

Benchmarking reservoir operation schemes for large-scale hydrological models

Jesús Casado-Rodríguez^{1, 2}, Juliana Disperati³, Stefania Grimaldi¹, and Peter Salamon¹

¹European Commission - Joint Research Centre, Ispra (Italy)

²Kajo s.r.o., Sládkovičova (Slovakia)

³Fincons Group, Vimercate (Italy)

Correspondence: Jesús Casado-Rodríguez (j.casado@kajoservices.com)

Abstract. There are approximately 62,000 large dams worldwide that significantly alter the hydrological regimes of most major rivers. Despite their importance, reservoirs remain poorly represented in Large-Scale Hydrological Models (LSHMs) due to the complexity of human-driven operations and a widespread lack of observational records. Consequently, reservoir routines in LSHMs must balance structural simplicity with limited data requirements. In this study, we utilize the ResOpsUS

5 dataset to benchmark four reservoir routines of increasing complexity: LISFLOOD, CaMa-Flood, mHM, and STARFIT. We evaluate these routines across 164 reservoirs in the United States and test which target variables are most informative for parameter estimation. Our results indicate that the mHM routine consistently achieves the highest performance; however,

its dependence on site-specific demand data limits its applicability at the global scale. In contrast, the CaMa-Flood routine provides a robust compromise, significantly outperforming the linear logic of LISFLOOD while maintaining parsimonious 10 data requirements. Crucially, we find that calibrating to reservoir storage is more informative than calibrating to outflow, as it effectively captures the dynamics of both state variables. This finding paves the way for the use of satellite-derived storage products in the calibration of LSHMs. The findings of this study have been implemented in the upcoming versions of the European and Global Flood Awareness Systems (EFAS v6 and GloFAS v5).

1 Introduction

15 The increasing global population and expanding economic activities have amplified the pressure on freshwater resources, making human interventions—such as the construction and operation of dams and reservoirs—fundamental elements of the terrestrial water cycle (Biemans et al., 2011; Busker et al., 2019; Moreno-Rodenas et al., 2025; Sadki et al., 2023). These structures are pivotal in fulfilling diverse societal needs, including domestic and industrial water supply, irrigation, hydropower production, and flood control (Moreno-Rodenas et al., 2025; Abeshu et al., 2023; Lehner et al., 2024; Mekonnen and Hoekstra, 2012; Sadki et al., 2023; Shrestha et al., 2024). The global proliferation of dams has been unprecedented over the last century (Lehner et al., 2024; Moreno-Rodenas et al., 2025). Globally, there are approximately 62,000 large dams (heights > 20 15 m), cataloged by the ICOLD (International Commission on Large Dams) (Moreno-Rodenas et al., 2025), with an estimated cumulative storage capacity ranging between 7,420 km³ (GDW v1.0) (Lehner et al., 2024) and 8,300 km³ (Sadki et al., 2023).

This volume represents roughly 20% of the average annual river discharge to the oceans and is four times the amount of water
25 stored in river channels worldwide (Sadki et al., 2023).

The immense scale of this intervention profoundly impacts natural hydrology by modifying the magnitude, timing, and duration of streamflows (Salwey et al., 2024; Shin et al., 2019), resulting in the fragmentation of over 60% of the world's largest rivers (Sadki et al., 2023). In the contiguous United States (CONUS), dams regulate all major rivers, providing a total storage capacity equivalent to roughly 75% of the mean annual CONUS runoff (Steyaert et al., 2022). While beneficial
30 for human use, these operations cause significant environmental trade-offs, including the homogenization of river dynamics (Lehner et al., 2024; Scherer and Pfister, 2016), alteration of sediment and nutrient transport (Shin et al., 2019; Yassin et al., 2019), and increased open-water evaporation, which can exacerbate water scarcity (Yassin et al., 2019; Scherer and Pfister, 2016).

Despite their importance, the accurate representation of complex reservoir operations remains a challenge in large-scale
35 hydrological models (LSHMs) and land-surface models (LSMs) (Sadki et al., 2023; Salwey et al., 2024; Turner et al., 2021). Reservoir operations are intrinsically non-linear and governed by complex, human-driven decisions related to water demand, flood risk, and environmental constraints (Shrestha et al., 2024; Coerver et al., 2018; Hanazaki et al., 2022; Salwey et al., 2024;
40 Steyaert et al., 2022). Consequently, models lacking adequate reservoir schemes often exhibit degraded predictive performance, especially in highly regulated basins (Salwey et al., 2024; Turner et al., 2021; Steyaert et al., 2025). Historically, global models have relied on generic operational rules utilizing static characteristics (e.g., capacity, primary purpose) from datasets like the Global Reservoir and Dam (GRanD) database (Lehner et al., 2011). However, these generic approaches often fail to capture the local operating behaviors necessary for simulating realistic daily releases and storage dynamics (Turner et al., 2021). Recent research has therefore pivoted toward exploiting newly available observational data and integrating data-driven techniques, such as machine learning or statistically inferred policies, to better reflect historical decision-making (Turner et al., 2021;
45 Coerver et al., 2018; Steyaert et al., 2025).

This research aims to improve the reservoir module within the **LISFLOOD** hydrological model (Roo et al., 2000; van der Knijff et al., 2010; Burek et al., 2013), a core component of the Global and European Flood Awareness Systems (GloFAS and EFAS) (Joint Research Centre - European Commission, 2025; Commission, 2025). Current LISFLOOD parameterizations typically employ a single general reservoir operation scheme based on three fixed storage limits (minimum, normal, and
50 non-damaging outflow) without explicit classification of reservoir purpose (Zajac et al., 2017). Our objective is to identify a reservoir routine suitable for global application—one that balances portability with efficacy in data-scarce environments.

Specifically, this study benchmarks the performance of four distinct methodologies—ranging from classical parameterizations to data-driven approaches—against observed records. The methodologies include the current LISFLOOD parameterization, a recently published evolution (CaMa-Flood), a demand-driven approach (mHM), and a seasonality-driven approach
55 (STARFIT). By assessing their capability to simulate observed flow regulation and storage dynamics, this research provides critical guidance for the selection and development of human-water interaction modules in the next generation of operational flood systems and large-scale hydrological models.

A novel aspect of this study is the exploration of a decoupled calibration strategy. Rather than calibrating reservoir parameters simultaneously with the rest of the hydrological model against downstream river discharge—a process often plagued by equifinality (Beven and Freer, 2001)—we define reservoir parameters in advance using specific reservoir records. These pre-determined parameters are then integrated into the hydrological model for the subsequent calibration of the remaining parameters. This approach ensures that reservoir parameters are sensitive to actual reservoir dynamics rather than being masked by gauge observations far downstream.

However, the practical implementation of this strategy faces a significant data bottleneck. Often, river discharge in gauging stations downstream the reservoir has been used as the variable informing about reservoir operations. While ground-based records or reservoir outflow provide a more direct target, this data is frequently inaccessible or classified as private. To overcome these limitations, satellite-derived products offer a promising alternative (Pekel et al., 2016; Schwatke et al., 2015, 2020; Donchyts et al., 2022; Khandelwal et al., 2022; Hou et al., 2024). These products offer the potential to monitor volume fluctuations in ungauged regions, making them crucial for future global model calibration and data assimilation.

A calibration based on satellite products would target reservoir level or storage, instead of outflow (or river discharge as a proxy variable). To assess the feasibility of using satellite products, we compare the performance of the model when calibrated against three distinct targets: (i) reservoir release only, (ii) reservoir storage only, and (iii) a combined objective function utilizing both variables. By evaluating these configurations, this study aims to assess whether satellite-inferred storage can serve as a robust alternative to ground-based outflow records, thereby enabling improved reservoir modeling in ungauged or data-scarce regions.

The benchmarking is supported by the ResOpsUS dataset (Steyaert et al., 2022; Steyaert and Condon, 2024), the first comprehensive, national-scale inventory providing historical daily reservoir operations for 679 major dams across the CONUS. ResOpsUS includes records of reservoir inflow, level, storage and outflow (not all variables for all reservoirs), providing the necessary variables for our benchmark. We leverage this data-rich environment to test the models and calibration targets, establishing a proof-of-concept for operational environments where such granular data is unavailable.

In summary, this study benchmarks four reservoir modeling paradigms across the CONUS domain to address three primary objectives: (i) to identify the most robust reservoir routine for the next operational versions of GloFAS and EFAS; (ii) to determine the most suitable target variable for calibration—storage, outflow, or both; and (iii) to evaluate whether remotely sensed storage fluctuations are sufficiently informative to calibrate large-scale models.

2 Data

2.1 Reservoir Selection and Filtering

To evaluate reservoir routines in a data-rich context, we utilized the ResOpsUS dataset (Steyaert et al., 2022; Steyaert and Condon, 2024), which provides daily time series of reservoir operations for 679 reservoirs across the conterminous United States (CONUS). From this initial pool, we applied a rigorous filtering process to ensure the suitability of the reservoirs for large-scale hydrological modeling:

1. Data Availability: We selected only those reservoirs containing concurrent daily records for all three primary variables: inflow (I), storage (V), and outflow (Q).

2. Physical Dimensions: Reservoirs were required to have a minimum catchment area of 50 km² and a minimum storage capacity (V_{max}) of 10 hm³.

95 3. Hydrological Impact: Following Shrestha et al. (2024), we excluded "non-disruptive" reservoirs—those that do not significantly alter the natural flow regime. This was quantified using the Degree of Regulation (DOR; Eq. 1) and the Degree of Disruptivity (DOD; Eq. 2). Reservoirs with DOR < 0.08 or DOD < 0.06 m were discarded:

$$DOR = \frac{V_{max}}{\bar{I}_{yr}} \quad [-] \quad (1)$$

100 $DOD = \frac{V_{max}}{A_c} \quad [m] \quad (2)$

where V_{max} is the reservoir storage capacity [m³], \bar{I}_{yr} is the average annual inflow [m³/year], and A_c is the catchment area [m²].

4. Record Length: To ensure sufficient data for robust model calibration, we removed reservoirs with time series shorter than four years.

105 5. Water Balance Integrity: Preliminary analysis revealed large biases between average inflow and outflow in several reservoirs. Such discrepancies may stem from poor data quality or significant consumptive use (e.g., water diversions) not explicitly represented in our modeling framework. To ensure a closed water balance for benchmarking, we removed reservoirs where the absolute bias exceeded 30%.

110 Applying these criteria resulted in a final selection of 164 reservoirs. Their geographical distribution, categorized by catchment area and storage capacity, is depicted in Fig. 1.

2.2 Dataset Integration and Attribute Enrichment

To satisfy the input requirements of the diverse reservoir routines, we enriched the ResOpsUS records with supplementary hydrometeorological time series, and catchment and reservoir attributes.

115 As detailed in Sect. 3.1, the routines require time series of direct precipitation and evaporation. These were extracted for each reservoir location from the ERA5 reanalysis dataset (Hersbach et al., 2020).

Static reservoir and dam attributes were primarily sourced from the Global Reservoir and Dam dataset (GRanD) (Lehner et al., 2011). While we included all GRanD attributes in our dataset, four were critical for the simulations: storage capacity (V_{max}), maximum surface area (A_{max}), dam crest elevation (Z_{max}), and dam height (H_{dam}). A comparison of these attributes against observed time series revealed inconsistencies in some attribute values. To rectify this, we cross-referenced GRanD

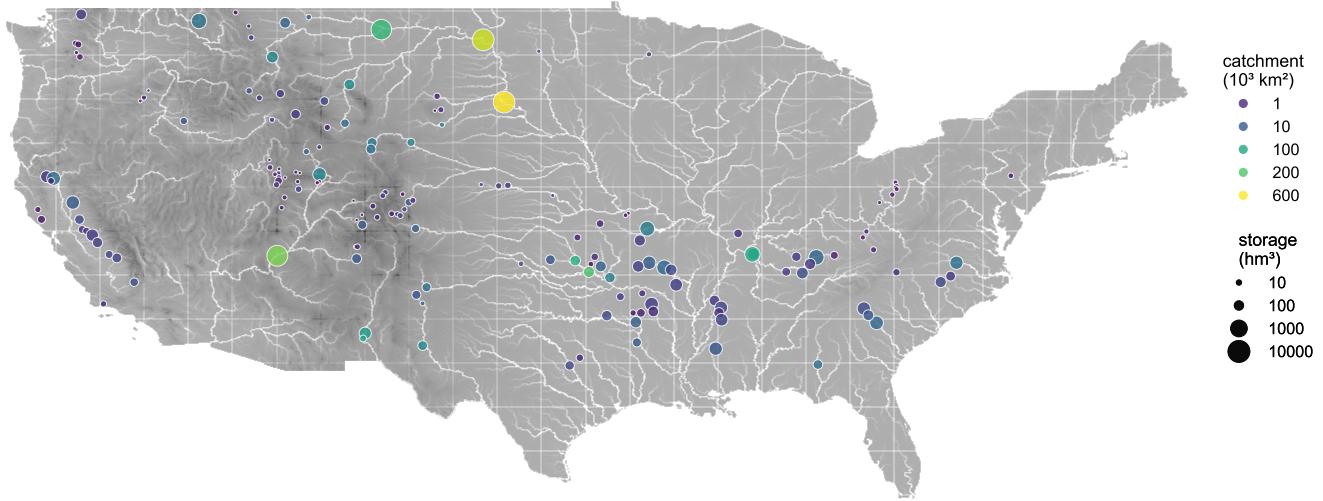


Figure 1. Distribution of the 164 reservoirs from the ResOpsUS dataset used in the benchmarking. Colors represent the catchment area and point size represents the storage capacity.

120 with the Global Dam Watch (GDW) dataset (Lehner et al., 2024). For instances where the attributes remained inconsistent, we manually updated the records using official data from U.S. authorities, including the U.S. Army Corps of Engineers' National Inventory of Dams (NID), the U.S. Bureau of Reclamation (USBR), and the U.S. Geological Survey (USGS) (US Army Corps of Engineers, 2025; US Bureau of Reclamation, 2025; US Geological Survey, 2025). Similar errors were reported by Shrestha et al. (2024).

125 For the purpose of parameter regionalization or the development of data-driven models, we further included catchment-scale characteristics derived from GloFAS surface fields (Choulga et al., 2024) and climate indices derived from ERA5 (Hersbach et al., 2020).

We named the enriched dataset *Reservoir Operations US and CAatchment and Reservoir Static attributes* (ResOpsUS+CARS). It is available via Zenodo. The dataset structure follows the standards defined by the CARAVAN initiative (Kratzert et al., 2023),
130 ensuring it is formatted for immediate use in deep learning and large-sample hydrological studies.

3 Methods

3.1 Reservoir models

Each of the reservoir routines evaluated in this study is governed by the water mass balance equation (Eq. 3). At each time step Δt , the model calculates the reservoir storage at the end of the interval (V_{t+1}) based on the preceding storage state (V_t), the
135 water surface area (A_t), and three forcing variables: inflow (I_t), precipitation (P_t), and evaporation (E_t):

$$V_{t+1} = V_t + (P_t - E_t) \cdot A_t + (I_t - Q_t) \cdot \Delta t \quad (3)$$

where Q_t is the reservoir outflow (release), the fundamental difference between the four reservoir schemes.

To represent the reservoir's bathymetry, we assume a simplified triangular pyramid geometry, following the approach of Liebe et al. (2005) and subsequent studies (Shin et al., 2019; Sadki et al., 2023; Shrestha et al., 2024). This geometric approximation allows for the continuous estimation of the water surface area (Eq. 4), which is required for the mass balance, and the water level (Z_t ; Eq. 5), provided that the dam crest elevation (Z_{max}) and the dam height (H_{dam}) are known. We evaluate the validity of this geometric assumption by comparing simulated levels against observed records in a subset of reservoirs where such data is available.

$$A_t = A_{max} \left(\frac{V_t}{V_{max}} \right)^{\frac{2}{3}} \quad (4)$$

145

$$h_t = H_{dam} \left(\frac{V_t}{V_{max}} \right)^{\frac{1}{3}} \quad (5)$$

$$Z_t = Z_{max} - H_{dam} + h_t$$

where h_t represents the water depth at the dam.

3.1.1 LISFLOOD

The current LISFLOOD reservoir routine (Burek et al., 2013) operates as a piecewise linear storage-outflow relationship. The 150 release is determined by the storage zone in which the reservoir resides at any given time step: conservative, normal or flood.

$$Q_t = \begin{cases} Q_{min} & \text{if } V_t < 2 \cdot V_{min} \\ Q_{min} + (Q_n - Q_{min}) \frac{V_t - 2 \cdot V_{min}}{V_n - 2 \cdot V_{min}} & \text{if } 2 \cdot V_{min} \leq V_t < V_n \\ Q_n & \text{if } V_n \leq V_t < V_{n,adj} \\ Q_n + (Q_f - Q_n) \frac{V_t - V_{n,adj}}{V_f - V_{n,adj}} & \text{if } V_{n,adj} \leq V_t < V_f \\ \max \left(\frac{V_t - V_f}{\Delta t}, \min(Q_f, \max(k \cdot I_t, Q_n)) \right) & \text{if } V_t > V_f \end{cases} \quad (6)$$

where V_{min} , V_n , $V_{n,adj}$ and V_f are the minimum storage, the lower and upper bounds of the normal storage zone, and the flood storage limit, respectively. Q_{min} , Q_n and Q_f are the corresponding outflow values associated with these limits.

In the current versions of EFAS and GloFAS, only two parameters were tuned: the normal outflow (Q_n) and the upper limit 155 of the normal storage zone ($V_{n,adj}$). The remaining breakpoints are kept at their default values to limit the dimensionality of the global calibration (which currently involves 14 parameters per basin).

Table 1. Calibration parameters in the LISFLOOD reservoir module.

parameter	description	units	minimum	maximum	default
α	Fraction of the total storage corresponding to the flood limit (V_f)	-	0.2	0.99	0.97
β	Proportion between flood limit and minimum storage corresponding to the normal limit	-	0.001	0.999	0.655
γ	Proportion between flood and normal limits corresponding to the adjusted normal limit	-	0.001	0.999	0.633
δ	Factor multiplying the 100-year inflow that defines the flood outflow (Q_f)	-	0.1	0.5	0.3
ϵ	Ratio between normal and flood outflows	-	0.001	0.999	$\frac{Q_f}{I}$
k	Release coefficient	-	1	5	1.2

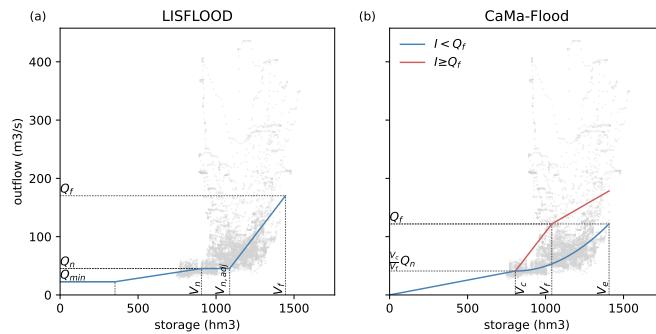


Figure 2. Storage–outflow relationships for the (a) LISFLOOD and (b) CaMa-Flood reservoir routines. The curves illustrate the calibrated operational rules for the Yellowtail dam (GRanD ID 355). Grey points represent observed daily records, highlighting the spread of historical operations.

However, in this benchmarking study, we allow maximum flexibility to the LISFLOOD routine by calibrating the six parameters listed in Table 1. This expanded parameterization serves two purposes. On one hand, it ensures a fair comparison with more flexible, data-driven schemes. On the other hand, because our decoupled calibration focuses exclusively on reservoir parameters rather than the entire hydrological model, we can afford the increased dimensionality to explore the routine’s maximum performance potential.

A fundamental characteristic of this scheme is that it creates a unique (one-to-one) relationship between storage and outflow. Regardless of seasonality, current inflow or water demand, the outflow is identical for any given storage state. The only exception occurs in the flood zone ($V_t \geq V_f$), where the inclusion of the inflow (I_t) in the release logic allows for deviations from this static behavior. This makes the LISFLOOD scheme an ideal baseline for assessing whether more complex, dynamic routines are necessary to capture observed reservoir behavior.

3.1.2 CaMa-Flood

The reservoir module in the Catchment-based Macro-scale Floodplain model (CaMa-Flood) (Hanazaki et al., 2022) represents an evolution of the LISFLOOD approach by introducing inflow-dependency into the operational logic. Unlike the static

170 LISFLOOD scheme, CaMa-Flood distinguishes between two operational modes based on the current inflow (I_t).

When I_t is below a threshold (Q_f), the outflow is modeled as a quadratic function of storage. This non-linear behavior restricts releases as the reservoir empties, effectively conserving water for future demand. Conversely, when I_t exceeds Q_f , the system transitions toward a linear reservoir behavior to manage high-flow conditions.

$$Q_t = \begin{cases} Q_n \frac{V_t}{V_f} & \text{if } V_t < V_{\min} \\ Q_n \frac{V_{\min}}{V_f} + \left(\frac{V_t - V_{\min}}{V_e - V_{\min}} \right)^2 \left(Q_f - Q_n \frac{V_{\min}}{V_f} \right) & \text{if } I_t < Q_f \text{ and } V_{\min} \leq V_t < V_e \\ Q_f & \text{if } I_t < Q_f \text{ and } V_t \geq V_e \\ Q_n \frac{V_{\min}}{V_f} + \frac{V_t - V_{\min}}{V_f - V_{\min}} \left(Q_f - Q_n \frac{V_{\min}}{V_f} \right) & \text{if } I_t \geq Q_f \text{ and } V_{\min} \leq V_t < V_f \\ Q_f + k \cdot \frac{V_t - V_f}{V_e - V_f} \cdot (I_t - Q_f) & \text{if } I_t \geq Q_f \text{ and } V_f \leq V_t < V_e \\ I_t & \text{if } I_t \geq Q_f \text{ and } V_t \geq V_e \end{cases} \quad (7)$$

175 where k is the release coefficient that limits reservoir releases under flood conditions based on available storage capacity (Eq. 8). In the original formulation (Hanazaki et al., 2022), k was a fixed value computed using the flood volume (V_f). In that definition, k becomes zero for high-regulation reservoirs, resulting in a constant release regardless of whether the reservoir actually has available storage at a given time step or not. To correct this unrealistic behavior, we introduced a dynamic calculation of k at each time step based on the instantaneous reservoir volume (V_t).

$$180 \quad k = \max \left(1 - \frac{V_{\max} - V_t}{A_c \cdot 0.2}, 0 \right) \quad (8)$$

While Hanazaki et al. (2022) utilized default parameters, we calibrated five parameters (Table 2) to ensure the routine has comparable degrees of freedom to the other routines. To maintain consistency with the LISFLOOD benchmark, three parameters (α, δ, ϵ) share identical definition across both models.

185 The primary advantage of the CaMa-Flood routine over LISFLOOD is its ability to represent a non-unique relationship between storage and outflow. By incorporating I_t into the operational rules, the routine can better capture the observed dispersion in the storage-outflow values, providing a more realistic representation of human-regulated flow variability.

3.1.3 mHM

The reservoir module within the Multiscale Hydrological Model (mHM) (Samaniego et al., 2010; Kumar et al., 2013; Thober et al., 2019) represents a fundamental shift in modeling philosophy compared to the LISFLOOD and CaMa-Flood routines.

190 Rather than treating outflow as a direct function of storage, the mHM routine determines releases primarily based on a water

Table 2. Calibration parameters in the CaMa-Flood reservoir module.

parameter	description	units	minimum	maximum	default
α	Fraction of the total storage corresponding to the flood limit (V_f)	-	0.2	0.99	0.97
β	Proportion between flood limit and total storage corresponding to the extreme limit	-	0.001	0.999	0.2
γ	Proportion of the flood limit corresponding to the normal limit	-	0.001	0.999	0.5
δ	Factor multiplying the 100-year inflow that defines the flood outflow (Q_f)	-	0.1	0.5	0.3
ϵ	Ratio between normal and flood outflows	-	0.001	0.999	$\frac{Q_f}{I}$

demand time series (D_t). While Shrestha et al. (2024) utilized random forests to predict D_t for individual reservoirs, we implement a seasonal approach to maintain model portability.

In our implementation, we derive an annual demand climatology by computing the mean observed outflow for each day of the year across the available record. To ensure this signal reflects target demand rather than total spill, we apply a factor 195 representing the proportion of outflow dedicated to supply and smooth the result using a 28-day moving average. This smoothed climatology serves as the seasonal demand signal D_t for each reservoir, introducing the temporal variability missing in static storage–discharge schemes.

3.1.4 Reservoir release

The total daily release Q_t is partitioned into a component satisfying the hedged demand (\hat{D}_t) and a component representing 200 the passthrough of inflow (I_t):

$$Q_t = \rho \cdot \kappa_t \cdot \hat{D}_t + (1 - \rho) \cdot I_t \quad (9)$$

The partition coefficient ρ is a function of the reservoir's degree of regulation (DOR). For highly regulated reservoirs ($DOR \geq \alpha$), the release is entirely demand-driven ($\rho = 1$). In all other cases, the influence of demand relative to inflow is governed by the parameters α and β :

$$205 \quad \rho = \min \left(1, \left(\frac{DOR}{\alpha} \right)^\beta \right) \quad (10)$$

The demand component is further moderated by a time-varying coefficient κ_t , which hedges the release based on the current storage level V_t relative to a target "normal" storage V_n :

$$\kappa_t = \left(\frac{V_t}{V_n} \right)^\lambda = \left(\frac{V_t}{\gamma \cdot V_{max}} \right)^\lambda \quad (11)$$

where γ defines the normal storage fraction and λ controls the sensitivity of the hedging to storage deficits.

Table 3. Calibration parameters in the mHM reservoir module.

parameter	description	units	minimum	maximum	default
α	Threshold in the degree of regulation that defines demand-controlled reservoirs (DOR > α)	-	0.0	5.0	0.5
β	It controls indirectly the proportion of inflow and demand in the releases	-	0.5	3.0	1.0
γ	Ration between normal and total storage	-	0.0	1.0	$\frac{V_{0.9}}{V_{tot}}$
λ	It further controls hedging based on the current reservoir storage	-	0.25	3.0	1.0
ω	It controls hedging based on the ratio between the current and average demand	-	0.0	1.0	0.1

210 3.1.5 Demand hedging

The hedged demand \hat{D}_t is calculated based on the ratio between average annual demand (\bar{D}) and inflow (\bar{I}). The parameter ω represents the expected annual water surplus and is used to categorize the reservoir's water stress. If there is no water stress ($\bar{D}/\bar{I} < 1 - \omega$), the daily demand (D_t) is supplied in addition to the average water surplus. In there is stress, the hedged demand is a combination of a fraction of the average inflow and a fraction of the daily demand.

$$215 \quad \hat{D}_t = \begin{cases} \bar{I} - \bar{D} + D_t & \text{if } \frac{\bar{D}}{\bar{I}} < 1 - \omega \\ \omega \cdot \bar{I} + (1 - \omega) \frac{\bar{I}}{\bar{D}} \cdot D_t & \text{otherwise} \end{cases} \quad (12)$$

3.1.6 Calibration

Following the procedure established by Shrestha et al. (2024), we calibrated five parameters (Table 3). Other authors used default values of some of the parameters (Hanasaki et al., 2006; Shin et al., 2019; Sadki et al., 2023), which we have used for the simulation with default parameters.

220 Unlike the rule-based schemes, the mHM routine produces a more realistic storage–outflow relationship, as the release is decoupled from the instantaneous storage volume and instead follows the seasonal cycle of demand. While more flexible, the routine requires the estimation of demand signals.

3.1.7 STARFIT

225 The Storage Targets And Release Function Inference Tool (STARFIT) framework represents a data-driven paradigm that infers operational rules by learning the seasonality of reservoir storage and outflows (Turner et al., 2021). Unlike previous rule-based models, STARFIT employs harmonic functions to represent seasonal cycles. A key advantage of this formulation is its scale-independence; while the model parameters are typically fitted using weekly data to reduce noise, the harmonic frequency (ω) can be adjusted for daily simulations ($\omega = 1/365$), ensuring consistency with the hydrological model time step.

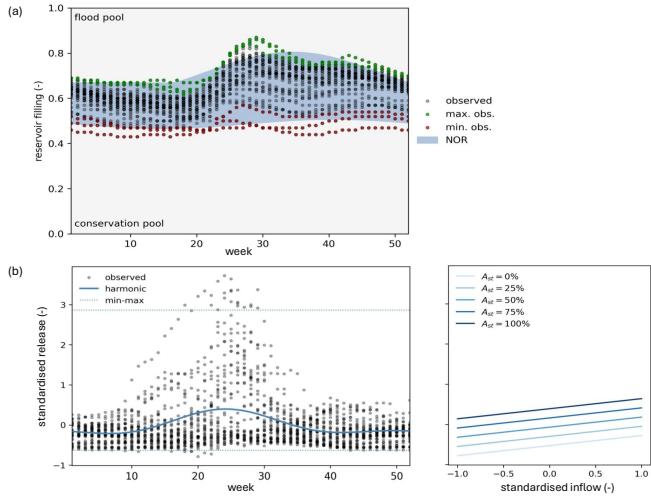


Figure 3. The STARFIT model for the Yellowtail dam (GRanD ID 355): (a) seasonal harmonic curves defining the storage normal operating range (NOR) for each calendar week; (b) the reservoir release model partitioned into the deterministic harmonic weekly release (left) and the linear regression model of release residuals (right).

All variables are standardized to facilitate regionalization and comparison across basins. Storage is converted to reservoir
230 filling, while inflow and outflow are expressed as normalized anomalies relative to the mean annual inflow (\bar{I}):

$$\begin{aligned}\hat{V}_t &= \frac{V_t}{V_{max}} \\ \hat{Q}_t &= \frac{Q_t - \bar{I}}{\bar{I}} \\ \hat{I}_t &= \frac{I_t - \bar{I}}{\bar{I}}\end{aligned}\quad (13)$$

3.1.8 Storage Normal Operating Range (NOR)

STARFIT defines the operational target of a reservoir through a Normal Operating Range (NOR), bounded by upper (NOR_{up}) and lower (NOR_{low}) seasonal harmonics. These bounds represent the typical filling levels for any given time of the year:

$$\begin{aligned}235 \quad NOR_{up,t} &= \min \left(\max \left(A + B \cdot \sin 2\pi\omega t + C \cdot \cos 2\pi\omega t, \hat{V}_{min} \right), \hat{V}_{max} \right) \\ NOR_{low,t} &= \min \left(\max \left(a + b \cdot \sin 2\pi\omega t + c \cdot \cos 2\pi\omega t, \hat{v}_{min} \right), \hat{v}_{max} \right)\end{aligned}\quad (14)$$

where t is the time (day or week) of the year. The NOR has 10 parameters: six defining the harmonic curves and four capping the maximum and minimum filling. To fit these curves, the model uses only the three most extreme (highest and lowest) observed storage for each calendar week (Fig. 3), ensuring the NOR captures the envelope of historical operations.

3.1.9 Release function

240 The release policy is conditioned on the reservoir's position relative to the NOR. To maintain the reservoir within its seasonal bounds, STARTIF applies a three-zone operational policy:

$$Q_t = \begin{cases} \min\left(Q_{min} + (Q_{NOR_{low},t} - Q_{min}) \cdot \frac{\hat{V}_t}{NOR_{low,t}}, I_t\right) & \text{if } \hat{V}_t < NOR_{low,t} \\ \max(\min(Q_{NOR,t}, Q_{max}), Q_{min}) & \text{if } NOR_{low,t} \leq \hat{V}_t \leq NOR_{up,t} \\ Q_{NOR_{up},t} + (Q_{max} - Q_{NOR,t}) \cdot \frac{\hat{V}_t - NOR_{up,t}}{1 - NOR_{up,t}} & \text{if } \hat{V}_t > NOR_{up,t} \end{cases} \quad (15)$$

We modified the definitions of the outflow for both below and above the NOR, as the original definitions produced in some cases noisy releases. In both cases, we have introduced a linear interpolation between a fixed minimum or maximum outflow (Q_{min} or Q_{max}) and the outflow associated with the respective NOR boundary (Q_{NOR_{low},t} or Q_{NOR_{up},t}) depending on the proximity of the current storage to the NOR limits. 245 Within the NOR, the standardized outflow (\hat{Q}_t) is modeled as the sum of an harmonic component ($\hat{Q}_{harm,t}$) and a linear component ($\hat{Q}_{lin,t}$) that accounts for the current storage and inflow:

$$\begin{aligned} \hat{Q}_{NOR,t} &= \hat{Q}_{harm,t} + \hat{Q}_{lin,t} \\ \hat{Q}_{harm,t} &= d \cdot \sin 2\pi\omega t + e \cdot \cos 2\pi\omega t + f \cdot \sin 4\pi\omega t + g \cdot \cos 4\pi\omega t \\ \hat{Q}_{lin,t} &= h + i \cdot \frac{\hat{V}_t - NOR_{low,t}}{NOR_{up,t} - NOR_{low,t}} + j \cdot \hat{I}_t \end{aligned} \quad (16)$$

The release harmonic allows for bimodal release patterns (e.g., two release peaks per year). The linear term is discarded if 250 the linearity is weak ($R^2 < 0.2$), reverting to a purely seasonal policy.

The complete release model comprises 7 parameters, fitted using the full historical record of weekly releases (Fig. 3). Additionally, the minimum and maximum releases (Q_{min} and Q_{max}) are estimated as quantiles of the observed daily inflows; we have changed the original values of these quantiles to 1% and 99%, respectively.

3.2 Experimental design

255 To address our research questions, we conducted four distinct simulation experiments designed to evaluate the routines under varying levels of data availability and calibration complexity.

The first run evaluates model performance using default parameters sourced from the literature (see Tables 1, 2, and 3). This experiment assesses the "out-of-the-box" reliability of the routines, which is particularly relevant for continental or global applications where historical operational data are unavailable.

260 The remaining three experiments involve local parameter optimization targeting different target variables:

- SCEUA-Q: Univariate calibration of reservoir outflow. This mirrors the standard calibration procedure for operational systems like EFAS/GloFAS, where model parameters are tuned to match observed streamflow.

– SCEUA-V: Univariate calibration of reservoir storage. This experiment evaluates the potential for training routines using satellite-derived storage products, which is vital for ungauged basins.

265 – SCEUA-QV: Bivariate calibration of both storage and outflow. This multi-objective approach analyzes the trade-offs in model fidelity and identifies the information loss associated with single-variable calibration.

Parameter optimization was performed using the Shuffled Complex Evolution (SCEUA) algorithm (Duan et al., 1993, 1994), implemented via the *Spotpy* Python library (Houska et al., 2015). For each reservoir, the algorithm was executed for up to 1,000 iterations using four complexes. Following the recommendations of Shen et al. (2022) for large-sample hydrology, we used the
270 full length of the available daily time series for calibration to capture the widest possible range of hydrological conditions.

Performance was quantified using the modified Kling-Gupta Efficiency (KGE') (Kling et al., 2012), which decomposes performance into correlation, bias, and variability components. For the bivariate experiment, we integrated the individual KGE' scores into a single multi-objective metric:

$$KGE' = 1 - \sqrt{(r - 1)^2 + \left(\frac{\mu_{\text{sim}}}{\mu_{\text{obs}}} - 1\right)^2 + \left(\frac{CV_{\text{sim}}}{CV_{\text{obs}}} - 1\right)^2} \quad (17)$$

275

$$KGE'_{\text{bivariate}} = 1 - \sqrt{(1 - KGE'_Q)^2 + (1 - KGE'_V)^2} \quad (18)$$

where r is the Pearson correlation coefficient, μ is the mean, and CV is the coefficient of variation.

While the LISFLOOD, CaMa-Flood, and mHM routines are optimized using the iterative SCEUA algorithm, STARFIT employs a direct statistical fitting procedure. This approach does not require a heuristic search. Instead, the storage model
280 parameters are estimated in a two-step process: an initial ordinary least squares (OLS) regression provides a first guess for the three harmonic parameters, followed by a non-linear optimization using the L-BFGS-B algorithm to finalize all five parameters, including the upper and lower capping values ($\hat{V}_{\max}, \hat{V}_{\min}$). In contrast, the parameters for the release model are derived exclusively through OLS regression.

Because the original STARFIT formulation inherently requires observations of both storage (to define the NOR) and outflow
285 (to define the release policy), it is only evaluated within the SCEUA-QV bivariate experiment. This ensures a consistent comparison between the statistically-derived STARFIT policies and the calibrated rule-based routines.

4 Results

4.1 Which target variable is most relevant for reservoir calibration)

Figure 4 illustrates the performance of the three rule-based reservoir routines (LISFLOOD, CaMa-Flood, and mHM) across the
290 four experimental configurations. The results highlight a critical dependency between the calibration objective and the model's ability to represent the dual states of discharge and storage.

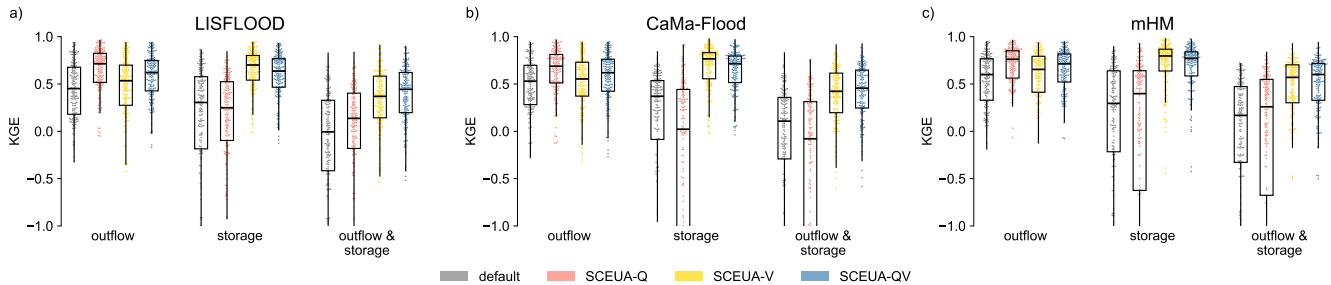


Figure 4. Comparison of the model performance depending on the variable(s) targeted in the calibration. The panels corresponds to the three reservoirs routines that were calibrated: a) LISFLOOD, b) CaMa-Flood and c) mHM. Within each panel, three groups of box plots show the performance in terms of outflow, storage, and both variables together. Four parameter sets are compared —default values, calibration of outflow (SCEUA-Q), storage (SCEUA-V) and both variables (SCEUA-QV)— indicated by colors. The dots represent individual reservoirs, indicating the distribution of the overall performance.

The default parameterization (gray) provides a baseline that is reasonably effective for reservoir outflow, yielding median KGE' values near 0.5. Among the three routines, mHM performs slightly better. However, default parameters fail to capture storage dynamics effectively; median KGE' scores for storage drop to approximately 0.3, with a large interquartile range (IQR) indicating high variability in performance across the 164 reservoirs.

As expected, each routine achieves its peak performance for a specific variable when that variable is the sole target of the calibration. However, there is a clear trade-off when we look at how the models perform on variables they weren't trained on. Calibrating outflow (SCEUA-Q) severely degrades storage performance; in LISFLOOD and CaMa-Flood, this configuration yields the lowest median KGE' for storage among all experiments; in mHM, while the median is slightly higher than the default parameters, the dispersion between reservoirs is remarkably large. Conversely, calibrating to storage does not result in a proportional degradation of outflow performance. In all three routines, the SCEUA-V run maintains outflow KGE' values that are superior to the default parameterization.

The bivariate calibration (SCEUA-QV) consistently achieves the highest overall (combined) performance across all routines. Notably, the performance of the SCEUA-QV run is closely mirrored by the storage-only calibration (SCEUA-V). These findings suggest that reservoir storage is a more informative variable than outflow for identifying model parameters.

While operational hydrology has traditionally relied on streamflow (outflow) for calibration, our results indicate that storage time series provide a more robust constraint on the internal state of the reservoir. This has significant implications for global hydrology, as it supports the feasibility of using satellite-derived storage products to calibrate reservoir routines in ungauged or data-poor regions.

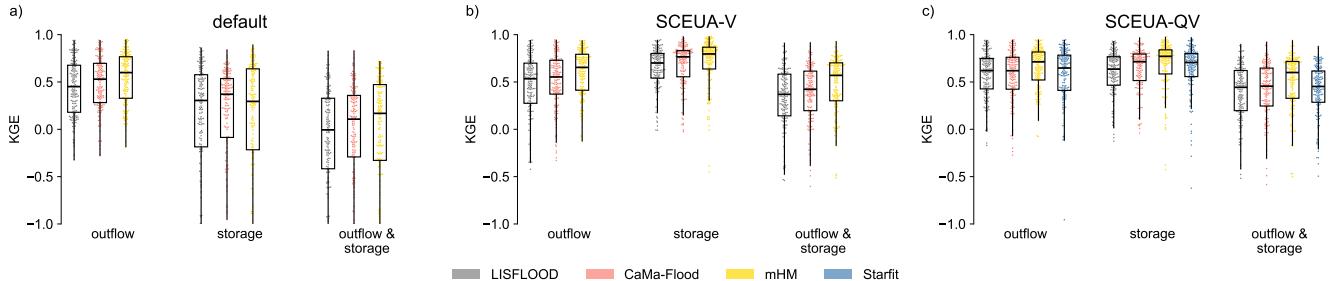


Figure 5. Comparison of the model performance depending on the reservoir routine. The panels corresponds to three of the experiments: a) default values, b) calibration of storage (SCEUA-V), c) bivariate calibration of outflow and storage (SCEUA-QV). Within each panel, three groups of box plots show the performance in terms of outflow, storage, and both variables together. Four reservoir routines are compared—LISFLOOD, CaMa-Flood, mHM and STARFIT—indicated by colors. The dots represent individual reservoirs, indicating the distribution of overall performance.

310 4.2 Which is the best-performing reservoir routine?

Figure 5 compares the performance of the four routines across the different experimental configurations. Note that the outflow-only calibration (SCEUA-Q) is excluded here for simplicity, as the previous section demonstrated that it is a suboptimal strategy for capturing overall reservoir behavior.

Among the rule-based schemes, the mHM routine consistently achieves the highest performance across nearly all scenarios. 315 Its primary strength lies in its ability to represent seasonal operations through demand-driven logic. The only exception is the uncalibrated (default) simulation of storage, where CaMa-Flood shows a slight advantage. Note that the CaMa-Flood routine was specifically designed for global applications without site-specific tuning (Hanazaki et al., 2022).

The results also highlight the benefits of model complexity: the CaMa-Flood routine, which evolved from the original LISFLOOD scheme by adding quadratic storage logic and inflow-dependency, systematically outperforms LISFLOOD in 320 every experiment. This suggests that even small increases in the physical complexity of the storage-discharge relationship can significantly improve model fidelity.

Finally, the bivariate calibration results provide the first direct comparison with the STARFIT model. Despite its high parameterization (up to 19 parameters) and seasonal flexibility, STARFIT does not outperform mHM. Its combined performance is generally comparable to that of CaMa-Flood. This suggests that while seasonal harmonics are valuable, the structured rules 325 in mHM and CaMa-Flood may provide a more robust framework for continuous daily simulations in large-scale hydrological models.

4.3 How accurate is the reservoir shape approximation?

To represent the elevation–storage–area relationship in our modeling, we adopted a simplified geometric assumption where all reservoirs are treated as an inverted half-pyramid (Liebe et al., 2005). Under this assumption, the reservoir geometry is

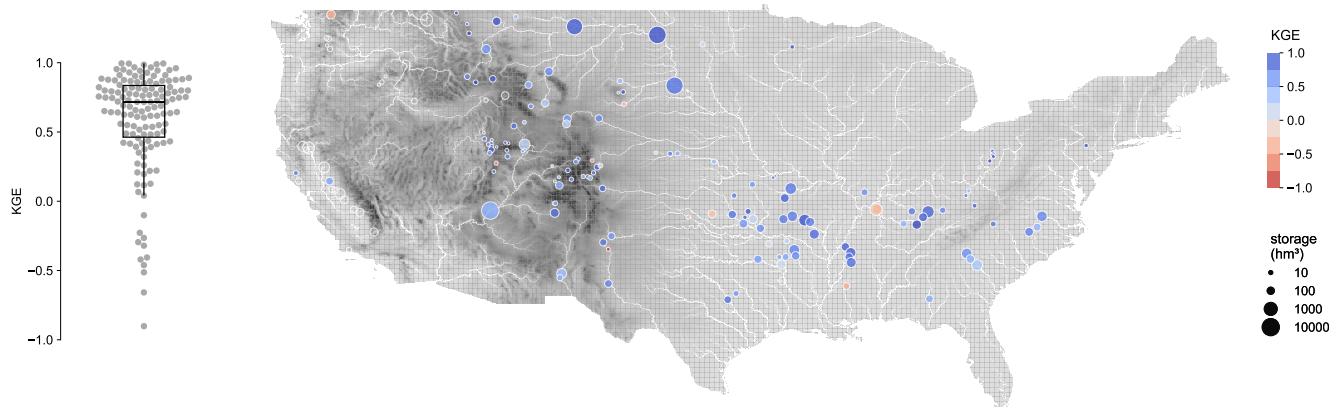


Figure 6. Performance of the reservoir level estimation assuming a half-pyramid reservoir shape. The swarm plot on the left shows the overall distribution, where each point represents a reservoir and the box plot the quartiles. The map on the right shows the geographical distribution of performance; the dot size indicates reservoir storage.

330 defined by the maximum surface area (A_{max}) and the dam height (H_{dam}) (see (4)). This approximation directly influences the calculation of the instantaneous surface area, which is a critical term for estimating direct precipitation and open-water evaporation.

While direct observations of time-varying reservoir area are often unavailable, we validated this geometric assumption by comparing simulated water levels against observed elevation time series. A subset of 138 reservoirs from our original selection 335 provided consistent level data in addition to inflow, outflow, and storage. The performance of this elevation estimation is summarized in Figure 6.

The half-pyramid assumption proved to be a robust approximation for the majority of the study sites. The median KGE' for reservoir level estimation is 0.72, with 75% of the reservoirs achieving a KGE' at least 0.46. However, the accuracy of the level calculation is highly sensitive to the quality of the static attributes, specifically the crest elevation (Z_{max}) and dam 340 height (H_{dam}) (see (5)). As discussed in Section 2.2, errors in global databases such as GRanD or GDW can propagate into these level estimations; therefore, data cleaning and the use of locally validated attributes are essential for accurate water-level modeling.

5 Discussion

5.1 Balancing model complexity and operational feasibility

345 This study compared reservoir routines of varying complexity, from the simple storage-dependent logic of LISFLOOD to the demand-driven and statistically-inferred policies of mHM and STARFIT. Our results indicate that model performance generally scales with the number of environmental covariates included in the routine (e.g., inflow or demand).

While mHM emerged as the best-performing routine (Section 4.2), its dependence on a demand time series poses a significant operational challenge. In data-rich environments, demand can be inferred via machine learning (Shrestha et al., 2024), but in global systems like EFAS or GloFAS, such data are often unavailable.

A similar constraint applies to STARFIT, although its performance was not markedly superior to the more parsimonious CaMa-Flood routine. Despite this, STARFIT offers a unique advantage for global regionalization. For instance, Steyaert et al. (2025) recently demonstrated a global application of the STARFIT routine within the PCR-GLOBWB2 hydrological model by leveraging satellite-derived storage from the GloLAKES dataset (Hou et al., 2024) to define the storage NOR.

We have selected the CaMa-Flood routine for the implementation in the new version of the LISFLOOD hydrological model that will be the core of the upcoming EFASv6 and GloFASv5 systems. By evolving the current LISFLOOD logic to include inflow-dependency and quadratic storage-outflow relationships, we achieve a measurable gain in performances without increasing the data requirements for calibration. Furthermore, CaMa-Flood's robust default performance makes it an ideal candidate for global applications where historical records are absent.

360 5.2 Current limitations and the forecast challenge

A notable limitation across all benchmarked routines is the absence of an explicit use of the reservoir purpose (e.g., irrigation, hydropower, flood control...). While we hypothesize that local calibration allows these routines to learn the specific operations of different reservoir types, the lack of a purpose-aware routine remains a weakness for global modeling.

Furthermore, these routines lack a forward-looking component. In practice, reservoir managers operate based on meteorological and hydrological forecasts across multiple timescales. While replacing current inflow with forecasted inflow as a covariate could address this, it introduces a significant computational bottleneck: the hydrological model would need to simulate the entire upstream catchment for the full forecast horizon before the reservoir release can be determined at any given node.

5.3 The potential of satellite products

370 The scarcity of in situ records remains the primary hurdle for large-scale reservoir modeling. Satellite-derived products offering reservoir area, elevation, and storage anomalies provide an opportunity to bridge this gap (Zhao and Gao, 2018; Schwatke et al., 2020; Donchyts et al., 2022; Khandelwal et al., 2022; Hou et al., 2024). Two of the outcomes of this study specifically support the transition toward using satellite products. First, we found that reservoir storage is a more informative calibration target than reservoir outflow, as storage-based calibration yields models that perform well across both state variables (Section 4.1). This 375 represents a "lucky coincidence" for the community, as many available satellite products provide direct estimates of reservoir storage or storage variations rather than discharge. Second, our validation of the inverted half-pyramid shape (Section 4.3) suggests that even simplified geometric assumptions are sufficient to translate satellite-derived levels into volumes with reasonable accuracy.

The flexibility of STARFIT's harmonic formulation is particularly well-suited for satellite data, as emphasized by Steyaert et al. (2025). Because the frequency term in the harmonic curves allows the model to be fitted at coarser temporal resolutions

(e.g., monthly satellite passes) and then run at a daily time step, it overcomes the temporal sampling limitations of many satellite products.

Furthermore, we have conducted a preliminary analysis of the potential use of the Global Water Watch dataset (Donchyts et al., 2022). Our findings indicate that while satellite-derived storage follows observed historical trends, a persistent bias often exists between in situ and remotely-sensed values. The primary bottleneck for the quality of these products is the availability of accurate elevation-storage-area curves. While several efforts have attempted to provide these curves globally (Yigzaw et al., 2018; Busker et al., 2019; Khazaei et al., 2022), they frequently show shortcomings when validated against in situ data. This underscores why curated datasets of reservoir operations are so vital: they provide the ground-truth necessary to test and refine these global remotely-sensed approaches.

390 5.4 Toward data-driven reservoir modeling

The recent revolution in deep learning has demonstrated that data-driven models can outperform traditional physical models in streamflow prediction (Nearing et al., 2024). To support this transition in reservoir modeling, we have formatted our ResOpsUS+CARS dataset—along with new national datasets for Brazil (Rodríguez et al., 2025), Mexico, and Spain—in the CARAVAN standard.

395 Reservoir operations are the product of complex human-natural interactions that process-based routines struggle to capture entirely. While we acknowledge that deep learning offers a path toward superior performance, these models are notoriously data-hungry. This highlights the ongoing necessity of curated, multi-national observational datasets—not just for training, but for the rigorous benchmarking of the hybrid physical-statistical models.

6 Conclusions

400 This study provided a comprehensive benchmarking of four reservoir routines (LISFLOOD, CaMa-Flood, mHM, and STARFIT) evaluated across 164 reservoirs in the United States using the ResOpsUS and GRanD datasets. By comparing different calibration strategies and model structures, we draw the following conclusions.

Reservoir storage is the most informative variable for calibration. Our results demonstrate a significant asymmetry in model identifiability. While calibrating routines to match observed outflow (the standard practice in many hydrological models) leads 405 to a poor representation of internal reservoir states, calibrating to storage anomalies yields robust performance for both storage and discharge. This finding supports the shift toward using satellite-derived storage products for model parameterization, particularly in ungauged basins.

mHM offers superior performance, but CaMa-Flood provides the best operational balance. The demand-driven logic of the mHM routine consistently outperformed the other schemes. However, its reliance on site-specific demand data limits its 410 current applicability in global systems like EFAS or GloFAS. The CaMa-Flood routine, which we have now integrated into the LISFLOOD model, represents an optimal compromise: it provides significant improvements over the original linear storage logic without requiring additional input variables or data-intensive calibration.

The half-pyramid geometric assumption is a robust approximation. Validating reservoir level estimations against in situ observations confirmed that a simplified inverted half-pyramid shape is sufficient for large-scale modeling. This validates the
415 use of maximum area and dam height as sufficient proxies to link storage, area, and elevation, which is critical for estimating evaporative losses and integrating altimetry data.

STARFIT's flexibility is a key asset for remote sensing integration. Although STARFIT did not outperform the rule-based mHM or CaMa-Flood routines in daily simulations, its harmonic formulation is uniquely capable of bridging temporal gaps in observations. Its ability to be fitted at monthly scales and run at daily resolutions makes it a powerful tool for leveraging
420 current and future satellite altimetry missions.

Looking forward, the integration of reservoir modeling into global hydrology must move beyond the "data-poor" paradigm. The creation of standardized, multi-national datasets like the ResOpsUS+CARS and our new national datasets for Brazil, Mexico, and Spain is a step toward this goal. By providing these data in the CARAVAN format, we aim to facilitate the development of hybrid models that combine the physical consistency of process-based routines with the predictive power of
425 deep learning. Such advancements will be essential for improving the accuracy of operational flood and drought forecasting systems in an increasingly managed global water cycle.

Code and data availability. All the code used to generate the reservoir dataset, implement the four reservoir schemes, and perform the model calibrations is available in the following GitHub repository: <https://github.com/casadoj/reservoirs-LSHM>. The benchmarking dataset generated for this study can be accessed via Zenodo at <https://doi.org/10.5281/zenodo.15978041>.

430 *Author contributions.* TEXT

Competing interests. TEXT

Acknowledgements. TEXT

References

- Abeshu, G. W., Tian, F., Wild, T., Zhao, M., Turner, S., Chowdhury, A. F., Vernon, C. R., Hu, H., Zhuang, Y., Hejazi, M., and Li, H. Y.:
435 Enhancing the representation of water management in global hydrological models, *Geoscientific Model Development*, 16, 5449–5472,
<https://doi.org/10.5194/gmd-16-5449-2023>, 2023.
- Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *Journal of Hydrology*, 249, 11–29, [https://doi.org/10.1016/S0022-1694\(01\)00421-8](https://doi.org/10.1016/S0022-1694(01)00421-8), 2001.
- Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R. W., Heinke, J., Bloh, W. V., and Gerten, D.: Impact of reservoirs on river
440 discharge and irrigation water supply during the 20th century, *Water Resources Research*, 47, <https://doi.org/10.1029/2009WR008929>, 2011.
- Burek, P., van der Knijff, J., and de Roo, A.: LISFLOOD. Distributed Water Balance and Flood Simulation Model, Tech. rep., European Commission - Joint Research Centre, Luxembourg, ISBN 879-92-79-33190-9, ISSN 1831-9424, <https://doi.org/10.2788/24719>, 2013.
- Busker, T., De Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., Pekel, J. F., and Cottam, A.: A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry, *Hydrology and Earth System Sciences*, 23, 669–690,
445 <https://doi.org/10.5194/hess-23-669-2019>, 2019.
- Choulga, M., Moschini, F., Mazzetti, C., Grimaldi, S., Disperati, J., Beck, H., Salamon, P., and Prudhomme, C.: Technical note: Surface fields for global environmental modelling, *Hydrology and Earth System Sciences*, 28, 2991–3036, <https://doi.org/10.5194/hess-28-2991-2024>, 2024.
- 450 Coerver, H. M., Rutten, M. M., and Giesen, N. C. V. D.: Deduction of reservoir operating rules for application in global hydrological models, *Hydrology and Earth System Sciences*, 22, 831–851, <https://doi.org/10.5194/hess-22-831-2018>, 2018.
- Commission, J. R. C. E.: European Flood Awareness System, <https://european-flood.emergency.copernicus.eu/en>, 2025.
- Donchyts, G., Winsemius, H., Baart, F., Dahm, R., Schellekens, J., Gorelick, N., Iceland, C., and Schmeier, S.: High-resolution surface water dynamics in Earth's small and medium-sized reservoirs, *Scientific Reports*, 12, <https://doi.org/10.1038/s41598-022-17074-6>, 2022.
- 455 Duan, Q., Gupta, V. K., and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, *Journal of optimization theory and applications*, 76, 501–521, 1993.
- Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, *Journal of Hydrology*, 158, 265–284, 1994.
- Hanasaki, N., Kanae, S., and Oki, T.: A reservoir operation scheme for global river routing models, *Journal of Hydrology*, 327, 22–41,
460 <https://doi.org/10.1016/j.jhydrol.2005.11.011>, 2006.
- Hanazaki, R., Yamazaki, D., and Yoshimura, K.: Development of a Reservoir Flood Control Scheme for Global Flood Models, *Journal of Advances in Modeling Earth Systems*, 14, <https://doi.org/10.1029/2021MS002944>, 2022.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellán, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049,
465 <https://doi.org/10.1002/qj.3803>, 2020.

- Hou, J., Dijk, A. I. V., Renzullo, L. J., and Larraondo, P. R.: GloLakes: Water storage dynamics for 27000 lakes globally from 1984 to present
470 derived from satellite altimetry and optical imaging, *Earth System Science Data*, 16, 201–218, <https://doi.org/10.5194/essd-16-201-2024>, 2024.
- Houska, T., Kraft, P., Chamorro-Chavez, A., and Breuer, L.: SPOTting model parameters using a ready-made python package, *PLoS ONE*, 10, 1–22, <https://doi.org/10.1371/journal.pone.0145180>, 2015.
- Joint Research Centre - European Commission: Global Flood Awareness System, <https://global-flood.emergency.copernicus.eu/>, 2025.
- 475 Khandelwal, A., Karpatne, A., Ravirathinam, P., Ghosh, R., Wei, Z., Dugan, H. A., Hanson, P. C., and Kumar, V.: ReaLSAT, a global dataset of reservoir and lake surface area variations, *Scientific Data*, 9, <https://doi.org/10.1038/s41597-022-01449-5>, 2022.
- Khazaei, B., Read, L. K., Casali, M., Sampson, K. M., and Yates, D. N.: GLOBathy, the global lakes bathymetry dataset, *Scientific Data*, 9, <https://doi.org/10.1038/s41597-022-01132-9>, 2022.
- Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, *Journal 480 of Hydrology*, 424–425, 264–277, <https://doi.org/10.1016/j.jhydrol.2012.01.011>, 2012.
- Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and Matias, Y.: Caravan - A global community dataset for large-sample hydrology, *Scientific Data*, 10, <https://doi.org/10.1038/s41597-023-01975-w>, 2023.
- Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales 485 and locations, *Water Resources Research*, 49, 360–379, <https://doi.org/10.1029/2012WR012195>, 2013.
- Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, *Frontiers in Ecology and the Environment*, 9, 494–502, <https://doi.org/10.1890/100125>, 2011.
- Lehner, B., Beames, P., Mulligan, M., Zarfl, C., De Felice, L., van Soesbergen, A., Thieme, M., Garcia de Leaniz, C., Anand, M., Belletti, B., 490 Brauman, K. A., Januchowski-Hartley, S. R., Lyon, K., Mandle, L., Mazany-Wright, N., Messager, M. L., Pavelsky, T., Pekel, J.-F., Wang, J., Wen, Q., Wishart, M., Xing, T., Yang, X., and Higgins, J.: The Global Dam Watch database of river barrier and reservoir information for large-scale applications, *Scientific Data*, 11, 1069, <https://doi.org/10.1038/s41597-024-03752-9>, 2024.
- Liebe, J., van de Giesen, N., and Andreini, M.: Estimation of small reservoir storage capacities in a semi-arid environment, *Physics and Chemistry of the Earth*, 30, 448–454, <https://doi.org/10.1016/j.pce.2005.06.011>, 2005.
- 495 Mekonnen, M. M. and Hoekstra, A. Y.: The blue water footprint of electricity from hydropower, *Hydrology and Earth System Sciences*, 16, 179–187, <https://doi.org/10.5194/hess-16-179-2012>, 2012.
- Moreno-Rodenas, A., Mantilla-Jones, J. D., and Valero, D.: Age, climate and economic disparities drive the current state of global dam safety, *Nature Water*, <https://doi.org/10.1038/s44221-025-00402-1>, 2025.
- Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T. Y., Weitzner, D., and Matias, Y.: Global prediction of extreme floods in 500 ungauged watersheds, *Nature*, 627, 559–563, <https://doi.org/10.1038/s41586-024-07145-1>, 2024.
- Pekel, J. F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, *Nature*, 540, 418–422, <https://doi.org/10.1038/nature20584>, the article presents the results of a comprehensive analysis of surface water occurrence in the complete LANDSAT archive (1984–2015).
Main findings
* 162,000 km² of permanent water bodies have either turned into seasonal (72,000 km²) or completely vanished (90,000 km²). /0

- Rodríguez, J. C., Disperati, J., and Salamon, P.: ResOpsBR+CARS: Reservoir Operations Brazil and CAtchment and Reservoir Static attributes, <https://doi.org/10.5281/zenodo.16096623>, 2025.
- Roo, A. P. D., Wesseling, C. G., and Deursen, W. P. V.: Physically based river basin modelling within a GIS: The LISFLOOD model, *Hydrological Processes*, 14, 1981–1992, [https://doi.org/10.1002/1099-1085\(20000815/30\)14:11/12<1981::aid-hyp49>3.0.co;2-f](https://doi.org/10.1002/1099-1085(20000815/30)14:11/12<1981::aid-hyp49>3.0.co;2-f), 2000.
- 510 Sadki, M., Munier, S., Boone, A., and Ricci, S.: Implementation and sensitivity analysis of the Dam-Reservoir OPeration model (DROP v1.0) over Spain, *Geoscientific Model Development*, 16, 427–448, <https://doi.org/10.5194/gmd-16-427-2023>, 2023.
- Salwey, S., Coxon, G., Pianosi, F., Lane, R., Hutton, C., Bliss Singer, M., McMillan, H., and Freer, J.: Developing water supply reservoir operating rules for large-scale hydrological modelling, *Hydrology and Earth System Sciences*, 28, 4203–4218, <https://doi.org/10.5194/hess-28-4203-2024>, 2024.
- 515 Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resources Research*, 46, 1–25, <https://doi.org/10.1029/2008WR007327>, 2010.
- Scherer, L. and Pfister, S.: Global water footprint assessment of hydropower, *Renewable Energy*, 99, 711–720, <https://doi.org/10.1016/j.renene.2016.07.021>, 2016.
- Schwatke, C., Dettmering, D., Bosch, W., and Seitz, F.: DAHITI - An innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry, *Hydrology and Earth System Sciences*, 19, 4345–4364, <https://doi.org/10.5194/hess-19-4345-2015>, 2015.
- 520 Schwatke, C., Dettmering, D., and Seitz, F.: Volume variations of small inland water bodies from a combination of satellite altimetry and optical imagery, *Remote Sensing*, 12, <https://doi.org/10.3390/rs12101606>, 2020.
- Shen, H., Tolson, B. A., and Mai, J.: Time to Update the Split-Sample Approach in Hydrological Model Calibration, *Water Resources Research*, 58, <https://doi.org/10.1029/2021WR031523>, 2022.
- 525 Shin, S., Pokhrel, Y., and Miguez-Macho, G.: High-Resolution Modeling of Reservoir Release and Storage Dynamics at the Continental Scale, *Water Resources Research*, 55, 787–810, <https://doi.org/10.1029/2018WR023025>, 2019.
- Shrestha, P. K., Samaniego, L., Rakovec, O., Kumar, R., Mi, C., Rinke, K., and Thober, S.: Toward Improved Simulations of Disruptive Reservoirs in Global Hydrological Modeling, *Water Resources Research*, 60, <https://doi.org/10.1029/2023WR035433>, 2024.
- 530 Steyaert, J. C. and Condon, L. E.: Synthesis of historical reservoir operations from 1980 to 2020 for the evaluation of reservoir representation in large-scale hydrologic models, *Hydrology and Earth System Sciences*, 28, 1071–1088, <https://doi.org/10.5194/hess-28-1071-2024>, 2024.
- Steyaert, J. C., Condon, L. E., W.D. Turner, S., and Voisin, N.: ResOpsUS, a dataset of historical reservoir operations in the contiguous United States, *Scientific Data*, 9, <https://doi.org/10.1038/s41597-022-01134-7>, 2022.
- 535 Steyaert, J. C., Sutanudjaja, E. H., Bierkens, M., and Wanders, N.: Data derived reservoir operations simulated in a global hydrologic model, *Hydrology and Earth System Sciences*, 29, 6499–6527, <https://doi.org/10.5194/hess-29-6499-2025>, 2025.
- Thober, S., Cuntz, M., Kelbling, M., Kumar, R., Mai, J., and Samaniego, L.: The multiscale routing model mRM v1.0: Simple river routing at resolutions from 1 to 50 km, *Geoscientific Model Development*, 12, 2501–2521, <https://doi.org/10.5194/GMD-12-2501-2019>, 2019.
- Turner, S. W., Steyaert, J. C., Condon, L., and Voisin, N.: Water storage and release policies for all large reservoirs of conterminous United 540 States, *Journal of Hydrology*, 603, <https://doi.org/10.1016/j.jhydrol.2021.126843>, 2021.
- US Army Corps of Engineers: National Inventory of Dams, <https://nid.sec.usace.army.mil/#/>, 2025.
- US Bureau of Reclamation: Projects & Facilities, <https://www.usbr.gov/projects/facilities.php?type=Dam>, 2025.
- US Geological Survey: Water Data for the Nation, <https://waterdata.usgs.gov/>, 2025.

van der Knijff, J. M., Younis, J., and de Roo, A. P.: LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood
545 simulation, International Journal of Geographical Information Science, 24, 189–212, <https://doi.org/10.1080/13658810802549154>, 2010.

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G., and Wheater, H.: Representation and improved parameterization of reservoir operation in hydrological and land-surface models, Hydrology and Earth System Sciences, 23, 3735–3764,
550 <https://doi.org/10.5194/hess-23-3735-2019>, 2019.

Yigzaw, W., Li, H. Y., Demissie, Y., Hejazi, M. I., Leung, L. R., Voisin, N., and Payn, R.: A New Global Storage-Area-Depth
555 Data Set for Modeling Reservoirs in Land Surface and Earth System Models, Water Resources Research, 54, 10,372–10,386,
<https://doi.org/10.1029/2017WR022040>, 2018.

Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F., and Beck, H.: The impact of lake and reservoir parameterization on global
streamflow simulation, Journal of Hydrology, 548, 552–568, <https://doi.org/10.1016/j.jhydrol.2017.03.022>, 2017.

Zhao, G. and Gao, H.: Automatic Correction of Contaminated Images for Assessment of Reservoir Surface Area Dynamics, Geophysical
555 Research Letters, pp. 6092–6099, <https://doi.org/10.1038/nature20584>, 2018.