

Ribagua

Revista Iberoamericana del Agua

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/trib20

Simulación del proceso precipitación-escorrentía con paso diario: comparación de los modelos GR4J, SWAT y random forest

Federico Vilaseca, Santiago Narbondo, Christian Chreties, Alberto Castro & Angela Gorgoglione

To cite this article: Federico Vilaseca, Santiago Narbondo, Christian Chreties, Alberto Castro & Angela Gorgoglione (2023) Simulación del proceso precipitación-escorrentía con paso diario: comparación de los modelos GR4J, SWAT y random forest, Ribagua, 10:1, 31-47, DOI: [10.1080/23863781.2023.2238127](https://doi.org/10.1080/23863781.2023.2238127)

To link to this article: <https://doi.org/10.1080/23863781.2023.2238127>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 04 Aug 2023.



Submit your article to this journal [↗](#)



Article views: 985



View related articles [↗](#)



View Crossmark data [↗](#)

Simulación del proceso precipitación-escorrentía con paso diario: comparación de los modelos GR4J, SWAT y random forest

Federico Vilaseca ^a, Santiago Narbondo^a, Christian Chreties^a, Alberto Castro ^{b,c} and Angela Gorgoglione ^a

^aInstituto de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA), Facultad de Ingeniería, Universidad de la República, Uruguay; ^bInstituto de Computación (InCo), Facultad de Ingeniería, Universidad de la República, Uruguay; ^cInstituto de Ingeniería Eléctrica (IIE), Facultad de Ingeniería, Universidad de la República, Uruguay

RESUMEN

Un sólido estudio hidrológico diario es una tarea desafiante en regiones caracterizadas por una alta variabilidad hidro-climática, como Uruguay. Por esta razón, los modelos hidrológicos de base física de diferentes escalas temporales y espaciales (concentrados, semi-distribuidos y distribuidos) han pasado por un largo período de desarrollo y aplicación local. En los últimos años, los modelos basados en datos se están usando con éxito para resolver problemas hidrológicos. Hasta ahora, estos diferentes tipos de modelos se han estudiado individualmente para evaluar su capacidad para simular el proceso diario de precipitación-escorrentía. Este trabajo proporciona una profunda comparación entre un modelo agregado (GR4J), un modelo semi-distribuido (SWAT) y otro basado en datos (Random Forest (RF)) para simular el proceso diario de precipitación-escorrentía de dos cuencas hidrográficas ubicadas en Uruguay (una con reservorio y la otra sin). El rendimiento de cada modelo se analizó comparando numéricamente y gráficamente el caudal observado versus el simulado en términos de correspondencia temporal y cuantiles. En general, RF presenta un mejor rendimiento en comparación con los otros modelos físicamente basados. Sin embargo, carece de la capacidad de generalización que caracterizó a los otros dos enfoques. GR4J y SWAT logran un desempeño similar en nuestros casos de estudio.

Simulation of the daily rainfall-runoff process: comparison among GR4J, SWAT, and Random Forest

ABSTRACT

A robust daily hydrological assessment is a challenging task in regions characterized by high hydro-climatic variability, like Uruguay. For this reason, physically-based hydrological models of different temporal and spatial scales (lumped, semi-distributed, and distributed) have undergone a long period of development and local application. In the last years, data-driven models have been successfully adopted to solve hydrological problems. So far, these different types of models have been studied individually to test their capability in simulating the daily precipitation-runoff process. This study provides an in-depth comparison among a lumped (GR4J), semi-distributed (SWAT), and data-driven (Random Forest (RF)) model to simulate the daily rainfall-runoff process of two watersheds located in Uruguay (with and without a reservoir). The model performance was analyzed by comparing numerically and graphically observed and simulated streamflow in terms of temporal correspondence and quantiles. Overall, RF shows better performance compared to the other physically-based models. However, it lacks the generalization capability that characterized the other two approaches. GR4J and SWAT achieve similar performance at our case studies.

ARTICLE HISTORY

Received 6 September 2022
Revised 5 June 2023
Accepted 10 July 2023

PALABRAS CLAVE

GR4J; SWAT; Random Forest; Santa Lucía Chico; Escorrentía diaria

KEYWORDS

GR4J; SWAT; random forest; santa lucía chico; daily runoff

1. Introducción

Los modelos hidrológicos confiables son indispensables para la hidrología operativa y la gestión de los recursos hídricos [1]. En la literatura científica, existen varios tipos de modelos que simulan el proceso de precipitación-escorrentía. Los modelos basados en hipótesis físicas utilizan

representaciones matemáticas de diferente complejidad de los procesos hidrológicos. Estos pueden ser concentrados, semi-distribuidos o distribuidos [2–4]. En contraposición, los modelos basados en datos consisten en regresores potentes que no incorporan previamente ninguna información acerca de los procesos involucrados, sino que los interpretan a partir de los datos que le son proporcionados.

CONTACT Angela Gorgoglione  agorgoglione@fing.edu.uy  Instituto de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA), Facultad de Ingeniería, Universidad de la República

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Encontrar un modelo apropiado para cada caso de estudio todavía se considera una tarea desafiante para los hidrólogos, ya que los principales mecanismos de control del sistema hidrológico a menudo varían de un lugar a otro bajo el alcance de las escalas espaciales y temporales en uso [5].

Los modelos concentrados consideran la cuenca de estudio como una sola unidad, promediando las características espaciales relacionadas con la respuesta precipitación-escorrentía [6]. Uno de los modelos concentrados más adoptados en Uruguay es el modelo Génie Rural à 4 paramètres Journalier (GR4J) [7–9], que pertenece a la familia de modelos basados en la cuantificación de la humedad del suelo.

Los modelos distribuidos dividen la cuenca en unidades elementales, como celdas de cuadrícula, y los flujos se enrutan de una celda a otra a medida que el agua drena a través de la cuenca. Esto permite representar la heterogeneidad de la cuenca hidrográfica. La resolución de la cuadrícula generalmente se elige de modo que sea del tamaño adecuado para representar la variación espacial de los principales procesos de escorrentía (como precipitación, infiltración, etc.) [10]. También existen modelos semi-distribuidos, que discretizan la cuenca en subáreas o subcuencas homogéneas según la topografía, las características físicas de la cuenca o el área de drenaje. Los parámetros de infiltración o precipitación se tratan como homogéneos dentro de cada subcuenca [10]. El modelo Soil & Water Assessment Tool (SWAT) es uno de los modelos semi-distribuidos más adoptados en Uruguay [11–15] y es capaz de simular el proceso precipitación-escorrentía para pasos de tiempo mensuales y diarios.

A nivel internacional, se han realizado varias investigaciones comparando modelos hidrológicos de distinto tipo [16–19]. En el caso de estudio de Krajewski et al. (1991), el modelo concentrado tendía a subestimar severamente los picos de inundación en comparación con el modelo distribuido [16]. Refsgaard y Knudsen (1996) realizaron una comparación entre modelos conceptuales concentrados, semi-distribuidos y totalmente distribuidos para cuencas con escasez de datos en Zimbabue. Sus resultados no pudieron justificar el uso del modelo completamente distribuido [17]. También Anshuman et al. (2019) propusieron considerar modelos concentrados en lugar de los modelos físicamente basados cuando se analizan cuencas hidrográficas con baja disponibilidad de datos [20]. Sin embargo, en la mayoría de estos estudios se informó que los modelos distribuidos pueden en algunos casos proporcionar mejoras en comparación con los concentrados [17,18].

Debido a su capacidad para modelar los sistemas no lineales complejos, los modelos basados en datos, como los de aprendizaje automático (AA), han surgido como

una poderosa alternativa a los modelos basados en hipótesis físicas [21,22]. A diferencia de dichos modelos, los modelos de AA usan funciones matemáticas para conectar las variables de entradas a las de salidas del modelo, ignorando la relación física entre las mismas [15]. Entre varias técnicas, Random Forest (RF) es una de las más adoptadas para resolver problemas ambientales ya que tiene la ventaja de ser un método de caja blanca y es capaz de cuantificar la sensibilidad del modelo a las variables de entrada a partir del muestreo ‘*out of the bag*’, lo que permite la interpretación del proceso de aprendizaje desde el punto de vista físico [21]. En los últimos años, en la literatura científica internacional se pueden encontrar muchas aplicaciones exitosas de RF para la simulación del proceso hidrológico de precipitación-escorrentía [22–24]. También a nivel nacional, están habiendo impulsos en el uso de estas técnicas de AA para abordar problemas ambientales [9,24–27].

Hasta el momento, se han estudiado individualmente los modelos GR4J, SWAT y RF para la simulación del proceso diario precipitación-escorrentía. Sin embargo, pocos estudios han comparado los tres enfoques [28,29]. Por lo tanto, se necesita un análisis riguroso y una discusión exhaustiva para comprender cuáles son las capacidades y los inconvenientes de los tres enfoques de modelación cuando se aplican al mismo caso de estudio.

Basándonos en lo anterior, el objetivo principal de este estudio es proporcionar una comparación profunda entre un modelo agregado, GR4J, un modelo semi-distribuido, SWAT, y un modelo de AA, RF, en la simulación del proceso diario de precipitación-escorrentía a escala de cuenca. En particular, la interrogante científica que se busca responder es qué lugar ocupa un modelo basado en AA, como RF, respecto a modelos de base física agregados (GR4J) y semi-distribuidos (SWAT) si se analiza su desempeño en términos de correspondencia temporal y cuantiles. Los tres modelos serán implementados en las mismas áreas de estudio para una mejor comparación.

2. Materiales y métodos

2.1. Área de estudio

La cuenca del río Santa Lucía Chico (SLC) (Figura 1) fue seleccionada como área de estudio debido a su importancia estratégica para Uruguay. Es la principal fuente de agua potable del país, ya que destina agua para el 60% de la población uruguaya y también sustenta varias actividades agroindustriales, siendo uno de los principales polos de producción de alimentos [30–32]. En la misma se ubica el reservorio de Paso Severino (PS), un embalse de 15 km² de superficie, 20 m de

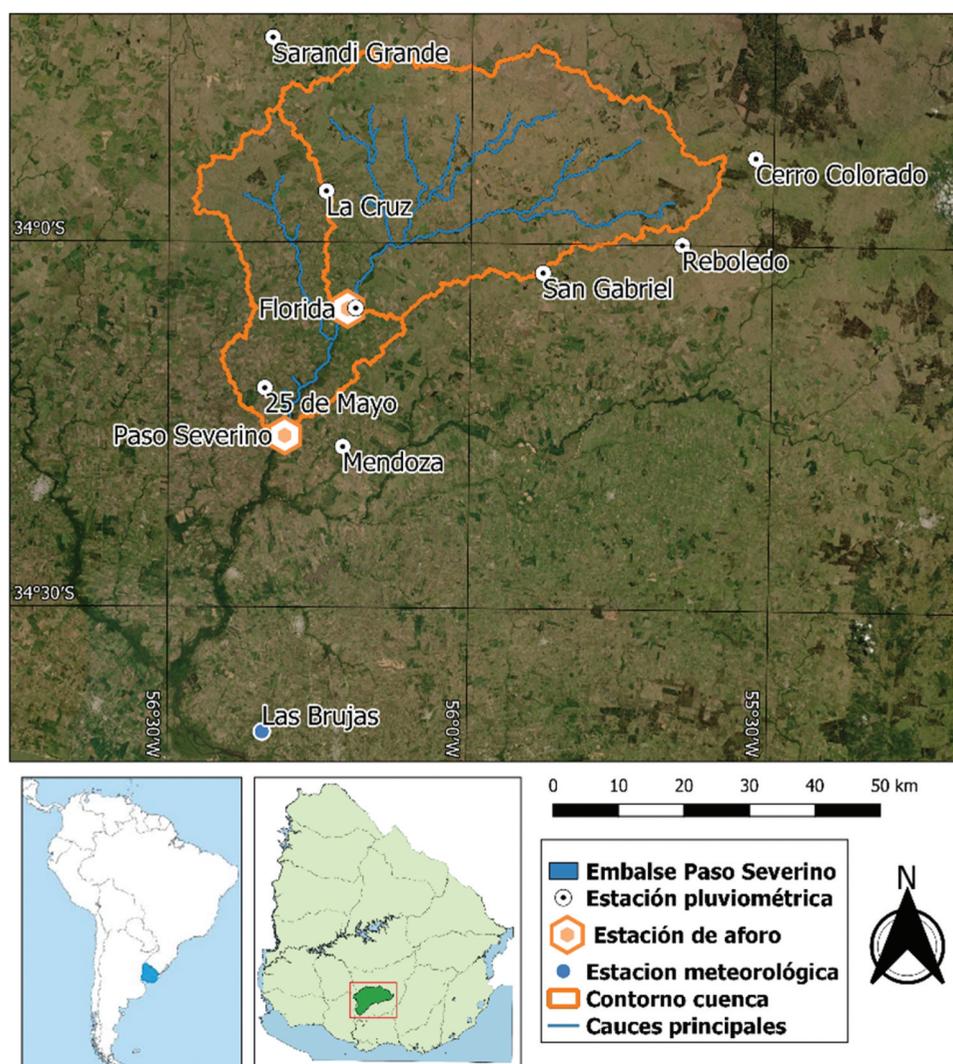


Figura 1. Cuencas de estudio con cierre en Florida (sin reservorio) y Paso Severino (con reservorio), ubicadas en la cuenca del río Santa Lucía Chico.

profundidad máxima y con una capacidad de almacenamiento de 65 hm^3 [30].

La cuenca se encuentra en la región centro-sur de Uruguay, cubriendo un área igual a 2478 km^2 . Su principal centro urbano es la ciudad de Florida y las principales coberturas del suelo son pastizales (82.4%) y agricultura (9.4%). El clima es templado con una precipitación total anual que oscila entre 1000 mm y 1500 mm y una temperatura del aire que varía entre 3°C y 30°C [33].

Esta cuenca fue elegida no solo por su importancia estratégica sino también porque es un sistema mixto lótico y léntico [34]. Resulta de interés llevar a cabo el objetivo principal de este estudio en dos subcuencas, que son hidrológicamente diferentes entre sí. Por esta razón, se consideraron como puntos de cierre las dos estaciones de monitoreo de caudales ubicadas en la cuenca del SLC, los cuales definen una cuenca aguas arriba con punto de cierre en la ciudad de Florida (FL), caracterizada por un

sistema lótico con una respuesta hidrológica más rápida; y toda el área de estudio, con punto de cierre en la desembocadura de la presa PS, que también incluye el sistema léntico del embalse de PS (Figura 1).

2.2. Datos disponibles

2.2.1. Variables meteorológicas

Los registros de precipitación diaria acumulada (P) fueron proporcionados por el Instituto Uruguayo de Meteorología (INUMET). Se consideró la P registrada en ocho estaciones pluviométricas ubicadas cerca de la cuenca en La Cruz, Sarandí Grande, San Gabriel, 25 de Mayo, Reboledo, Cerro Colorado, Florida y Mendoza (Figura 1). Florida es la única estación meteorológica con registros desde el 01/05/1989. El resto de las estaciones son pluviométricas y tienen observaciones que cubren el período 01/01/1980 – 30/06/2020.

También se consideraron series temporales de humedad relativa (HR), velocidad del viento (WS), radiación solar (SR), temperaturas diarias mínimas y máximas (T_{min} y T_{max}), y evapotranspiración potencial (ETP) calculada con el método de Penman-Monteith (período del 01/01/1980 al 30/06/2020). Dichas variables fueron proporcionadas por el Instituto Nacional de Investigaciones Agropecuarias (INIA). En particular, se adoptaron las observaciones registradas en la estación meteorológica Las Brujas, ya que es la más cercana al sitio en estudio entre las disponibles (Lat: -34.67 , Lon: -56.34).

2.2.2. Caudal

El caudal medio diario (Q) considerado en este trabajo corresponde al período comprendido entre el 01/01/1990 y el 31/12/2015. Fue registrado en la estación ubicada en la ciudad de Florida por la Dirección Nacional de Aguas (DINAGUA) de Uruguay. Además, se utilizó la información de la descarga media diaria de la represa PS que cubre el período 01/01/1990 – 31/12/2015 (Figura 1). Los datos fueron registrados y proporcionados por la empresa nacional responsable de las operaciones de gestión de presas y distribución de agua potable (Obras Sanitarias del Estado, OSE).

2.2.3. Información espacial

La información espacial de topografía, uso/cobertura del suelo y tipo de suelo fue considerada para este estudio, particularmente para el modelo SWAT. El modelo digital del terreno (MDT) se obtuvo de un mapa topográfico nacional (escala 1:50,000). El mismo indica que las pendientes de la cuenca son bajas, con menos del 5% para el 77% de la región y más del 10% para sólo el 2% de la misma. El mapa actual del uso/cobertura del suelo (año 2018) fue proporcionado por el Ministerio de Ganadería, Agricultura y Pesca (MGAP) (escala 1:50,000). Los usos del suelo incluyen pastizales (82.4%), agricultura (9.4%), silvicultura (4.9%), ganadería lechera (1.2%), áreas urbanas (0.9%), cuerpos de agua (0.7%) y humedales (0.5%). La clasificación de suelos que considera el mapa de suelos tiene en cuenta el CONEAT, que es el índice de productividad nacional [33]. En la cuenca de estudio se pueden encontrar diecinueve tipos de suelo, principalmente suelos arcillosos pertenecientes a los grupos hidrológicos B y C. En la Figura MS1, y la Figura MS2, ubicadas en el Material Suplementario, se pueden encontrar graficados los mapas de usos y tipos de suelo respectivamente.

2.3. Modelos

2.3.1. Modelo conceptual agregado: GR4J

El modelo GR4J es un modelo conceptual de precipitación-escorrentía propuesto por Perrin et al. [34], con paso

diario, caracterizado por el cálculo de la humedad del suelo. El mismo, toma como datos de entrada la P y la evapotranspiración potencial (ETP) para luego modelar el almacenamiento en el suelo teniendo en cuenta dos reservorios. En este caso se ingresa la serie de P promediada para el área de la cuenca con el método de Thiessen, mientras que la ETP es la determinada por INIA para la estación de Las Brujas. A su vez, el modelo transita una parte del escurrimiento mediante un hidrograma unitario simple y el resto a través de un hidrograma unitario que se almacena en un reservorio no lineal para finalmente con ambas formar la escorrentía total. En la Figura 2, se presenta el esquema del modelo y a continuación se lo describe brevemente.

Primero, con un reservorio de intercepción de capacidad cero, en el modelo se calcula la P efectiva (P_n) o la evapotranspiración neta (E_n), dependiendo de si la P es mayor a la E o viceversa:

$$\begin{cases} \text{Si } P \geq E; P_n = P - E \text{ y } E_n = 0 \\ \text{Si } P < E; P_n = 0 \text{ y } E_n = E - P \end{cases} \quad (1)$$

En caso de que P_n o E_n sea mayor a 0, una fracción de la misma (P_s o E_s respectivamente) aporta al reservorio de

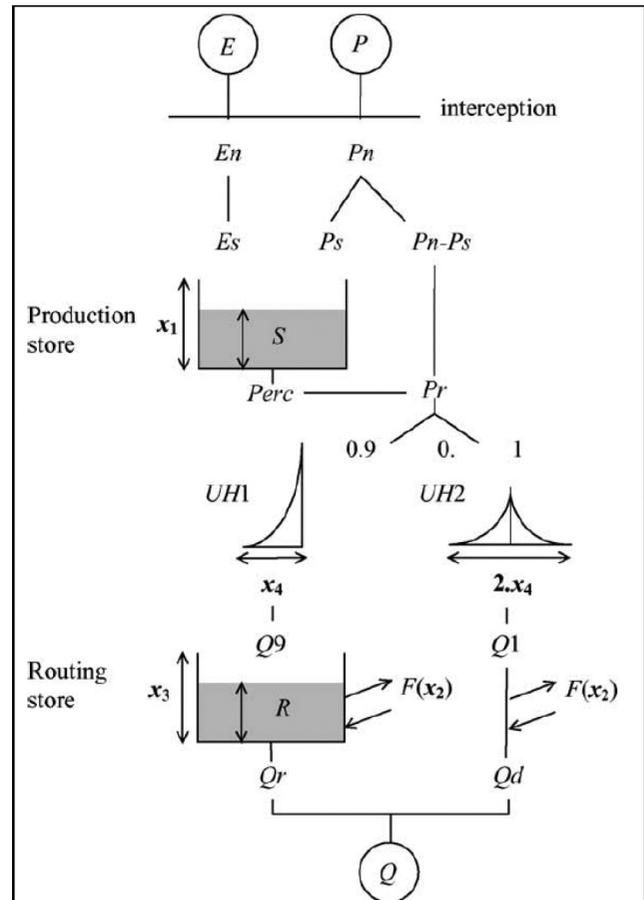


Figura 2. Esquema del modelo GR4J [34].

producción (capa sub-superficial del suelo), la cual es función del contenido de agua en el suelo (S) y de x_1 (mm), un parámetro del modelo que representa la capacidad máxima de almacenamiento del suelo (es decir, del reservorio de producción).

Una vez calculado P_s o E_s , el contenido de agua en el suelo es actualizado con: $S = S - E_s + P_s$. A continuación, se estima la percolación ($Perc$) hacia la zona saturada del suelo y se actualiza nuevamente el contenido de agua en el suelo ($S = S - Perc$).

Luego, la suma del exceso de almacenamiento de humedad del suelo y la P_n genera la escorrentía total ($Q_t = Perc + [P_n - P_s]$) que se divide en dos componentes: una correspondiente a la escorrentía subsuperficial o flujo lento, que transita el 90% de la lluvia efectiva con un hidrograma unitario y un reservorio no lineal; la otra correspondiente a la escorrentía directa, que transita el 10% restante con un segundo hidrograma unitario.

El tiempo base de ambos hidrogramas unitarios se representa a través de un único parámetro x_4 (días) y la capacidad de almacenamiento del reservorio no lineal constituye otro parámetro, x_3 (mm).

Finalmente, el modelo utiliza una función de flujo de agua subterránea entre cuencas para calcular las ganancias o pérdidas derivadas de la interacción entre la cuenca y las cuencas vecinas, a través de una función dependiente de un cuarto parámetro x_2 (mm). El caudal total en el punto de descarga es la suma de los componentes antes mencionados [35]. Para este estudio, el modelo fue implementado en MATLAB.

2.3.2. Modelo distribuido físicamente basado: SWAT

El modelo SWAT, propuesto por Arnold et al. [3], es un modelo hidrológico semi-distribuido ampliamente utilizado, que ha demostrado ser preciso en la representación del proceso precipitación-escorrentía para paso de tiempo diario y mensual. Tiene un esquema de cálculo complejo ya que considera la distribución espacial de suelos, usos de suelo y topografía, y caracteriza el proceso lluvia-escorrentía a través de más de 30 parámetros. Consecuentemente, la superficie de la cuenca se divide en dos escalas: *i) subcuencas*, que se determinan a partir de la distribución de la red de drenaje superficial, y *ii) unidades de respuesta hidrológica* (HRU por su sigla en inglés), que son una subdivisión de las subcuencas, determinada por la intersección de los mapas de uso de suelo, tipo de suelo y pendiente. Se calcula la escorrentía para cada HRU y luego se agrega el resultado para cada subcuenca y posteriormente para toda la cuenca. Además, acopla modelos para simular la generación, movilización y transporte de nutrientes y sedimentos [9,36]. El balance hidrológico se calcula con la siguiente ecuación:

$$S_t = S_0 + \sum_{i=1}^t (P_{día} - Q_{sup} - E - Perc - Q_{gw}) \quad (3)$$

donde S_t representa el contenido final de agua (mm), S_0 es el contenido inicial de agua en el suelo en el día i (mm), t es el tiempo (días), $P_{día}$ es la precipitación en el día i (mm), Q_{sup} representa la escorrentía superficial en el día i (mm), E es la evapotranspiración en el día i (mm), $Perc$ es la cantidad de agua que ingresa a la zona vadosa desde el suelo (mm), Q_{gw} es la cantidad de agua subterránea en el día i (mm).

Para la implementación del modelo se utilizó QSWAT (herramienta basada en QGIS). La configuración del modelo final consta de 20 subcuencas y 317 HRU. SWAT comprende varios métodos para calcular los procesos hidrológicos. En este estudio, el cálculo de la escorrentía superficial diaria se basó en el método del Número de Curva (CN) del Servicio de Conservación de Suelos (SCS). Para calcular la evapotranspiración, se utilizó la opción de Penman-Monteith, y se eligió el método de Muskingum para representar el tránsito fluvial desde la salida de las subcuencas hasta el punto de descarga de la cuenca. La evaporación potencial del agua del suelo se estimó en función de la evapotranspiración potencial y del índice de área foliar. La evaporación real del suelo se calculó usando funciones exponenciales de profundidad del suelo y contenido de agua. La evaporación del agua de las plantas se simuló con una función lineal de la evapotranspiración potencial, el índice de área foliar y la profundidad de la raíz, y está limitada por el contenido de agua del suelo. La información sobre el manejo del suelo se recuperó de un modelo SWAT implementado previamente en una cuenca limítrofe [12]. Los datos meteorológicos de entrada incluyen HR , SR , T_{max} , T_{min} y WS de la estación INIA Las Brujas y P en las ocho estaciones mencionadas anteriormente (Figura 1). El embalse PS, ubicado cerca del punto de cierre de la cuenca, fue considerado en el modelo como un embalse no controlado con una tasa de descarga anual promedio, estimada a partir de los datos del Q erogado por la represa.

2.3.3. Modelo de aprendizaje automático: random forest

El modelo RF, introducido por Brieman [37], es un algoritmo de AA supervisado capaz de representar relaciones no lineales entre las variables de entrada y salida. Es un método de ensamble, es decir, se compone de muchos árboles de decisión (*weak learners*) que se utilizan para predecir un valor (regresión) o una clase (clasificación). Su respuesta es el

promedio de los valores predichos (para regresión) o la clase más predicha (para clasificación). En este trabajo, se explotó la capacidad de este algoritmo como regresor.

Muchos métodos estadísticos clásicos (incluidos los métodos de árboles de decisión) suelen presentar una tendencia a sobreajustarse (*overfitting*). Esto quiere decir que tienden a aprender muy bien los datos de entrenamiento, pero sin lograr una buena generalización. Una forma de mejorar dicha generalización (o sea, bajar el *overfitting*), es combinar modelos diferentes, de modo que cada uno funcione de forma distinta y sus errores tiendan a compensarse. Como se mencionó anteriormente, un RF es un ensamble de árboles de decisión. Para conseguir que los errores se compensen entre sí y la probabilidad de sobreajuste sea baja, se emplea el método *bootstrap aggregation (bagging)* [38]. Con este método, cada árbol de decisión se entrena con un subconjunto del conjunto de entrenamiento. Estos subconjuntos se forman eligiendo muestras aleatoriamente (con repetición) del conjunto de entrenamiento. Luego la predicción final es el promedio de las predicciones de los árboles del bosque. Está comprobado que este método otorga una muy buena precisión al tiempo que reduce las posibilidades de sobreajuste o la introducción de sesgos. A continuación, se proporciona una breve descripción de este enfoque [38].

Dado un conjunto de entrenamiento E de n muestras (filas) y m variables (columnas), el algoritmo de RF para regresión es el siguiente:

- *Bagging*: Se crean b conjuntos de muestras aleatorias E^* de tamaño n con reemplazo de todo el conjunto de entrenamiento.
- Entrenamiento: se entrenan b árboles de decisión, cada uno con un conjunto de muestras aleatorias distinto, repitiendo los siguientes pasos para cada nodo del árbol de decisión hasta que se alcance el tamaño mínimo de nodo:
 - seleccionar un subconjunto de variables de tamaño m^* del conjunto completo;
 - elegir la mejor variable de división junto con el mejor valor de división en el sentido de minimizar el error cuadrático medio del subconjunto de variables seleccionado en el último subpaso;
 - dividir el nodo en dos ramas según la mejor variable/punto de división seleccionado en el último subpaso.
- Predicción: El valor de la predicción del RF es el promedio de los valores observados en los nodos terminales de los árboles de decisión, es decir, el

promedio de las predicciones de los árboles de decisión que forman parte del ensamble.

Se consideraron las mismas variables hidrometeorológicas de entrada que requiere el modelo SWAT, para que la comparación sea válida. Además, se incorporaron otras variables artificiales que derivan de las primeras, que permitieron facilitar al algoritmo la interpretación de los datos y de los procesos que se quieren simular. Dichas variables artificiales fueron: el acumulado de P en los últimos 7 días (P_{acum}), Q desfasado de 1, 2 y 3 días (Q_{t-1} , Q_{t-2} y Q_{t-3} , respectivamente) y la media móvil ponderada exponencialmente (EWMA por sus siglas en inglés) del resto de las variables climáticas (p.e., HR_{ewma}). El modelo fue implementado en *Python*, utilizando la biblioteca *scikit-learn* [39].

2.4. Calibración/entrenamiento y validación/testeo de los modelos

2.4.1. Desempeño de los modelos

Se comparó el desempeño de los tres modelos en términos de correspondencia temporal y cuantiles. Lograr una buena correspondencia temporal es importante para validar la robustez del modelo hidrológico pensando en posibles aplicaciones de operación en tiempo real. Esto se evaluó con una representación gráfica de las series temporales observadas y simuladas, así como también con gráficos de dispersión de caudales observados vs. caudales simulados, para cada uno de los tres modelos, y en los dos sitios de interés. Esto fue realizado para las salidas de paso diario y también para sus acumulados mensuales en el período de estudio. Una buena simulación de los cuantiles correspondientes a los caudales altos y bajos es crítica para evaluar la precisión del modelo en eventuales aplicaciones para evaluación del impacto del cambio climático o gestión de los recursos hídricos. Para identificar diferentes intervalos de caudales, se utilizaron las curvas de permanencia, que pueden considerarse como indicadores generales de las condiciones hidrológicas. Los intervalos de las curvas de permanencia se pueden agrupar en varias categorías [21]. Un enfoque estándar es dividir dichas curvas en cinco zonas: caudales pico (0–10%), caudales altos (10–40%), caudales medios (40–60%), caudales bajos (60–90%) y caudales de sequía o estiaje (90–100%) [40]. Por lo tanto, se evaluó el desempeño de los tres modelos, en términos de cuantiles, para estas cinco condiciones de caudal.

Para proceder con la evaluación del desempeño global (correspondencia temporal y cuantiles) de los modelos, se calcularon la eficiencia de Nash-Sutcliffe (NSE), el porcentaje de sesgo (PBIAS) y el estadístico conocido como RSR, que surge de dividir la raíz del error

cuadrático medio entre la desviación estándar de los datos observados. Estos indicadores reportan el desempeño general de los modelos, ponderando equitativamente todo el rango de caudales. A ellos se agregaron dos indicadores más, la eficiencia de Nash-Sutcliffe logarítmica (LNSE) y la eficiencia de Nash-Sutcliffe de picos (PNSE), que otorgan una ponderación mayor a los caudales de estiaje y caudales pico respectivamente, permitiendo evaluar con mayor precisión la performance de los modelos en esos rangos.

La función objetivo utilizada para la calibración/entrenamiento y validación/testeo de los tres modelos fue NSE:

$$NSE = 1 - \frac{\sum_{i=1}^n n(O_i - P_i)^2}{\sum_{i=1}^n n(O_i - \bar{O})^2} \quad (4)$$

donde P_i representa el valor simulado, O_i es el valor observado y \bar{O} es el promedio de los valores observados.

NSE es una estadística normalizada que expresa la magnitud relativa de la varianza de los residuos en comparación con la varianza de los datos medidos. NSE varía entre $-\infty$ y 1: NSE = 1 es el valor óptimo, los valores negativos significan que la media de los datos observados es un mejor predictor que los datos simulados. Es conocido que el valor de este indicador puede tener alta incertidumbre cuando se utiliza para la comparación de series hidrológicas, ya que el mismo puede verse influenciado significativamente por unos pocos valores altos de caudal [28]. Es por esto que se determinó su incertidumbre de muestreo mediante el método de *jackknife-after-bootstrap* propuesto por [28]. Se hizo únicamente con el NSE por tratarse de la función objetivo de calibración. Para el cálculo se utilizó el paquete de código 'gumboot' de R [41].

El desempeño global de los modelos también fue evaluado utilizando el PBIAS y el RSR:

$$PBIAS = 100 * \frac{\sum_{i=1}^n n(O_i - P_i)}{\sum_{i=1}^n nO_i} \quad (5)$$

$$RSR = \frac{\sqrt{\sum_{i=1}^n n(O_i - P_i)^2}}{\sqrt{\sum_{i=1}^n n(O_i - \bar{O})^2}} \quad (6)$$

PBIAS mide la tendencia promedio de los datos simulados a ser más grandes o más pequeños que los respectivos datos observados. PBIAS = 0 representa el valor

óptimo. Valores bajos indican una simulación más precisa del modelo. Los valores positivos indican un sesgo de subestimación del modelo y los valores negativos indican un sesgo de sobreestimación del mismo.

RSR incluye los beneficios de los índices estadísticos de error (RMSE) incluyendo el ajuste de escala/normalización (al dividir por la desviación estándar). Su valor óptimo es RSR = 0 que implica cero variación residual. Entre menor sea su valor, mejor el desempeño del modelo.

El desempeño para caudales bajos se evaluó mediante LNSE, que consiste en el cálculo del NSE, tal como se indica en la Ecuación 4, pero aplicando el logaritmo natural a las series de caudales observados y simulados.

Como indicador de caudales altos, se utilizó PNSE, el cual también se calcula de forma idéntica al NSE (Ecuación 4), salvo que se considera O_i como los valores observados de caudal pico (máximos locales de la serie) y P_i sus correspondientes temporales en la serie simulada. Cabe destacar que, para que sea un indicador fiel del desempeño para caudales elevados, se consideraron únicamente los valores que superaran el percentil 90 de la serie observada (este valor es aproximadamente 61 m³/s en Paso Severino y 33 m³/s en Florida).

La evaluación de los tres modelos se basó sobre las calificaciones generales de desempeño de un modelo hidrológico de paso mensual introducidas por Moriasi et al. [42] (Tabla 1). Para LNSE y PNSE se utilizó la misma calificación de NSE [43–45]. Los índices de desempeño de los modelos se resumen en la Tabla 2.

2.4.2. Calibración y validación del modelo GR4J

La calibración del modelo GR4J se realizó de forma manual, empleando un procedimiento iterativo, que consiste en ir cambiando los cuatro parámetros del modelo dentro de su rango convencional de variación y comparando (numérica y gráficamente) el hidrograma simulado con el observado tanto en Florida como en PS. Si bien este procedimiento tiene sus limitaciones, ya que no explora en su totalidad el espacio de parámetros, se entendió adecuada su utilización en este caso en que el modelo utilizado es sencillo (solamente tiene 4 parámetros) y se encuentra regionalizado para todo el Uruguay [7]. Los cuatro parámetros del modelo con los respectivos rangos de variación se resumen en la Tabla 3. Se consideró el NSE como función objetivo y se evaluaron también PBIAS y RSR. El período de calibración fue del

Tabla 1. Calificaciones generales de desempeño de un modelo hidrológico de paso mensual [28].

Nivel de desempeño	NSE	PBIAS	RSR
Muy bueno	0.75 < NSE ≤ 1.00	PBIAS < ±10	0.00 < RSR ≤ 0.50
Bueno	0.65 < NSE ≤ 0.75	±10 ≤ PBIAS < ±15	0.50 < RSR ≤ 0.60
Satisfactorio	0.50 < NSE ≤ 0.65	±15 ≤ PBIAS < ±25	0.60 < RSR ≤ 0.70
Insatisfactorio	NSE ≤ 0.50	PBIAS ≥ ±25	RSR > 0.70

Tabla 2. Índices de desempeño de los modelos para cada criterio de comparación.

Indicador de performance	Criterios de comparación	
	Correspondencia temporal	Cuantiles
NSE	X	X
PBIAS	X	X
RSR	X	X
LNSE	X	X
PNSE	X	X
Series temporales	X	
Gráficos de dispersión	X	
Curvas de permanencia		X

17/01/1989 al 31/12/2008. Después del proceso de calibración, el modelo GR4J fue validado del 01/01/2009 al 06/05/2016, considerando el mejor conjunto de parámetros seleccionados durante la calibración.

2.4.3. Calibración y validación del modelo SWAT

Se utilizó el software SWAT-CUP para el proceso de calibración y validación del modelo SWAT. El mismo dispone del algoritmo SUFI-2 [2], una técnica de optimización que permite calibrar el modelo y cuyos resultados se pueden usar para calcular indicadores de sensibilidad para cada uno de los parámetros utilizados. Se aplicó el mismo inicialmente con una amplia selección de parámetros. Luego se eligieron los 10 de mayor sensibilidad para continuar trabajando (Tabla 4). La descripción del análisis de sensibilidad, junto con sus resultados, se presentan en el Material Suplementario (Tabla MS 1).

La variable de respuesta del modelo fue el caudal observado tanto en Florida como en PS. Se consideró el NSE como función objetivo y se calcularon también PBIAS y RSR. La calibración fue multiobjetivo,

buscando optimizar el conjunto de parámetros seleccionados para que el desempeño conjunto sea óptimo en ambos puntos de cierre (PS y FL). Para esto se establecieron dos tipos de variación de los parámetros: 'Valor,' para la cual el algoritmo de optimización (SUFI-2) itera cambiando el valor del parámetro y 'Porcentual,' en la cual se modifica el valor del parámetro sumando un porcentaje (positivo o negativo) por sobre su valor inicial. El primer método de variación se utilizó en parámetros que son representativos de toda la cuenca, mientras que el segundo se aplicó a aquellos que varían espacialmente. El período de calibración fue el mismo usado para el modelo GR4J (del 17/01/1989 al 31/12/2008). Posteriormente, utilizando los parámetros óptimos resultantes de la calibración, SWAT fue validado con una sola corrida para el período 01/01/2009 – 06/05/2016.

2.4.4. Entrenamiento y testeo del modelo random forest

Para el desarrollo del modelo RF, el conjunto de datos se dividió en subconjuntos de entrenamiento (75% del total de los datos) y de evaluación (el restante 25%), correspondientes respectivamente a los datos considerados para la calibración y validación de los modelos GR4J y SWAT. Se adoptó una validación cruzada de 5 veces (*5-fold cross validation*) para el entrenamiento. Durante la validación cruzada se ejecutó un proceso de optimización de los hiperparámetros del algoritmo utilizando la biblioteca *Optuna*, implementada en *Python* [41]. Se consideraron dos hiperparámetros para la validación cruzada del modelo: *max_features*, que representa el número máximo de variables (*features*) que

Tabla 3. Parámetros de calibración del modelo GR4J.

Parámetro	Descripción	Rango de variación
x_1 (mm)	Capacidad máxima del primer reservorio	[50, 130]
x_2 (mm)	Coefficiente de intercambio con el agua subterránea	[-3, 2]
x_3 (mm)	Capacidad máxima del segundo reservorio	[30, 90]
x_4 (días)	Tiempo base del hidrograma unitario	[1.7, 3.3]

Tabla 4. Parámetros de calibración del modelo SWAT.

Parámetro	Descripción	Tipo de variación	Rango de variación
ALPHA_BF.gw	Constante de recesión del flujo base (1/días)	Valor	0.1 a 1
GW_DELAY.gw	Tiempo de desfase del agua subterránea (días)	Valor	0 a 75
GWQMN.gw	Umbral de contenido de agua en el depósito subterráneo para que se produzca el flujo de retorno (mm)	Valor	1500 a 15,000
CH_N2.rte	Coefficiente de rugosidad de Manning en cauces (fuera de las subcuencas) ($s/m^{1/3}$)	Valor	0.05 a 0.15
CH_N1.sub	Coefficiente de rugosidad de Manning en cauces (dentro de las subcuencas) ($s/m^{1/3}$)	Valor	0.02 a 0.1
OV_N.hru	Coefficiente de rugosidad de Manning de la superficie para flujos no concentrados ($s/m^{1/3}$)	Valor	0.04 a 0.4
SURLAG.bsn	Tiempo de desfase del agua superficial (días)	Valor	2 a 12
CN2.mgt	Número de curva asociado a la cobertura del suelo	Porcentual	-50% a 50%
SOL_AWC.sol	Valor de Agua Disponible del suelo (mm H ₂ O/mm suelo)	Porcentual	-50% a 100%
SOL_K.sol	Conductividad hidráulica saturada del suelo (mm/h)	Porcentual	-50% a 50%

Tabla 5. Parámetros de entrenamiento del modelo RF.

Parámetro	Descripción	Rango de variación
max_features	Número máximo de variables que puede considerar un árbol (porcentaje del total)	10% – 100%
max_depth	Profundidad máxima de cada árbol (cantidad de niveles)	2 a 20

puede considerar un árbol y *max_depth*, que corresponde a la profundidad máxima de cada árbol (Tabla 5). *max_features* controla la cantidad de variables que se seleccionan en busca de la mejor división, mientras que *max_depth* controla el crecimiento del árbol de decisiones: cuanto más profundo es el árbol, más divisiones tiene y captura más información sobre los datos. Después de identificar el mejor conjunto de hiperparámetros, se generó un modelo utilizando sus valores óptimos. Luego se calculó el desempeño del mismo para el conjunto de datos de testeo. Así como para los modelos anteriores, se utilizó NSE como función objetivo y se calcularon también PBIAS y RSR como indicadores de desempeño.

3. Resultados

3.1. Optimización de los parámetros de calibración/entrenamiento

Los valores óptimos de los parámetros de calibración/entrenamiento de los tres modelos se muestran en la

Tabla 6. Para el modelo SWAT, los parámetros que representan la rugosidad de los cauces (CH_N1 y CH_N2) y el número de curva (CN2) asumen valores razonables en comparación a los valores propuestos en la literatura científica. Se nota también que los parámetros que se refieren al contenido de agua en el suelo, al agua subterránea y al intercambio entre agua subterránea y escorrentía superficial (SOL_AWC, ALPHA_BF, GW_DELAY, GWQMN) tienden a valores altos del rango de variación reportado en la Tabla 4.

Considerando los parámetros del modelo GR4J, se puede observar que los rangos de variación mostrados en la Tabla 3 son similares a los rangos de variación propuestos por Perrin et al. ($x_1 \in [100, -1200]$; $x_2 \in [-5, 3]$; $x_3 \in [20, 300]$; $x_4 \in [1.1, 2.9]$) [3]. En particular, los valores de x_1 son bajos para ambas cuencas. Esto se justifica por el hecho de que los suelos uruguayos no son espesos y se caracterizan por arcillas expandidas formadas por minerales que con el agua aumentan su volumen y reducen la capacidad de almacenamiento del suelo. Consecuentemente, los suelos uruguayos se

Tabla 6. Parámetros óptimos de calibración/entrenamiento de los tres modelos.

		Parámetro	Unidad	Valor óptimo	
Paso Severino	SWAT	ALPHA_BF.gw	1/días	0.83	
		GW_DELAY.gw	días	69	
		GWQMN.gw	mm	14,175	
		CH_N2.rte	s/m ^{1/3}	0.1	
		CH_N1.sub	s/m ^{1/3}	0.027	
		OV_N.hru	s/m ^{1/3}	0.086	
		SURLAG.bsn	días	8.3	
		CN2.mgt		-0.16	
		SOL_AWC.sol	mm H ₂ O/ mm suelo	0.99	
		SOL_K.sol	mm/h	0.36	
		GR4J	x1	mm	120
			x2	mm	-3
			x3	mm	30
			x4	días	3.3
RF	max_features		1		
	max_depth		9		
Florida	SWAT	ALPHA_BF.gw	1/días	0.83	
		GW_DELAY.gw	días	69	
		GWQMN.gw	mm	14,175	
		CH_N2.rte	s/m ^{1/3}	0.1	
		CH_N1.sub	s/m ^{1/3}	0.027	
		OV_N.hru	s/m ^{1/3}	0.09	
		SURLAG.bsn	días	8.3	
		CN2.mgt		-0.16	
		SOL_AWC.sol	mm H ₂ O/ mm suelo	0.99	
		SOL_K.sol	mm/h	0.36	
		GR4J	x1	mm	99
			x2	mm	-3
			x3	mm	40
			x4	días	2.1
RF	max_features		1		
	max_depth		12		

caracterizan por una baja disponibilidad agua en el suelo [7]. Los valores de x_4 están en el borde superior de su rango (Tabla 3). Esto depende de la superficie de la cuenca: cuanto mayor es el área, mayor es la influencia de los pequeños almacenamientos de la cuenca, mayor es el valor de x_4 .

Enfocándonos en los parámetros del modelo RF, parece que $max_features$ depende más de los datos de entrada del modelo que de las características de la cuenca de estudio, en cuanto asume el mismo valor para las dos

cuenas consideradas en este trabajo. Diferente comportamiento parece tener el parámetro max_depth , el cual asume valores diferentes para las dos cuencas.

3.2. Desempeño de los modelos en términos de correspondencia temporal

A continuación, se presentan en forma comparativa los resultados obtenidos para los tres modelos implementados. En particular, en la Figura 3 y la Figura 4 se puede ver

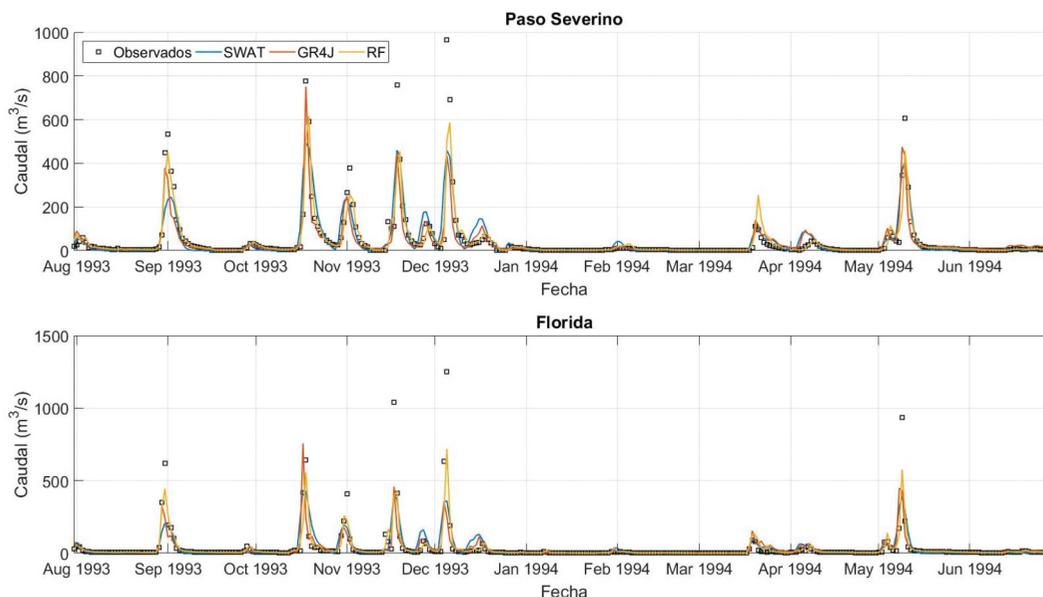


Figura 3. Comparación entre series temporales de caudales observados y simulados con los tres modelos durante la etapa de calibración/entrenamiento. Se presenta, a modo de ejemplo, un período de un año, desde agosto de 1993, en las dos cuencas.

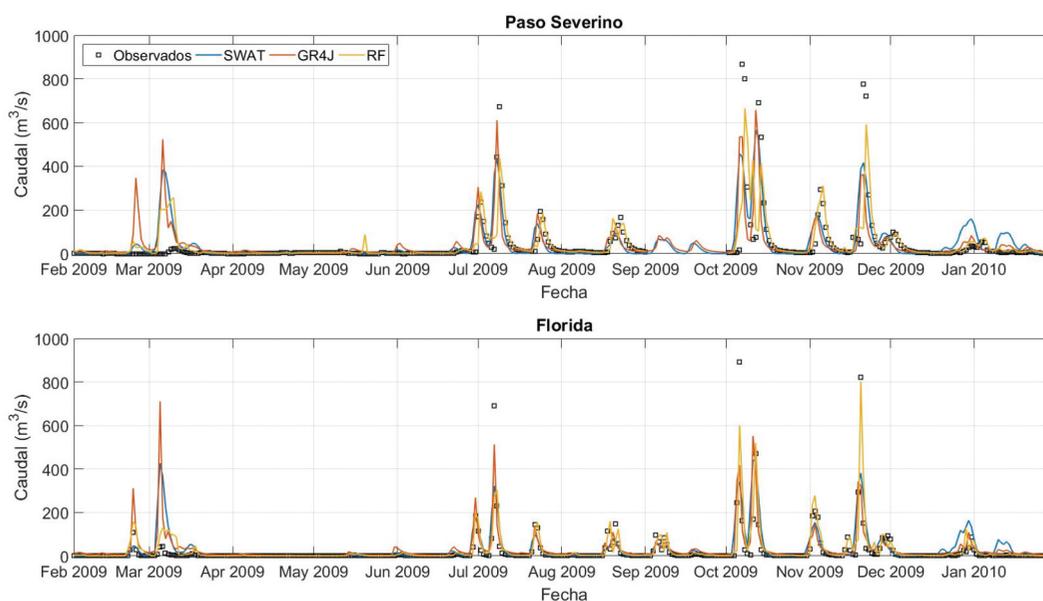


Figura 4. Comparación entre series temporales de caudales observados y simulados con los tres modelos durante la etapa de validación/testeo. Se presenta, a modo de ejemplo, un período de un año, desde febrero de 2009, en las dos cuencas.

una comparación entre series temporales observadas y simuladas, para las etapas de calibración/entrenamiento y validación/testeo respectivamente en las dos cuencas de estudio. En ellas se presenta, a modo de ejemplo, un período de un año de cada una de las etapas. En ambos casos se seleccionaron años con variedad de eventos hidrológicos. En las Figuras MS3 a MS16 del material suplementario se presentan gráficas para la totalidad del período simulado. Complementariamente, en la [Figura 5](#)

y [Figura 6](#) se presenta la comparación entre series temporales de acumulados mensuales de caudal.

La [Figura 7](#) contiene los gráficos de dispersión que comparan los datos modelados con los observados para las dos cuencas de estudio. En cada uno se presenta también el valor de coeficiente de determinación para facilitar la interpretación de los resultados.

Para una evaluación global del desempeño de los modelos, en la [Tabla 7](#) y la [Tabla 8](#), se pueden ver los

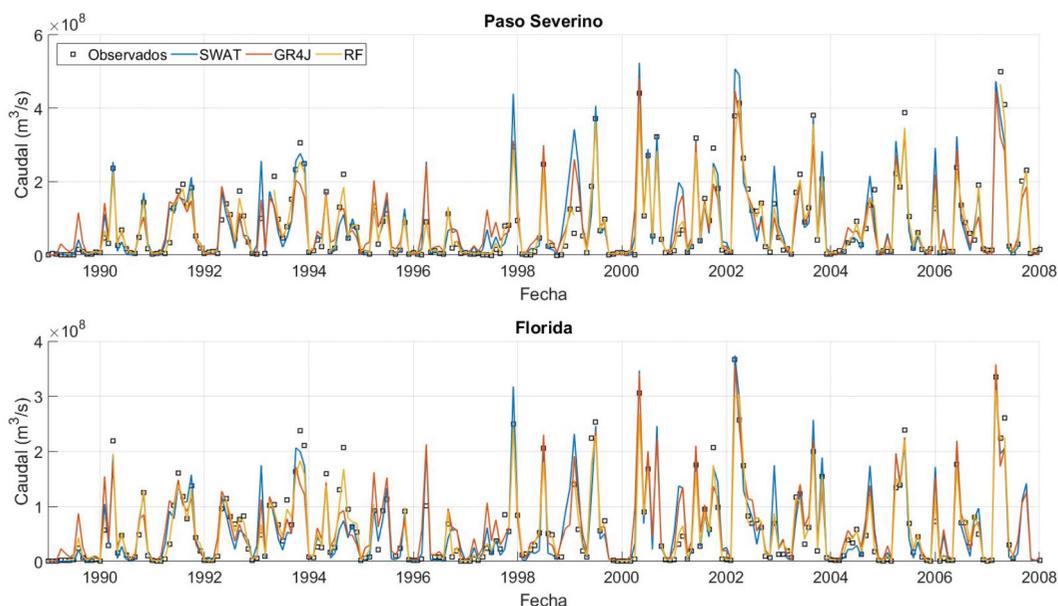


Figura 5. Comparación entre series temporales de acumulados mensuales de caudales observados y simulados con los tres modelos durante la etapa de calibración/entrenamiento.

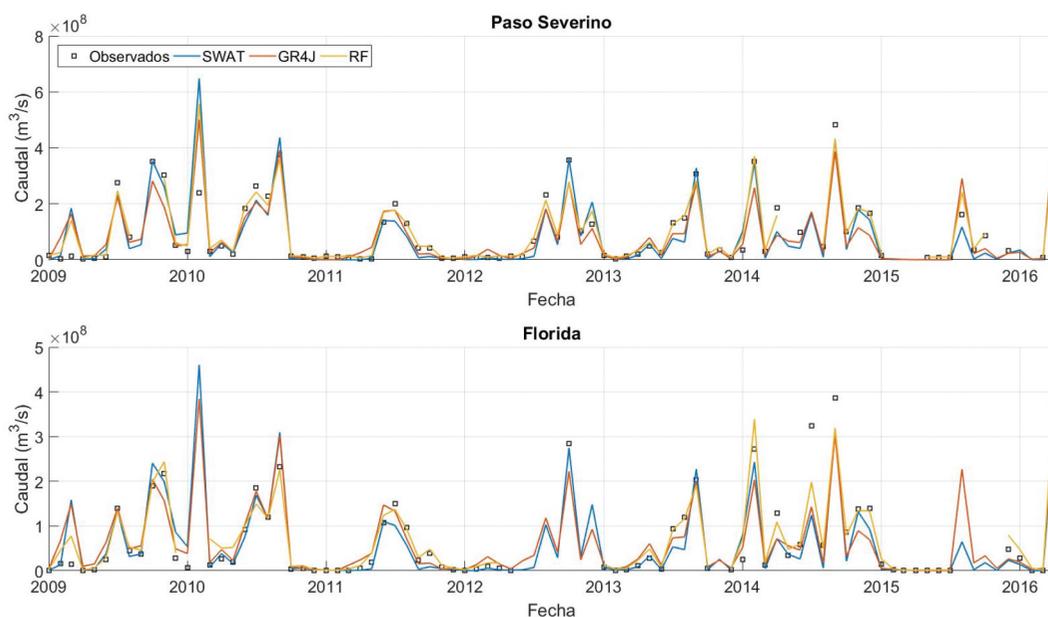


Figura 6. Comparación entre series temporales de acumulados mensuales de caudales observados y simulados con los tres modelos durante la etapa de validación/testeo.

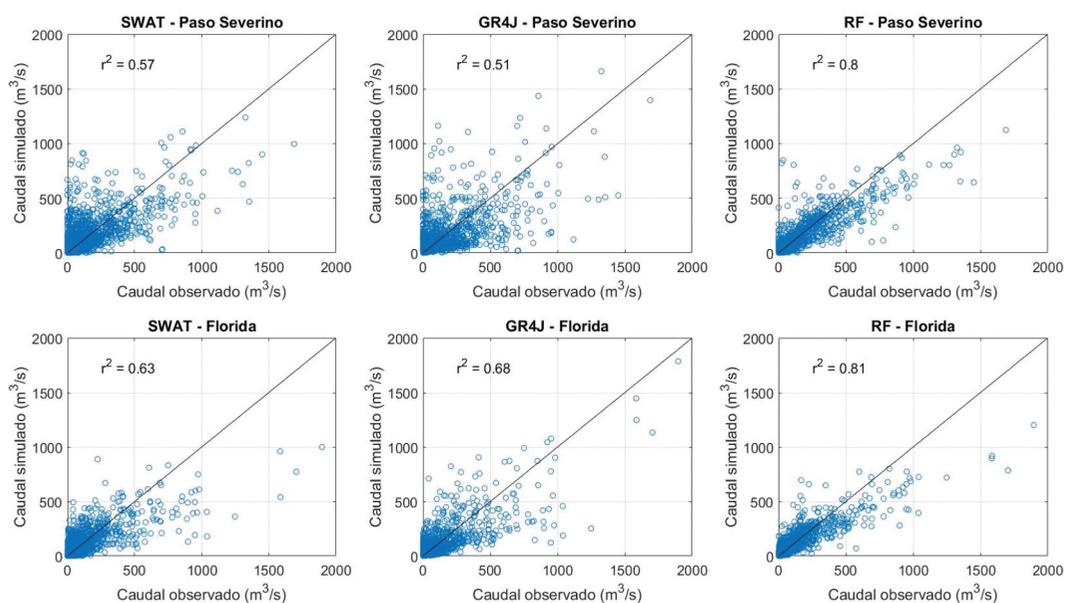


Figura 7. Gráficos de dispersión comparando caudales observados y simulados para los tres modelos implementados, en las dos cuencas.

Tabla 7. Indicadores de desempeño de los tres modelos, correspondientes a la fase de calibración/entrenamiento. Los mejores valores están en negrita.

		NSE	Incert. NSE	Valoración	RSR	Valoración	PBIAS	Valoración	LNSE	Valoración	PNSE	Valoración
Paso Severino	SWAT	0.58	0.06 (10%)	Satisfactorio	0.65	Satisfactorio	-9.0	Muy bueno	0.01	No satisfactorio	0.45	No satisfactorio
	GR4J	0.56	0.08 (15%)	Satisfactorio	0.66	Satisfactorio	-6.4	Muy bueno	0.40	No satisfactorio	0.44	No satisfactorio
	RF	0.86	0.01 (1%)	Muy bueno	0.37	Muy bueno	-0.4	Muy bueno	0.85	Muy bueno	0.73	Bueno
Florida	SWAT	0.62	0.03 (5%)	Satisfactorio	0.62	Satisfactorio	-5.7	Muy bueno	0.46	No satisfactorio	0.45	No satisfactorio
	GR4J	0.68	0.05 (8%)	Bueno	0.57	Bueno	-14.7	Bueno	0.53	Satisfactorio	0.59	Satisfactorio
	RF	0.84	0.01 (1%)	Muy bueno	0.40	Muy bueno	-1.5	Muy bueno	0.89	Muy bueno	0.76	Muy bueno

valores de la función objetivo (NSE) y de los indicadores de desempeño seleccionados (RSR, PBIAS, LNSE y PNSE) para todos los modelos, en las fases de calibración/entrenamiento y validación/testeo respectivamente.

3.3. Desempeño de los modelos en términos de cuantiles

En la Figure 9, se comparan las curvas de permanencia para los datos modelados y observados, teniendo en cuenta la subdivisión del rango de caudales en: caudales

pico, altos, medios, bajos y de estiaje, tal como se detalló anteriormente en la sección 2.4.1.

4. Discusión

4.1. Desempeño de los modelos en términos de correspondencia temporal

Las series temporales graficadas en la Figura 3 y la Figura 4 indican que todos los modelos logran reproducir las distintas etapas del hidrograma de forma adecuada. En todos los casos se observan ascensos

Tabla 8. Indicadores de desempeño de los tres modelos, correspondientes a la fase de validación/testeo. Los mejores valores están en negrita.

		NSE	Incert. NSE	Valoración	RSR	Valoración	PBIAS	Valoración	LNSE	Valoración	PNSE	Valoración
Paso Severino	SWAT	0.45	0.14 (32%)	No satisfactorio	0.74	No satisfactorio	6.0	Muy bueno	-0.88	No satisfactorio	0.45	No satisfactorio
	GR4J	0.26	0.11 (41%)	No satisfactorio	0.86	No satisfactorio	8.0	Muy bueno	0.25	No satisfactorio	0.12	No satisfactorio
	RF	0.64	0.13 (20%)	Satisfactorio	0.60	Bueno	-8.4	Muy bueno	0.86	Muy bueno	0.19	No satisfactorio
Florida	SWAT	0.66	0.03 (4%)	Bueno	0.59	Bueno	10.7	Bueno	0.53	Satisfactorio	0.45	No satisfactorio
	GR4J	0.69	0.05 (7%)	Bueno	0.56	Bueno	3.5	Muy bueno	0.49	No satisfactorio	0.61	Satisfactorio
	RF	0.69	0.03 (4%)	Bueno	0.56	Bueno	-9.8	Muy bueno	0.76	Muy bueno	0.57	Satisfactorio

y descensos del mismo que son coherentes con los observados, así como también sucede con los períodos de caudal bajo o estiaje. Lo mismo puede observarse para las series de acumulados mensuales presentadas en la [Figura 5](#) y la [Figura 6](#). No obstante, en términos cuantitativos, las series temporales de paso diario evidencian dificultades por parte de todos los modelos para reproducir el caudal pico en eventos de crecida. De los tres modelos, RF parece ser el que mejor se desempeña en este sentido.

En los gráficos de dispersión de paso diario ([Figura 7](#)), lo primero que queda de manifiesto es la dificultad que presentan los tres modelos para reproducir el valor pico de los caudales durante crecidas, que ya fue observada en los gráficos de series temporales. Esto se aprecia claramente en los eventos en que el caudal observado supera los $1000 \text{ m}^3/\text{s}$, en donde todos los modelos tienden a la subestimación. En el caso del modelo RF, que es basado en datos, esto se explica por la poca cantidad de eventos de tal magnitud, que hace que el algoritmo no tenga información suficiente para poder aprender su comportamiento. En el caso de los modelos basados en procesos físicos, la explicación para este sesgo no es tan clara dado que en parte puede deberse a que la calibración se realizó para todos los rangos de caudal (no específicamente para caudales altos) y a su vez ambos modelos tienen diferentes limitantes debidas a la estructura propia de cada uno. Por ejemplo, el tránsito de los hidrogramas que para el modelo SWAT se simula con Muskingum, mientras que para GR4J se utilizan simplemente un hidrograma unitario y un reservorio no lineal (sin tránsito

hidrodinámico). Este tipo de simplificaciones podría llegar a afectar la representación de los caudales pico. En términos generales, los gráficos de la [Figura 7](#) presentan una menor dispersión de los puntos para el modelo RF. A su vez, es posible ver que, si bien el modelo GR4J es el que presenta una mayor dispersión de los puntos en términos generales, es también el que presenta un menor sesgo respecto a los caudales de crecida, indicando una buena capacidad de generalización, a pesar de su menor desempeño global. Los gráficos de dispersión a partir de acumulados mensuales, visibles en la [Figura 8](#), muestran de forma consecuente una menor dispersión para el modelo RF respecto a los restantes.

La [Tabla 7](#) indica que el modelo RF sobresale en cuanto a desempeño en la etapa de calibración/entrenamiento, alcanzando en ambas cuencas el mayor valor de la función objetivo (NSE), así como también valores óptimos para el resto de los indicadores. Sin embargo, el valor del NSE disminuye considerablemente durante el testeo ([Tabla 8](#)), al igual que ocurre con varios de los indicadores de desempeño, evidenciando que el modelo presenta una leve tendencia al sobreajuste. Los modelos restantes presentan un desempeño similar entre sí durante la etapa de calibración/entrenamiento para Paso Severino, con la excepción del LNSE, que es significativamente menor para el modelo SWAT, evidenciando una mala representación de los caudales bajos. En términos generales, los dos simulan satisfactoriamente el comportamiento de la cuenca, pero presentan problemas para estiaje (indicado por el LNSE) y crecidas (indicado por el PNSE), siendo su desempeño *no*

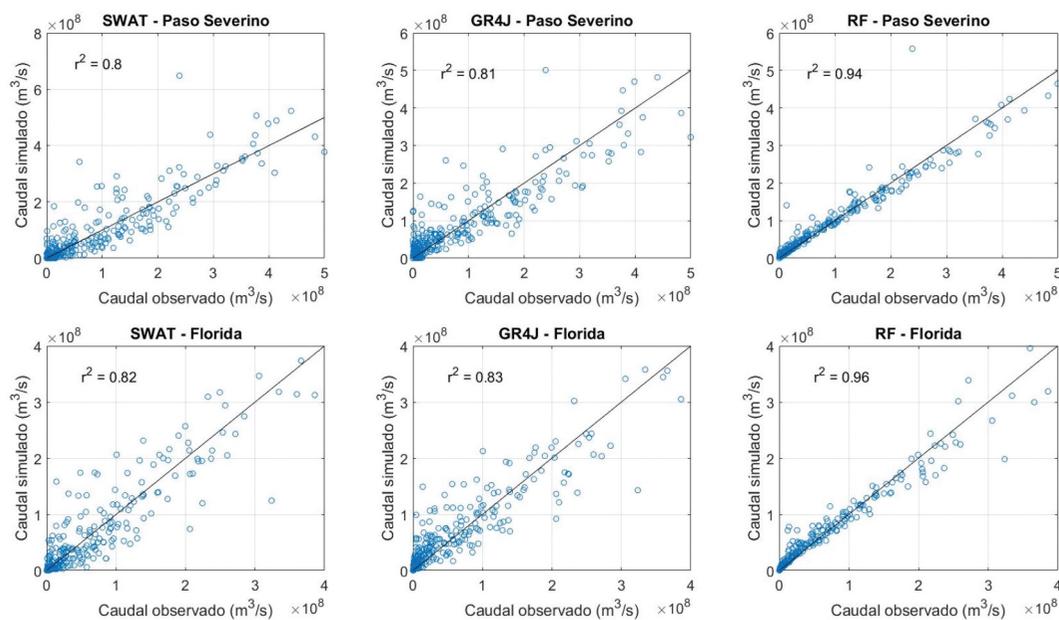


Figura 8. Gráficos de dispersión comparando acumulados mensuales de caudales observados y simulados para los tres modelos implementados, en las dos cuencas.

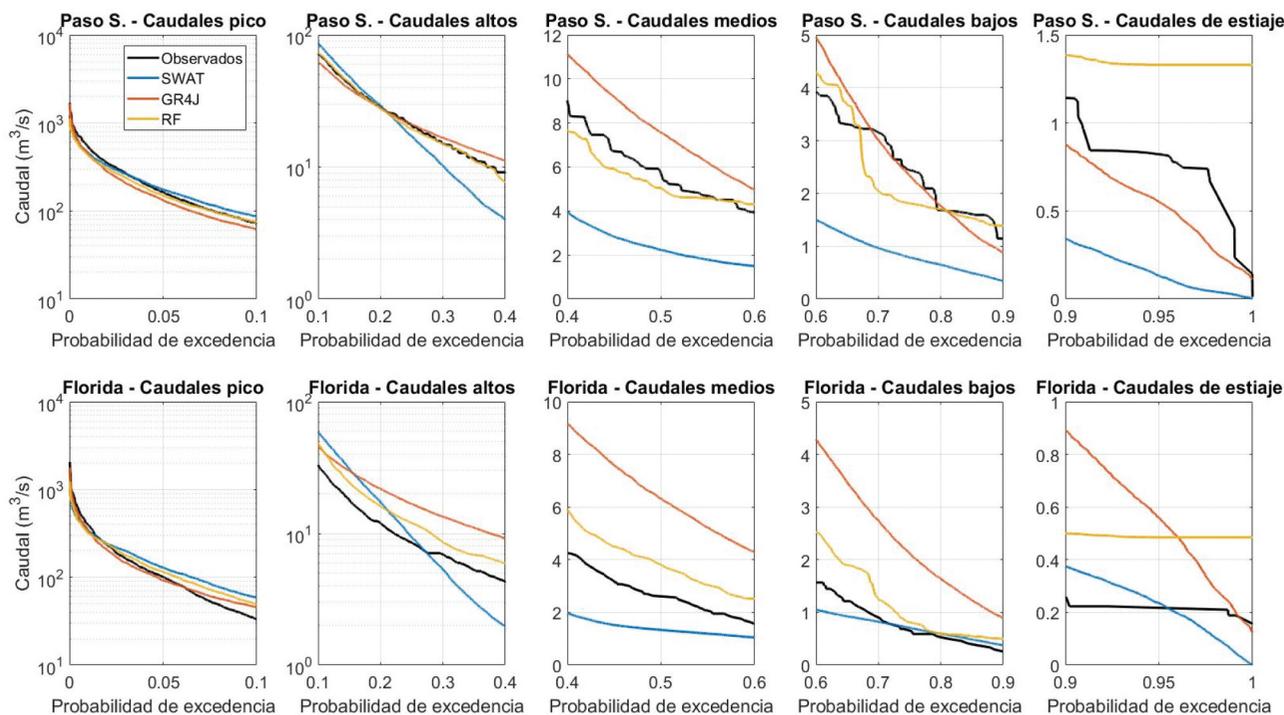


Figura 9. Comparación entre curvas de permanencia de caudales promedio diarios para los caudales observados y simulados con los tres modelos implementados, en las dos cuencas.

satisfactorio en esos rangos. Esto podría justificarse con el hecho de que los caudales observados provienen de una curva de aforo, la cual fue calibrada mayoritariamente con datos de caudales medios. Esto hace que los valores extremos altos y bajos (obtenidos mediante extrapolación de dicha curva) presenten incertidumbre.

En el caso de Florida, es posible ver que se obtienen resultados levemente mejores con el modelo GR4J para todos los indicadores, con excepción de PBIAS, que menor en comparación con SWAT. La mejora en la performance en la cuenca con cierre en Florida, para el caso del GR4J, puede explicarse en parte a que: *i*) esta cuenca no posee ningún gran embalse como Paso Severino (no simulado en GR4J); *ii*) cubre una superficie menor de la otra cuenca (los procesos hidrodinámicos son menos relevantes en la representación hidrológica).

En la etapa de validación/testeo, es posible observar que los modelos físicamente basados (SWAT y GR4J) presentan un desempeño llamativamente bajo en la cuenca de Paso Severino, que no les permite llegar a valores *satisfactorios* de NSE y RSR, lo cual no ocurre con el modelo RF. Esto puede deberse en parte a la forma en que los modelos simulan el embalse, el mismo no se simula en GR4J y sí se representa, pero de forma simplificada en SWAT, con lo cual se explica el mejor desempeño que se obtiene con este último en términos generales y para

caudales de crecida (expresado por PNSE). Mientras tanto, el modelo RF, al tomar el dato de caudal del día anterior ya transitado en el embalse entre sus inputs no acarrea ese error. A su vez en este punto, sobre todo en estiaje, cabe la posibilidad que cambios no registrados en la operación del embalse afecten el régimen de caudales, lo cual no sería captado por los modelos de base física y si por RF. Este último punto explica el pobre desempeño que se obtiene con los modelos de base física para caudales bajos, indicado por los valores de LNSE. Esto no ocurre en la cuenca con cierre en Florida, para la cual los modelos basados en procesos físicos presentan buenos indicadores. Los mismos son similares (e incluso mejores en el caso de SWAT) a los obtenidos en la etapa de calibración para dicha cuenca, por lo que ambos modelos tienen una buena capacidad de generalización de los procesos que pretenden representar. Este comportamiento (aumento del NSE en la validación respecto a la calibración) resulta llamativo, pero puede explicarse por las características de la serie observada en el período de validación. Se trata de un período mayormente de caudales bajos con poca ocurrencia de crecidas, y a su vez hay una cantidad significativa de datos faltantes, lo que puede ser la causa del comportamiento mencionado. Entre ambos, se destaca el desempeño alcanzado por el modelo GR4J, que tiene mejores valores que SWAT en cuatro de los cinco

indicadores que se evalúan. Esto deja de manifiesto que, en cuencas como la de Florida, que no tienen intervenciones significativas (como el embalse), no se justifica el esfuerzo adicional que implica la implementación de un modelo semi-distribuido como SWAT exclusivamente para la simulación de la serie de caudales. En cambio, en Paso Severino, donde el embalse evidentemente influye sobre el régimen de caudales, SWAT produce resultados mejores, aunque con limitaciones. Como se menciona anteriormente, el modelo RF disminuye su performance en la etapa de testeo respecto a la de entrenamiento. No obstante, en Paso Severino el mismo no presenta las mismas dificultades que los restantes durante la etapa de validación/testeo, alcanzando un desempeño *satisfactorio* (al menos), y en Florida se iguala al modelo GR4J en términos generales, siendo mejor en la representación de caudales bajos.

4.2. Desempeño de los modelos en términos de cuantiles

Los gráficos de permanencia permiten tener una idea global de la capacidad de los modelos para reproducir los caudales en función de su magnitud. En este caso, los mismos se dividieron en cinco, para tener una idea del comportamiento de los modelos por rango de caudales.

En el rango de los caudales pico (permanencia entre 0% y 10%), todos los modelos presentan buena aproximación a los caudales observados para ambas cuencas, siendo RF el que más se asemeja en Paso Severino y GR4J en Florida. Este resultado no se corresponde con los valores obtenidos de PNSE, que se utilizó como indicador de caudales de crecida ya que en muchas ocasiones este indicador resulta en valores *no satisfactorios*. Esto se debe a que PNSE no cuantifica únicamente el error en la magnitud del caudal, sino que también evalúa la simultaneidad de los eventos entre lo observado y lo simulado. Esto indica que los modelos presentan una deficiencia en ese sentido, particularmente en Paso Severino, donde el embalse incide sobre el tránsito de las crecidas.

Para caudales altos, RF y GR4J presentan una tendencia que se asemeja a la de los caudales observados, siendo esta mejor en Florida que en Paso Severino. Esto no ocurre con SWAT, cuya curva de permanencia tiene una pendiente significativamente distinta. Este es un comportamiento que se observa también para caudales medios, bajos y de estiaje, lo que coincide con el peor desempeño de SWAT en términos generales para los diferentes indicadores, y que fue comentado en la sección 4.1.

Para los caudales medios y bajos se observan comportamientos similares, en ambas cuencas. En todos los

casos, RF es el que reproduce de mejor manera el comportamiento de los caudales observados. Puede verse además que el modelo SWAT tiende en todos los casos a la subestimación de los caudales, mientras que GR4J tiende a la sobreestimación, particularmente en Florida.

En los caudales de estiaje se observan las mayores diferencias de comportamiento entre observaciones y simulaciones. En particular RF presenta un valor de caudal constante para esta condición, cercano a $1.5 \text{ m}^3/\text{s}$ en Paso Severino y a $0.5 \text{ m}^3/\text{s}$ en Florida. En el caso de Florida, la serie observada tiene un comportamiento similar (lo cual resulta llamativo) aunque con un caudal menor, del entorno de los $0.2 \text{ m}^3/\text{s}$. En Paso Severino esto no se observa, y es GR4J quien acompaña mejor la tendencia de los datos observados. Nuevamente, este comportamiento difiere del de LNSE, indicador que fue utilizado para representar el rango menor de caudales. Esto seguramente se deba a que dicho indicador se ve influenciado también por el rango de caudales bajos (no solo los de estiaje), en donde RF presenta una mejor representación que GR4J.

4.3. Información requerida, parámetros y tiempo de computación

Desde el punto de vista de la información necesaria, SWAT es el modelo que requiere un número mayor de datos de ingreso, ya que no solo necesita datos meteorológicos (HR , SR , T_{max} , T_{min} , WS y P) sino también información espacial (topografía, uso/cobertura del suelo y tipo de suelo). El modelo GR4J solo necesita P y ETP como datos de ingreso. Para nuestro estudio en particular, para el modelo RF, decidimos utilizar los mismos datos meteorológicos de input del modelo SWAT, con la adición de algunas variables artificiales (p.e., P_{acum} , Q_{t-1} , HR_{EWMA} , etc.) que demostraron mejorar el rendimiento y la precisión predictiva del modelo.

En cuanto a los parámetros de calibración de los modelos, se eligieron diez para el modelo SWAT, cuatro para el modelo GR4J y dos para el RF. Esto se traduce en que los tiempos de computación necesarios para el modelo SWAT son mayores en comparación con los otros dos modelos.

5. Conclusiones

En este trabajo se presenta una comparación profunda entre tres modelos hidrológicos de paso diario (SWAT, GR4J y RF) con diferentes características, que fueron implementados en dos cuencas, con cierres en Paso Severino y Florida. Se alcanzaron las siguientes conclusiones:

- El modelo RF presenta, en términos generales, un mejor desempeño que los modelos basados en hipótesis físicas.
- Los dos modelos físicamente basados alcanzan una performance similar, por lo que no parece justificarse el trabajo adicional que involucra la implementación de un modelo semi-distribuido (SWAT) respecto a un concentrado (GR4J) para este caso de estudio.
- El modelo RF presenta problemas de sobreajuste. Si bien el modelo GR4J tiene un desempeño algo menor, tiene mejor capacidad de generalización, ya que sus resultados son similares en las etapas de calibración y validación.

Estos resultados permitieron realizar un avance hacia el entendimiento del lugar que ocupan los modelos basados en AA, como RF, respecto a los tradicionales modelos basados en procesos físicos concentrados y semi-distribuidos. En este estudio, quedó de manifiesto que los primeros presentan ventajas en cuanto a desempeño y velocidad de implementación respecto a los de base física, siendo aún necesario avanzar en su capacidad de generalización para dotarlos de confiabilidad al aplicarlos en escenarios fuera de los que se utilizaron para su entrenamiento.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Comisión Académica de Posgrado (CAP), Universidad de la República [Beca de Doctorado].

ORCID

Federico Vilaseca  <http://orcid.org/0000-0002-5471-6165>
 Alberto Castro  <http://orcid.org/0000-0002-9174-398X>
 Angela Gorgoglione  <http://orcid.org/0000-0002-2476-2339>

Financiación

Este trabajo fue financiado en parte por una beca de Doctorado de la Comisión Académica de Posgrado (CAP) Universidad de la República.

References

- [1] Razavi T, Coulibaly P. An evaluation of regionalization and watershed classification schemes for continuous daily streamflow prediction in ungauged watersheds. *Can Water Resour J Rev Can Ressour Hydr.* 2017;42(1):2–20.
- [2] Abbaspour KC, Johnson CA, Genuchten MT. Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone J.* 2004;3(4):1340–1352.
- [3] Perrin C, Michel C, Andréassian V. Improvement of a parsimonious model for streamflow simulation. *J Hydrol.* 2003;279(1–4):275–289.
- [4] Rossman LA. Storm water management model user's manual version 5.1. Cincinnati: OH: U.S. Environmental Protection Agency (EPA), National Risk Management Research Laboratory Office of Research and Development U.S. Environmental Protection Agency 2015 .
- [5] Phandee W, Jothityangkoon C, Dasananda S. Development of distributed conceptual hydrological model for forest watershed in Northern Thailand: a downward approach. *Eur Sci J.* 2014;10(8):219–229.
- [6] Brirhet H, Benaabidate L. Comparison of two hydrological models (lumped and distributed) over a pilot area of the issen watershed in the Souss Basin, Morocco. *Eur Sci J.* 2016;12(18):1857–7431.
- [7] Narbondo S, Gorgoglione A, Crisci M, et al. Enhancing physical similarity approach to predict runoff in ungauged watersheds in sub-tropical regions. *Water.* 2020;12(2):528.
- [8] De Vera A, Alfaro P, Terra R. Operational implementation of satellite-rain gauge data merging for hydrological modeling. *Water.* 2021;13(4):533.
- [9] Vilaseca F, Narbondo S, Chreties C, et al. A comparison between lumped and distributed hydrological models for daily rainfall-runoff simulation. *IOP Conf Ser: Earth Environ Sci.* 2022;958(1):012016.
- [10] Paudel M, Nelson EJ, Downer CW, et al. Comparing the capability of distributed and lumped hydrologic models for analyzing the effects of land use change. *J Hydroinforma.* 2011;13(3):461–473.
- [11] Ashraf Vaghefi S, Abbaspour KC, Faramarzi M, et al. Modeling crop water productivity using a coupled SWAT–MODSIM model. *Water.* 2017;9(3):157.
- [12] Mer F, Badano L, Neighbor N, et al. SWAT subcuenca Santa Lucia, 2020, doi: [10.17605/OSF.IO/UQB5J](https://doi.org/10.17605/OSF.IO/UQB5J)
- [13] Mer F, Baethgen W, Vervoort RW. Building trust in SWAT model scenarios through a multi-institutional approach in Uruguay. *Socio-Environl Sys Model.* 2020;2:17892.
- [14] Aznarez C, Jimeno-Sáez P, López-Ballesteros A, et al. Analysing the impact of climate change on hydrological ecosystem services in Laguna del Sauce (Uruguay) using the SWAT model and remote sensing data. *Remote Sens.* 2021;13(10):2014.
- [15] Vilaseca F, Castro A, Chreties C, et al. Daily rainfall-runoff modeling at watershed scale: a comparison between physically-based and data-driven models. In: Gervasi O, editor. *Computational Science and Its Applications – ICCSA 2021.* ICCSA 2021. Lecture Notes in Computer Science. Vol. 12955. Cham: Springer; 2021. p. 18–33.
- [16] Krajewski WF, Lakshmi V, Georgakakos KP, et al. A Monte Carlo study of rainfall sampling effect on a distributed catchment model. *Water Resources Research.* 1991;27(1):119–128.

- [17] Refsgaard JC, Knudsen J. Operational validation and intercomparison of different types of hydrological models. *Water Resour Res.* 1996;32(7):2189–2202.
- [18] Reed S, Koren V, Smith M, et al. Overall distributed model intercomparison project results. *J Hydrol.* 2004;298(1–4):27–60.
- [19] Carpenter TM, Georgakakos KP. Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *J Hydrol.* 2006;329(1–2):174–185.
- [20] Anshuman A, Kunnath-Poovakka A, Eldho TI. Towards the use of conceptual models for water resource assessment in Indian tropical watersheds under monsoon-driven climatic conditions. *Environ Earth Sci.* 2019;78(9):282.
- [21] Li M, Zhang Y, Wallace J, et al. Estimating annual runoff in response to forest change: a statistical method based on random forest. *J Hydrol.* 2020;589:125168.
- [22] Desai S, Ouarda T. Regional hydrological frequency analysis at ungauged sites with random forest regression. *J Hydrol.* 2021;594:125861.
- [23] Muñoz P, Orellana-Alvear J, Willems P, et al. Flash-flood forecasting in an Andean mountain catchment – development of a step-wise methodology based on the random forest algorithm. *Water.* 2018;10(11):1519.
- [24] Pini M, Scalvini A, Liaqat MU, et al. Evaluation of machine learning techniques for inflow prediction in Lake Como, Italy. *Procedia Comput Sci.* 2020;176:918–927.
- [25] Díaz I, Levrini P, Achkar M, et al. Empirical modeling of stream nutrients for countries without robust water quality monitoring systems. *Environments.* 2021;8(11):129.
- [26] Bouriel M, Segura AM, Crisci C, et al. Machine learning methods for imbalanced data set for prediction of faecal contamination in beach waters. *Water Res.* 2021;202:117450.
- [27] Crisci C, Terra R, Pacheco JP, et al. Multi-model approach to predict phytoplankton biomass and composition dynamics in a eutrophic shallow lake governed by extreme meteorological events. *Ecol Modell.* 2017;360(80):93.
- [28] Tegegne G, Park DK, Kim Y-O. Comparison of hydrological models for the assessment of water resources in a data-scarce region, the Upper Blue Nile River Basin. *J Hydrol.* 2017;14:49–66.
- [29] Jimeno-Sáez P, Martínez-España R, Casali J, et al. A comparison of performance of SWAT and machine learning models for predicting sediment load in a forested Basin, Northern Spain. *Catena.* 2022;212:105953.
- [30] Gorgoglione A, Gregorio J, Ríos A, et al. Influence of land use/land cover on surface-water quality of Santa Lucía river, Uruguay. *Sustainability.* 2020;12(11):4692.
- [31] Navas R, Alonso J, Gorgoglione A, et al. Identifying climate and human impact trends in streamflow: a case study in Uruguay. *Water.* 2019;11(7):1433.
- [32] Ríos A. Implementación de un modelo hidrodinámico tridimensional en el embalse de Paso Severino. Aportes para la modelación de calidad de agua. Tesis de Maestría en Ingeniería Ambiental. Instituto de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA), Facultad de Ingeniería (FIing). Montevideo, Uruguay: Universidad de la República (UdelaR); 2019.
- [33] Gorgoglione A, Alonso J, Chreties C, et al. Assessing temporal and spatial patterns of surface-water quality with a multivariate approach: a case study in Uruguay. *IOP Conf Series: Earth Environ Sci.* 2020;612:012002.
- [34] Rodríguez R, Pastorini M, Etcheverry L, et al. Water-quality data imputation with a high percentage of missing values: a machine learning approach. *Sustainability.* 2021;13(11):6318.
- [35] MGAP. Clasificación de suelos con índice de productividad CONEAT para suelos uruguayos. Disponible en: (consultado el 21 de julio de 2022) <https://www.gub.uy/ministerio-ganaderia-agricultura-pesca/politicas-y-gestion/coneat>
- [36] Narbondo S. Incorporación de información satelital de humedad de suelo en modelos hidrológicos para pronóstico de inundaciones en cuencas de Uruguay. Tesis de Maestría en Ingeniería Ambiental. Instituto de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA), Facultad de Ingeniería (FIing). Montevideo, Uruguay: Universidad de la República (UdelaR); 2021.
- [37] Arnold JG, Srinivasan R, Mutiah RS, et al. Large area hydrologic modeling and assessment Part I: model development. *J Am Water Resources Assoc.* 1998;34(1):73–89.
- [38] Abbaspour KC, Yang J, Maximov I, et al. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT J. *Journal of Hydrology.* 2007;333(2–4):413–430.
- [39] Brieman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
- [40] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12(85):2825–2830.
- [41] Kannan N, Jeong J. An approach for estimating stream-flow health using flow duration curves and indices of hydrologic alteration. EPA region 6 water quality protection division U.S. Dallas: TX: Environmental Protection Agency; 2011.
- [42] Clark MP, Vogel RM, Lamontagne JR, et al. The abuse of popular performance metrics in hydrologic modeling. *Water Resour Res.* 2021;57(9). DOI:10.1029/2020WR029001
- [43] Clark MP, Shook KR, 2021. Gumbot: bootstrap analyses of sampling uncertainty in goodness-of-fit statistics. <https://github.com/CH-Earth/gumbot>
- [44] Moriasi DN, Arnold JG, Van Liew MW, et al. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE.* 2007;50(3):885–900.
- [45] Akiba T, Sano S, Yanase T, et al. Optuna: a next-generation hyperparameter optimization framework.” *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining; Anchorage AK USA.* 2019. p. 2623–2631.