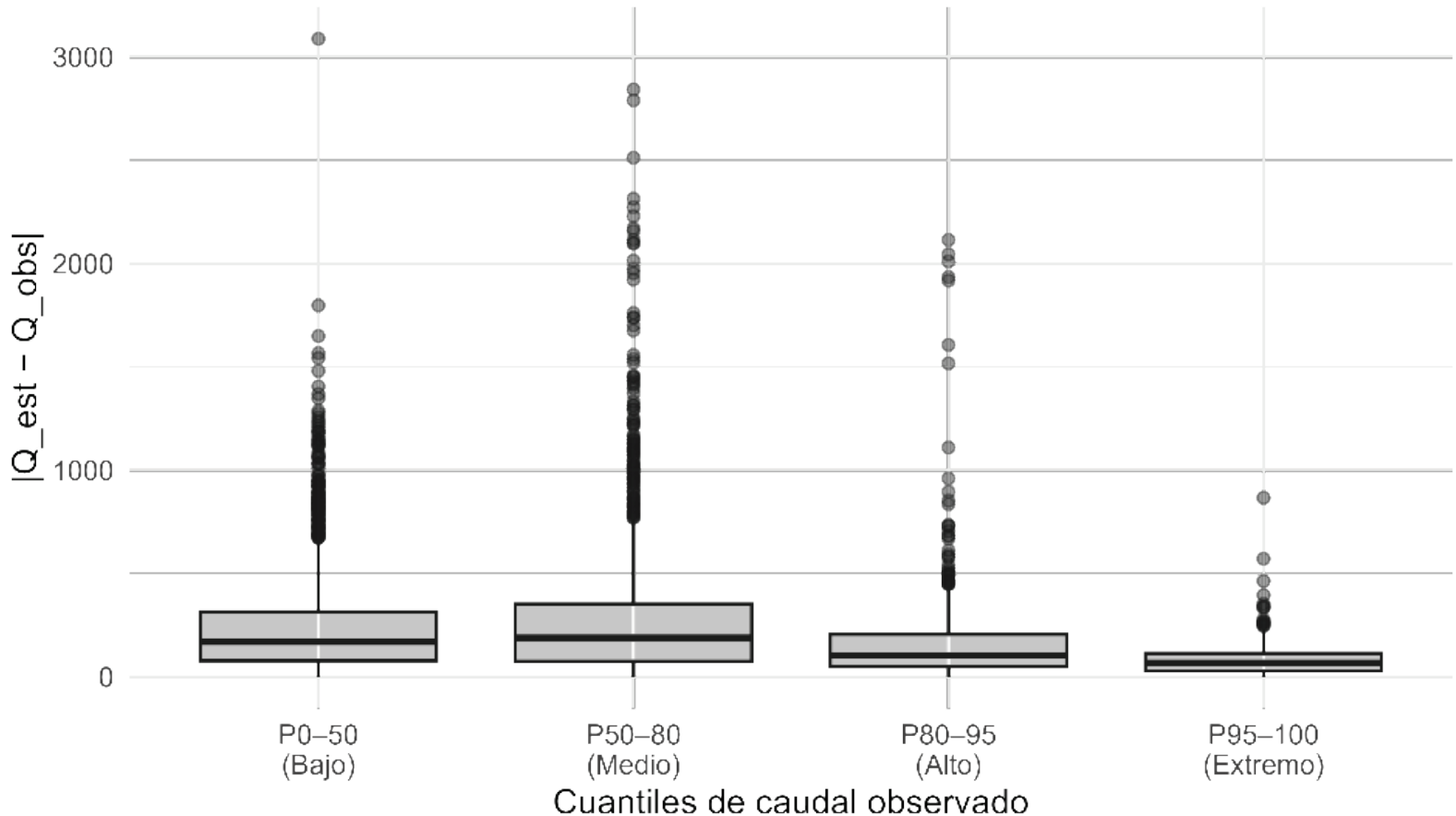


Figura 3. Distribución del error absoluto del caudal estimado por régimen hidrológico para la estación Bellavista bajo el escenario esc50.

La Figura 3 evidencia que los errores absolutos no se distribuyen uniformemente entre los distintos regímenes hidrológicos. En la estación Bellavista bajo el escenario esc50, los mayores errores se concentran principalmente en caudales bajos y medios, mientras que los eventos extremos presentan una dispersión relativamente menor. Este comportamiento sugiere que el modelo presenta mayores dificultades para representar la variabilidad cotidiana e intraestacional del sistema que

para reproducir tendencias generales asociadas a eventos de mayor magnitud. Una posible explicación es que, en sistemas amazónicos con alta variabilidad temporal y fuerte influencia de procesos hidrológicos no lineales, los predictores temporales locales utilizados por el modelo pueden no capturar completamente las fluctuaciones de corto plazo, especialmente bajo escenarios severos de pérdida de información.

4. Discusión

Los resultados obtenidos muestran que el desempeño de los métodos de relleno de datos faltantes no depende únicamente del algoritmo empleado, sino de la interacción entre el tipo de sistema hidrológico, el régimen de caudal predominante y el nivel de información disponible. Este hallazgo contrasta con la percepción generalizada de una superioridad sistemática de los enfoques de machine learning y es consistente con estudios previos que advierten sobre la variabilidad contextual del desempeño de estas técnicas (Baddoo et al., 2021; Hamzah et al., 2021).

La mayor sensibilidad de los modelos de machine learning frente al incremento de datos faltantes puede explicarse por su dependencia estructural de patrones estadísticos aprendidos durante el entrenamiento. Cuando el porcentaje de datos eliminados aumenta, el conjunto de entrenamiento se reduce y los predictores temporales utilizados, como rezagos de caudal, promedios móviles y variables de calendario, pierden representatividad respecto a la dinámica real del sistema, limitando la capacidad de generalización del modelo. Este comportamiento es particularmente pronunciado en cuencas con alta variabilidad intraestacional, como las de la Selva, donde las fluctuaciones de corto plazo no quedan bien capturadas por predictores locales cuando la

información disponible es escasa. En contraste, los métodos tradicionales se apoyan en supuestos más simples y explícitos, como continuidad temporal, estacionalidad o persistencia, que no dependen de la densidad del conjunto de entrenamiento y, por ello, mantienen mayor estabilidad bajo escenarios severos de pérdida de información (Hirsch, 1982; Moriasi et al., 2007).

Este comportamiento también explica los patrones de error observados en el análisis visual. En la estación Bellavista, bajo el escenario esc50, los mayores errores absolutos se concentraron en caudales bajos y medios, y no en los eventos extremos. Una posible causa es que los algoritmos basados en árboles tienden a interpolar adecuadamente dentro del rango de valores frecuentes en el entrenamiento, pero pierden precisión en la variabilidad cotidiana e intraestacional cuando ese rango queda submuestreado por la eliminación de datos. Los eventos extremos, al ser estadísticamente más distintivos, son identificados con mayor facilidad por los predictores disponibles, mientras que la variabilidad de caudales medios, más continua y dependiente del contexto temporal inmediato, resulta más difícil de reconstruir con información reducida.

Las diferencias entre regiones naturales no son atribuibles exclusivamente al algoritmo, sino a la estructura del régimen hidrológico y al tipo de sistema analizado. En cuencas de la Sierra, la marcada estacionalidad y la alta continuidad temporal del caudal favorecen métodos que explotan esa estructura, como la interpolación y la persistencia, independientemente de su complejidad. En la Costa, los regímenes intermitentes con eventos concentrados limitan la utilidad de los predictores de rezago empleados por los modelos de machine learning, ya que en períodos de estiaje los valores previos aportan poca información

sobre la dinámica futura. En sistemas intercuenca, como Bocatoma San Pablo, el bajo desempeño generalizado de todos los métodos refleja una limitación estructural. Cuando el caudal registrado está condicionado por derivaciones y operaciones hidráulicas externas, ningún método basado únicamente en la dinámica temporal local del caudal puede representar adecuadamente el sistema, independientemente de su sofisticación algorítmica.

Respecto a los enfoques de ensamble, la ausencia de mejoras consistentes frente a modelos individuales sugiere que el stacking y la combinación de modelos base no resuelven errores estructurales compartidos. Cuando los modelos base fallan por las mismas razones, como escasez de información o predictores insuficientes para capturar la dinámica del sistema, el meta-modelo aprende a combinar predicciones igualmente sesgadas, sin que la agregación aporte una corrección sustancial (Zounemat-Kermani et al., 2020; Zhang et al., 2018). Esto sugiere que la utilidad de los enfoques de ensamble en el relleno de datos hidrológicos está condicionada a que los modelos base capturen aspectos complementarios y no redundantes del sistema.

El aporte principal de este estudio reside en evidenciar, de manera sistemática y bajo escenarios controlados de pérdida progresiva de información, que la selección del método de relleno debe estar guiada por el contexto hidrológico y no por la complejidad del algoritmo. A diferencia de estudios que evalúan estos métodos bajo condiciones ideales o en regiones homogéneas, este trabajo incorpora la heterogeneidad hidroclimática del Perú como variable de análisis, mostrando que las conclusiones sobre el desempeño relativo de los métodos varían sustancialmente entre regiones con regímenes contrastantes. En ese sentido, los resultados aportan criterios prácticos para la selección de

técnicas para completar datos faltantes en redes hidrométricas operativas, donde la robustez y la coherencia hidrológica del método son más relevantes que su sofisticación algorítmica.

5. Conclusiones

Este estudio demuestra que la efectividad de los métodos de relleno de datos faltantes en series de caudales diarios está determinada por la interacción entre el régimen hidrológico regional, el tipo de sistema analizado y el nivel de pérdida de información, y no únicamente por la complejidad del algoritmo empleado.

Un hallazgo relevante es que métricas globales de eficiencia, como el NSE, pueden enmascarar errores sistemáticos en rangos específicos de caudal, particularmente en caudales bajos y medios, donde la variabilidad cotidiana resulta más difícil de reconstruir. Esto subraya la necesidad de evaluar los métodos de relleno mediante un conjunto complementario de métricas de eficiencia, error y sesgo, en lugar de depender de un único indicador agregado.

Desde una perspectiva aplicada, los resultados sugieren que, en redes hidrométricas operativas con recursos limitados, los métodos tradicionales adecuadamente seleccionados constituyen alternativas robustas y competitivas, cuya implementación no requiere infraestructura computacional compleja ni grandes volúmenes de datos para entrenamiento. En este contexto, la adopción de métodos de machine learning debería estar condicionada a una evaluación previa del régimen hidrológico y a la disponibilidad de información suficiente para garantizar una representación adecuada de la variabilidad del sistema.

Como limitación del estudio, los resultados se basan en series de una sola estación por cuenca, sin incorporar información espacialmente distribuida de estaciones vecinas ni variables meteorológicas complementarias, lo que representa una oportunidad de mejora para investigaciones futuras. Asimismo, la extensión de este análisis a sistemas con mayor densidad de red hidrométrica, así como la incorporación de variables fisiográficas y climáticas como factores explicativos del desempeño de los métodos, constituyen líneas de investigación que podrían contribuir a generalizar y robustecer las conclusiones aquí presentadas.

6. Agradecimientos

Se agradece a la Autoridad Nacional del Agua (ANA) por la disponibilidad pública de la información hidrométrica utilizada en este estudio, la cual se obtuvo a través de su portal institucional. Asimismo, se reconoce al Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) y a las entidades administradoras de infraestructura hidráulica —Junta de Usuarios Huallaga Central, Kallpa Generación, Proyecto Especial Chavimochic, Proyecto Especial Olmos Tinajones y Proyecto Especial Chira Piura— como fuentes originales de los registros hidrométricos que integran la base de datos pública administrada por la ANA.

7. Financiamiento

La presente investigación no recibió financiamiento externo de entidades públicas, privadas ni organismos de cooperación. Los autores desarrollaron el estudio con recursos propios y apoyo institucional correspondiente a sus actividades académicas y de investigación.



8. Declaración de conflicto de interés

Los autores declaran que no existen conflictos de interés financieros, personales, académicos ni institucionales que puedan haber influido en el desarrollo, análisis o publicación de la presente investigación.

9. Declaración de uso de IA

El coautor Jamil Ramirez utilizó herramientas de inteligencia artificial como apoyo en tareas de revisión gramatical, organización del contenido y asistencia en el desarrollo y revisión de código en R durante la elaboración del presente artículo. El uso de dichas herramientas se realizó bajo supervisión y validación académica, contando con el conocimiento y conformidad del autor principal. Tanto la metodología aplicada como el análisis, interpretación de resultados y conclusiones finales fueron revisados y aprobados íntegramente por ambos autores, quienes asumen plena responsabilidad sobre el contenido del manuscrito.

10. Referencias

- Addi, M., Gyasi-Agyei, Y., Obuobie, E., & Amekudzi, L. K. (2022). Evaluation of imputation techniques for infilling missing daily rainfall records on river basins in Ghana. *Hydrological Sciences Journal*, 67(4), 613–627. <https://doi.org/10.1080/02626667.2022.2030868>
- Arriagada, P., Karelovic, B., & Link, O. (2021). Automatic gap-filling of daily streamflow time series in data-scarce regions using a machine learning algorithm. *Journal of Hydrology*, 598, 126454. <https://doi.org/10.1016/j.jhydrol.2021.126454>

- Baddoo, T. D., Li, Z., Odai, S. N., Boni, K. R. C., Nooni, I. K., & Andam-Akorful, S. A. (2021). Comparison of missing data infilling mechanisms for recovering a real-world single station streamflow observation. *International Journal of Environmental Research and Public Health*, *18*(16), 8375. <https://doi.org/10.3390/ijerph18168375>
- Coulibaly, P., & Evora, N. D. (2007). Comparison of neural network methods for infilling missing daily weather records. *Journal of Hydrology*, *341*(1-2), 27-41. <https://doi.org/10.1016/j.jhydrol.2007.04.020>
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1-2), 80-91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Gupta, H. V., & Kling, H. (2011). On typical range, sensitivity, and normalization of mean squared error and Nash-Sutcliffe efficiency type metrics. *Water Resources Research*, *47*, W10601. <https://doi.org/10.1029/2011WR010962>
- Hamzah, F. B., Hamzah, F. M., Mohd Razali, S. F., & Samad, H. (2021). A comparison of multiple imputation methods for recovering missing data in hydrological studies. *Civil Engineering Journal*, *7*(9), 1474-1491. <https://doi.org/10.28991/cej-2021-03091747>
- Hirsch, R. M. (1982). A comparison of four streamflow record extension techniques. *Water Resources Research*, *18*(4), 1081-1088. <https://doi.org/10.1029/WR018i004p01081>

- Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.*, 5, 89–97. <https://doi.org/10.5194/adgeo-5-89-2005>
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). John Wiley & Sons. <https://doi.org/10.1002/9781119013563>
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>
- Papacharalampous, G., & Tyralis, H. (2020). Hydrological time series forecasting using simple combinations: Big data testing and investigations on one-year-ahead river flow predictability. *Journal of Hydrology*, 590, 125205. <https://doi.org/10.1016/j.jhydrol.2020.125205>
- Tyralis, H., & Papacharalampous, G. (2017). Variable selection in time series forecasting using random forests. *Algorithms*, 10(4), 114. <https://doi.org/10.3390/a10040114>
- Zhang, Z., Zhang, Q., & Singh, V. P. (2018). *Univariate streamflow forecasting using commonly used data-driven models: Literature review and case study*. *Hydrological Sciences Journal*, 63(7), 1091–1111. <https://doi.org/10.1080/02626667.2018.1469756>
- Zhou, Y., Tang, Q., & Zhao, G. (2023). Gap infilling of daily streamflow data using a machine learning algorithm (MissForest) for impact assessment of human activities. *Journal of Hydrology*, 627, 130404. <https://doi.org/10.1016/j.jhydrol.2023.130404>

Zounemat-Kermani, M., Matta, E., Cominola, A., Xia, X., Zhang, Q., Liang, Q., & Hinkelmann, R. (2020). Neurocomputing in surface water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects. *Journal of Hydrology*, 590, 125085. <https://doi.org/10.1016/j.jhydrol.2020.125085>