

SOCIEDAD CHILENA DE INGENIERÍA HIDRÁULICA
XXVI CONGRESO CHILENO DE INGENIERÍA HIDRÁULICA

**MODELACIÓN HIDROLÓGICA HORARIA UTILIZANDO CALIBRACIÓN
ENFOCADA EN EVENTOS DE CRECIDA EN CUENCAS CHILENAS**

DIEGO PARRA V.¹
ALVARO OSSANDÓN A.²
JOAQUIN MEZA A.²

RESUMEN

Las inundaciones representan una de las amenazas naturales más frecuentes y dañinas a nivel mundial, cuyos impactos se ven incrementados por el cambio climático y la expansión urbana. La modelación hidrológica es fundamental para anticipar eventos extremos y apoyar sistemas de alerta temprana. En este estudio se evaluó la capacidad de los modelos conceptuales a escala horaria GR4H, SAC-SMA y TUV para simular crecidas extremas en 43 cuencas chilenas, utilizando siete funciones objetivo y dos enfoques de calibración: el tradicional y el centrado en crecidas. SAC-SMA presentó el desempeño más robusto y consistente, TUV mostró resultados satisfactorios, aunque con menor estabilidad, y GR4H fue el más sensible a la función objetivo, con alta variabilidad entre cuencas. Entre las funciones objetivo, LCE destacó por ofrecer el mejor rendimiento en todas las métricas. La calibración centrada en crecidas mejora la estimación de la magnitud de los eventos extremos, pero tiende a generar más falsas alarmas. Se propone una estrategia combinada que utilice funciones enfocadas en extremos para caracterizar los hidrogramas y funciones generales para activar alertas, potenciando así la confiabilidad de los sistemas de monitoreo a escala horaria.

¹ Alumno, Depto. de Obras Civiles, Universidad Técnica Federico Santa María – diego.parrava@usm.cl

² Profesor Asistente, Depto. de Obras Civiles, Universidad Técnica Federico Santa María –
alvaro.ossandon@usm.cl; joaquin.meza@usm.cl

1. INTRODUCCIÓN

Las inundaciones representan una de las amenazas naturales más frecuentes y destructivas a nivel global, con impactos significativos tanto en zonas urbanas como rurales. El aumento en su recurrencia e intensidad, impulsado por el cambio climático y la expansión urbana, ha elevado la vulnerabilidad de diversos territorios (Swain et al., 2020). En este contexto, la modelación hidrológica ha cobrado relevancia como herramienta clave para estimar escorrentías y anticipar eventos extremos mediante sistemas de alerta temprana y mapas de inundación (Mudashiru et al., 2021). Entre los enfoques disponibles, los modelos conceptuales destacan por su equilibrio entre simplicidad estructural y capacidad predictiva, lo que facilita su implementación en contextos con información limitada (Jaiswal et al., 2020). A nivel internacional, modelos como SAC-SMA, HBV y GR4J han sido ampliamente utilizados para la gestión de crecidas, mostrando un rendimiento adecuado en la simulación de caudales extremos (Emerton et al., 2016; Bergström et al., 1992; Cantoni et al., 2022). Sin embargo, pese a su amplio uso, todos los modelos hidrológicos suelen subestimar los caudales máximos (Lane et al., 2019).

Una de las explicaciones para esta limitante se encuentra en la función objetivo utilizada en la calibración, la cual evalúa la bondad de ajuste entre el caudal simulado y el observado. Entre estas funciones, el coeficiente de eficiencia de Nash-Sutcliffe (NSE; Nash & Sutcliffe, 1970) y la eficiencia de Kling-Gupta (KGE; Gupta et al., 2009) se destacan como las más utilizadas. Sin embargo, ambas presentan limitaciones: generan resultados variables incluso dentro de una misma cuenca y tienen dificultades para reproducir caudales máximos y modelar la variabilidad temporal (Mizukami et al., 2019; Liu, 2020). Para abordar estas deficiencias, se han desarrollado estrategias de calibración más avanzadas. Una de ellas consiste en combinar múltiples funciones objetivo de forma simultánea, lo que permite capturar distintos aspectos del régimen hidrológico; por ejemplo, en Chile se ha observado que la combinación $KGE(Q) + NSE(\log Q)$ muestra buen rendimiento en cuencas con régimen nival (Araya et al., 2023). También se han propuesto y utilizado nuevas métricas individuales orientadas a mejorar el ajuste en eventos extremos, como LME y KMoments (Liu, 2020; Pizarro et al., 2024). Además, se han implementado estrategias específicas según el contexto, como la calibración usando únicamente años húmedos y la calibración basada exclusivamente en los caudales máximos anuales (Sun et al., 2017; Mizukami et al., 2019).

Este estudio tiene como objetivo analizar la capacidad de distintas métricas de calibración y de modelos hidrológicos conceptuales —GR4H, TUW y SAC-SMA en sus versiones horarias— para simular caudales de crecida a escala horaria en cuencas chilenas, así como evaluar la sensibilidad de estos modelos a las distintas métricas empleadas. Para ello, se utilizan registros horarios de caudal obtenidos desde la Dirección General de Aguas (DGA) y datos meteorológicos horarios provenientes del producto de reanálisis ERA5-Land (Muñoz-Sabater, 2021). La calibración de los modelos se realiza empleando tanto métricas tradicionales como versiones modificadas orientadas específicamente a periodos de crecidas, con el propósito de comparar su desempeño, analizar su sensibilidad y determinar cuáles resultan más adecuadas para la simulación de caudales extremos a escala horaria. Este análisis busca aportar conocimientos fundamentales para mejorar la precisión de los pronósticos de crecidas y fortalecer las herramientas disponibles para la gestión del riesgo de inundaciones en Chile y regiones con características similares.

2. ZONA DE ESTUDIO Y DATOS

El área de estudio abarca 43 cuencas en la zona central de Chile (32–39°S, 70–73°W); estas se pueden visualizar en la Figura 1. En esta zona se concentra más del 70 % de la población y se desarrollan importantes actividades económicas ligadas al agua, por lo que las crecidas representan una seria amenaza para el bienestar social y económico del país. Las cuencas seleccionadas pertenecen al conjunto de datos CAMELS-CL (Alvarez-Garreton et al., 2018) y cumplen con los siguientes criterios: (i) contar con al menos el 90 % de observaciones horarias en crecidas durante 14 años continuos dentro del periodo 1980–2023; los días de grandes crecidas se identifican agrupando cuencas cercanas y de regímenes de grandes crecidas similares, calculando un umbral común (percentil 95 de caudales específicos) y registrando las fechas en el grupo para verificar la cobertura de datos en todas las cuencas; (ii) un bajo índice de intervención humana (<0.05) o índice desconocido, evitando excluir cuencas posiblemente no intervenidas; y (iii) ausencia de grandes embalses. Las definiciones de índice de intervención y de gran embalse se adoptan según CAMELS-CL.

Para el análisis, se utilizaron series horarias de caudales observados de la Dirección General de Aguas (DGA) para el periodo 1980–2023. Del producto de reanálisis ERA5-Land (Muñoz-Sabater, 2021) se obtuvieron datos horarios grillados de precipitación, temperatura y otras variables necesarias para calcular la evapotranspiración potencial mediante la fórmula de Thornthwaite, considerando los 14 años continuos de cada cuenca más dos años previos para el calentamiento del modelo. Las variables se agregaron por cuenca mediante un promedio aritmético simple. Adicionalmente, para construir la curva hipsométrica se empleó el modelo digital de elevación (DEM) ASTER versión 3.0, desarrollado por el equipo científico de ASTER de EE. UU. y Japón.

La Figura 1 muestra la zona de estudio y las 43 cuencas analizadas, con elevación creciente hacia el este y decreciente hacia el sur. Los regímenes de grandes crecidas, clasificados mediante el índice de Pardé (Pardé, 1933), adaptado al percentil 95 estacional, incluyen 5 cuencas nivales, 16 mixtas y 22 pluviales. Hacia el sur aumentan el caudal de excedencia del 95 % y la precipitación, mientras que las temperaturas más altas se concentran en zonas bajas. Se calcularon los valores del estadístico U del Test U de Buishand (1984), donde valores más altos sugieren posibles intervenciones no registradas en CAMELS-CL.

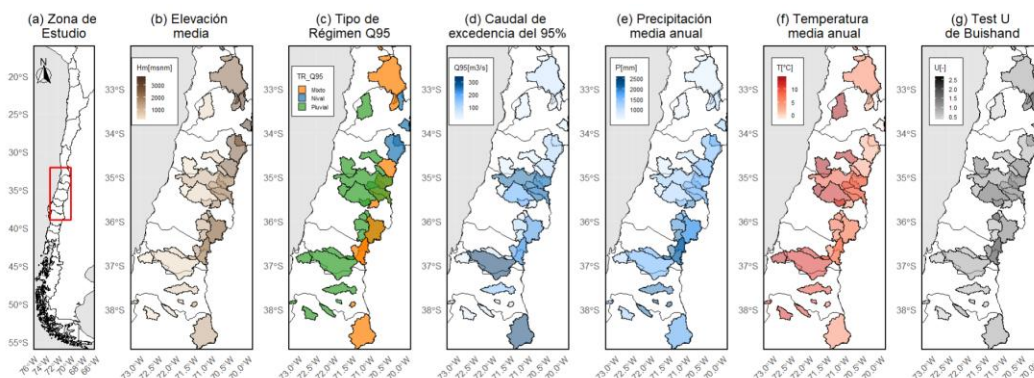


Figura 1. Zona de estudio y variabilidad espacial de las características de las cuencas analizadas: (a) zona de estudio, (b) elevación media, (c) tipo de régimen de grandes crecidas, (d) caudal de excedencia del 95 %, (e) precipitación media anual, (f) temperatura media anual, y (g) Estadístico U del Test U de Buishand.

3. METODOLOGIA

3.1 MODELOS HIDROLOGICOS

En este estudio se implementan tres modelos hidrológicos conceptuales con resolución horaria: (i) GR4H (Perrin et al., 2003), acoplado al módulo nival CemaNeige (Valéry et al., 2014); (ii) TUW (Parajka et al., 2007), basado en la estructura del modelo HBV desarrollado por Bergström (1976); y (iii) SAC-SMA (Burnash et al., 1973), combinado con el módulo de nieve SNOW-17 (Anderson, 1973) y con enrutamiento tipo Saint-Venant linealizado (Lohmann et al., 1996). Estos modelos han sido aplicados en Chile a escala diaria (Araya et al., 2023) y en operaciones de gestión de inundaciones en EE.UU., Europa y África (Emerton et al., 2016; Bergström et al., 1992; Cantoni et al., 2022).

- GR4H funciona con dos reservorios y seis parámetros libres, incluyendo el módulo CemaNeige, que particiona lluvia y nieve mediante un esquema de grado-día y genera escurrimiento en dos componentes con intercambio subterráneo (Perrin et al., 2003; Valéry et al., 2014).
- TUW integra tres reservorios y quince parámetros libres, usando un esquema de grado-día para la nieve y un módulo de humedad del suelo; la escorrentía superficial, intermedia y base se propaga mediante una función de transferencia triangular (Parajka et al., 2007).

SAC-SMA opera con cinco reservorios y veinticinco parámetros libres, más tres fijados. El módulo de nieve SNOW-17 calcula acumulación y fusión nival, los reservorios generan escorrentía directa, intermedia y flujo base mediante convolución con funciones gamma, y el enrutamiento final se aplica mediante un esquema linealizado de Saint-Venant (Burnash et al., 1973; Anderson, 1973; Lohmann et al., 1996).

La implementación se realiza en el entorno R, mediante los paquetes “airGR”, “TUWmodel” y “sacsmaR”. Este último fue modificado para operar a escala horaria, adaptando su estructura originalmente diaria. Los tres modelos requieren como forzantes datos horarios de precipitación acumulada, temperatura media y evapotranspiración potencial. Adicionalmente, GR4H necesita la curva hipsométrica para representar la distribución de nieve por bandas de elevación.

3.2 CALIBRACIÓN

Los parámetros de los modelos se calibraron mediante el algoritmo de optimización global Shuffled Complex Evolution (SCE-UA; Duan et al., 1992), implementado en el paquete R “rtop” (Skøien et al., 2014), empleando siete funciones objetivo (ver sección 3.2.1), las cuales miden la bondad del ajuste entre el caudal simulado y observado, y dos enfoques de calibración: uno sobre la serie completa de caudales y otro centrado únicamente en los caudales de crecida (ver sección 3.2.2). Cada cuenca cuenta con un periodo continuo de 14 años diferentes, determinado según la cobertura de datos horarios de crecida. De este periodo, se destinan 8 años a la calibración y 6 a la validación, considerando un calentamiento previo de 2 años. Si bien la búsqueda de ventanas se realizó entre 1980 y 2024, todas se ubicaron finalmente entre 1997 y 2024, reflejando la mayor disponibilidad de datos en años recientes.

3.2.1 FUNCIONES OBJETIVO

Las funciones objetivo se maximizan en cada iteración de calibración con el fin de obtener la combinación óptima de parámetros. Sus valores van de $-\infty$ y 1, donde 1 representa un ajuste perfecto, y se construyen a partir de las series de caudales observados (x_s) y simulados (x_o) y un conjunto común de indicadores estadísticos:

- Razón entre desviación estándar simulada y observada ($\alpha = \sigma_s/\sigma_o$)
- Razón entre media simulada y observada ($\beta = \mu_s/\mu_o$)
- Razón entre coeficiente de variación simulado y observado ($\gamma = CV_s/CV_o$, donde $CV_s = \sigma_s/\mu_s$ y $CV_o = \sigma_o/\mu_o$)
- Coeficiente de correlación de Pearson (r) entre valores simulados y observados.

La primera función objetivo, NSE (Nash & Sutcliffe, 1970), cuantifica la eficiencia del modelo como la razón entre el error cuadrático medio y la varianza de los caudales observados:

$$NSE = 1 - \frac{\sum(x_s - x_o)^2}{\sum(x_o - \mu_o)^2} \quad (1)$$

Aunque NSE es ampliamente utilizada, tiende a subestimar la variabilidad y los caudales extremos. Para superar estas limitaciones, Kling et al. (2009) propusieron la eficiencia de Kling–Gupta (KGE), que equilibra correlación, sesgo y variabilidad:

$$KGE = 1 - \sqrt{(\alpha - 1)^2 + (\beta - 1)^2 + (r - 1)^2} \quad (2)$$

Posteriormente, Kling et al. (2012) propusieron una versión modificada de KGE, reemplazando la desviación estándar por coeficientes de variación (γ) para reducir la correlación entre sesgo y variabilidad:

$$KGEp = 1 - \sqrt{(\gamma - 1)^2 + (\beta - 1)^2 + (r - 1)^2} \quad (3)$$

Aun así, KGE y KGEp presentan cierta subestimación de caudales máximos. Para mejorar la simulación de eventos extremos, Liu (2020) introdujo la métrica LME:

$$LME = 1 - \sqrt{(r\alpha - 1)^2 + (\beta - 1)^2} \quad (4)$$

Lee & Choi (2022) propusieron LCE, corrigiendo limitaciones de LME, como la existencia de múltiples soluciones máximas y la tendencia a sobreestimar caudales altos y subestimar los bajos:

$$LCE = 1 - \sqrt{(r\alpha - 1)^2 + (r/\alpha - 1)^2 + (\beta - 1)^2} \quad (5)$$

El uso combinado de métricas permite una evaluación integral del desempeño del modelo. Por ejemplo, KGE junto con NSE aplicada a caudales log-normalizados ha mostrado buen desempeño en cuencas chilenas (Araya et al., 2023):

$$KGE + NSE(\log(Q)) = 0.5 KGE + 0.5 NSE(\log(Q)) \quad (6)$$

Finalmente, la métrica KMoments utiliza momentos estadísticos de orden superior para evaluar la precisión de los modelos hidrológicos (Pizarro et al., 2024):

$$KMoments = 1 - \sqrt{(C_1 - 1)^2 + (C_2 - 1)^2 + (r - 1)^2} \quad (7)$$

$$\text{donde: } C_1 = \frac{\widehat{K}_{1s}}{\widehat{K}_{1o}} = \frac{\mu_s}{\mu_o} \quad \text{y} \quad C_2 = \frac{\sqrt{\widehat{K}_{2s}/\widehat{K}_{1s}}}{\sqrt{\widehat{K}_{2o}/\widehat{K}_{1o}}} \quad \text{con} \quad \widehat{K}_{2(s,o)} = 2 \sum_{i=1}^n \left(\frac{2(i-1)}{n(n-1)} * x_{i(s,o)} \right)$$

Aquí $x_{i(s,o)}$ son los caudales simulados (s) u observados (o) ordenados de menor a mayor, i representa la posición en la serie y n es el número total de datos. Los términos $\widehat{K}_{1(s,o)}$ y $\widehat{K}_{2(s,o)}$ corresponden al primer y segundo momento estadístico de orden superior.

3.2.2 ENFOQUES DE CALIBRACIÓN

En este estudio se aplican dos enfoques de calibración: el primero corresponde al enfoque tradicional, donde la función objetivo se calcula considerando toda la serie de caudales observados y simulados; el segundo, denominado calibración enfocada en crecidas (indicada anteponiendo “W95” al nombre de la función), evalúa la función objetivo únicamente durante las horas asociadas a hidrogramas de crecida, identificadas seleccionando los caudales que superan el percentil 95, agrupándolos en ventanas temporales y ampliando cada ventana hacia atrás y hacia adelante según el tiempo de concentración de la cuenca, estimado mediante la fórmula empírica española (CEDEX, 1994), ampliamente utilizada por su confiabilidad.

$$tc_{Spain} = 18 \left(\frac{L_{km}^{0,76}}{S^{0,19}} \right) \quad (8)$$

Donde tc es el tiempo de concentración, L_{km} es la longitud del cauce principal en kilómetros y S es la pendiente media.

3.3 MÉTRICAS DE DESEMPEÑO

Para evaluar el desempeño de las simulaciones se utilizó el porcentaje de sesgo (PBIAS), que mide la diferencia porcentual entre los caudales simulados y observados, indicando si la simulación tiende a sobreestimar o subestimar. Además, se calculó el sesgo porcentual específicamente en crecidas (Q95_PBIAS), considerando solo las horas en que el caudal observado supera su percentil 95:

$$PBIAS = \frac{\sum(x_s - x_o)}{\sum x_o} \quad (8)$$

Además, se emplearon tres métricas de desempeño categórico. Para ello, se definió un umbral para identificar crecidas, utilizando los percentiles 95 y 98 de los caudales observados. Luego, cada hora se clasificó según la comparación entre caudal simulado y observado: aciertos (a), cuando el modelo predice una crecida que realmente ocurre; falsas alarmas (b), cuando se simula una crecida que no ocurre; omisiones (c), cuando se simula una no crecida frente a una crecida real; y rechazos correctos (d), cuando se predice correctamente la ausencia de crecida.

Las métricas calculadas fueron: el sesgo categórico (B), que mide la proporción de crecidas simuladas respecto a las observadas, donde $B = 1$ indica igual número de crecidas simuladas y observadas, $B < 1$ subestimación y $B > 1$ sobreestimación; la tasa de aciertos (H), que indica la proporción de crecidas correctamente simuladas, con valores cercanos a 0 si no se modeló ninguna crecida observada y 1 si todas las crecidas se modelaron; y la tasa de falsas alarmas (F), que evalúa la proporción de no crecidas simuladas como crecidas, siendo 0 si no hubo falsas alarmas y 1 si todas las simulaciones de crecidas fueron falsas:

$$B = \frac{a + b}{a + c} , \quad H = \frac{a}{a + c} , \quad F = \frac{b}{b + d} \quad (9 - 11)$$

4. RESULTADOS Y DISCUSIÓN

4.1 EJEMPLO: MODELACIÓN DE CAUDALES EN CUENCA RÍO ÑUBLE

La Figura 2 muestra la crecida más significativa registrada en la cuenca Río Ñuble en La Puntilla (1.250 km²), junto con las simulaciones horarias de GR4H, SAC-SMA y TUW. Se presentan los caudales simulados con cuatro funciones objetivo, tanto con el enfoque tradicional como con el enfocado en crecidas (W95), el caudal observado (en negro), el umbral del percentil 95 de excedencia (línea segmentada roja) y el hietograma horario de precipitaciones. Se observa que el reanálisis ERA5-Land tiende a distribuir las precipitaciones menores al inicio del día y las mayores al final, generando discontinuidades en la agregación y un pre-peak simulado por los tres modelos, así como un desfase de 17 horas en el caudal máximo simulado. Respecto a la forma de los hidrogramas, GR4H produce respuestas más abruptas, SAC-SMA respuestas más suaves debido a su esquema de convolución de funciones gamma, y TUW hidrogramas triangulares. Al tratarse de una única crecida, no es posible determinar qué función objetivo es superior, aunque se observa que LME genera las simulaciones más deficientes, mostrando resultados notablemente distintos a los de las demás funciones objetivo.

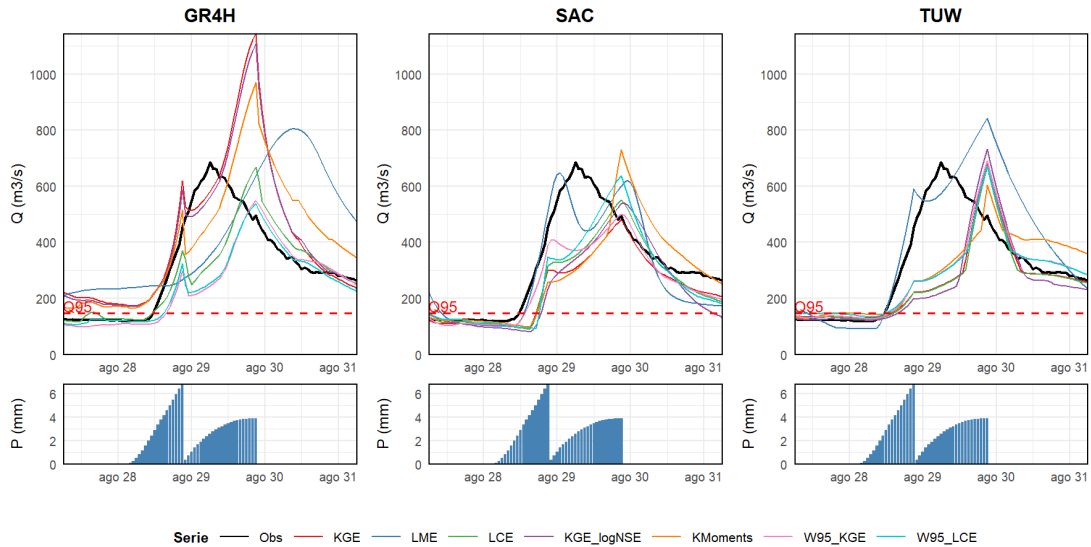


Figura 2. Crecida observada y simulada más importante del período de validación en la cuenca Río Ñuble en La Puntilla, utilizando los modelos: GR4H, SAC-SMA y TUW, junto a la precipitación en cada hora.

4.2 CALIDAD DE SIMULACIÓN EN CAUDALES GENERALES Y EXTREMOS

La Figura 3 presenta los valores del sesgo porcentual entre caudales observados y simulados, evaluado tanto para la serie completa de caudales (PBIAS, en gris) como solo para caudales de crecida (Q95_PBIAS, en verde). Las métricas se calcularon para las 43 cuencas consideradas en este trabajo (Figura 1) y se representan mediante boxplots según el modelo hidrológico (GR4H, SAC-SMA y TUW), la función objetivo utilizada (siete en total), el tipo de calibración (tradicional o centrada en crecidas) y el periodo de análisis (seis años de calibración y ocho de validación). Se observan las siguientes características:

- Al pasar del periodo de calibración al de validación, todos los modelos y funciones objetivo muestran una disminución del rendimiento, evidenciada tanto en PBIAS como en Q95_PBIAS; las medianas varían hasta 46 % y los rangos intercuartil (RIQ) hasta 96 %.
- Al analizar el sesgo porcentual para todos los caudales (PBIAS) en validación, las funciones objetivo calibradas con la serie completa presentan, en promedio, medianas de 7 %, 3 % y 6 % y RIQ de 31 %, 29 % y 28 % para GR4H, SAC-SMA y TUW, respectivamente. En cambio, las calibradas solo con caudales de crecida muestran en promedio medianas de 85 %, -2 % y 9 % y rangos de 210 %, 83 % y 108 %, evidenciando que la calibración enfocada en crecidas produce un desempeño bajo e inconsistente para la totalidad de caudales, siendo SAC-SMA el mejor modelo y GR4H el peor, especialmente con calibración centrada en crecidas.
- Al analizar el sesgo porcentual en caudales de crecida (Q95_PBIAS) en validación, las funciones objetivo calibradas con la serie completa presentan, en promedio, medianas de -19 %, -10 % y -11 % y RIQ de 37 %, 34 % y 33 % para GR4H, SAC-SMA y TUW. Las calibradas solo con caudales de crecida muestran en promedio medianas de 0 %, 2 % y 2 % y rangos de 49 %, 34 % y 31 %, evidenciando que la calibración enfocada en crecidas mejora la reproducción de los caudales de crecida,

mientras que el enfoque tradicional los subestima, y GR4H se mantiene como el modelo con mayor dispersión y subestimación en este enfoque.

- Bajo el enfoque tradicional, las funciones objetivo LME y KMoments presentan los RIQ más elevados, especialmente en Q95_PBIAS, indicando mayor dispersión de resultados entre cuencas y, por tanto, menor consistencia en el desempeño.
- La combinación que evidencia mejor equilibrio entre PBIAS y Q95_PBIAS corresponde al modelo SAC-SMA con la función objetivo LCE. En ambos enfoques, esta configuración logra medianas de sesgo más cercanas a 0 % y rangos intercuartílicos más reducidos, destacándose como la opción más robusta y estable.

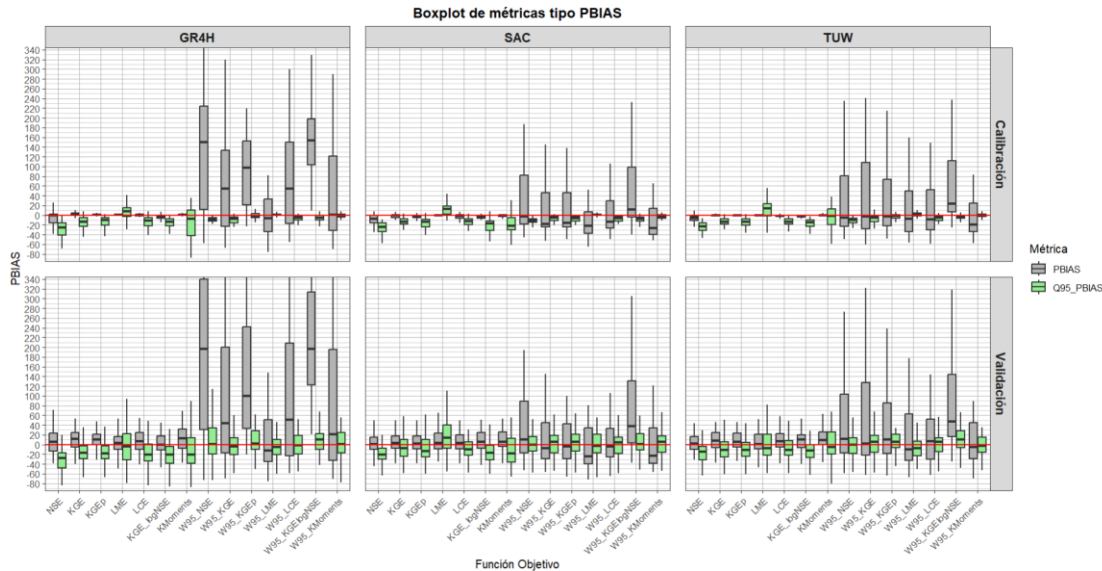


Figura 3. Comparación del sesgo porcentual para todos los caudales (PBIAS, gris) y de crecida (Q95_PBIAS, verde) entre caudales observados y simulados en 43 cuencas, usando los modelos GR4H, SAC-SMA y TUW, siete funciones objetivo y dos enfoques de calibración (tradicional y centrada en crecidas, anteponiendo “W95” a la función objetivo), evaluados en los periodos de calibración y validación.

4.3 CALIDAD DE SIMULACIÓN EN CAUDALES EXTREMOS SEGÚN EL TIPO DE CUENCA

La Figura 4 muestra el sesgo porcentual entre caudales observados y simulados durante el periodo de validación, considerando únicamente los caudales de crecida (Q95_PBIAS, en verde). Los resultados se presentan en boxplots agrupados según el régimen dominante de crecida en cada cuenca, con 5 cuencas nivales, 16 mixtas y 22 pluviales. Se destacan las siguientes características:

- Al utilizar las funciones objetivo calibradas con la serie completa de caudales, se obtienen en promedio medianas de -57% , -31% y -29% para GR4H, SAC-SMA y TUW en cuencas nivales; -19% , -5% y -9% en cuencas mixtas; y -14% , -10% y -6% en cuencas pluviales, mostrando mayor subestimación de caudales a medida que aumenta la componente nival, probablemente por la mayor complejidad de los procesos nivales y porque estas cuencas presentan los valores más elevados del test U de Buishand (sobre 1,7), lo que puede indicar intervención humana, variación en la masa glaciar o alteración de otra variable hidrometeorológica.

- Al analizar las funciones objetivo calibradas únicamente sobre caudales de crecida, se obtienen en promedio medianas de -8% , -15% y -6% para GR4H, SAC-SMA y TUW en cuencas nivales; 6% , 7% y 6% en cuencas mixtas; y -2% , 1% y 0% en cuencas pluviales, indicando que la calibración centrada en crecidas mejora la reproducción de los caudales de crecida en todos los tipos de cuenca, logrando valores de sesgo cercanos a cero o con ligera sobreestimación.
- Al observar los RIQ, se aprecia que las cuencas nivales muestran valores muy diferentes entre sí debido a la complejidad de los procesos nivales; al pasar de cuencas mixtas a pluviales, los rangos aumentan hasta un 10% . Entre las funciones objetivo, LME y KMoments generan la mayor dispersión, evidenciando menor consistencia de resultados entre cuencas para estas configuraciones.

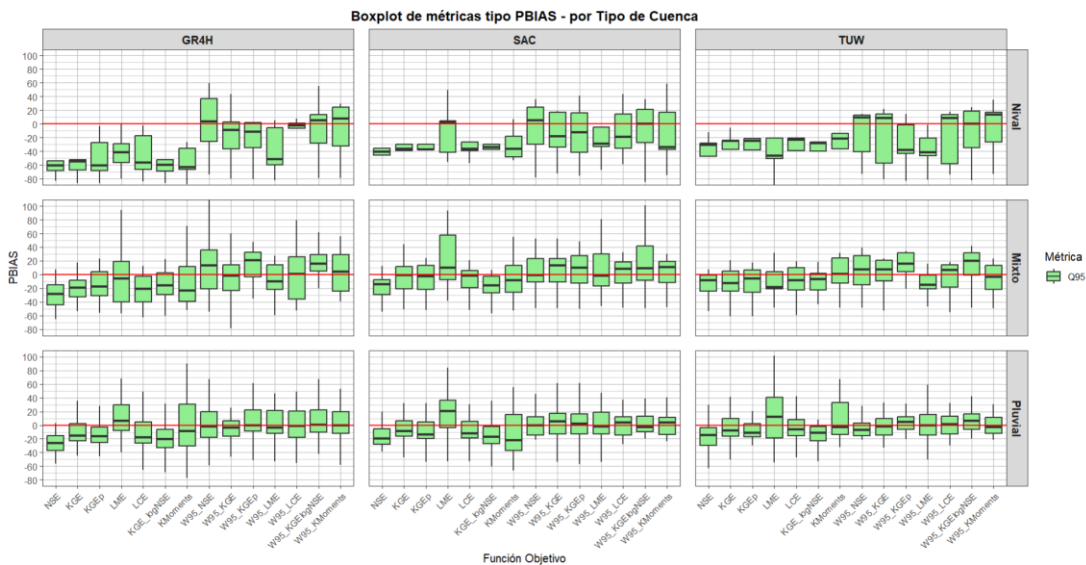


Figura 4. Sesgo porcentual en caudales de crecida (Q95_PBIAS, verde) durante el periodo de validación para combinaciones de modelo y función objetivo, considerando dos enfoques de calibración (tradicional y centrada en crecidas, esta última identificada con “W95”), agrupadas por régimen de gran crecida: 5 cuencas nivales, 16 mixtas y 22 pluviales.

4.4 CAPACIDAD DE DETECCIÓN DE CRECIDAS

La Figura 5 presenta el sesgo categórico (B), la tasa de aciertos (H) y la tasa de falsas alarmas (F) entre caudales simulados y observados en el periodo de validación, considerando los umbrales de crecida definidos por los percentiles 95 (Q95, verde) y 98 (Q98, celeste). Cada boxplot resume el desempeño en las 43 cuencas analizadas, e incluye líneas horizontales rojas que indican los valores óptimos o límites de referencia de cada métrica. Se observan las siguientes características:

- Al analizar el sesgo categórico (B), considerando el umbral de crecida en el percentil 95, las funciones objetivo calibradas con la serie completa presentan, en promedio, medianas de 1.04, 1.32 y 1.05 y RIQ de 0.96, 1.07 y 0.99 para los modelos GR4H, SAC-SMA y TUW, respectivamente. Las calibraciones enfocadas únicamente en eventos de crecida muestran medianas de 4.19, 1.53 y 1.69 y RIQ de 8.31, 1.96 y 2.23 para GR4H, SAC-SMA y TUW, respectivamente, evidenciando que, bajo este

enfoque, se sobreestima el número de crecidas: GR4H simula aproximadamente cuatro veces más crecidas, mientras que SAC-SMA y TUW cerca de una vez y media; además, la variabilidad del sesgo entre cuencas aumenta, especialmente en GR4H.

- Al evaluar la tasa de aciertos (H) con el umbral de crecida en el percentil 95, las funciones objetivo calibradas con la serie completa presentan medianas de 0.47, 0.56 y 0.52 y RIQ de 0.28, 0.32 y 0.33 para GR4H, SAC-SMA y TUW, respectivamente, lo que indica que aproximadamente el 47–56 % de las crecidas fueron simuladas correctamente. Las calibraciones enfocadas solo en eventos de crecida muestran medianas de 0.67, 0.61 y 0.66 y rangos de 0.39, 0.26 y 0.32. Este aumento aparente en H se explica por la sobreestimación del número de crecidas observada en la métrica B, especialmente en GR4H.
- Al considerar la tasa de falsas alarmas (F) con el umbral de crecida en el percentil 95, las funciones objetivo calibradas con la serie completa presentan medianas de 0.03, 0.04 y 0.03 y RIQ de 0.05, 0.05 y 0.04 para GR4H, SAC-SMA y TUW, respectivamente, indicando que entre un 3 y 4 % de las alertas fueron falsas. Las calibraciones enfocadas en crecidas muestran medianas de 0.21, 0.05 y 0.06 y rangos de 0.44, 0.10 y 0.12, evidenciando que, bajo este enfoque, GR4H genera un exceso de alertas, mientras que SAC-SMA y TUW presentan incrementos moderados. Esto confirma que el mayor puntaje de H en GR4H refleja una sobreestimación general de crecidas más que una mejora real en la predicción.
- Al aumentar el umbral de crecida de Q95 a Q98, el sesgo categórico (B) tiende a aumentar, la tasa de falsas alarmas (F) disminuye y la tasa de aciertos (H) en algunos tiene una variación marginal. Esto ocurre porque un umbral más alto reduce la ocurrencia de horas clasificadas como crecida, disminuyendo las falsas alarmas, pero al mismo tiempo el número de crecidas observadas disminuye, lo que provoca una sobreestimación del número de crecidas simuladas respecto a las observadas. Al comparar las tres métricas categóricas, GR4H presenta la mayor variabilidad entre cuencas, atribuible a su estructura simple de seis parámetros. TUW ofrece resultados aceptables, pero con variaciones notables entre cuencas y funciones objetivo, especialmente en B y F, evidenciando cierta falta de robustez. En contraste, SAC-SMA muestra menor variabilidad y un desempeño más consistente al cambiar de enfoque, destacándose como el modelo más robusto y confiable.
- Estos resultados indican que la calibración enfocada en crecidas no es adecuada para predecir la ocurrencia de crecidas, pero sí resulta prometedora para estimar su magnitud y la forma del hidrograma.
- Con base en que la función objetivo LCE mostró el mejor equilibrio entre el sesgo porcentual general (PBIAS) y el de crecidas (Q95_PBIAS), y además alcanza la mayor tasa de aciertos manteniendo un sesgo categórico y una tasa de falsas alarmas comparables a los demás enfoques, se reafirma su potencial para la simulación de caudales máximos.

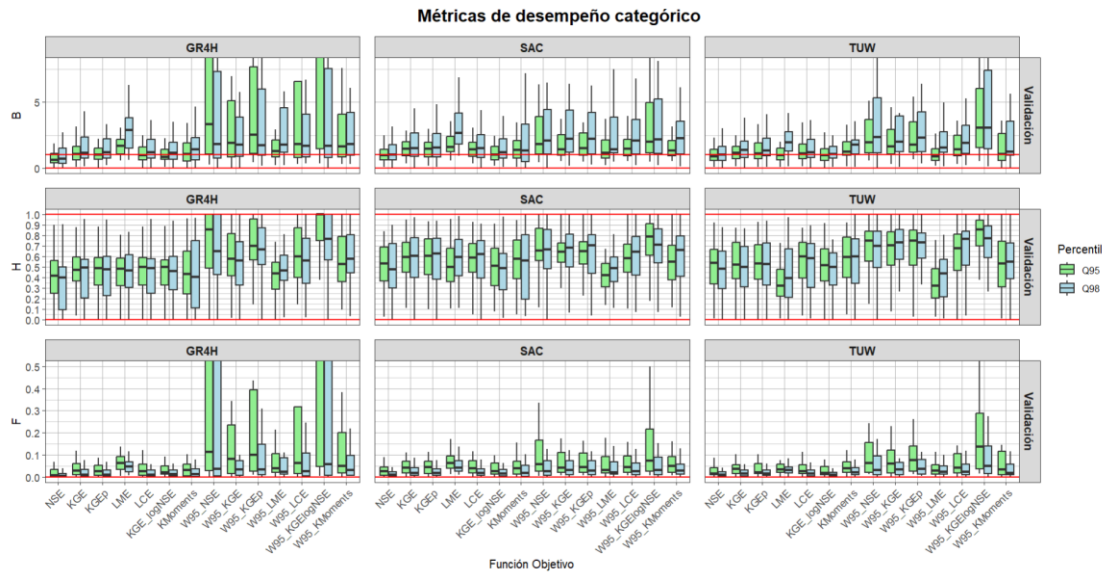


Figura 5. Comparación del sesgo categórico (B), tasa de aciertos (H) y tasa de falsas alarmas (F) durante el periodo de validación, considerando el umbral de crecida en los percentiles 95 (verde) y 98 (celeste) en 43 cuencas, usando los modelos GR4H, SAC-SMA y TUW, siete funciones objetivo y dos enfoques de calibración. Se incluyen líneas rojas que marcan los valores de referencia de cada métrica.

5. CONCLUSIÓN

En este artículo se calibraron tres modelos hidrológicos a escala horaria (GR4H, SAC-SMA y TUW) utilizando siete funciones objetivo y dos enfoques de calibración: el tradicional, que evalúa la función objetivo considerando toda la serie de caudales, y el enfoque centrado en crecidas, que se enfoca únicamente en los hidrogramas que superan el percentil 95. El estudio abarcó 43 cuencas chilenas, analizando la consistencia de los modelos y funciones objetivo entre distintas cuencas, su desempeño según el tipo de régimen hidrológico y su capacidad para reproducir eventos extremos, con especial atención a su aplicabilidad en sistemas de alerta temprana contra inundaciones. A partir de los resultados obtenidos, se concluye lo siguiente:

- Se identificaron limitaciones en los datos meteorológicos, especialmente en las precipitaciones ERA5, cuyas distribuciones poco realistas afectaron la simulación de crecidas. Las cuencas con componente nival mostraron menor desempeño y consistencia, debido a la complejidad de los procesos de fusión nival y glaciario.
- Entre los modelos, SAC-SMA presentó el mejor desempeño y mayor robustez, generando hidrogramas realistas y graduales, siendo la opción más recomendable para aplicaciones operativas. TUW mostró resultados prometedores, aunque menos consistentes, mientras que GR4H evidenció mayor sensibilidad a la función objetivo y mayor dispersión entre cuencas. LME y KMoments mostraron un desempeño inconsistente y variable entre cuencas. En cambio, LCE ofreció el mejor equilibrio en el sesgo porcentual y un rendimiento estable, siendo adecuada para su uso en sistemas de alerta temprana.
- La calibración centrada en extremos mejoró la representación de la magnitud de las crecidas, pero aumentó el número de falsas alarmas, lo que limita su aplicación

directa. El enfoque tradicional, aunque con mayor sesgo en la magnitud del caudal, resultó más confiable para detectar la superación de umbrales.

Finalmente, se propone una estrategia combinada que emplea la calibración centrada en extremos para representar con mayor precisión la magnitud de las crecidas y el enfoque tradicional para la activación de alertas, logrando así una mayor confiabilidad y eficacia del sistema de respuesta ante inundaciones, especialmente en cuencas con alta complejidad hidrológica.

AGRADECIMIENTOS

Esta investigación fue financiada por el proyecto ANID FONDECYT de Iniciación N.º 11250238, por el proyecto interno de investigación PI_LIR_24_01 de la Universidad Técnica Federico Santa María, y contó con el apoyo de la beca de asistencia al Congreso Chileno de Ingeniería Hidráulica 2025.

REFERENCIAS

- Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., ... & Ayala, A. (2018). The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies—Chile dataset. *Hydrology and Earth System Sciences*, 22(11), 5817-5846. <https://doi.org/10.5194/hess-22-5817-2018>
- Anderson, E. A. (1973). National Weather Service river forecast system: Snow accumulation and ablation model (Vol. 17). US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service. <https://repository.library.noaa.gov/view/noaa/13507>
- Araya, D., Mendoza, P. A., Muñoz-Castro, E., & McPhee, J. (2023). Towards robust seasonal streamflow forecasts in mountainous catchments: impact of calibration metric selection in hydrological modeling. *Hydrology and Earth System Sciences*, 27(24), 4385-4408. <https://doi.org/10.5194/hess-27-4385-2023>
- Bergström, S. (1976). Development and application of a conceptual runoff model for Scandinavian catchments, Report RHO 7, SMHI, Norrköping, Sweden. <http://www.diva-portal.org/smash/record.jsf?pid=diva2:1456191&dswid=4221>
- Bergström, S., Harlin, J., & Lindström, G. (1992). Spillway design floods in Sweden: I. New guidelines. *Hydrological Sciences Journal*, 37(5), 505-519. <https://doi.org/10.1080/02626669209492615>
- Burnash, R., Ferral, R., & McGuire, R. (1973). A generalized streamflow simulation system – Conceptual modeling for digital computers, Sacramento, California. <https://searchworks.stanford.edu/view/753303>
- Cantoni, E., Trambly, Y., Grimaldi, S., Salamon, P., Dakhlaoui, H., Dezetter, A., & Thiemi, V. (2022). Hydrological performance of the ERA5 reanalysis for flood modeling in Tunisia with the LISFLOOD and GR4J models. *Journal of Hydrology: Regional Studies*, 42, 101169. <https://doi.org/10.1016/j.ejrh.2022.101169>
- Centro de Estudios Hidrográficos (CEDEX). (1994). Métodos de estimación del tiempo de concentración en pequeñas cuencas naturales. Madrid, España: Ministerio de Obras Públicas, Transportes y Medio Ambiente, Dirección General de Obras Hidráulicas.
- Doswell, C. H., Davies-Jones, R., & Keller, D. L. (1990). On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting*, 5(4), 576-585.
- Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3), 501-521. <https://doi.org/10.1007/BF00939380>
- Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., ... & Cloke, H. L. (2016). Continental and global scale flood forecasting systems. *Wiley Interdisciplinary Reviews: Water*, 3(3), 391-418. <https://doi.org/10.1002/wat2.1137>

- 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>
- Jaiswal, R. K., Ali, S., & Bharti, B. (2020). Comparative evaluation of conceptual and physical rainfall–runoff models. *Applied Water Science*, 10(1), 48. <https://doi.org/10.1007/s13201-019-1122-6>
- Lane, R. A., Coxon, G., Freer, J. E., Wagener, T., Johnes, P. J., Bloomfield, J. P., ... & Reaney, S. M. (2019). Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. *Hydrology and Earth System Sciences*, 23(10), 4011-4032. <https://doi.org/10.5194/hess-23-4011-2019>
- Lee, J. S., & Choi, H. I. (2022). A rebalanced performance criterion for hydrological model calibration. *Journal of Hydrology*, 606, 127372. <https://doi.org/10.1016/j.jhydrol.2021.127372>
- Liu, D. (2020). A rational performance criterion for hydrological model. *Journal of Hydrology*, 590, 125488. <https://doi.org/10.1016/j.jhydrol.2020.125488>
- Lohmann, D. A. G., Nolte-Holube, R. A. L. P. H., & Raschke, E. (1996). A large-scale horizontal routing model to be coupled to land surface parametrization schemes. *Tellus A*, 48(5), 708-721. <https://doi.org/10.1034/j.1600-0870.1996.t01-3-00009.x>
- Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019). On the choice of calibration metrics for “high-flow” estimation using hydrologic models. *Hydrology and Earth System Sciences*, 23(6), 2601-2614. <https://doi.org/10.5194/hess-23-2601-2019>
- Mudashiru, R. B., Sabtu, N., Abustan, I., & Balogun, W. (2021). Flood hazard mapping methods: A review. *Journal of Hydrology*, 603, 126846. <https://doi.org/10.1016/j.jhydrol.2021.126846>
- J. Muñoz-Sabater, Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J.-N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data*, 13, 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, 10(3), 282-290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Parajka, J., Merz, R., & Blöschl, G. (2007). Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments. *Hydrological Processes: An International Journal*, 21(4), 435-446. <https://doi.org/10.1002/hyp.6253>
- Pardé, M. (1933). *Cours d'hydrologie appliquée*. Gauthier-Villars.
- Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279, 275–289. [https://doi.org/10.1016/S0022-1694\(03\)00225-7](https://doi.org/10.1016/S0022-1694(03)00225-7)
- Pizarro, A., & Jorquera, J. (2024). Advancing objective functions in hydrological modelling: Integrating knowable moments for improved simulation accuracy. *Journal of Hydrology*, 634, 131071. <https://doi.org/10.1016/j.jhydrol.2024.131071>
- Skøien, J. O., Blöschl, G., Laaha, G., Pebesma, E., Parajka, J., & Viglione, A. (2014). rtop: An R package for interpolation of data with a variable spatial support, with an example from river networks. *Computers & Geosciences*, 67, 180-190. <https://doi.org/10.1016/j.cageo.2014.02.009>
- Sun, W., Wang, Y., Wang, G., Cui, X., Yu, J., Zuo, D., & Xu, Z. (2017). Physically based distributed hydrological model calibration based on a short period of streamflow data: case studies in four Chinese basins. *Hydrology and Earth System Sciences*, 21(1), 251-265. <https://doi.org/10.5194/hess-21-251-2017>
- Swain, D. L., Wing, O. E., Bates, P. D., Done, J. M., Johnson, K. A., & Cameron, D. R. (2020). Increased flood exposure due to climate change and population growth in the United States. *Earth's Future*, 8(11), e2020EF001778. <https://doi.org/10.1029/2020EF001778>
- Valéry, A., Andréassian, V., & Perrin, C. (2014). ‘As simple as possible but not simpler’: What is useful in a temperature-based snow-accounting routine? Part 2–Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. *Journal of Hydrology*, 517, 1176-1187. <https://doi.org/10.1016/j.jhydrol.2014.04.058>