https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

Streamflow and satellite evapotranspiration are asymmetric calibration targets in water-limited catchments

Roland Yonaba[1,*], Tazen Fowé[1], Soumahila Sankandé[1], Nazira Tinto[1], Omar Goudiaby[2], Elias Nkiaka[3], Arsène Kiéma[4], Ansoumana Bodian[2], Harouna Karambiri[1]

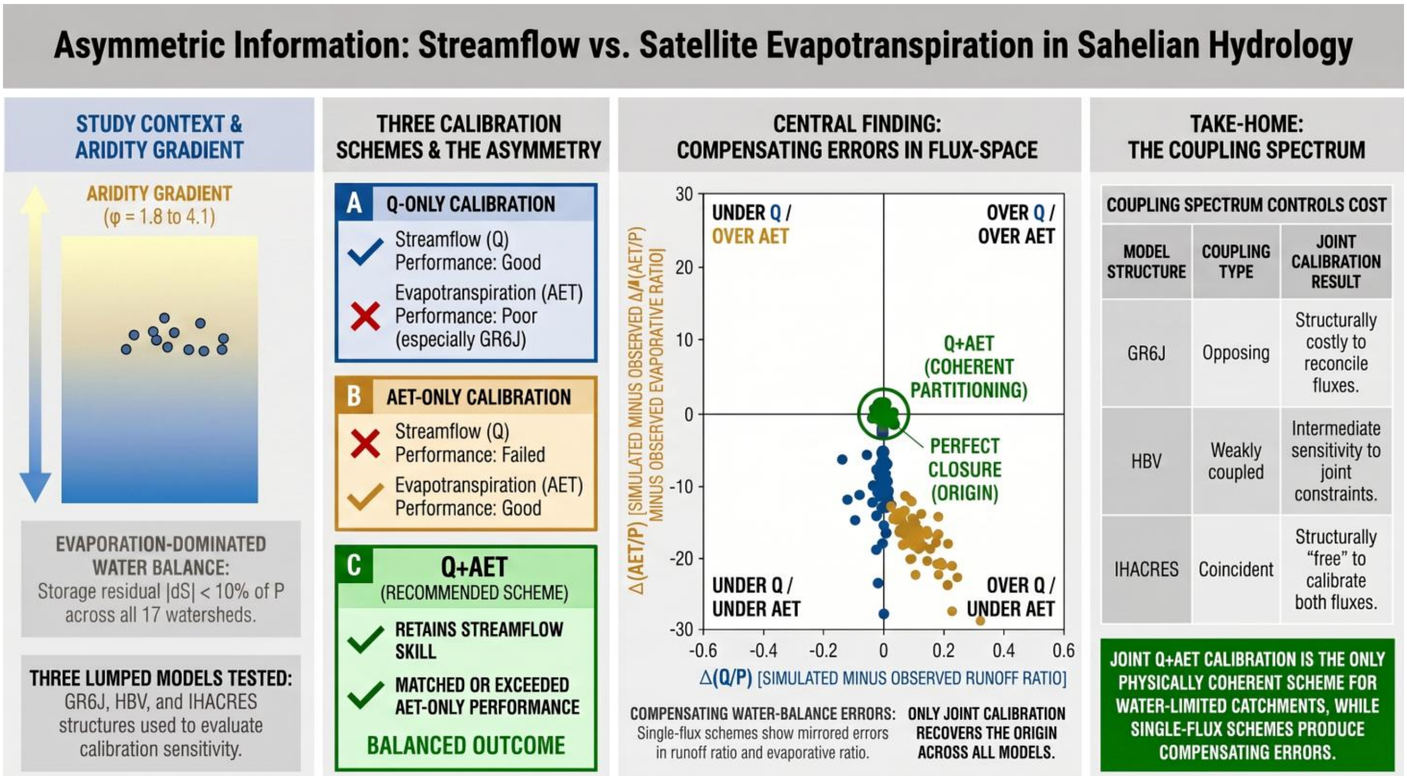
[1]Laboratoire Eaux, Hydro-Systèmes et Agriculture (LEHSA), Institut International d'Ingénierie de l'Eau et de l'Environnement (2iE), 01 BP 594, Ouagadougou, Burkina Faso
[2]Laboratoire Dynamique des Territoires et Développement (Leïdi), Université Gaston Berger (UGB), BP 2, Saint-Louis, Sénégal.
[3]Climate, Catchments, Coasts and Community Research Group, School of Natural Sciences, University of Lincoln, Lincoln, UK
[4]Département Climat-Eau-Sol, Centre Climatique Régional pour l'Afrique de l'Ouest et du Sahel (CCR-AOS), BP 11011, Niamey, Niger

Correspondence to: Roland Yonaba (ousmane.yonaba@2ie-edu.org)

Abstract. Reliable hydrological modelling in data-scarce regions is constrained by the scarcity of streamflow observations, which has driven increasing adoption of satellite-derived actual evapotranspiration (AET) as a supplementary calibration target. Whether AET provides parameter-constraining information complementary to streamflow, or merely trades discharge skill for evapotranspiration reproduction, remains contested. This study compared three calibration schemes: streamflow-only (Q-only), AET-only, and joint streamflow-AET calibration (Q+AET), applied to three conceptual rainfall-runoff models (GR6J, HBV, IHACRES) across 17 water-limited watersheds spanning the hydroclimatic gradient of the West African Sahel during 1991-2020. Parameter identifiability was assessed via variance reduction and distributional shift from prior to behavioural samples. Long-term water-balance estimates were evaluated against the Budyko-Fu framework, with the reference evaporative index anchored both on GLEAM AET and independently on water-balance closure. Results show that streamflow provided substantially more parameter-constraining information than AET. AET-only calibration reproduced evapotranspiration dynamics but left the production, soil-moisture and loss-module parameters governing runoff generation unconstrained. Q-only calibration constrained water partitioning sufficiently to recover AET dynamics in HBV and IHACRES, while AET in GR6J remained poorly constrained under streamflow-only conditioning. Joint calibration preserved streamflow skill of Q-only while matching or exceeding AET-only on AET simulation. It emerged as the only scheme to yield a simultaneously coherent long-term runoff ratio and evaporative index across all three model structures, which held under both Budyko reference anchors. The cost of multivariable calibration was governed by model structure: the direction of parameter displacement imposed by each variable (opposing in GR6J, weakly coupled in HBV and coincident in IHACRES) predicted whether joint calibration was structurally costly or essentially free. These results suggest that streamflow should remain the primary calibration target wherever observations are available, however, satellite-derived AET estimates remain valuable as an independent check on long-term water-balance plausibility rather than as a substitute calibration signal.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

Graphical abstract

35 

1 Introduction

Hydrological models are critical tools for quantifying the components of the terrestrial water balance and for supporting the sustainable management of water resources (Mei et al., 2023). The reliability of a simulation, however, does not rest solely on reproducing streamflow at a watershed outlet. It also depends on how accurately the model represents the internal partitioning of water within the watershed: how precipitation is divided between runoff, evapotranspiration and storage change across the range of hydroclimatic conditions under which the model is applied (Guo et al., 2024a; Pool et al., 2024). This distinction is especially consequential in water-limited environments, where evapotranspiration dominates the water balance and where the feedback between moisture availability and evaporative demand governs most of the interannual variability in streamflow (Pool et al., 2024; Hsu et al., 2025; Wu et al., 2026b).

Actual evapotranspiration (AET) is the dominant terrestrial water flux in water-limited watersheds, returning roughly 60-80% of annual precipitation to the atmosphere, therefore exerting a strong control over runoff generation, soil moisture dynamics and groundwater recharge (Jin and Jin, 2020; Adeyeri and Ishola, 2021; Do et al., 2024). Representing AET realistically is therefore a prerequisite for maintaining the physical plausibility of a model, particularly when that model is transferred beyond the conditions under which it has been calibrated (Abate et al., 2023). Conventional calibration, however, relies almost exclusively on gauged discharge at the watershed outlet, a variable which is strongly integrated in space and time and also records the cumulative response of the watershed to multiple forcings (Dangol et al., 2023; Guo et al., 2024a). This makes calibration an ill-posed inverse problem: the integrated streamflow signal cannot uniquely constrain the spatiotemporal structure of the processes that generate it (Dembélé et al., 2020a; Abate et al., 2023). Relying on a single aggregated variable therefore masks the equifinality problem (Beven, 2006), wherein physically inconsistent parameter sets may reproduce the observed hydrograph with statistically similar skill (Rane and Jayaraj, 2023; Taia et al., 2023). The result is that models appear accurate against standard benchmarks while carrying structural errors and parameter uncertainty forward (Rajib et al., 2018; Zhang et al., 2021).

Streamflow-only (Q-only) calibration remains the dominant practice, largely because discharge data are more widely available than other hydrological observations (Poméon et al., 2018). A satisfactory discharge simulation, however, offers no guarantee that the internal water balance is correctly represented (Mostafaie et al., 2018; Hui et al., 2020). Q-only calibration places weak constraints on vertical fluxes and on internal storage, leaving substantial uncertainty in AET, soil moisture, and groundwater (Dembélé et al., 2020a). Parameters governing infiltration, percolation and recharge are free to absorb errors in the forcing, making predictions less robust in the case of hydroclimatic shifts (Jiang et al., 2020; Yonaba et al., 2025b; Zhang et al., 2025). A model can therefore score well on discharge while partitioning the water balance unrealistically and producing implausible AET estimates (Pan et al., 2022; Kumar and Mishra, 2025; Wu et al., 2026b). Equifinality compounds this, since many parameter sets yield similar hydrographs while implying fundamentally different internal dynamics (Shah et al., 2021).

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

In response, a growing number of studies now incorporate satellite-derived AET as a second calibration target, with the aim of simultaneously constraining both horizontal and vertical water fluxes. Two strategies have emerged. The first calibrates against AET alone and is commonly applied in ungauged basins (Huang et al., 2021; Ding and Zhu, 2022; Taia et al., 2023). The second combines discharge and AET in a joint objective function (Q+AET) to impose simultaneous constraints on runoff generation and evaporative fluxes (Herman et al., 2020; Guo et al., 2024a). Several studies report that adding AET reduces equifinality and improves the representation of soil moisture and groundwater storage, increasing overall physical consistency (Poméon et al., 2018; Abate et al., 2023; Abdallah et al., 2025; Kalura et al., 2025). However, the extent to which these gains translate into improved physical realism remains an open question. Some studies find that discharge and internal processes improve together; others report clear trade-offs, with better AET at the cost of streamflow performance (Dangol et al., 2023; Pool et al., 2024). Additional uncertainty is introduced by the satellite AET products themselves, which carry systematic biases in water-limited settings and propagate forcing errors into the simulation (Koltsida and Kallioras, 2022). It therefore remains unclear whether multi-variable calibration genuinely improves the plausibility of a model outputs or instead masks structural weaknesses in the modelling framework (Mei et al., 2023; Pool et al., 2024).
Two deeper questions underlie this ambiguity and have received less attention in the literature. The first concerns the asymmetry of information content between the two fluxes: does streamflow, when used alone as a calibration target, implicitly constrain AET to a degree that makes satellite AET redundant; or does AET provide genuine independent parameter constraint? The second concerns the role of model structure in mediating that relationship: because AET and discharge are computed from shared state variables in conceptual rainfall-runoff models, the degree to which one variable constrains parameters sensitive to the other is governed by the architecture of the coupling between them, not only by the data. These two questions together determine what "adding AET to calibration" means in practice, which cannot be answered by performance metrics alone. Addressing them requires an identifiability strategy, which maps how the prior parameter distribution collapses under each calibration scheme and links that collapse explicitly to the model nodes through which AET and discharge share information. However, such analyses are largely absent from the literature.

These questions are particularly pressing in sparsely monitored, water-limited watersheds, such as those across semi-arid West Africa, where declining hydrometric networks have pushed modelling practice towards the use of satellite- and reanalysis-based AET products (Poméon et al., 2018; Jung et al., 2019; Dembélé et al., 2020a; Adeyeri and Ishola, 2021; Odusanya et al., 2021; Mohamed et al., 2025). However, in these environments, the physical plausibility of calibrated models against long-term ecohydrological constraints has rarely been assessed. Yet, the interaction between a highly seasonal, sometimes ephemeral streamflow regime and the structural assumptions embedded in conceptual models makes the equifinality and identifiability problems especially acute, and the value of satellite AET as an independent constraint especially uncertain.

This study addresses these questions through a systematic comparison of conceptual rainfall-runoff models calibrated under three schemes: streamflow-only (Q-only), satellite-derived AET-only (AET-only) and combined (Q+AET) across 17 water-

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

limited watersheds spanning a north-to-south hydroclimatic gradient from semi-arid to sub-humid conditions. The study objectives are threefold, as follows:

i. To quantify, via prior-to-posterior analysis across the three schemes, how asymmetrically AET and streamflow constrain model parameters; in particular, whether Q-only yields a tolerable AET simulation while AET-only leaves streamflow essentially unconstrained.

ii. To explain these calibration differences in terms of model structure, by showing how the number of AET-exclusive parameters and the sensitivity of the AET flux to the soil storage coupling parameter under water-limited conditions govern the information content of each scheme.

iii. To assess whether Q+AET calibration yields more physically consistent water-balance partitioning across the gradient than single-variable calibration.

2 Material and methods

2.1 Study area

Burkina Faso is a landlocked West African country (9-15° N, 5° W-2° E; **Fig. 1**) covering 274,200 km² within the tropical belt, where the climate is governed by the seasonal migration of the Inter-Tropical Convergence Zone (Dembélé et al., 2020b; Yonaba et al., 2024). A single wet season (June to September in the Sahelian north and May to October in the southern Sudanian zone) is followed by a 6- to 8-month dry season. Mean annual precipitation increases along a pronounced north-south gradient, from below 400 mm in the semi-arid north to about 1,100 mm in the south and southwest (Yonaba et al., 2024). Potential evapotranspiration (PET) exceeds precipitation across most of the country (Yonaba et al., 2023, 2025a), with the aridity index (φ = PET/P) ranging from about 1.8 in the south to 3.7 in the north.

The drainage network is organised into four major river basins: the Niger, Nakanbe (White Volta), Mouhoun (Black Volta), and Comoé (**Fig. 1**). Hydrological behaviour is shaped by shallow crystalline basement aquifers of limited storage capacity (Kafando et al., 2022; Coulidiati et al., 2025) and shifts from intermittent to ephemeral regimes in the north to more perennial conditions in the humid south and west (Belemtougri et al., 2021). Runoff is generated predominantly by Hortonian (infiltration-excess) overland flow over degraded, crusted soils (Mounirou et al., 2020, 2022).

Seventeen gauged watersheds across the four basins were selected because of the availability of continuous daily streamflow records over 1991-2020 (**Fig. 1**). Morphometric characteristics of these watersheds were derived from the HydroSHEDS v1 digital elevation model (Lehner et al., 2008; **Supplement Fig. S1**). The watersheds are generally low relief, with median area, elevation, slope and drainage density of 3,754 km², 321 m, 2.1 %, and 0.3 km^{-1}, respectively (**Table 1**) and areas spanning 1,113 to 75,907 km².

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

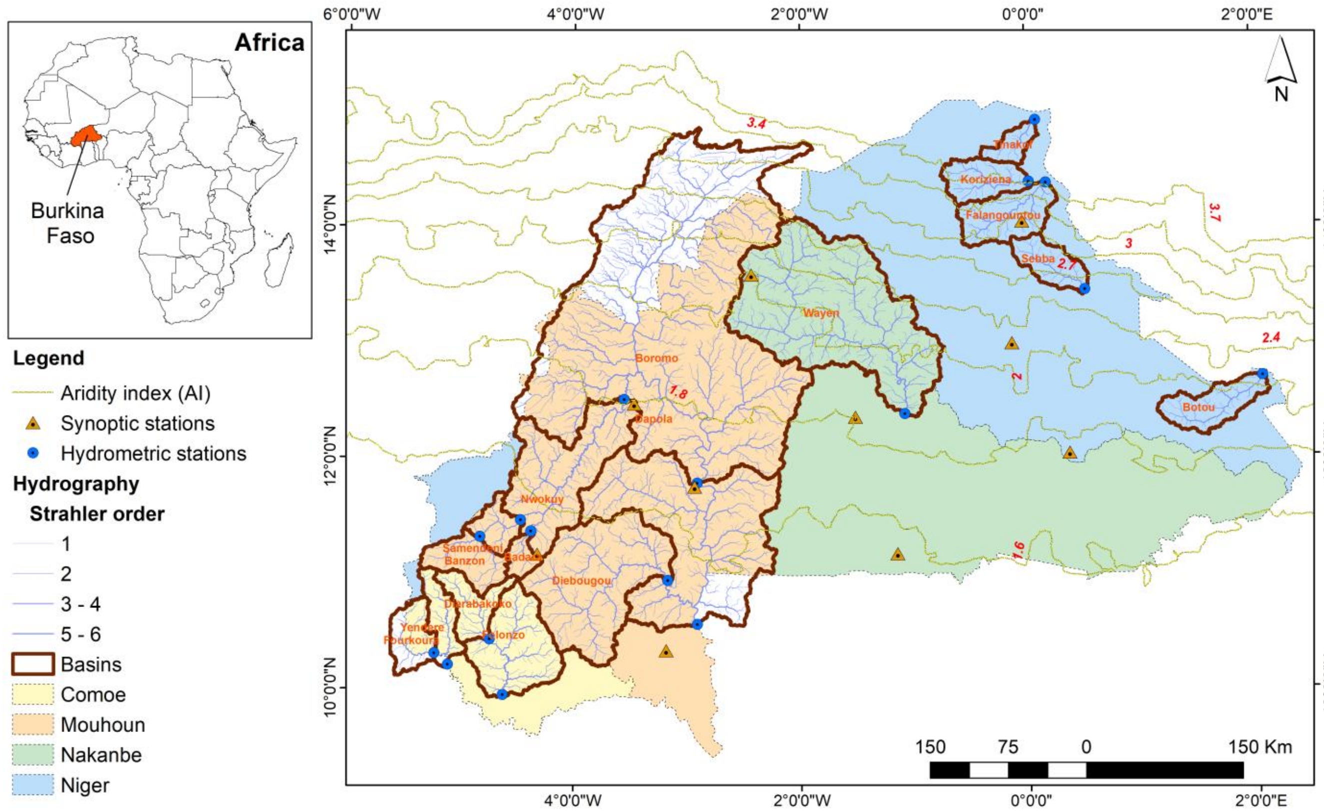


Figure 1: Location of the study area. The map shows the four major river basins of Burkina Faso (Niger, Nakanbe, Mouhoun, Comoe), the 17 selected gauged watersheds, the location of synoptic meteorological stations and hydrometric stations used in this study.

Table 1: Summary statistics of the morphometric characteristics of the 17 study watersheds in Burkina Faso.

Morphometric characteristic	Median	5th quantile	95th quantile
Area [km²]	3754.3	1112.5	75907.0
Perimeter [km]	362.4	208.9	2323.6
Watershed Equivalent Length [km]	122.9	68.8	597.6
Median Elevation [m]	320.8	271.4	424.7
Hypsometric Integral [-]	0.3	0.2	0.5
Median Slope [%]	2.1	0.9	3.0
Elongatio Ratio [-]	0.6	0.5	0.6
Longest Stream Path [km]	63.8	8.2	197.7
Drainage Density [km^{-1}]	0.3	0.1	0.9
Time of Concentration - Kirpich [h]	34.2	8.1	190.8

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

2.2 Data

2.2.1 Precipitation (P)

Daily precipitation for 1989-2020 (including the 1989-1990 warm-up period) was obtained from the synoptic station network of the national meteorological agency (ANAM) in Burkina Faso. Areal precipitation over each watershed was estimated by inverse-distance weighting (IDW, power 2), with station daily totals weighted by the inverse squared distance between each station and the watershed centroid; on days with missing records, weights were renormalised over the available stations. Mean annual precipitation ranges from 516 mm (Falangountou, in the north) to 1050 mm (Badara, in the south).

2.2.2 Potential (PET) and actual evapotranspiration (AET)

Daily potential evapotranspiration (PET) was estimated from synoptic stations from the same network used for precipitation over 1989-2020. Daily PET was computed at each station using the FAO-56 Penman-Monteith equation (Allen et al., 1998; Yonaba et al., 2023, 2025a), driven by measured air temperature, relative humidity, wind speed, and daily minimum and maximum temperature. Areal PET over each watershed was estimated by inverse-distance weighting (IDW, power 2), following the same procedure applied to precipitation.

Daily actual evapotranspiration (AET) was obtained from the Global Land Evaporation Amsterdam Model (GLEAM v4.3a; Miralles et al., 2025). GLEAM v4.3a is a global dataset at 0.1° resolution (spanning 1980 to present), produced by combining satellite observations with reanalysis forcing. GLEAM does not measure evaporation directly; rather, it estimates AET from environmental drivers by scaling the potential evaporation rates of the bare-soil and vegetated fractions by a multiplicative evaporative-stress factor (a function of root-zone and surface soil moisture, vapour pressure deficit, air temperature, atmospheric CO_2, and vegetation state) and adding a separately modelled interception flux (Miralles et al., 2025). GLEAM has been reported as a reliable evapotranspiration source over West Africa (Jung et al., 2019). AET was retrieved for 1991-2020, covering the full extent of the study watersheds. Areal AET over each watershed was computed as a spatial average of all GLEAM pixels within the watershed boundary.

PET and AET were therefore obtained from independent sources, rather than both from GLEAM. In GLEAM, potential and actual evaporation are produced by the same model through a single evaporative-stress formulation, so the AET/PET ratio is partly an internal property of GLEAM rather than an externally observed quantity (Miralles et al., 2025). Using GLEAM data for both the model forcing (PET) and the calibration target (AET) would make model calibration schemes using AET information partly self-referential, since the models would be conditioned toward a stress ratio embedded in their own forcing. The use of station-based PET breaks this dependency, with PET (forcing) and AET (calibration target) derived from independent sources. Any constraint AET imposes on the simulated water balance reflects information external to the forcing data.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

Because PET and AET originate from independent products, the constraint AET ≤ PET was checked. Exceedances (AET > PET) occurred on approximately 2% of the combined record across all watersheds; on these days, PET was capped to the corresponding AET value to preserve physical consistency for the calibration targets.

2.2.3 Streamflow (Q)

Daily discharge records for the 17 gauging stations were obtained from the national hydrological services of Burkina Faso (DGRE) for the 1991-2020 reference period. The discharge observations were converted to depth equivalents (mm d^{-1}) by normalising each daily value by the corresponding watershed area. Record completeness was assessed prior to calibration, and stations were retained if their overall data-gap rate did not exceed 50% (**Supplement Fig. S2**). This threshold was set to preserve the geographic and hydroclimatic extent of the network rather than to maximise per-station record length. In fact, the Sahelian hydrometric network has declined over the study period (Belemtougri et al., 2021; Yonaba et al., 2025b), and a stricter completeness criterion would disproportionately exclude stations rather than reduce gaps at any particular point along the gradient, narrowing the range of conditions the study is designed to sample. We verified that gap rate was not clustered at specific ends of the network: it appears negatively associated with latitude across the 17 retained stations (Spearman's ρ = -0.68, p = 0.003; see **Supplement Fig. S3**), which indicates that record completeness does not coincide with any single climatic zone. Because objective functions and performance metrics are further evaluated only on days with valid observations, missing data reduce the number of evaluation days without biasing the criteria; and because all three calibration schemes are applied to the identical record at a given watershed, gap rate cannot differentially favour one scheme over another at any site.

The mean annual regime is unimodal, with peak flows in August-September and near-zero discharge through the dry season, consistent with the Sahelian context. Some exceptions are the more perennial Mouhoun basin stations (Nwokuy, Dapola), where moderate baseflow persists into the early dry season (see **Supplement Fig. S4**).

2.3 Hydrological models

Three lumped conceptual rainfall-runoff models were used: "*Génie Rural à 6 paramètres Journalier*" (GR6J) (Pushpalatha et al., 2011), Hydrologiska Byråns Vattenbalansavdelning (HBV) (Bergström, 1992; Bergström and Lindström, 2015) and IHACRES-CMD (Croke and Jakeman, 2004, 2005). All operate at a daily time step and are driven by watershed-averaged precipitation (P) and potential evapotranspiration (PET), consistent with the available data. Their parsimonious lumped structures are well suited to the data-scarce conditions of the study region, where limited information on soils, land cover and subsurface properties limits reliable application of physically distributed models and where such models have shown robust performance (in tropical and semi-arid catchments) while limiting the equifinality associated with highly parameterised structures (Kodja et al., 2020; Mbaye et al., 2020; Ndiaye et al., 2024; Goudiaby et al., 2025). The three models share a bucket-type architecture but differ specifically in how the actual evapotranspiration (AET) flux is formulated and how it is coupled to runoff generation processes. Using three structurally distinct models allows the results obtained under a given calibration

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

scheme to be assessed as either model-dependent or reflective of broader constraints on the runoff-evapotranspiration partitioning. The calibrated parameters of all three models are listed in **Table 2**.

GR6J is a 6-parameter model of the GR family (Coron et al., 2017b), structured around three stores: a production (soil moisture) store, a routing store and an exponential store. It also uses two-unit hydrographs. At each step, P and PET are first netted: when PET exceeds P, AET is withdrawn from the production store as an increasing, saturating (hyperbolic-tangent) function of its relative filling, so that AET approaches PET as the store fills and declines toward zero as it empties; when P exceeds PET, PET is fully met and the net rainfall partly refills the store, the remainder becoming effective rainfall. Effective rainfall, augmented by a power-law percolation leak from the production store, is split between the two-unit hydrographs and routed through the routing and exponential stores, the latter improving low-flow recession. Streamflow is the sum of the routed, exponential and direct components. An inter-catchment groundwater exchange term, applied to the routing branches, allows non-conservative gains or losses (Pushpalatha et al., 2011). The production store is the single state from which both AET and the runoff-generating effective rainfall are drawn; the routing and exchange parameters act on streamflow alone.

IHACRES-CMD (hereafter IHACRES) couples a non-linear soil-moisture-accounting module to a linear routing structure of two parallel exponential reservoirs (Croke and Jakeman, 2004, 2005). Catchment wetness is tracked as a moisture deficit (CMD), which increases with evapotranspiration and decreases with rainfall drainage. AET is computed as a moisture-limited fraction of PET, scaled by an efficiency factor (e) and reduced as the deficit exceeds a stress threshold set by the reference deficit (d) and the threshold modifier (f). The drainage that reduces the deficit becomes effective rainfall, which is partitioned between a quick-flow and a slow-flow exponential reservoir with distinct recession time constants; total discharge is their sum. The deficit is the state shared by both fluxes: the reference deficit d governs both effective-rainfall production and the evapotranspiration threshold, whereas the efficiency factor e enters the AET expression only, and the two routing time constants and the slow-flow fraction act on streamflow alone.

The HBV model is implemented in lumped form via the HBV.IANIGLA package (Toum, 2019). A single soil-moisture store is controlled by a maximum capacity (FC), an evapotranspiration threshold (LP), and a non-linear recharge exponent (β). In contrast to GR6J and IHACRES models, AET in HBV follows a threshold relationship with soil moisture: it equals PET once soil moisture exceeds the fraction LP of FC and decreases linearly with soil moisture below that threshold. Recharge leaving the soil store is a non-linear (β-controlled) function of its relative filling and feeds three linear reservoirs in series: an upper store releasing quick flow above a threshold (UZL, recession $K0$), an intermediate store generating interflow with capped percolation ($K1$, $PERC$) and a lower baseflow store ($K2$), followed by a triangular transfer function ($Bmax$); total streamflow is the sum of the three routing components. The soil store is the node shared by the two fluxes, with AET set through FC and LP and recharge through FC and β. The routing parameters act on streamflow alone.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

Table 2: Parameters of the three hydrological models and their prior ranges used for calibration.

Model	Parameter	Description	Unit	Range
GR6J[1]	$X1$	Maximum production store capacity	mm	[0, 3000]
	$X2$	Inter-catchment groundwater exchange coefficient	mm d^{-1}	[-5, 3]
	$X3$	Maximum routing store capacity	mm	[1, 1000]
	$X4$	Unit hydrograph time base	d	[0.01, 2.90]
	$X5$	Inter-catchment exchange threshold	-	[-2, 2]
	$X6$	Exponential store depletion coefficient	mm	[0, 1000]
HBV[2]	FC	Maximum soil moisture storage capacity	mm	[50, 700]
	LP	Soil moisture threshold for full PET extraction (SM/FC)	-	[0.3, 1.0]
	$BETA$	Non-linearity exponent for soil recharge	-	[1.0, 6.0]
	$K0$	Recession coefficient - upper (fast) reservoir	d^{-1}	[0.05, 0.99]
	$K1$	Recession coefficient - intermediate reservoir	d^{-1}	[0.01, 0.80]
	$K2$	Recession coefficient - lower (slow) reservoir	d^{-1}	[0.001, 0.150]
	UZL	Overflow threshold - upper to intermediate reservoir	mm	[0, 100]
	$PERC$	Maximum percolation - intermediate to lower reservoir	mm d^{-1}	[0.0, 6.0]
	$Bmax$	Base of triangular unit hydrograph	mm	[1.0, 3.0]
IHACRES	f	Soil-moisture sensitivity of ET (threshold modifier)	-	[0.01, 3.00]
	e	Evapotranspiration efficiency	-	[0.01, 1.50]
	d	Reference moisture deficit	mm	[50, 550]
	τ_s	Slow-flow recession time constant	d	[1, 200]
	τ_q	Quick-flow recession time constant	d	[0, 10]
	v_s	Fraction of volume routed to slow store	-	[0.01, 0.99]

[1] GR6J parameter ranges were defined based on the airGR R package documentation (Coron et al., 2017a) and were further refined after Perrin et al. (2003) and Pushpalatha et al. (2011). Then, the production- and routing-store capacities (X1, X3) were extended upward and the unit-hydrograph time base (X4) lowered to accommodate the deep soil stores and rapid runoff response of semi-arid Sahelian watershed, following Park et al. (2025).

[2] With HBV, six snow and glacier parameters were fixed to deactivate the snow routine, which is not relevant to the tropical Sahelian climate: SFCF = 1.0; Tr = 0.0 °C; Tt = 10.0 °C; fm = 0.0; fi = 0.0; fic = 0.0 mm °C^{-1} d^{-1}. All precipitation is then treated as rainfall and fed directly into the soil-moisture routine.

2.4 Calibration framework

Each of the three models was calibrated under three schemes and with two complementary approaches, applied identically across all 17 watersheds. A global Latin-hypercube sample (LHS) combined with Generalized Likelihood Uncertainty Estimation (GLUE) was used to characterise parameter identifiability and equifinality and the Shuffled Complex Evolution

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

algorithm (SCE-UA) was used to locate a single best parameter set per watershed for the performance and water-balance analyses. The two approaches share the same objective functions, parameter ranges (**Table 2**) and temporal framework, resulting in identifiability (from GLUE) and optimal performance (from SCE-UA) are addressed on a common basis.

2.4.1 Calibration schemes and temporal framework

Three calibration schemes were defined according to the variable(s) used to condition the parameters: a streamflow-only scheme (Q-only), in which parameters are constrained against gauged discharge alone; a satellite-derived actual evapotranspiration-only scheme (AET-only), constrained against GLEAM AET alone; and a combined scheme (Q+AET), in which both fluxes jointly condition the parameters. The three schemes were applied to each model with identical priors and an identical sampling, so that any difference in the resulting parameter distributions or simulated fluxes is attributable to the conditioning variable rather than to the calibration setup. Model performance was evaluated for both Q and AET under each scheme, including the variable not used for calibration, which provides the basis for assessing the asymmetry of information transfer between the two fluxes.

A single, fixed temporal framework was applied to all watersheds, models and schemes. The 2-year period 1989-1990 was used as a warm-up to remove the influence of initial storage states and were excluded from all objective-functions and evaluation metrics. That 2-year warm-up was judged sufficient to dissipate initial-condition effects given the strong seasonal, single-wet-season regime and the rapid within-year depletion of soil storage characteristic of these watersheds (Ruelland et al., 2012; Zouré et al., 2023). The period 1991-2010 was used for calibration (2/3 of the available data) and 2011-2020 for independent validation (1/3 of the available data). Objective functions and performance metrics were computed only over days with valid observations within each window, so that data gaps in the streamflow records reduce the number of evaluation days but do not bias the criteria.

2.4.2 Objective functions

The objective functions used in this study are based on the modified Kling-Gupta efficiency (KGE; Kling et al., 2012), defined as in **Eq. (1)** :

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

where r is the linear Pearson correlation coefficient, α the ratio of the simulated to observed coefficient of variation, and β the ratio of simulated to observed mean. KGE equals 1 for a perfect simulation, while a simulation reproducing only the observed mean yields KGE ≈ -0.41 (Knoben et al., 2019; Yonaba et al., 2025b).

The Q-only calibration scheme objective function OF_Q combines KGE on raw and log-transformed streamflow, giving balanced weight to high and low flows, the latter to prevent the small number of high-flow events from dominating the raw-

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

KGE term and to give proportional weight to the near-zero, intermittent low flows characteristic of these catchments, as in **Eq. (2)** (Santos et al., 2018; Ndiaye et al., 2025):

$$OF_Q = {}^1\!/_2\,[KGE(Q) + KGE(\log(\mathrm{Q} + \varepsilon))]\text{ , where } \varepsilon = 0.01\text{ mm d}^{-1} \tag{2}$$

The ε = 0.01 mm d^{-1} value was selected as a small offset lower than the smallest non-zero observed daily flow, thereby preventing logarithmic singularities while preserving the relative ordering of low-flow values.

The AET-only scheme objective function OF_{AET} combines KGE on square-root-transformed daily AET with KGE on square-root-transformed monthly AET totals. Here, the square-root transformation was applied because AET does not span the multi-order-of-magnitude range, nor does it contain the near-zero values, that justify the need for a log-transformation. The monthly component was included to constrain the objective function with the seasonal AET signal, which is less affected by day-to-day satellite retrieval noise than daily AET estimates. The OF_{AET} objective function was defined as in **Eq. (3)** (Trotter et al., 2023; Hsu et al., 2025):

$$OF_{AET} = {}^1\!/_2\,\left[KGE\left(AET^{0.5}_{daily}\right) + KGE\left(\mathrm{AET}^{0.5}_{\mathrm{monthly}}\right)\right] , \tag{3}$$

The combined objective function OF_{Q+AET} used for the the Q+AET calibration scheme is the unweighted mean of OF_Q and OF_{AET}, as in **Eq. (4)**:

$$OF_{Q+AET} = {}^1\!/_2\,[OF_Q + OF_{AET}]. \tag{4}$$

2.4.3 GLUE with Latin Hypercube Sampling

The GLUE framework (Beven and Binley, 1992) was used to propagate parameter uncertainty and quantify identifiability. For each model, a single LHS of 250,000 parameter sets is drawn once over the prior ranges defined in **Table 2** (with a fixed seed for reproducibility) and applied to each watershed and calibration scheme, so that prior coverage remains identical. Each parameter set was run over the full record and OF_Q, OF_{AET} and OF_{Q+AET} together with KGE, R^2, $RMSE$ and $PBIAS$ for both Q and AET in calibration and validation, were evaluated for every valid run.

Behavioural sets were defined per model, calibration scheme and watershed. The prior ensemble comprises all parameter sets yielding a finite objective (i.e. valid model runs); the behavioural (posterior) ensemble comprises those with a scheme-specific likelihood exceeding 0.3 (OF > 0.3), and the remaining valid runs (OF ≤ 0.3) form the non-behavioural pool used to construct the null distributions. This threshold aligns with the minimum KGE threshold deemed "*satisfactory*" established by Knoben et al. (2019) and Yoshida et al. (2022), and also corresponds to NSE = 0.5, commonly referred to as "*satisfactory*" at the daily timestep in the hydrological modelling practice (Moriasi et al., 2007, 2015). Behavioural retention varied across models, schemes, and watersheds but remained well above the minimum required for the constraint analysis in most cases (see **Supplement Fig. S5**); notably, AET-only generally yielded retention rates comparable to or higher than Q-only. Cases with

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

fewer than 50 behavioural sets were flagged as unidentifiable and excluded from the permutation-based constraint analysis, since the null distribution becomes unreliable at such small posterior sizes (Beven and Binley, 1992, 2014; Yoshida et al., 2022).

2.4.4 SCE-UA optimization

To obtain a single optimal parameter set per model, scheme and watershed for the performance and water-balance analyses, the SCE-UA algorithm (Duan et al., 1992) was applied using the same objective functions, priors as in the GLUE approach (Sect. 2.4.3). The search minimised a scheme-specific loss function derived from the objective, $1 - OF$ augmented by a soft volumetric bias penalty defined to constrain an absolute volume bias below 25%. The SCE-UA search was configured with N complexes (N = 5 for GR6J and HBV; N = 7 for IHACRES, because of the larger number of parameters), a maximum of 10,000 objective-function evaluations and convergence declared after 10 shuffling loops without at least a 0.01% improvement; the seed was fixed per watershed for reproducibility. These control settings were chosen to allow sufficient exploration of the comparatively flat response surfaces encountered in semi-arid catchments (Duan et al., 1992; Chen et al., 2015). The optimal parameter sets and their simulated Q and AET series were then used as deterministic simulations used for further performance metrics and Budyko diagnostics evaluation.

2.5 Evaluation and diagnostics

2.5.1 Parameter identifiability and constraint

Parameter identifiability under each calibration scheme was assessed from the GLUE behavioural samples (Sect. 2.4.3) by comparing, for every parameter, its behavioural (posterior) distribution against its prior. Two complementary, scale-independent metrics were used. The variance ratio (VR), which measures the reduction in spread from prior to posterior, defined as in **Eq. (5)** (Thorndahl et al., 2008; Fang et al., 2020):

$$VR = \sigma^2_{posterior} / \sigma^2_{prior} \tag{5}$$

where σ^2_{prior} and $\sigma^2_{posterior}$ are the variances of the parameter across the prior and behavioural samples, respectively. VR ranges from 0 to 1, with VR = 0 indicating that conditioning strongly collapses the parameter range (high identifiability), whereas VR = 1 indicates that the behavioural sample retains the prior spread (no constraint).

The second metric is the Kolmogorov-Smirnov distance (KS) which measures the displacement and reshaping of the distribution as the maximum absolute difference between the prior and posterior empirical cumulative distribution functions, as in **Eq. (6)** (Thorndahl et al., 2008; Pianosi and Wagener, 2015; Xu et al., 2024):

$$KS = \sup_x |F_{posterior}(x) - F_{prior}(x)| \tag{6}$$

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

with F_{prior} and $F_{posterior}$ the empirical CDFs over the prior and behavioural samples. KS ranges from 0 to 1, with values near 0 indicating that the posterior is indistinguishable from the prior, and values approaching 1 indicate a strong shift in the location or shape of the distribution. The two metrics are complementary because a parameter may narrow without relocating (low VR, low-to-moderate KS) or relocate without appreciable narrowing (VR = 1, larger KS).

2.5.2 Performance metrics for optimal parameter sets

The performance of all simulations driven by the optimal parameter sets obtained using SCE-UA optimization search was assessed with four complementary metrics: the KGE (Kling et al., 2012), the coefficient of determination (R^2), the root-mean-square error (RMSE) and the percent bias (PBIAS). Each metric was computed for both streamflow (Q) and actual evapotranspiration (AET), separately over the calibration (1991-2010) and validation (2011-2020) periods, and evaluated only on days with valid observations within each window. The formulations of all the performance metrics are given in **Table 3**.

Table 3: Performance metrics used to evaluate simulated streamflow and actual evapotranspiration.

Metric	Equation	Range	Optimal value	Interpretation
KGE	$1-\sqrt{(r-1)^2+(\alpha-1)^2+(\beta-1)^2}$	$[-\infty, 1]$	1	Evaluates overall agreement between observed and simulated dynamics
R^2	$\left[\frac{\sum_{i=1}^{n}(O_i-\bar{O})(S_i-\bar{S})}{\sqrt{\sum_{i=1}^{n}(O_i-\bar{O})^2}\sqrt{\sum_{i=1}^{n}(S_i-\bar{S})^2}}\right]^2$	[0,1]	1	Quantifies the proportion of observed variance explained by simulations
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i-O_i)^2}$	$[0, +\infty]$	0	Measures the average magnitude of simulation errors, with greater sensitivity to large deviations
PBIAS	$100\times\frac{\sum_{i=1}^{n}(S_i-O_i)}{\sum_{i=1}^{n}O_i}$	$[-\infty, +\infty]$	0	Evaluates long-term volumetric bias; positive values indicate overestimation and negative values indicate underestimation

where $\boldsymbol{r}$ is the Pearson correlation coefficient, $\boldsymbol{\alpha}$ the ratio of coefficient of variation of simulated to observed ($\boldsymbol{\alpha = CV_s/CV_o}$), β is the ratio of means of simulated to observed ($\boldsymbol{\beta = \mu_s/\mu_o}$), $\boldsymbol{O_i}$ is the observed series, $\boldsymbol{S_i}$ the simulated series, $\boldsymbol{\bar{O}}$ is the mean of observed values, $\boldsymbol{\bar{S}}$ the mean of simulated values, $\boldsymbol{n}$ the length of the observed or simulated series.

2.5.3 Water-balance consistency

Each scheme was assessed for the physical plausibility of the long-term water partitioning using the Budyko framework (Budyko, 1958). The analyses presented in this section were derived from the deterministic SCE-UA optima (Sect. 2.4.4). Long-term means were computed from annual totals over 1991-2020. The Budyko framework assumes that, over a sufficiently long period, changes in catchment storage are negligible and precipitation is partitioned entirely into streamflow and evapotranspiration, i.e., $P \approx AET + Q$ (Budyko, 1958). This assumption was checked beforehand by evaluating the long-term storage residual $dS = P - AET - Q$ for each watershed from the observed long-term means, using gauged Q and satellite-derived AET (see **Supplement Table S1** and **Figure S6**). The residuals ranged from -6.4% to +8.9% relative to P ($|\overline{dS}|$ = 4.7% of P) and remained well within ±10% of P in all 17 watersheds, confirming approximate long-term closure (Koppa et al., 2021; Cheng et al., 2025; Ibrahim et al., 2025). The small, predominantly positive residuals reflects biases in satellite AET, gauged P, or groundwater contributions (Han et al., 2020; Gordon et al., 2022).

Each watershed was placed in Budyko space through its aridity index $\varphi = PET/P$ and evaporative index $\eta = AET/P$, both formed from long-term mean annual totals. Because P and PET are model inputs, φ is identical for observed and simulated points. The observed evaporative index $\eta_{obs} = AET_{GLEAM}/P$ was taken directly from GLEAM, anchoring the physical reference on the satellite-observed partitioning rather than on streamflow. However, we acknowledge that such approach is not free of circularity, since AET_{GLEAM} is simultaneously the calibration target for the AET-only and Q+AET schemes and the observed reference against which those same schemes are evaluated for plausibility. Therefore, as a robustness check, the analysis was repeated using a closure-based reference, $\eta'_{obs} = 1 - Q_{obs}/P$, which involves no AET product and is therefore immune to this circularity. This complementary check was meant to put into perspective the conclusions drawn from the GLEAM-anchored analysis and ensure that the results are not contingent to the choice of the AET source (i.e., GLEAM). The expected partitioning was described by the Fu (1981) curve, as in **Eq. (7)**:

$$\eta = 1 + \varphi - (1 + \varphi^{\omega})^{1/\omega} \tag{7}$$

The Fu (1981) curve formulation was selected in this study because it has been shown to represent the long-term water balance of West African Sahelian watersheds well, including those in the Nakanbe basin, where its shape parameter ω was found to be a meaningful indicator of catchment soil water-holding capacity (Gbohoui et al., 2021; Yonaba et al., 2021a, b). The single parameter ω was fitted per watershed by minimising the squared difference between the curve and the observed point (φ, η_{obs}), with η_{obs} clamped to the physically admissible range ($0 < \eta < \min(1, \varphi)$). To characterise the climatic uncertainty of the reference, ω was refitted on 15-year rolling windows within 1991-2020, and the 10th-90th percentiles of the resulting distribution were used to define an uncertainty envelope around each watershed Fu (1981) curve.

The plausibility of each calibrated simulation was then quantified by its Budyko distance, i.e., the signed vertical departure of the simulated evaporative index from the observed Fu (1981) curve at the aridity index of the watershed, as in **Eq. (8)**:

$$d_{Fu} = \eta_{sim} - \eta_{Fu}(\varphi, \omega_{obs}) \tag{8}$$

with positive values indicating overestimation of $\eta = AET/P$ and negative values underestimation. The absolute value $|d_{Fu}|$ was used as a scalar plausibility metric, with smaller values denoting closer agreement with the expected partitioning. The schemes were compared per model through the distribution of $|d_{Fu}|$ across the 17 watersheds.

The directional nature of any partitioning error was examined through a complementary consistency diagnostic based on the water-balance fractions Q/P and AET/P. For each simulation, the signed deviations of both fractions from their observed

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

counterparts, $\Delta(Q/P) = (Q/P)_{sim} - (Q/P)_{obs}$ and $\Delta(AET/P) = (AET/P)_{sim} - (AET/P)_{obs}$ were plotted against one another, partitioning the plane into four quadrants according to whether each flux is over- or underestimated. This representation distinguishes simulations that reproduce the observed partition from those that achieve an apparent fit to one flux by compensating in the other, therefore highlighting whether a calibration scheme attains water-balance consistency in both fluxes simultaneously.

3 Results

3.1 Parameter identifiability and asymmetry of constraint

3.1.1 Prior to posterior constraint across schemes

Parameter constraint was quantified by two complementary measures of the change from the prior to the behavioural sample (Sect. 2.5): the variance ratio (VR), which captures the reduction in spread, and the Kolmogorov-Smirnov distance (KS), which captures the shift in distributional shape. **Figures 2** and **3** map both metrics per parameter and watershed for the three models and calibration schemes. The VR (**Fig. 2**) outline a clear contrast between schemes. Under Q-only (**Fig. 2a**), narrowing is widespread across all three models, concentrated in the production and soil-moisture parameters and, for GR6J, the inter-catchment exchange terms, while most routing store and recession parameters remain closer to their prior spread. Under AET-only (**Fig. 2b**) this narrowing is largely reduced: for GR6J the parameters remain close to their prior variance across all watersheds, whereas for HBV the reduction is concentrated in the soil-moisture parameters (FC, LP), with a comparable but more localised, watershed-dependent narrowing also evident in $BETA$; for IHACRES, narrowing is concentrated in the loss-module parameters (f, e, d). Routing parameters of all three models are left largely unaffected. The combined scheme (**Fig. 2c**) closely reproduces the Q-only pattern for GR6J and reinforces the constraint on the HBV and IHACRES soil and loss parameters. These panels also show that the AET-only constraint, where it exists, is spatially uneven, most visibly for HBV-FC and HBV-$BETA$, whose narrowing varies strongly between watersheds and is absent at several sites (e.g. Tinakof, Koriziena, Falangountou).

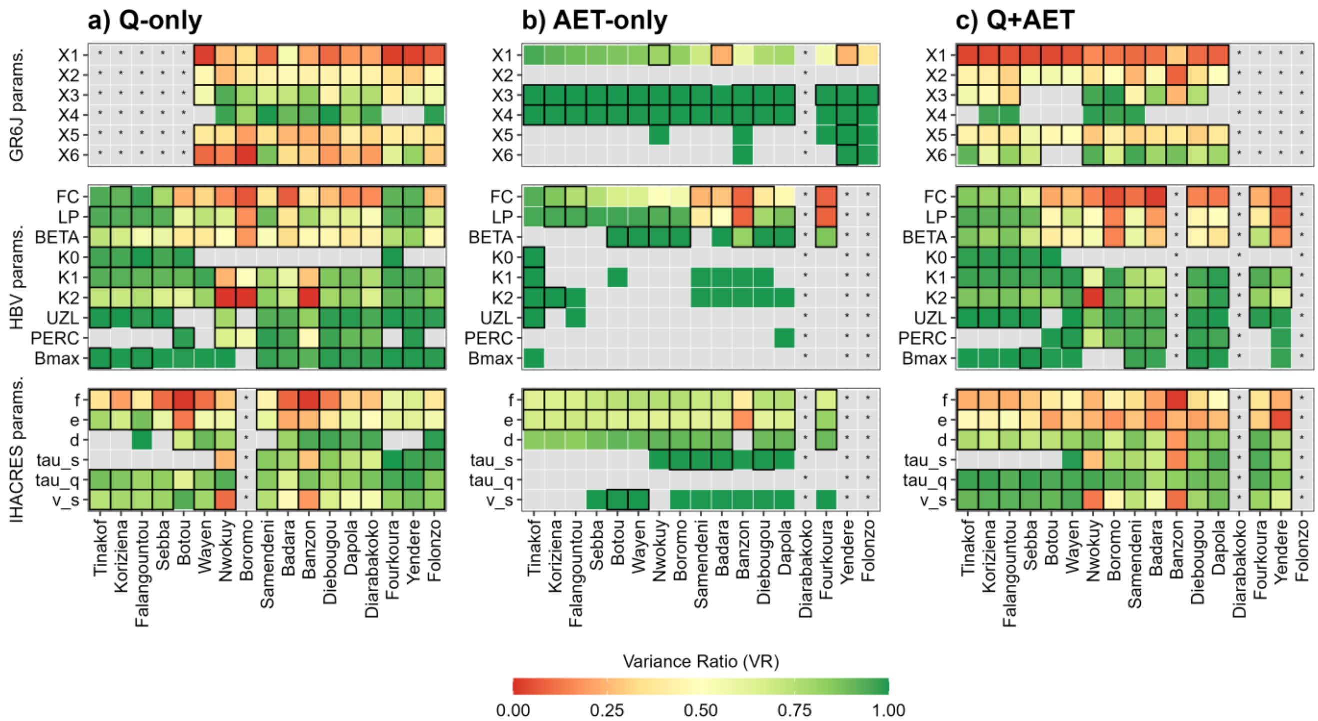
The KS distances (**Fig. 3**) align with the above, while adding distributional displacement. The largest relocations under Q-only (**Fig. 3a**) coincide with the narrowed parameters, most clearly GR6J-$X1$ and HBV-FC. Under AET-only (**Fig. 3b**), GR6J shows essentially no displacement of any parameter at any watershed, indicating that the satellite-AET carries no identifying information for the model parameters; HBV again shows displacement concentrated in FC, LP and, at a subset of watersheds, $BETA$, while IHACRES shows displacement restricted to the loss-module parameters. The weak cross-flux information transfer appearing in **Figs. 2-3** reflects a more fundamental tension between the two objective functions at the level of the prior sample itself: parameter sets behavioural under one flux rarely satisfy the behavioural threshold under the other, and the two

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

objective functions show no positive association within either behavioural set (see **Supplement Fig. S7**), most markedly for GR6J.

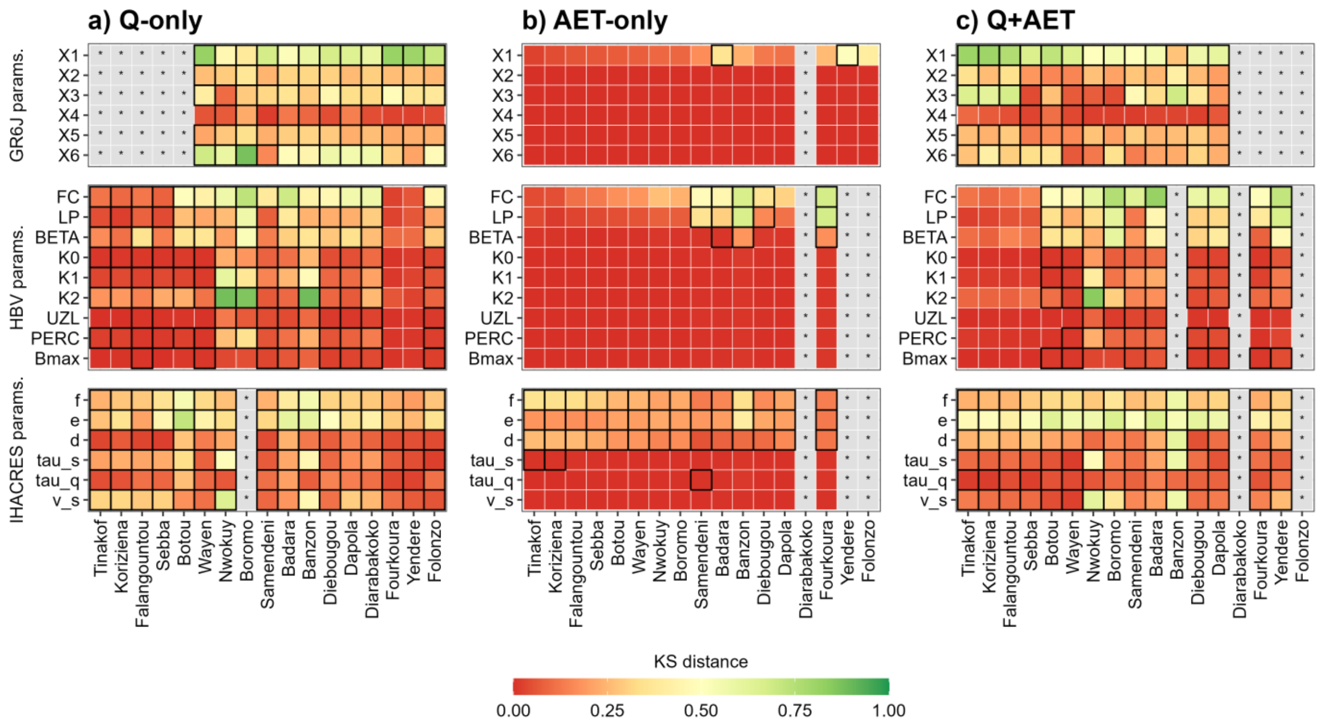
Aggregated across watersheds (**Fig. 4**), the asymmetry between the two conditioning fluxes becomes explicit, and the watershed-level $BETA$ signal which appears in **Figs. 2-3** does not survive aggregation: under AET-only, the median $BETA$ in VR and KS is close to the unconstrained values (VR ≈ 1, KS ≈ 0), confirming that its narrowing under AET-only is a localised, watershed-specific effect. For GR6J (**Fig. 4a**), the Q-based schemes constrain $X1$ most strongly, together with the exchange parameters $X2$ and $X5$ and, to a lesser extent, $X6$. The parameter $X4$ remains unidentified; under AET-only, every parameter lies at VR ≈ 1 and KS ≈ 0, with the production-store capacity $X1$ showing the only (yet, marginal) departure. For HBV (**Fig. 4b**), FC is the most constrained parameter overall and, together with LP, the only parameters appreciably conditioned by AET-only; $BETA$ responds essentially to Q, and the recession, threshold and transfer parameters ($K0$, $K1$, $K2$, UZL, $PERC$, $Bmax$) remain unidentified under all schemes. For IHACRES (**Fig. 4c**), the loss-module parameters f, e and d carry almost all the constraint while the routing parameters (τ_s, τ_q, v_s) are weakly constrained at best; notably, e attains its strongest constraint under the combined Q+AET scheme, exceeding that of either single-flux (Q-only, AET-only) scheme. Across all three models, the Q+AET scheme matches or slightly exceeds the constraint achieved by Q-only and never declines to the level of AET-only.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026



420 **Figure 2: Prior-to-posterior variance ratio (VR) of each model parameter (rows) at the 17 watersheds (columns), under the (a) Q-only, (b) AET-only and (c) Q+AET calibration schemes. Low VR (red) denotes a strong reduction of parameter variance from the prior to the behavioural sample (strong constraint); VR = 1 (green) denotes no reduction. Black outlines mark cells where the behavioural sample differs significantly from the prior (permutation test at 5% significance level); grey cells (asterisked, *) indicate watershed-model-scheme combinations with fewer than 50 behavioural sets (n_post < 50), excluded as unidentifiable (Sect. 2.4.3). Rows are grouped by model (GR6J, HBV, IHACRES) and watersheds are ordered from north (left) to south (right).**

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026



425 **Figure 3. Prior-to-posterior Kolmogorov-Smirnov (KS) distance of each model parameter (rows) at the 17 watersheds (columns), under the (a) Q-only, (b) AET-only and (c) Q+AET schemes. KS = 1 (green) denotes a strong shift of the parameter distribution from the prior to the behavioural sample; KS = 0 (red) denotes no shift. Outlines, grey cells (n_post < 50) and watershed ordering as in Fig. 2.**

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

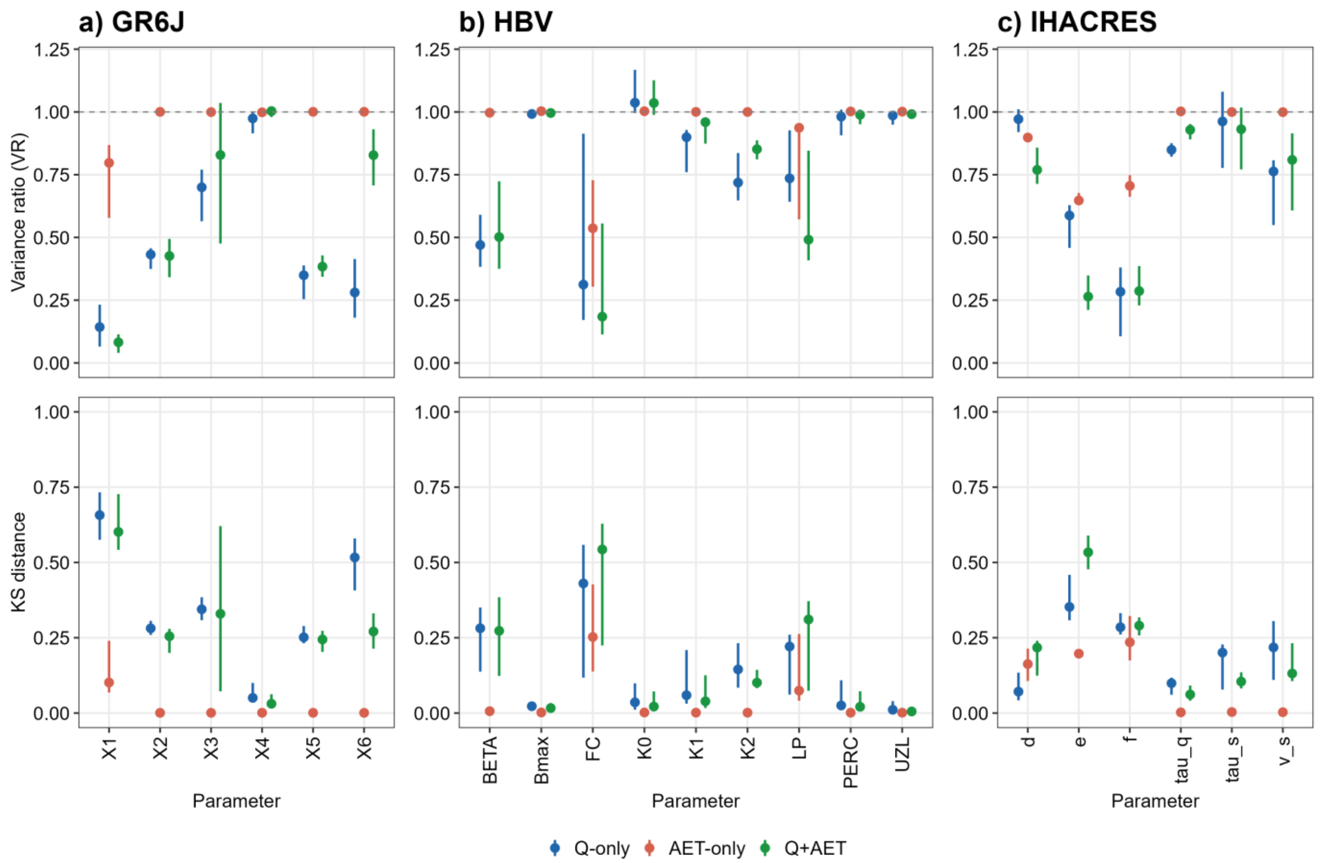


Figure 4: Parameter constraint summarised across the 17 watersheds for (a) GR6J, (b) HBV and (c) IHACRES: variance ratio (VR, top row) and KS distance (bottom row) for each parameter under the three calibration schemes (Q-only, AET-only, Q+AET). Symbols denote the median across watersheds and whiskers the interquartile range. Strong constraint corresponds to low VR and high KS; the dashed black line highlights VR = 1 (no variance reduction).

3.1.2 Behavioural parameter distributions by model

For GR6J (**Fig. 5**), the streamflow-based schemes (Q-only, Q+AET) redistribute the production and exchange parameters toward well-defined regions: under Q-only the production-store capacity $X1$ concentrates at the lower end of its range, the two exchange parameters $X2$ and $X5$ collapse onto sharp modes at zero, and $X3$ and $X6$ are drawn toward low values, while the unit-hydrograph base $X4$ retains the prior (**Fig. 5**, top row). Under AET-only, all six densities essentially coincide with the prior, the sole exception being $X1$, whose behavioural sample is depleted at very low values and displaced toward higher capacities (the opposite to the low-value preference imposed by Q). The Q+AET scheme reproduces the Q-only mode locations for $X1$, $X2$ and $X5$, with the posterior of $X1$ sharply peaked under the combined scheme, highlighting that the streamflow signal governs these parameters when both fluxes are used jointly.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

For HBV (**Fig. 6**), the displacement is confined to the soil-moisture parameters, the routing reservoirs ($K0$, $K1$, $K2$, UZL, $PERC$, $Bmax$) remaining near the prior under every scheme. The maximum soil-moisture capacity FC is displaced toward the upper end of its range by all three schemes, its behavioural values accumulating against the upper bound, indicating that the prior limit is restrictive for this parameter. The PET-extraction threshold LP shows a directional contrast analogous to GR6J-$X1$, drawn toward low values under Q-only and toward high values under AET-only, while the combined scheme lying between the two. The recharge exponent $BETA$ is displaced toward higher values under the streamflow-based schemes but coincides with the prior under AET-only, identifying it as a streamflow-informed parameter.

For IHACRES (**Fig. 7**), the three loss-module parameters f, e and d are conditioned under all schemes, while the routing parameters τ_s and v_s remain close to the prior throughout; τ_q shows a modest displacement toward higher values under the streamflow-based schemes (Q-only, Q+AET), but its posterior never departs from the prior in comparison to the loss-module parameters. In contrast to the opposing displacements seen for GR6J-$X1$ and HBV-LP, both fluxes act on the loss module: the efficiency factor e is displaced toward the upper part of its range by both Q-only and AET-only, and most strongly under the combined scheme, while the reference deficit d is drawn toward low values and f toward moderate values under all schemes. Because both fluxes pull these parameters in the same direction, the combined scheme reinforces their conditioning.

Two findings emerge across the three models. First, the displacement of behavioural values is restricted to the production, soil-moisture and loss modules, with the routing and recession parameters retaining their prior distribution under every scheme. A partial exception is the HBV-$K0$ and IHACRES-τ_q parameters, both showing a modest, scheme-dependent displacement without approaching the conditioning seen in the production and loss parameters. Second, where a single parameter is conditioned by both fluxes, the two either oppose (the combined behavioural distribution lies between the single-flux positions as in GR6J-$X1$, HBV-LP) or coincide (the combined distribution is the one strongly displaced as in IHACRES-e). These cross-scheme contrasts are statistically confirmed by pairwise comparison on the behavioural values (see **Supplement Fig. S8**), which are significant for the conditioned soil and loss parameters and non-significant for the routing parameters, with the partial exception of HBV-$K0$, which differs significantly between Q-only and AET-only but not between either single-flux scheme and Q+AET.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

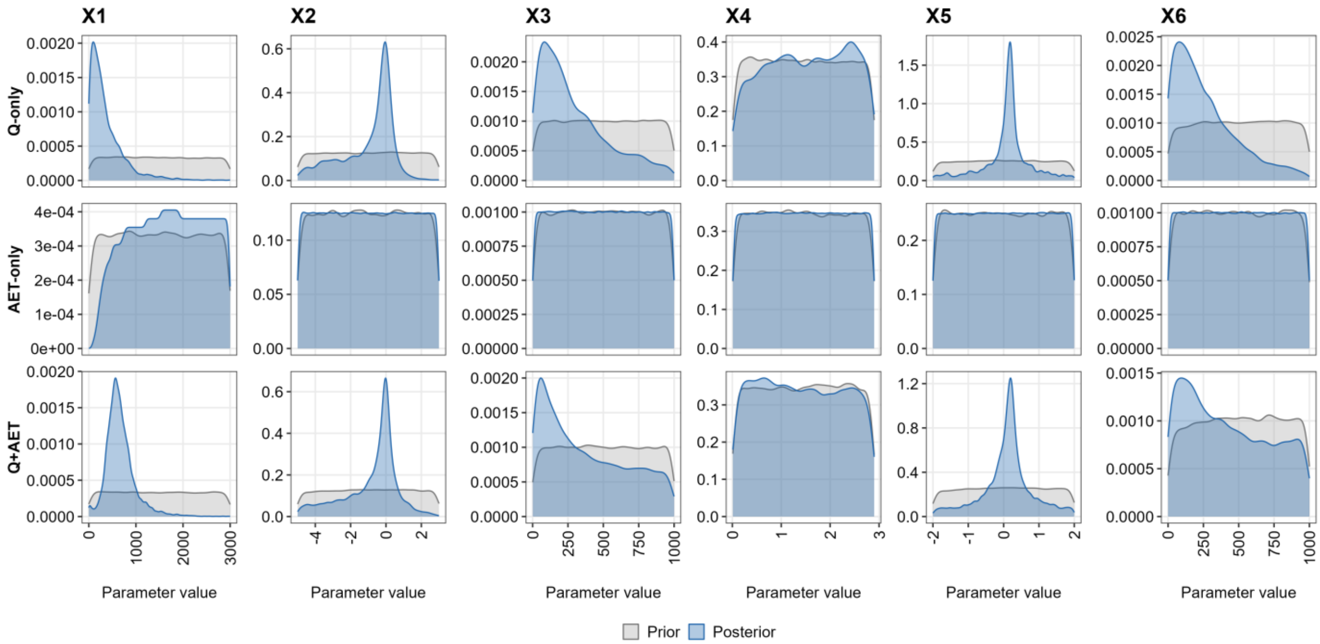


Figure 5: Prior and behavioural-posterior parameter densities for GR6J, by calibration scheme and parameter, pooling the behavioural parameter sets of all 17 watersheds. Overlap of the posterior with the prior indicates no conditioning; a narrowing or shift indicates constraint and its position along the x-axis indicates the preferred parameter region.

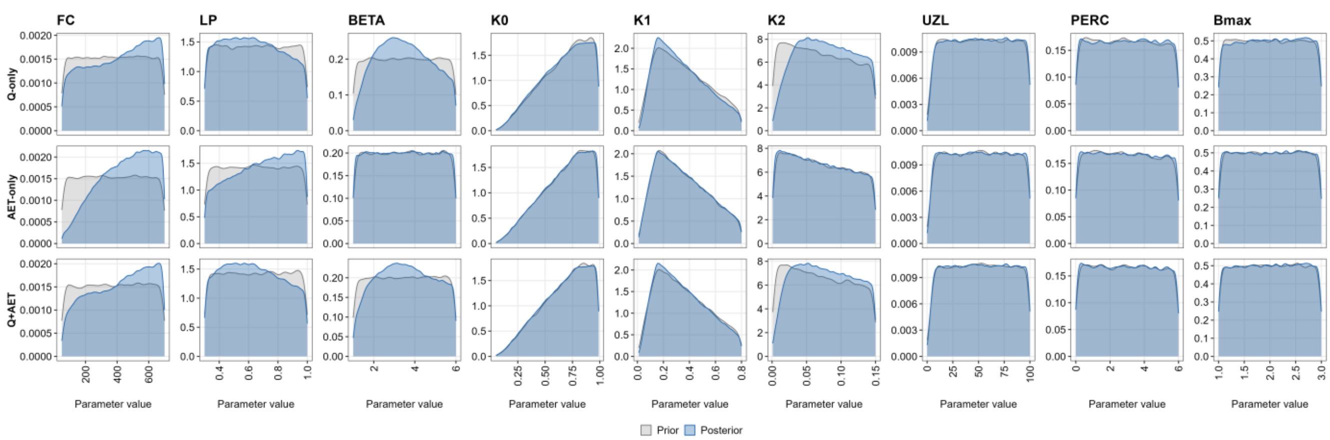


Figure 6: Prior and behavioural-posterior parameter densities for HBV, by scheme and parameter. Conventions as in Fig. 5.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

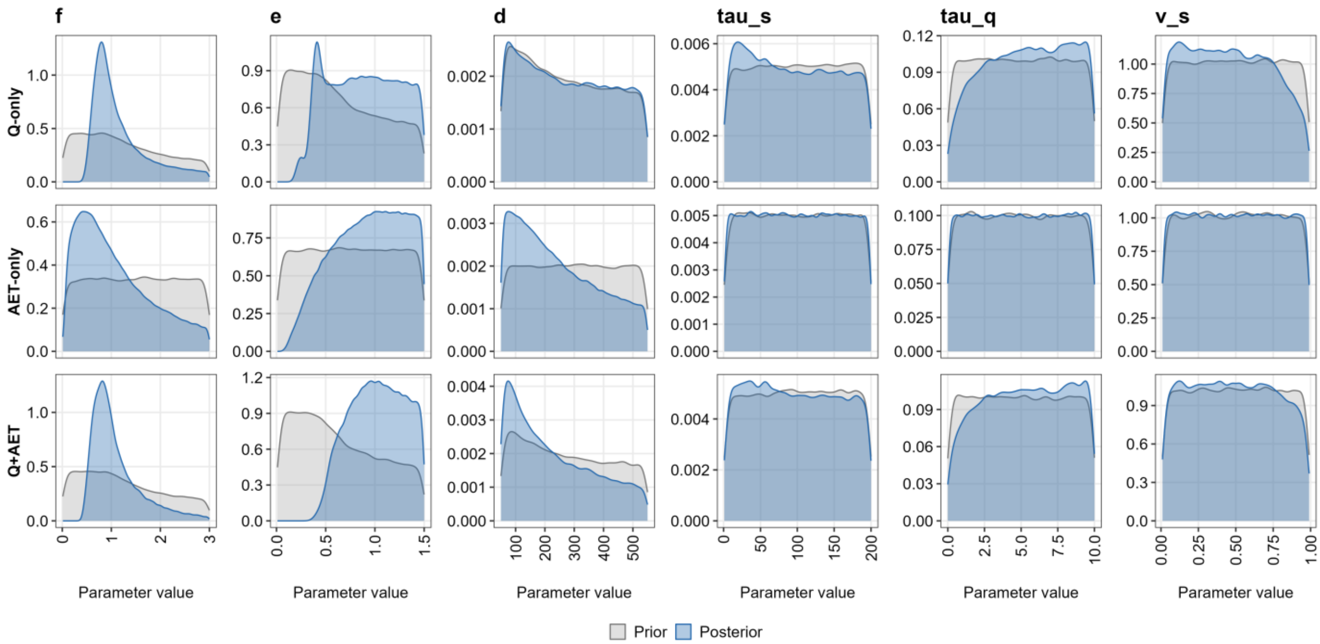


Figure 7: Prior and behavioural-posterior parameter densities for IHACRES, by scheme and parameter. Conventions as in Fig. 5.

475 ### 3.1.3 Constraint along the hydroclimatic gradient

Figure 8 resolves the constraint metrics along the north-south latitudinal gradient, averaging across parameters within each model and scheme. The streamflow-based schemes (Q-only, Q+AET) show a broadly coherent latitudinal trend, with parameter constraint generally strongest in the semi-arid north (14-15°N) and weakest in the sub-humid south (~10°N). The trend is not strictly monotonic, however, as both schemes show a local reversal around 12.5-13.5°N, where constraint temporarily weakens before strengthening again toward the northernmost stations. Also, the two schemes diverge below 11°N, where Q+AET shows stronger constraint (lower VR) than Q-only at the southernmost watersheds. Within the streamflow-based schemes, the three models separate consistently along the gradient: GR6J parameters sit at higher VR and lower KS than HBV and IHACRES at almost every latitude, reproducing in spatial terms the weaker overall constraint already identified for GR6J in Sect. 3.1.1.

The AET-only scheme shows uniformly a weak signal. Its mean VR remains close to unity and mean KS close to zero at the northern and southern extremes of the gradient, but both metrics show a modest, double-peaked departure from this baseline at intermediate latitudes, around 11-11.5°N and near 13-13.5°N. This indicates that such constraint is not organised along the hydroclimatic gradient in a simple northward- or southward-strengthening pattern, in contrast to the streamflow-based schemes.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

This finding reinforces, in spatial terms, the asymmetry already established (Sects. 3.1.1 and 3.1.2): the streamflow signal is not only the more informative target overall but also the one whose information content is organised along the hydroclimatic gradient, broadly strengthening toward the semi-arid north, whereas the AET-only signal, where present, is not. This latitudinal trend is not attributable to differences in record completeness across the network, since gap rates are negatively associated with latitude among the 17 retained watersheds (see **Supplement Fig. S3**), therefore the stronger constraint observed toward the semi-arid north coincides with more complete (and not more degraded) discharge records. The convergence of the Q-only and Q+AET trends across most of the gradient, and their divergence at the southernmost watersheds, further indicates that the combined scheme inherits the spatial behaviour of the streamflow component while occasionally exceeding it. The significant scatter around the smoothed trends, reflecting both between-model and between-watershed variability at a given latitude, indicates that latitude organises the constraint only in aggregate and does not override the model- and parameter-specific patterns described earlier.

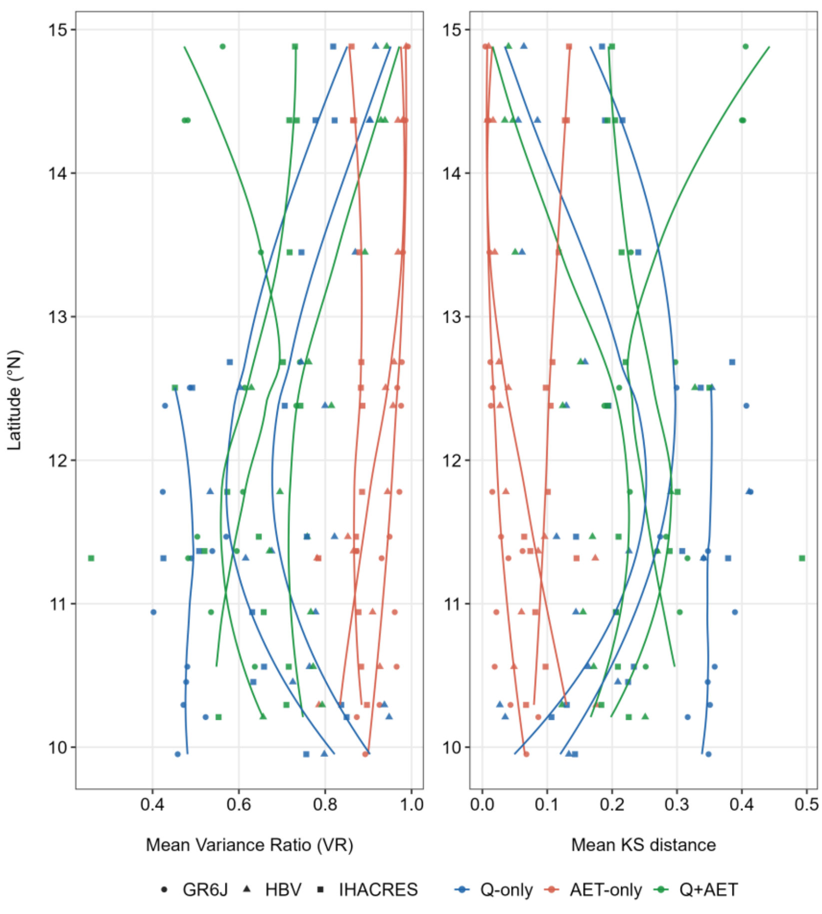


Figure 8: Parameter constraint along the latitudinal gradient: watershed-mean variance ratio VR (left) and watershed-mean KS distance (right) against latitude, for each model and calibration scheme, each point averaging the metric over all parameters of a

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

given model and scheme at that watershed. Curves are LOESS-smoothed trends per scheme, pooling all three models. Strong constraint corresponds to low VR and high KS.

3.2 Model performance under the three calibration schemes

3.2.1 Streamflow and evapotranspiration skill

We investigate the model and scheme performance obtained from the single SCE-UA optima (Sect. 2.4.4), evaluated for both fluxes over the calibration and validation periods. **Figure 9** summarises four performance metrics across the 17 watersheds, by model and scheme, for streamflow and AET. The corresponding detailed per-watershed values are given in **Supplement Figs. S9-S11**.

Streamflow skill exposes the asymmetry between the two single-flux schemes. Calibrating on Q yields good discharge simulations across all three models (**Fig. 9a**, median KGE 0.62-0.70), whereas calibrating on AET alone degrades streamflow severely: AET-only KGE collapses to low or negative medians (**Fig. 9a**) and is accompanied by large and erratic volume errors, most extreme for HBV (median PBIAS 59.9%, **Fig. 9m**). This degradation persists in validation for GR6J and IHACRES, while HBV median Q-KGE under AET-only improves (0.27 in calibration to 0.37 in validation, **Fig. 9a, b**), though it remains well below its Q-only. The combined scheme retains almost all the Q-only discharge skill (**Fig. 9a, b**), implying that adding the AET constraint to streamflow does not significantly compromise the hydrograph.

Evapotranspiration skill shows the opposite, in less proportions. Calibrating on AET improves AET reproduction over Q-only, most clearly for GR6J, whose Q-only AET skill is poor (**Fig. 9c**, median KGE -0.14) and whose RMSE and bias are large (**Fig. 9k, o**). Critically, however, the combined scheme matches or exceeds AET-only on AET in every model (**Fig. 9c**), with the Q+AET median equalling or surpassing AET-only for all three, while preserving discharge skill. The asymmetry is therefore not symmetric in cost: streamflow carries enough information to yield an acceptable AET simulation, whereas AET alone does not yield an acceptable streamflow simulation.

A consistent feature across panels is the bias signature of each single-flux scheme. AET-only systematically underestimates AET (negative PBIAS in **Fig. 9o, p** for HBV and IHACRES) while over estimating discharge volume in every model (**Fig. 9m, n**). Q-only carries the larger AET bias of the two single-flux schemes (**Fig. 9o)**, median PBIAS -33.3% for GR6J) and Q+AET keeps both volume errors moderate for both fluxes (**Fig. 9m-p**). The per-watershed heatmaps (see **Supplement Figs. S9-S11**) confirm that these patterns hold across the gradient rather than being driven by a few watersheds in particular, the AET-only Q columns in particular show poor scores and physically implausible volume errors at most sites and for every model. Differences between models are secondary to the scheme effect: GR6J and HBV attain marginally higher median Q skill than IHACRES under the Q-based schemes (**Fig. 9a, b**), but the scheme-driven contrasts dominate the metric distributions throughout.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

These results reinforce, at the level of simulated fluxes, the asymmetry already identified in parameter space: streamflow is
535 the more informative single target, AET alone is insufficient to constrain discharge, while the combined scheme is the only one acceptable for both fluxes simultaneously.

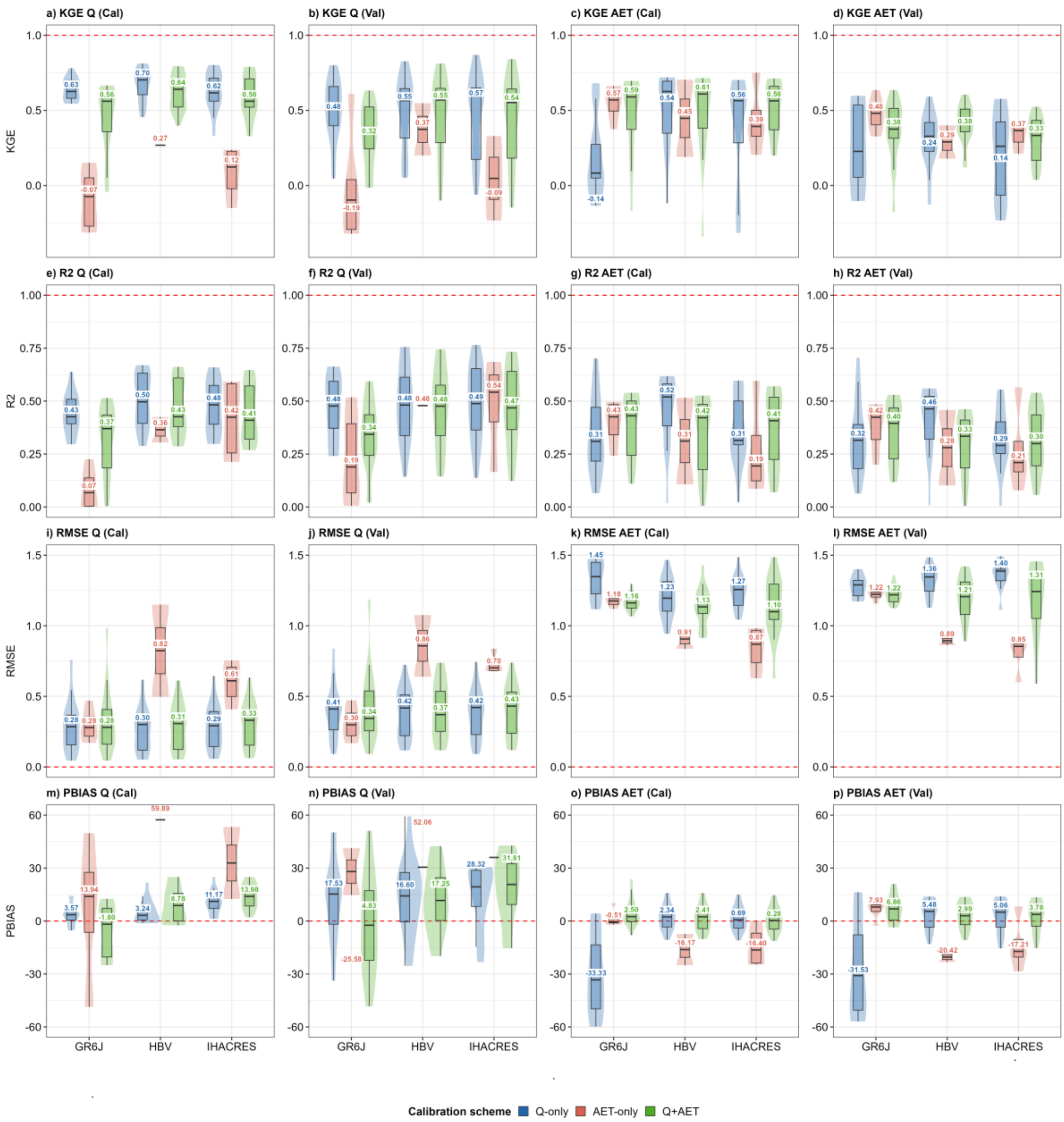


Figure 9: Performance of the calibrated models (SCE-UA optima) under the three calibration schemes, summarised across the 17 watersheds. Rows give the four metrics (KGE, R², RMSE, PBIAS) and columns give the evaluated variable and period: streamflow
540 **in calibration (a, e, i, m) and validation (b, f, j, n), and AET in calibration (c, g, k, o) and validation (d, h, l, p). The annotated value on all boxplots is the median. Red dashed lines indicate the optimum of each metric (KGE = 1, R² = 1, RMSE = 0, PBIAS = 0).**

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

3.2.2 Simulated seasonal dynamics

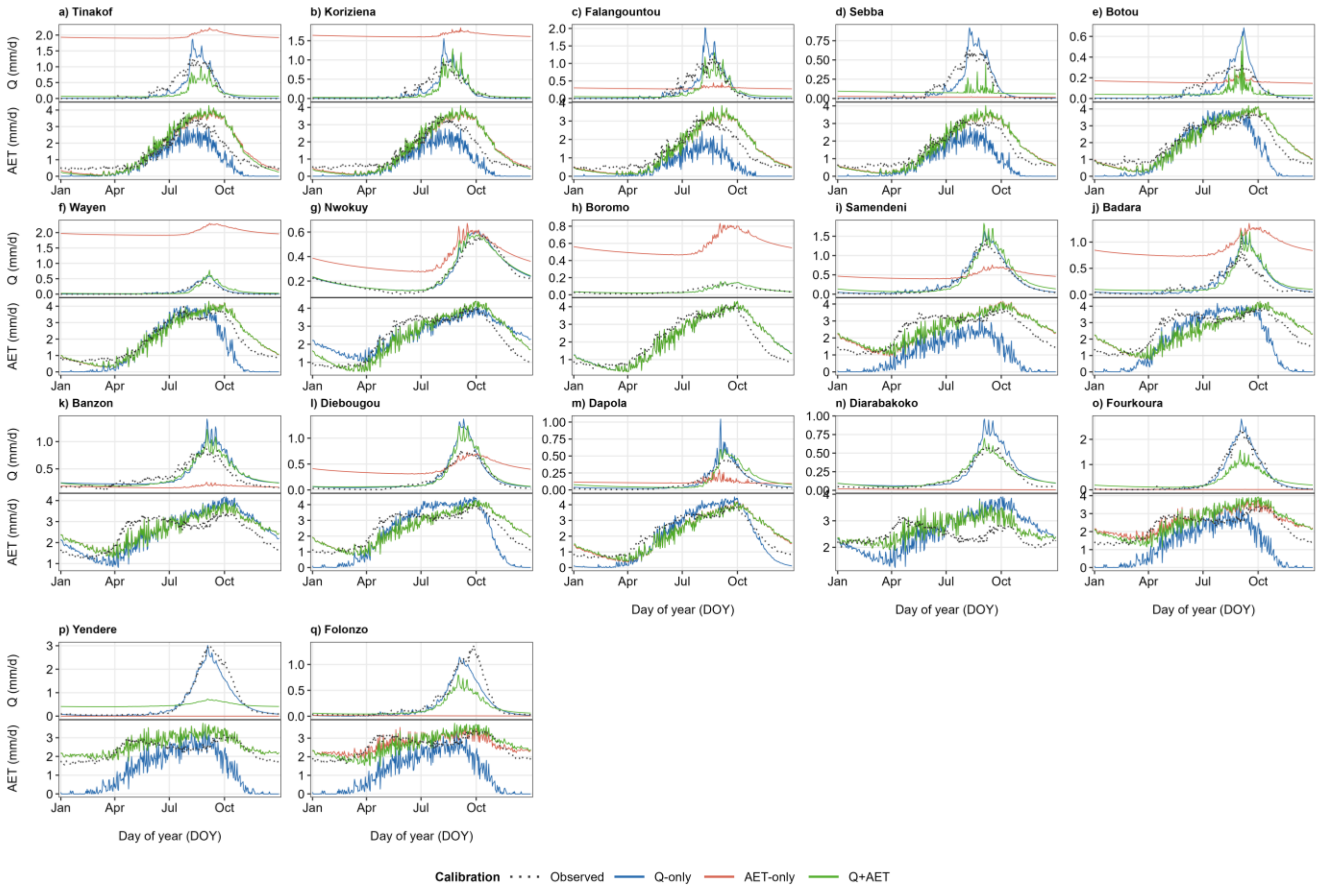
Figures 10-12 show the interannual-mean daily regimes of Q and AET for GR6J, HBV and IHACRES respectively, by watershed and scheme. The corresponding full timeseries, aggregated at the monthly timescale, are provided in **Supplement Figs. S12-S14**.

Under Q-only and Q+AET, all three models reproduce the unimodal August-September streamflow peak and the near-zero dry-season baseflow, the two streamflow-based regimes being nearly indistinguishable at most watersheds (**Figs. 10-12**, upper Q sub-panels). Yendere and Diarabakoko (**Fig. 10p, n**) are the clearest exceptions, where Q+AET peak flow under GR6J model falls short compared to Q-only. AET-only, by contrast, fails to reproduce the streamflow regime in a model-specific manner. For GR6J (**Fig. 10**), AET-only produces a nearly flat, non-seasonal discharge: either a low but non-negligible constant flow (e.g. Koriziena, Boromo, Dapola, Nwokuy, Wayen) or essentially zero streamflow (e.g. Sebba, Botou, Banzon, Diarabakoko, Fourkoura, Folonzo), consistent with a streamflow output left structurally unconstrained by the AET objective. For HBV and IHACRES (**Figs. 11-12**), AET-only instead generates a streamflow peak in the correct season but with excessive amplitude, while retaining the unimodal shape.

The evapotranspiration regimes highlight the converse limitation of Q-only and a clear model contrast. Under Q-only, the simulated AET regime departs systematically from the observed seasonal shape, and the nature of the departure differs between models. For GR6J (**Fig. 10**, lower AET sub-panels), Q-only produces a strongly flattened AET cycle, with elevated dry-season values and a suppressed wet-season rise (e.g. Boromo, Samendeni, Banzon), indicating that the streamflow objective leaves the GR6J AET poorly determined. For HBV and IHACRES (**Figs. 11-12**), the Q-only AET regime retains a more realistic seasonal shape but typically overestimating wet-season AET. Under AET-only and Q+AET, all three models track the observed AET regime closely, the two schemes being nearly coincident for AET in most watersheds, which shows that adding the streamflow constraint to AET does not distort the seasonal evapotranspiration cycle.

These results clarify the sharp model contrast under the single-flux schemes: GR6J is the most vulnerable to the conditioning variable, its unconstrained flux (AET under Q-only, Q under AET-only) collapsing to a non-seasonal or degenerated form, whereas HBV and IHACRES preserve the correct seasonal phase under all schemes - while failing mainly in amplitude. The full timeseries (see **Supplement Figs. S12-S14**) further confirm that these mean-regime features reflect the behaviour of individual years rather than averaging artifacts.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026



570 **Figure 10: Interannual average daily regimes of streamflow (upper sub-panel) and AET (lower sub-panel) simulated by GR6J under the three calibration schemes, compared with observations, for each of the 17 watersheds (a-q). Watersheds are ordered from north (a) to south (q).**

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

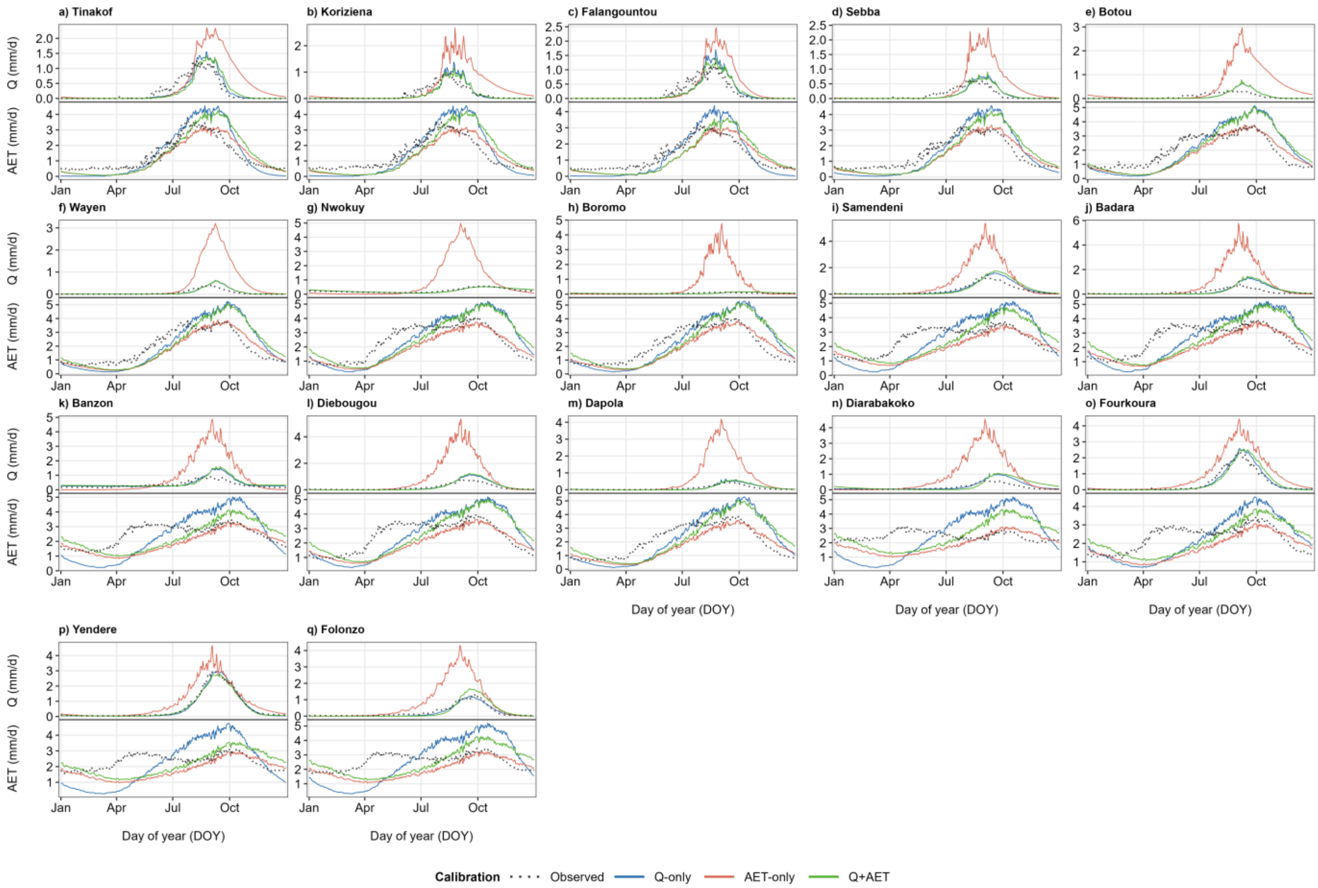


Figure 11: Interannual average daily regimes of streamflow (upper sub-panel) and AET (lower sub-panel) simulated by HBV under the three calibration schemes, compared with observations, for each of the 17 watersheds (a-q). Watersheds are ordered from north (a) to south (q).

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

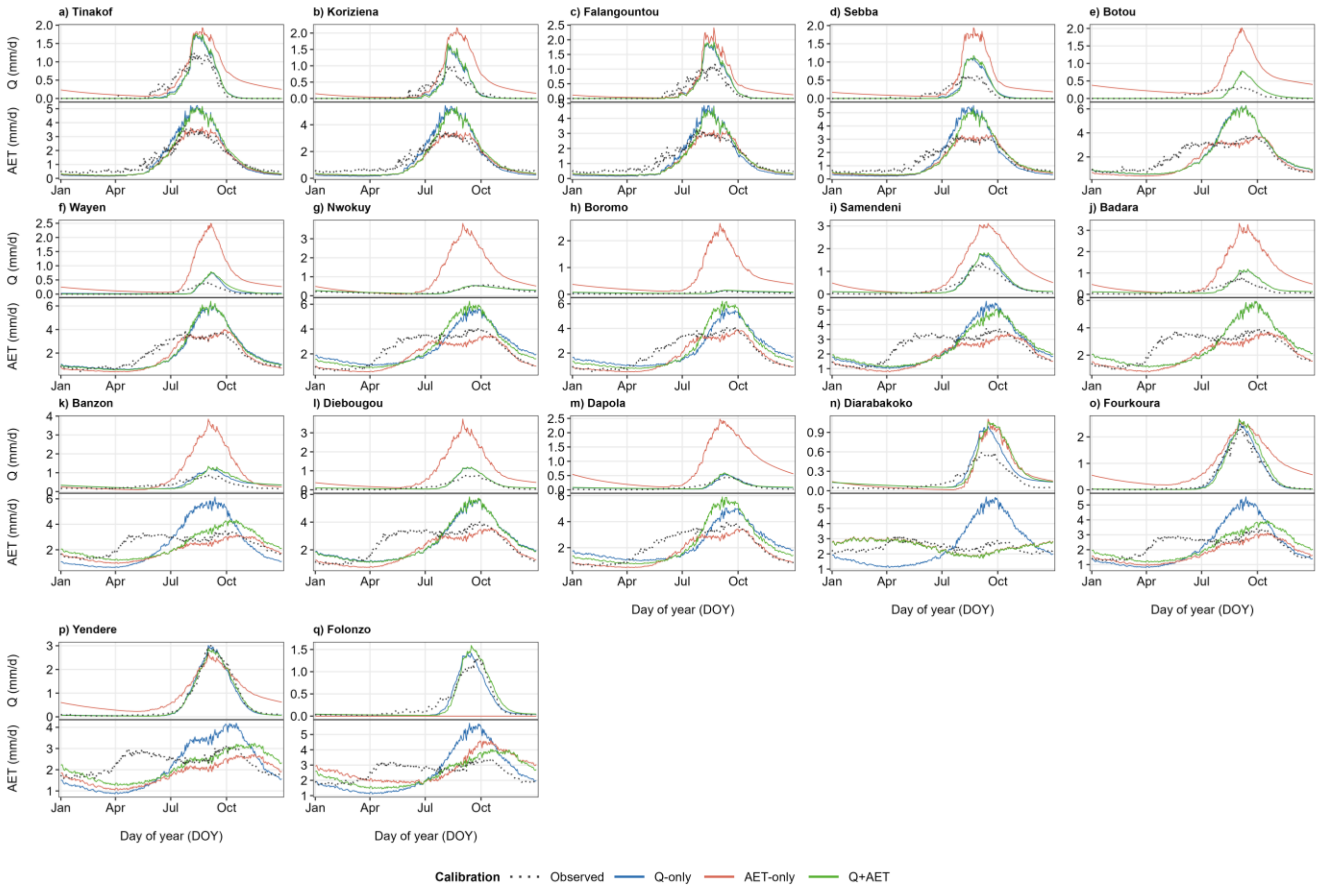


Figure 12: Interannual average daily regimes of streamflow (upper sub-panel) and AET (lower sub-panel) simulated by IHACRES under the three calibration schemes, compared with observations, for each of the 17 watersheds (a-q). Watersheds are ordered from north (a) to south (q).

3.3 Water-balance consistency

3.3.1 Budyko-space position and distance to the Fu curve

The Budyko framework provides an independent, climatologically constrained evaluation of whether the calibrated long-term partitioning between evapotranspiration and runoff is physically plausible. **Figure 13** summarises, per model, the absolute Budyko distance $|d_{Fu}|$ of the SCE-UA optima across the 17 watersheds, while **Figure 14** locates the individual simulations within Budyko space relative to the basin-specific Fu (1981) curves.

The distribution of $|d_{Fu}|$ (**Fig. 13**) appears to be model-dependent and discriminates between schemes. For the IHACRES and HBV models (**Fig. 13b, c**), the streamflow-based schemes lie close to the Fu (1981) reference, whereas AET-only is displaced further from it, with its distribution shifted above those of Q-only and Q+AET. For GR6J (**Fig. 13a**), the ordering is reversed and the spread is larger: Q-only attains the greatest and most variable departure from the curve, while AET-only and Q+AET

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

remain close to it. In every model, however, the combined scheme yields a tight distribution of $|d_{Fu}|$ clustered near the reference, matching the optimal of the two single-flux schemes. Q+AET is therefore the only scheme that is consistently close to the Fu (1981) curve across all three models.

The Budyko-space positions (**Fig. 14**) highlight the direction of these departures and confirm that the observed watersheds are well described by the Budyko-Fu framework: the observed points lie on their fitted curves, within the energy and water limits, along an aridity gradient spanning $\varphi \approx 1.8$ in the south to $\varphi \approx 4.1$ in the north. Against this reference, the AET-only optima of IHACRES and HBV fall systematically below the curve, i.e. they underestimate the evaporative index $\eta = AET/P$, which occurs mostly in the wetter southern and central basins (e.g. Nwokuy, Samendeni, Diebougou, Yendere; **Fig. 14g, i, l, p**). Because φ is fixed by the climate, a deficit in η necessarily implies a compensating excess in the runoff ratio, which is the Budyko-space expression of the excessive AET-only discharge identified in the seasonal regimes (Sect. 3.2.2). The Q-only and Q+AET optima of these two models instead cluster on or just above the curve, near the observed point. For GR6J, the simulated points lie close to the curve under all three schemes in Budyko space, with its large Q-only $|d_{Fu}|$ in **Fig. 13a** driven by a subset of strongly deviating watersheds.

To test whether this pattern depends on AET_{GLEAM} serving simultaneously as calibration target and as Budyko reference (Sect. 2.5.3), the $|d_{Fu}|$ and Budyko-space diagnostics were repeated using the closure-based reference $\eta'_{obs} = 1 - Q_{obs}/P$, which involves no satellite-AET product (Supplement **Figs. S15, S16**). The asymmetry reported above is preserved under this independent anchor and, for IHACRES and HBV, is sharper: AET-only remains the most displaced scheme from the Fu (1981) curve, with $|d_{Fu}|$ shifted further from zero than under the GLEAM anchor, while Q-only and Q+AET stay tightly clustered near it (see **Supplement Fig. S15b, c**). For GR6J, the large, dispersed Q-only $|d_{Fu}|$ persists under the closure anchor (see **Supplement Fig. S15a**), consistent with a structural rather than data source-driven origin. Across all three models, Q+AET remains the scheme closest to the reference partition under both anchors. Since the calibration-scheme asymmetry is preserved, and for two of the three models amplified, when judged against a reference independent of GLEAM, we conclude that it is not an artefact of the GLEAM AET forcing.

The signed distances (see **Supplement Fig. S17**) make these tendencies explicit and ordered by latitude. For IHACRES and HBV, AET-only is negative at nearly every watershed (η underestimated), Q-only and Q+AET scattering narrowly about zero. For GR6J, Q-only produces large negative departures concentrated in particular basins (e.g. Samendeni, Falangountou, Sebba, Diebougou) while AET-only and Q+AET remain near zero, consistent with the heavy-tailed Q-only distribution in **Fig. 13a**. Across all three models, Q+AET is the scheme whose signed distances remain closest to zero with least scatter, indicating that joint calibration delivers a long-term water partition compatible with the climatological expectation regardless of model structure.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

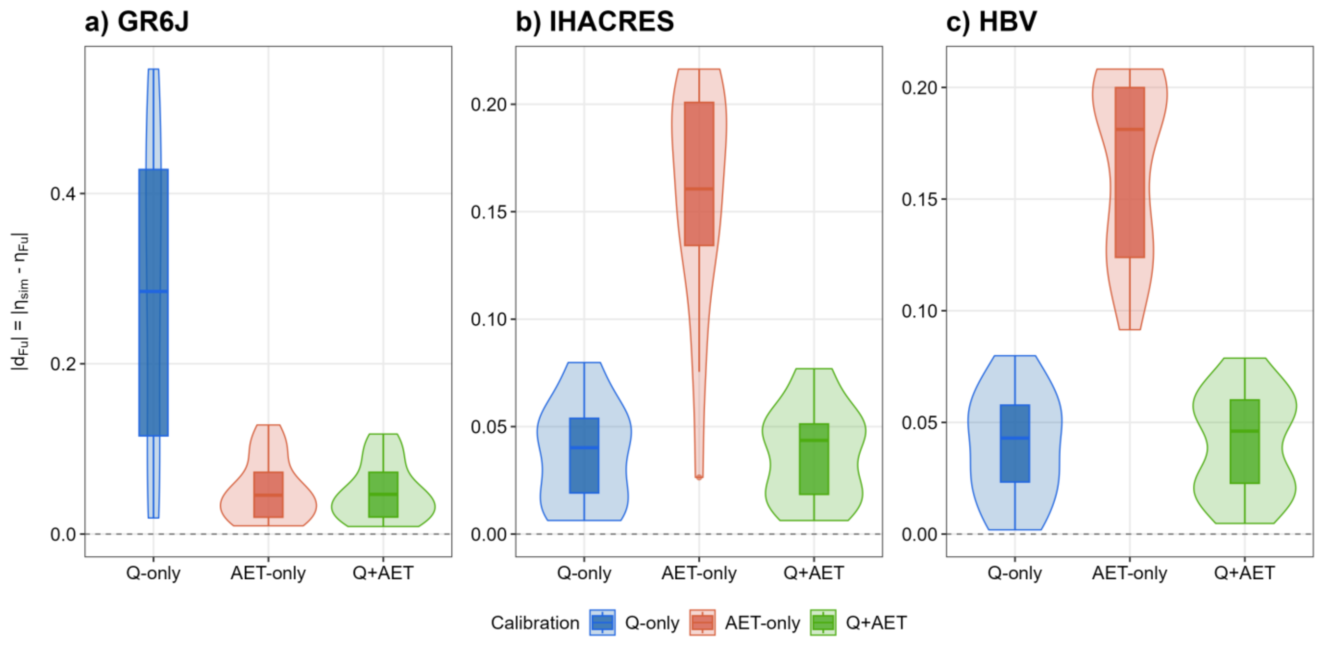
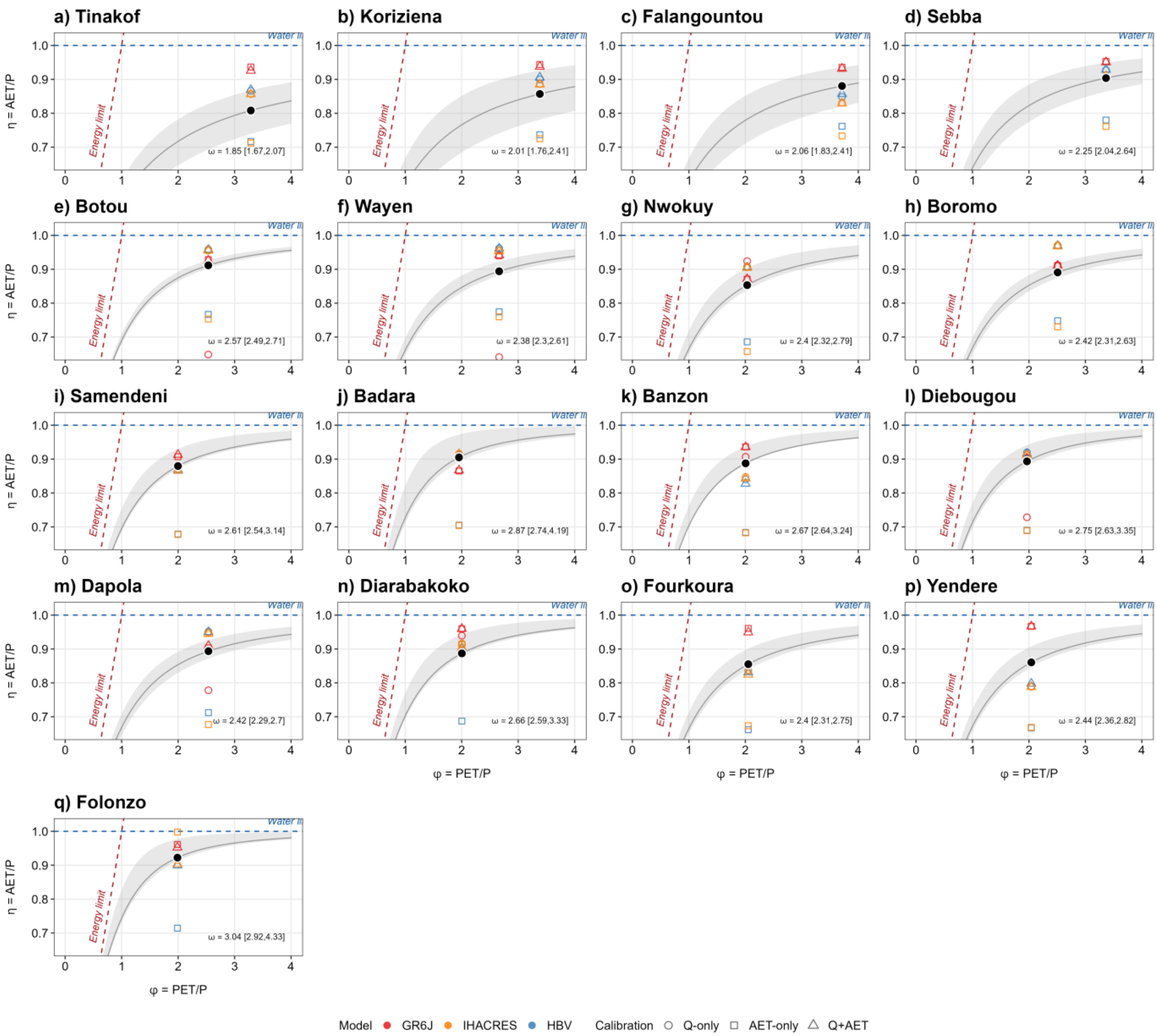


Figure 13: Absolute Budyko distance $|d_{Fu}| = |\eta_{sim} - \eta_{Fu}(\varphi, \omega_{obs})|$ of the calibrated (SCE-UA) optima from the basin-specific Fu (1981) curve, by calibration scheme for (a) GR6J, (b) IHACRES and (c) HBV. The dashed line marks $|d_{Fu}| = 0$ (exact agreement with the curve).

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026



625

Figure 14: Position of the calibrated optima in Budyko space (evaporative index $\eta = AET/P$ versus aridity index $\varphi = PET/P$) for each of the 17 watersheds (a-q, ordered north to south). The solid line is the watershed-specific Fu (1981) curve with the fitted ω (value and rolling-window uncertainty range); the grey band is the ω envelope. In each panel, the filled black point is the observed partition.

630

3.3.2 Joint Q-AET partitioning consistency

We analyse whether a model reproduces the two water-balance fluxes simultaneously or merely compensates an error in one through an offsetting error in the other. **Figure 15** presents, for each calibrated optimum, the joint deviation of the runoff and evaporative ratios from their observed values, $\Delta(Q/P)$ and $\Delta(AET/P)$, with the four quadrants distinguishing concurrent over- or underestimation of the two fluxes.

The single-flux schemes (Q-only, AET-only) occupy distinct and diagnostic regions of the plane. For the IHACRES and HBV models (**Fig. 15b, c**), the AET-only optima fall almost exclusively in the lower-right ("*over Q | under AET*") quadrant, forming a coherent cluster well displaced from the origin: matching the AET dynamics under this scheme is achieved at the cost of routing excess water to streamflow, the two errors being mirror images, as required by the fixed climatic forcing. One IHACRES watershed departs from this cluster into the opposite "*under Q | over AET*" quadrant, showing that the compensating pattern, while dominant, is not universal. The Q-only optima of the same two models lie near the origin along the $\Delta(AET/P)$ axis, deviating only modestly and without a systematic runoff bias.

For GR6J (**Fig. 15a**), the pattern is transposed and less uniform. At most watersheds the AET-only optima cluster close to the origin, but a small subset instead falls far into the "*over Q | over AET*" quadrant, with $\Delta(Q/P)$ exceeding 0.9: at these watersheds AET-only does not merely leave streamflow unconstrained (Sect. 3.2.2) but pushes it to a large positive bias alongside a smaller positive AET bias, rather than compensating between the two fluxes. The Q-only optima, by contrast, spread downward along the "*under AET*" axis at near-zero $\Delta(Q/P)$, reproducing the runoff ratio while substantially underestimating the evaporative ratio, which remains the dominant signature of Q-only for GR6J.

The combined scheme is distinguished by its position instead of its proximity to the origin. In all three models the Q+AET optima cluster tightly around the origin (**Fig. 15a-c**), departing little along either axis while avoiding the off-diagonal quadrants occupied by the single-flux schemes. Joint calibration therefore does not trade one flux against the other: it is the only scheme that reproduces the observed runoff and evapotranspiration ratios concurrently, and it does so consistently across the three model structures, including at the watersheds where AET-only fails most severely. This concurrent consistency, rather than performance on either flux alone, is the decisive evidence that the two targets, when imposed together, constrain a physically coherent water partition.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

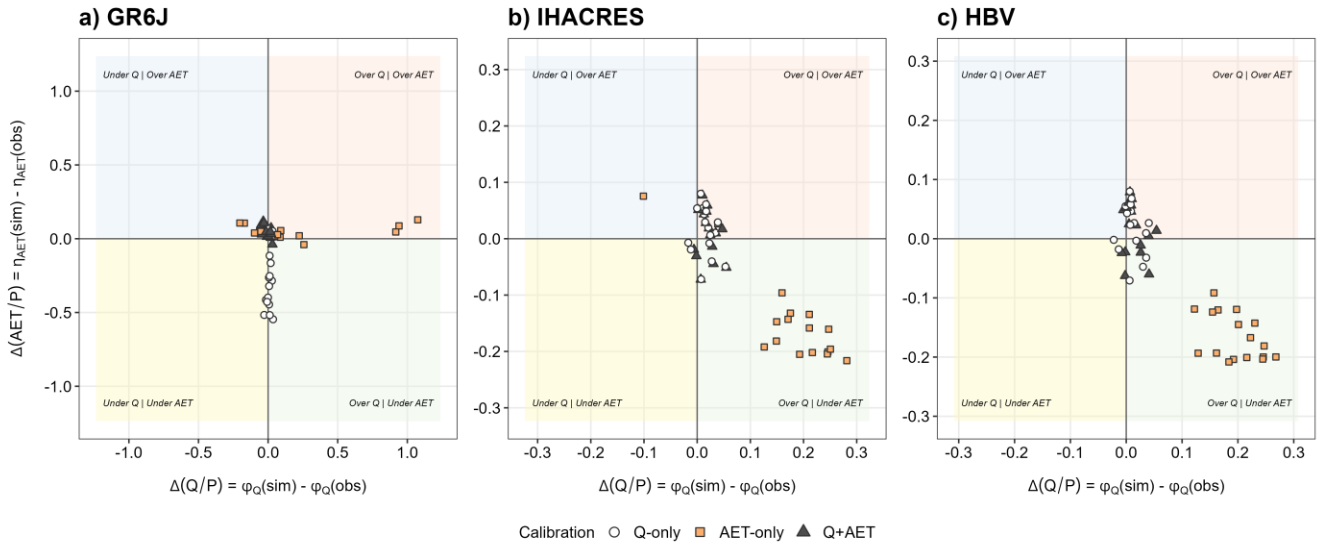


Figure 15: Joint consistency of the simulated long-term water partition for (a) GR6J, (b) IHACRES and (c) HBV. Each point is a calibrated (SCE-UA) optimum per watershed. Note the wider axis range in panel (a).

4 Discussion

This study highlights that the three calibration schemes (Q-only, AET-only and Q+AET) do not constrain the models symmetrically. Calibrating against actual evapotranspiration (AET) alone leaves the routing and recession parameters at their prior distributions, and in GR6J only the production-store capacity $X1$ departs from the prior, and only marginally. AET-only calibration reproduces AET while leaving streamflow essentially unconstrained, which indicates that the evaporative flux does not exercise the parameters governing hydrograph timing. Because AET integrates over the soil-moisture states that also generate runoff, it largely duplicates information that streamflow already carries while conveying less information on how that runoff is distributed in time. This redundancy explains why discharge is the more informative single target and why adding AET to a discharge objective preserves rather than improves the simulated hydrograph (Odusanya et al., 2021; Mei et al., 2023). Streamflow, in contrast, conditions a plausible AET in HBV and IHACRES but not in GR6J, whose Q-only AET is poorly skilled (median KGE = -0.14): the production-store capacity that streamflow favours being the opposite of the one that evaporation requires, no single parameter value therefore satisfies both fluxes.

The joint scheme is not a compromise between two competing fits. Q+AET retains nearly all the discharge skill of Q-only while matching or exceeding AET-only on evapotranspiration, and only the joint-scheme simulations cluster around the origin of the $\Delta(Q/P) - \Delta(AET/P)$ plane. It reproduces both water-balance components concurrently rather than trading one against the other, recovering the long-term partitioning that single-variable calibration distorts through compensating errors. This

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

concurrent closure matches the findings reported by Rientjes et al. (2013), specifically when discharge and satellite AET were combined.

This asymmetry is not a property of the data but a consequence of how each model couples the two fluxes. In all three structures AET and discharge are drawn from a shared soil-moisture or production state, so conditioning on one flux constrains the parameters governing the other, but only to the extent that the architecture couples them. The prior-to-posterior analysis shows that the parameters common to both pathways are displaced by streamflow and by AET in opposing directions, only weakly, or in the same direction. This distinction governs whether joint calibration is structurally costly or essentially free (Pool et al., 2024).

GR6J occupies one extreme. Its production store mediates both percolation to the routing module and the evaporative flux through a single capacity parameter, $X1$, via the same nonlinear filling function. The storage size that streamflow favours is therefore the opposite of the one AET favours: streamflow draws $X1$ towards lower values and AET towards higher values, leaving no single value to satisfy both constraints. This conflict explains why Q-only leaves AET unskilled in GR6J and why the model is the most sensitive to the conditioning flux (Perrin et al., 2003; Pushpalatha et al., 2011). The divergence is directional and reproducible across all watersheds, which distinguishes it from HBV: here the field-capacity parameter accumulates against its prior upper bound, pointing to an identifiability limit rather than a structural conflict (Koo et al., 2020).

HBV occupies an intermediate position. A dedicated soil-moisture threshold, partly separable from the response reservoirs, regulates its evaporative flux, so AET conditions the soil parameters while leaving the recession and routing parameters at their priors (Bergström, 1992). IHACRES lies at the opposite extreme: its evaporative efficiency and deficit threshold sit on the catchment-moisture-accounting pathway that also generates effective rainfall, so both fluxes exercise the same loss parameters in the same direction (Croke and Jakeman, 2004, 2005). The evaporative-efficiency parameter is consequently identified most sharply under the joint scheme, even though the model is the least responsive overall.

These three structures display a spectrum of coupling mechanisms that predicts the cost of combining the targets. This further relates to a concrete architectural cause: that simulated AET is sensitive to model structure (Zhang et al., 2020; Wu et al., 2026a) and that individual parameters can contribute oppositely to evaporative and streamflow skill (Do et al., 2024). The coupling node between the evaporative and runoff-generating pathways determines whether multivariable calibration reconciles the two fluxes or forces a trade-off (Pool et al., 2024). A larger-sample evidence of this structural control comes from Wu et al. (2026a), who compared joint Q-AET calibration against a Q-only benchmark across ten conceptual models and 467 CAMELS catchments and showed that models in which the evaporative flux varies linearly or non-linearly with soil-moisture storage generally preserved streamflow skill once AET was added to the objective; whereas models that equate evapotranspiration directly to its demand did not.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

Our results appear at first to contradict a body of work in which evapotranspiration-based calibration constrains discharge successfully. AET-only or streamflow-free calibration has recovered acceptable streamflow simulation in perennial Australian catchments (Jiang et al., 2020; Zhang et al., 2020), in Northern Africa (Taia et al., 2023) and, once the satellite product is bias-corrected, across a range of humid settings (Huang et al., 2020). These observations share a hydroclimatic setting, humid to sub-humid with sustained baseflow, in which the evaporative flux and the hydrograph are strongly coupled through a soil-moisture store that is rarely depleted. Across the semi-arid gradient examined here, streamflow is generated episodically by threshold exceedance under strong seasonal water limitation, conditions under which the slowly varying evaporative signal carries comparatively little information about discharge timing. Whether this reflects flow regime, aridity, or their covariation in this sample cannot be resolved with the present design. This interpretation is nonetheless broadly consistent with the reversed asymmetry reported by Do et al. (2024), where evapotranspiration was reported as the more transferable target in a largely humid sample. A more direct tension is reported in Wu et al. (2026a), who report that joint Q-AET calibration yields its largest streamflow improvements in arid catchments. Their mechanism also differs from the one tested in this study: improvements arose from correcting a winter-to-spring AET and soil-moisture overestimation under Q-only calibration and parameter regionalisation to nearby gauges, not from within-basin parameter identifiability under joint calibration.

Our conclusion that AET adds little parameter constraint applies to lumped, basin-averaged calibration against the magnitude of the flux. Where evapotranspiration is assimilated as a spatial pattern in distributed models, it constrains the internal distribution of states and can improve discharge (Rajib et al., 2018; Dembélé et al., 2020a; Kalura et al., 2025). This implies that the spatial dimension of satellite evapotranspiration carries information that its basin-mean magnitude does not while lumped multivariable calibration assesses only the latter. Therefore, two transferable propositions arise: firstly, the informational value of satellite evapotranspiration for discharge calibration appears contingent on the hydroclimatic setting examined, lowest here in a semi-arid, threshold-driven regime; and secondly, the cost of joint calibration is contingent on model structure.

These findings have direct implications for hydrological modelling practice in data-scarce watersheds. Where discharge observations are available, they should remain the primary calibration target. Satellite-derived actual evapotranspiration should not be treated as a substitute for streamflow observations, but rather as an independent diagnostic of simulated water-balance behaviour and as a constraint on physically coherent flux partitioning. Such coherence is only imperfectly achieved by single-variable calibration, often through compensating errors. Joint calibration is therefore recommended primarily for its contribution to hydrological plausibility.

Several limitations qualify these conclusions. Firstly, the analysis relies on a single evapotranspiration product; consequently, the poor performance of AET-only calibration cannot be fully disentangled from potential errors inherent in GLEAM (Jiang et al., 2020; Odusanya et al., 2021; Adeyeri and Ishola, 2021). Nevertheless, the parameter-space diagnostics, which show that runoff-generating parameters remain largely unconstrained, suggest that this limitation is fundamentally informational rather

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

than purely data-driven. Secondly, the primary Budyko reference, η_{obs}, is derived from GLEAM AET (the same product used as the AET-only and Q+AET calibration target) and the long-term evaluation window (1991-2020) overlaps the calibration period (1991-2010). We tested the former directly by repeating the $|d_{Fu}|$ and Budyko-space diagnostics against a closure-based reference, $\eta'_{obs} = 1 - Q_{obs}/P$, which involves no AET product and is therefore immune to this circularity. The calibration-scheme asymmetry persists under this independent anchor and, for HBV and IHACRES, is sharper than under the GLEAM anchor, indicating that it is not an artefact of the shared AET source. We agree that the temporal overlap between calibration and the long-term evaluation window is not removed by this check and remains a residual limitation of the plausibility assessment. More broadly, the conclusions drawn here apply to lumped-model calibration and should not be directly generalized to the spatial assimilation of evapotranspiration in distributed modelling frameworks (Immerzeel and Droogers, 2008; Jin and Jin, 2020; Dembélé et al., 2020a; Hulsman et al., 2021; Alemayehu et al., 2022; Guo et al., 2024b; Hsu et al., 2025).

These limitations point to three priorities for future research. Firstly, the use of an ensemble of independent evapotranspiration products such as WaPOR (Blatchford et al., 2020), SSEBop (Senay et al., 2013, 2020), MODIS (Mu et al., 2007) or the more recent VegET (Akpoti et al., 2026) would help separate structural limitations from product-specific data errors. Secondly, applications in spatially distributed modelling frameworks could test whether the spatial pattern of evapotranspiration provides discharge-relevant information where basin-integrated evapotranspiration magnitude does not. Lastly, controlled experiments that modify the evaporative formulation within a single model structure would allow the role of the identified coupling node to be tested more explicitly. Together, these extensions would build directly on the diagnostic framework developed in this study and further clarify the conditions under which evapotranspiration can provide useful constraints on hydrological model behaviour.

5 Conclusion

This study compared streamflow-only (Q-only), AET-only and joint (Q+AET) calibration schemes across three lumped hydrological models in 17 water-limited watersheds in West Africa, using satellite-derived AET as an additional calibration target and the Budyko-Fu framework to evaluate long-term water-balance consistency. The results demonstrate that the two variables provide fundamentally different information on the models. Discharge remains the most informative single calibration target. While AET-only calibration reproduces evapotranspiration satisfactorily, it fails to constrain the production, soil-moisture and loss-module parameters which influence streamflow generation. In contrast, discharge calibration constrains water partitioning sufficiently to recover plausible AET dynamics in two of the three model structures.

Joint calibration does not involve a trade-off between streamflow and evapotranspiration performance. Instead, it preserves the discharge skill achieved by Q-only calibration while matching or exceeding the AET performance of the AET-only scheme. More importantly, it yields a physically coherent long-term water balance, reducing the compensating errors that can arise

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

under single-variable calibration. The primary benefit of incorporating AET is therefore not improved streamflow prediction but improved hydrological realism and water-balance plausibility.

The effectiveness of multivariable calibration is strongly controlled by model structure. The magnitude of the trade-off between discharge and evapotranspiration depends on the degree to which runoff generation and evaporative processes share common parameters. Across the three models, this coupling ranged from weak (GR6J) to intermediate (HBV) and strong (IHACRES), and this structural gradient largely explained their contrasting responses to joint calibration.

These findings have practical implications for hydrological modelling in data-scarce water-limited watersheds. Streamflow should remain the primary calibration target where it is available, with satellite-derived AET serving as an independent check on water-balance plausibility rather than a substitute calibration target. Because the cost of combining the two targets is set by model structure, an a priori assessment of how a model's evaporative and runoff-generating pathways are coupled can indicate whether joint calibration is likely to improve physical realism without deteriorating discharge skill.

Code and data availability

The hydrometeorological data used in this study, including daily precipitation, potential evapotranspiration and streamflow records, were provided by the *Agence Nationale de la Météorologie* (ANAM), Burkina Faso. These data are proprietary and are subject to data-sharing restrictions imposed by the data provider. Consequently, they cannot be made publicly available by the authors. Researchers interested in accessing these data should contact ANAM directly and obtain permission from the agency. All code used for data processing, model calibration, performance assessment, uncertainty analysis and the generation of figures and tables presented in this study is publicly available in a GitHub repository: https://github.com/Yonaba/q_aet_constrained_calibration_sahel. Satellite-derived actual evapotranspiration data from the Global Land Evaporation Amsterdam Model (GLEAM) are publicly available from the GLEAM data portal (https://www.gleam.eu) and can be accessed following the instructions provided by the data distributor.

Author contributions

RY conceptualised the study, curated data, performed the formal analysis, developed the methodology and software, conducted investigation, acquired resources, validated results, produced all visualisations, and prepared the original draft. TF contributed to conceptualisation, formal analysis, methodology, and supervision, and reviewed and edited the manuscript. SS contributed to conceptualisation, data curation, formal analysis, investigation, software development, validation, and visualisation, and

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

reviewed and edited the manuscript. NT contributed to data curation, investigation, software, validation, and visualisation, and assisted with draft preparation. OG contributed to formal analysis, methodology, software, and visualisation, and reviewed and edited the manuscript. EN contributed to formal analysis, methodology, supervision, validation, and visualisation, and reviewed and edited the manuscript. AK contributed to formal analysis, methodology, resources, software, validation, and visualisation, and reviewed and edited the manuscript. AB contributed to conceptualisation, formal analysis, validation, and visualisation, and reviewed and edited the manuscript. HK contributed to formal analysis and validation and reviewed and edited the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements

The authors thank Posit Cloud and Google Cloud Platform for providing the computing infrastructure used to run the GLUE simulations underlying this study.

Financial support

This research has not received any external funding.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

References

Abate, B. Z., Assefa, T. T., Tigabu, T. B., Abebe, W. B., and He, L.: Hydrological Modeling of the Kobo-Golina River in the Data-Scarce Upper Danakil Basin, Ethiopia, Sustainability, 15, 3337, https://doi.org/10.3390/su15043337, 2023.

Abdallah, M., Zhang, K., Singh, S. K., Abdi, M. J., Elameen, A. M., Mohammed, A. A. A., and Elzain, H. E.: Hydrological insights: Comparative analysis of gridded potential evapotranspiration products for hydrological simulations and drought assessment, Journal of Hydrology: Regional Studies, 57, 102113, https://doi.org/10.1016/j.ejrh.2024.102113, 2025.

Adeyeri, O. E. and Ishola, K. A.: Variability and Trends of Actual Evapotranspiration over West Africa: The Role of Environmental Drivers., Agricultural and Forest Meteorology, 308–309, 108574, https://doi.org/10.1016/j.agrformet.2021.108574, 2021.

Akpoti, K., Velpuri, N. M., Leh, M., Kagone, S., Mekonnen, K., Owusu, A., Tadesse, M., Prabhath, P. T., Madushanka, L., Perera, T., Parrish, G. E. L., Nangia, V., Sy, S., Bliefernicht, J., Guug, S., Seid, A. H., and Senay, G. B.: VegET evapotranspiration for Africa: Continental-scale simulation, multi-product evaluation, and drought assessment, Journal of Hydrology: Regional Studies, 66, 103511, https://doi.org/10.1016/j.ejrh.2026.103511, 2026.

Alemayehu, T., Gupta, H. V., Van Griensven, A., and Bauwens, W.: On the Calibration of Spatially Distributed Hydrologic Models for Poorly Gauged Basins: Exploiting Information from Streamflow Signatures and Remote Sensing-Based Evapotranspiration Data, Water, 14, 1252, https://doi.org/10.3390/w14081252, 2022.

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., and others: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, Fao, Rome, 300, D05109, 1998.

Belemtougri, A. P., Ducharne, A., Tazen, F., Oudin, L., and Karambiri, H.: Understanding key factors controlling the duration of river flow intermittency: Case of Burkina Faso in West Africa, Journal of Hydrology: Regional Studies, 37, 100908, https://doi.org/10.1016/j.ejrh.2021.100908, 2021.

Bergström, S.: The HBV model–its structure and applications, SMHI, Norrkoping, Sweden, 32 pp., 1992.

Bergström, S. and Lindström, G.: Interpretation of runoff processes in hydrological modelling—experience from the HBV approach, Hydrological Processes, 29, 3535–3545, https://doi.org/10.1002/hyp.10510, 2015.

Beven, K.: A manifesto for the equifinality thesis, Journal of Hydrology, 320, 18–36, https://doi.org/10.1016/j.jhydrol.2005.07.007, 2006.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrological Processes, 6, 279–298, https://doi.org/10.1002/hyp.3360060305, 1992.

Beven, K. and Binley, A.: GLUE: 20 years on, Hydrological Processes, 28, 5897–5918, https://doi.org/10.1002/hyp.10082, 2014.

Blatchford, M. L., Mannaerts, C. M., Njuki, S. M., Nouri, H., Zeng, Y., Pelgrum, H., Wonink, S., and Karimi, P.: Evaluation of WaPOR V2 evapotranspiration products across Africa, Hydrological Processes, 34, 3200–3221, https://doi.org/10.1002/hyp.13791, 2020.

Budyko, M. I.: The heat balance of the earth's surface., U.S. Dept. of Commerce, Weather Bureau, Washington, 259 p. pp., 1958.

Chen, M., Willgoose, G. R., and Saco, P. M.: Evaluation of the hydrology of the IBIS land surface model in a semi-arid catchment, Hydrological Processes, 29, 653–670, https://doi.org/10.1002/hyp.10156, 2015.

Cheng, S., Hulsman, P., Koppa, A., Beck, H. E., Xia, J., Xu, J., Cheng, L., and Miralles, D. G.: Global Runoff Partitioning Based on Budyko-Constrained Machine Learning, Water Resources Research, 61, e2025WR039863, https://doi.org/10.1029/2025WR039863, 2025.

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., and Michel, C.: airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling, , 1.7.8, https://doi.org/10.32614/CRAN.package.airGR, 2017a.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V.: The suite of lumped GR hydrological models in an R package, Environmental Modelling & Software, 94, 166–171, https://doi.org/10.1016/j.envsoft.2017.05.002, 2017b.

Coulidiati, T. F. A., Biaou, A. C., Faye, M. D., Yonaba, R., Sauret, E. S. G., Compaoré, N. F., and Koïta, M.: Groundwater Vulnerability in the Kou Sub-Basin, Burkina Faso: A Critical Review of Hydrogeological Knowledge, Water, 17, 1317, https://doi.org/10.3390/w17091317, 2025.

Croke, B. F. W. and Jakeman, A. J.: A catchment moisture deficit module for the IHACRES rainfall-runoff model, Environmental Modelling & Software, 19, 1–5, https://doi.org/10.1016/j.envsoft.2003.09.001, 2004.

Croke, B. F. W. and Jakeman, A. J.: Corrigendum to “A Catchment Moisture Deficit module for the IHACRES rainfall-runoff model” [Environ. Model. Softw. 19 (1) (2004) 1–5], Environmental Modelling & Software, 20, 977, https://doi.org/10.1016/j.envsoft.2004.11.004, 2005.

Dangol, S., Zhang, X., Liang, X.-Z., Anderson, M., Crow, W., Lee, S., Moglen, G. E., and McCarty, G. W.: Multivariate Calibration of the SWAT Model Using Remotely Sensed Datasets, Remote Sensing, 15, 2417, https://doi.org/10.3390/rs15092417, 2023.

Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., and Schaefli, B.: Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets, Water Resources Research, 56, e2019WR026085, https://doi.org/10.1029/2019WR026085, 2020a.

Dembélé, M., Schaefli, B., Van De Giesen, N., and Mariéthoz, G.: Suitability of 17 gridded rainfall and temperature datasets for large-scale hydrological modelling in West Africa, Hydrol. Earth Syst. Sci., 24, 5379–5406, https://doi.org/10.5194/hess-24-5379-2020, 2020b.

Ding, J. and Zhu, Q.: The accuracy of multisource evapotranspiration products and their applicability in streamflow simulation over a large catchment of Southern China, Journal of Hydrology: Regional Studies, 41, 101092, https://doi.org/10.1016/j.ejrh.2022.101092, 2022.

Do, H. X., Nguyen, H. T. T., Tran, V. N., Le, M.-H., Nguyen, B. Q., Pham, H. T., Le, T. H., Van Binh, D., Dang, T. D., Tran, H., and Nguyen, T. V.: Uncertain Benefits of Using Remotely Sensed Evapotranspiration for Streamflow Estimation—Insights From a Randomized, Large-Sample Experiment, Water Resour Manage, 38, 3819–3835, https://doi.org/10.1007/s11269-024-03840-w, 2024.

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resources Research, 28, 1015–1031, https://doi.org/10.1029/91WR02985, 1992.

Fang, K., Kifer, D., Lawson, K., and Shen, C.: Evaluating the Potential and Challenges of an Uncertainty Quantification Method for Long Short-Term Memory Models for Soil Moisture Predictions, Water Resources Research, 56, e2020WR028095, https://doi.org/10.1029/2020WR028095, 2020.

Fu, B.: On the calculation of the evaporation from land surface, Scientia Atmospherica Sinica, 5, 23, 1981.

Gbohoui, Y. P., Paturel, J.-E., Fowe Tazen, Mounirou, L. A., Yonaba, R., Karambiri, H., and Yacouba, H.: Impacts of climate and environmental changes on water resources: A multi-scale study based on Nakanbé nested watersheds in West African Sahel, Journal of Hydrology: Regional Studies, 35, 100828, https://doi.org/10.1016/j.ejrh.2021.100828, 2021.

Gordon, B. L., Crow, W. T., Konings, A. G., Dralle, D. N., and Harpold, A. A.: Can We Use the Water Budget to Infer Upland Catchment Behavior? The Role of Data Set Error Estimation and Interbasin Groundwater Flow, Water Resources Research, 58, e2021WR030966, https://doi.org/10.1029/2021WR030966, 2022.

Goudiaby, O., Bodian, A., Dezetter, A., Ogilvie, A., Diouf, I., and Ndiaye, P. M.: Hydrological evaluation of gridded rainfall products for streamflow simulation in West Africa, Journal of Hydrology: Regional Studies, 62, 102984, https://doi.org/10.1016/j.ejrh.2025.102984, 2025.

Guo, N., Chen, H., Han, Q., and Wang, T.: Evaluating data-driven and hybrid modeling of terrestrial actual evapotranspiration based on an automatic machine learning approach, Journal of Hydrology, 628, 130594, https://doi.org/10.1016/j.jhydrol.2023.130594, 2024a.

Guo, X., Wu, Z., Fu, G., and He, H.: A multi-variable calibration framework at the grid scale for integrating streamflow with evapotranspiration data to improve the simulation of distributed hydrological model, Journal of Hydrology: Regional Studies, 55, 101944, https://doi.org/10.1016/j.ejrh.2024.101944, 2024b.

Han, J., Yang, Y., Roderick, M. L., McVicar, T. R., Yang, D., Zhang, S., and Beck, H. E.: Assessing the Steady-State Assumption in Water Balance Calculation Across Global Catchments, Water Resources Research, 56, e2020WR027392, https://doi.org/10.1029/2020WR027392, 2020.

Herman, M. R., Hernandez-Suarez, J. S., Nejadhashemi, A. P., Kropp, I., and Sadeghi, A. M.: Evaluation of Multi- and Many-Objective Optimization Techniques to Improve the Performance of a Hydrologic Model Using Evapotranspiration Remote-Sensing Data, J. Hydrol. Eng., 25, 04020006, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001896, 2020.

Hsu, S.-C., De Lavenne, A., Perrin, C., and Andréassian, V.: Extra constraint on actual evaporation in a semi-distributed conceptual model to improve model physical realism, Hydrological Sciences Journal, 70, 1143–1156, https://doi.org/10.1080/02626667.2025.2468846, 2025.

Huang, Q., Qin, G., Zhang, Y., Tang, Q., Liu, C., Xia, J., Chiew, F. H. S., and Post, D.: Using Remote Sensing Data-Based Hydrological Model Calibrations for Predicting Runoff in Ungauged or Poorly Gauged Catchments, Water Resources Research, 56, e2020WR028205, https://doi.org/10.1029/2020WR028205, 2020.

Huang, X., Wang, K.-R., Zou, Y., and Cao, X.-C.: Development of global soil erosion research at the watershed scale: a bibliometric analysis of the past decade, Environmental Science and Pollution Research, 28, 12232–12244, https://doi.org/10.1007/s11356-020-11888-5, 2021.

Hui, J., Wu, Y., Zhao, F., Lei, X., Sun, P., Singh, S. K., Liao, W., Qiu, L., and Li, J.: Parameter Optimization for Uncertainty Reduction and Simulation Improvement of Hydrological Modeling, Remote Sensing, 12, 4069, https://doi.org/10.3390/rs12244069, 2020.

Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Learning from satellite observations: increased understanding of catchment processes through stepwise model improvement, Hydrol. Earth Syst. Sci., 25, 957–982, https://doi.org/10.5194/hess-25-957-2021, 2021.

Ibrahim, M., Coenders-Gerrits, M., Van Der Ent, R., and Hrachowitz, M.: Catchments do not strictly follow Budyko curves over multiple decades, but deviations are minor and predictable, Hydrol. Earth Syst. Sci., 29, 1703–1723, https://doi.org/10.5194/hess-29-1703-2025, 2025.

Immerzeel, W. W. and Droogers, P.: Calibration of a distributed hydrological model based on satellite evapotranspiration, Journal of Hydrology, 349, 411–424, https://doi.org/10.1016/j.jhydrol.2007.11.017, 2008.

Jiang, L., Wu, H., Tao, J., Kimball, J. S., Alfieri, L., and Chen, X.: Satellite-Based Evapotranspiration in Hydrological Model Calibration, Remote Sensing, 12, 428, https://doi.org/10.3390/rs12030428, 2020.

Jin, X. and Jin, Y.: Calibration of a Distributed Hydrological Model in a Data-Scarce Basin Based on GLEAM Datasets, Water, 12, 897, https://doi.org/10.3390/w12030897, 2020.

Jung, H. C., Getirana, A., Arsenault, K. R., Holmes, T. R. H., and McNally, A.: Uncertainties in Evapotranspiration Estimates over West Africa, Remote Sensing, 11, 892, https://doi.org/10.3390/rs11080892, 2019.

Kafando, M. B., Koïta, M., Zouré, C. O., Yonaba, R., and Niang, D.: Quantification of Soil Deep Drainage and Aquifer Recharge Dynamics according to Land Use and Land Cover in the Basement Zone of Burkina Faso in West Africa, Sustainability, 14, 14687, https://doi.org/10.3390/su142214687, 2022.

Kalura, P., Pandey, A., Chowdary, V. M., and Dayal, D.: A TOPSIS -Based Multicriteria Assessment of Hydrologic Model Calibration Using Satellite-Derived Evapotranspiration and Streamflow Data, Hydrological Processes, 39, e70191, https://doi.org/10.1002/hyp.70191, 2025.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal of Hydrology, 424–425, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrology and Earth System Sciences, 23, 4323–4331, https://doi.org/10.5194/hess-23-4323-2019, 2019.

Kodja, D. J., Akognongbé, A. J. S., Amoussou, E., Mahé, G., Vissin, E. W., Paturel, J.-E., and Houndénou, C.: Calibration of the hydrological model GR4J from potential evapotranspiration estimates by the Penman-Monteith and Oudin methods in the Ouémé watershed (West Africa), in: Hydrological processes and water security in a changing world, 163–169, https://doi.org/10.5194/piahs-383-163-2020, 2020.

Koltsida, E. and Kallioras, A.: Multi-Variable SWAT Model Calibration Using Satellite-Based Evapotranspiration Data and Streamflow, Hydrology, 9, 112, https://doi.org/10.3390/hydrology9070112, 2022.

Koo, H., Chen, M., Jakeman, A. J., and Zhang, F.: A global sensitivity analysis approach for identifying critical sources of uncertainty in non-identifiable, spatially distributed environmental models: A holistic analysis applied to SWAT for input datasets and model parameters, Environmental Modelling & Software, 127, 104676, https://doi.org/10.1016/j.envsoft.2020.104676, 2020.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

Koppa, A., Alam, S., Miralles, D. G., and Gebremichael, M.: Budyko-Based Long-Term Water and Energy Balance Closure in Global Watersheds From Earth Observations, Water Resources Research, 57, e2020WR028658, https://doi.org/10.1029/2020WR028658, 2021.

Kumar, S. B. and Mishra, A.: Enhancing the Stability of Hydrological Modelling through Multivariable Calibration Schemes Using the Satellite-Based Soil Moisture and Evapotranspiration, Water Resour Manage, 39, 3213–3234, https://doi.org/10.1007/s11269-025-04104-x, 2025.

Lehner, B., Verdin, K., and Jarvis, A.: New Global Hydrography Derived From Spaceborne Elevation Data, EoS Transactions, 89, 93–94, https://doi.org/10.1029/2008EO100001, 2008.

Mbaye, M. L., Sy, K., Faty, B., and Sall, S. M.: Impact of 1.5 and 2.0 °C global warming on the hydrology of the Faleme river basin, Journal of Hydrology: Regional Studies, 31, 100719, https://doi.org/10.1016/j.ejrh.2020.100719, 2020.

Mei, Y., Mai, J., Do, H. X., Gronewold, A., Reeves, H., Eberts, S., Niswonger, R., Regan, R. S., and Hunt, R. J.: Can Hydrological Models Benefit From Using Global Soil Moisture, Evapotranspiration, and Runoff Products as Calibration Targets?, Water Resources Research, 59, e2022WR032064, https://doi.org/10.1029/2022WR032064, 2023.

Miralles, D. G., Bonte, O., Koppa, A., Baez-Villanueva, O. M., Tronquo, E., Zhong, F., Beck, H. E., Hulsman, P., Dorigo, W., Verhoest, N. E. C., and Haghdoost, S.: GLEAM4: global land evaporation and soil moisture dataset at 0.1° resolution from 1980 to near present, Sci Data, 12, 416, https://doi.org/10.1038/s41597-025-04610-y, 2025.

Mohamed, A., Werner, M., and Van Der Zaag, P.: Beyond streamflow: Plausible hydrological modelling for the Upper Blue Nile Basin, Ethiopia, Journal of Hydrology: Regional Studies, 59, 102290, https://doi.org/10.1016/j.ejrh.2025.102290, 2025.

Moriasi, D. N., Arnold, J. G., Liew, M. W. V., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, Transactions of the ASABE, 50, 885–900, https://doi.org/10.13031/2013.23153, 2007.

Moriasi, D. N., Gitau, M. W., Pai, N., and Daggupati, P.: Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria, Transactions of the ASABE, 58, 1763–1785, https://doi.org/10.13031/trans.58.10715, 2015.

Mostafaie, A., Forootan, E., Safari, A., and Schumacher, M.: Comparing multi-objective optimization techniques to calibrate a conceptual hydrological model using in situ runoff and daily GRACE data, Comput Geosci, 22, 789–814, https://doi.org/10.1007/s10596-018-9726-8, 2018.

Mounirou, L. A., Zouré, C. O., Yonaba, R., Paturel, J.-E., Mahé, G., Niang, D., Yacouba, H., and Karambiri, H.: Multi-scale analysis of runoff from a statistical perspective in a small Sahelian catchment under semi-arid climate, Arabian Journal of Geosciences, 13, https://doi.org/10.1007/s12517-020-5141-2, 2020.

Mounirou, L. A., Yonaba, R., Tazen, F., Ayele, G. T., Yaseen, Z. M., Karambiri, H., and Yacouba, H.: Soil Erosion across Scales: Assessing Its Sources of Variation in Sahelian Landscapes under Semi-Arid Climate, Land, 11, 2302, https://doi.org/10.3390/land11122302, 2022.

Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sensing of Environment, 111, 519–536, https://doi.org/10.1016/j.rse.2007.04.015, 2007.

Ndiaye, P. M., Bodian, A., Dezetter, A., Ogilvie, A., and Goudiaby, O.: Sensitivity of global hydrological models to potential evapotranspiration estimation methods in the Senegal River Basin (West Africa), Journal of Hydrology: Regional Studies, 53, 101823, https://doi.org/10.1016/j.ejrh.2024.101823, 2024.

Ndiaye, P. M., Ogilvie, A., Bodian, A., Descroix, L., Legay, T., and Guillet, R.: Hydrological variability of large rivers in West Africa: gap-filling with Earth observations and daily rainfall-runoff modelling, Hydrological Sciences Journal, 1–19, https://doi.org/10.1080/02626667.2025.2542477, 2025.

Odusanya, A. E., Schulz, K., Biao, E. I., Degan, B. A. S., and Mehdi-Schulz, B.: Evaluating the performance of streamflow simulated by an eco-hydrological model calibrated and validated with global land surface actual evapotranspiration from remote sensing at a catchment scale in West Africa, Journal of Hydrology: Regional Studies, 37, 100893, https://doi.org/10.1016/j.ejrh.2021.100893, 2021.

Pan, S., Xu, Y.-P., Gu, H., Yu, B., and Xuan, W.: Evaluation of Remote Sensing-Based Evapotranspiration Datasets for Improving Hydrological Model Simulation in Humid Region of East China, Remote Sensing, 14, 4546, https://doi.org/10.3390/rs14184546, 2022.

Park, J., Sung, J. H., Seo, Y.-H., and Kim, B.-S.: Applicability of the daily hydrological models (GR4J, GR5J, and GR6J) in the South Korean basins, Sci Rep, 15, 31912, https://doi.org/10.1038/s41598-025-16429-z, 2025.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275–289, https://doi.org/10.1016/S0022-1694(03)00225-7, 2003.

Pianosi, F. and Wagener, T.: A simple and efficient method for global sensitivity analysis based on cumulative distribution functions, Environmental Modelling & Software, 67, 1–11, https://doi.org/10.1016/j.envsoft.2015.01.004, 2015.

Poméon, T., Diekkrüger, B., and Kumar, R.: Computationally Efficient Multivariate Calibration and Validation of a Grid-Based Hydrologic Model in Sparsely Gauged West African River Basins, Water, 10, 1418, https://doi.org/10.3390/w10101418, 2018.

Pool, S., Fowler, K., and Peel, M.: Benefit of Multivariate Model Calibration for Different Climatic Regions, Water Resources Research, 60, e2023WR036364, https://doi.org/10.1029/2023WR036364, 2024.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V.: A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, Journal of Hydrology, 411, 66–76, https://doi.org/10.1016/j.jhydrol.2011.09.034, 2011.

Rajib, A., Evenson, G. R., Golden, H. E., and Lane, C. R.: Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters, Journal of Hydrology, 567, 668–683, https://doi.org/10.1016/j.jhydrol.2018.10.024, 2018.

Rane, N. L. and Jayaraj, G. K.: Enhancing SWAT model predictivity using multi-objective calibration: effects of integrating remotely sensed evapotranspiration and leaf area index, Int. J. Environ. Sci. Technol., 20, 6449–6468, https://doi.org/10.1007/s13762-022-04293-7, 2023.

Rientjes, T. H. M., Muthuwatta, L. P., Bos, M. G., Booij, M. J., and Bhatti, H. A.: Multi-variable calibration of a semi-distributed hydrological model using streamflow data and satellite-based evapotranspiration, Journal of Hydrology, 505, 276–290, https://doi.org/10.1016/j.jhydrol.2013.10.006, 2013.

Ruelland, D., Ardoin-Bardin, S., Collet, L., and Roucou, P.: Simulating future trends in hydrological regime of a large Sudano-Sahelian catchment under climate change, Journal of Hydrology, 424–425, 207–216, https://doi.org/10.1016/j.jhydrol.2012.01.002, 2012.

https://doi.org/10.5194/egusphere-2026-3981
Preprint. Discussion started: 22 July 2026

Santos, L., Thirel, G., and Perrin, C.: Technical note: Pitfalls in using log-transformed flows within the KGE criterion, Hydrol. Earth Syst. Sci., 22, 4583–4591, https://doi.org/10.5194/hess-22-4583-2018, 2018.

Senay, G. B., Bohms, S., Singh, R. K., Gowda, P. H., Velpuri, N. M., Alemu, H., and Verdin, J. P.: Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach, J American Water Resour Assoc, 49, 577–591, https://doi.org/10.1111/jawr.12057, 2013.

Senay, G. B., Kagone, S., and Velpuri, N. M.: Operational Global Actual Evapotranspiration: Development, Evaluation, and Dissemination, Sensors, 20, 1915, https://doi.org/10.3390/s20071915, 2020.

Shah, S., Duan, Z., Song, X., Li, R., Mao, H., Liu, J., Ma, T., and Wang, M.: Evaluating the added value of multi-variable calibration of SWAT with remotely sensed evapotranspiration data for improving hydrological modeling, Journal of Hydrology, 603, 127046, https://doi.org/10.1016/j.jhydrol.2021.127046, 2021.

Taia, S., Scozzari, A., Erraioui, L., Kili, M., Mridekh, A., Haida, S., Chao, J., and El Mansouri, B.: Comparing the ability of different remotely sensed evapotranspiration products in enhancing hydrological model performance and reducing prediction uncertainty, Ecological Informatics, 78, 102352, https://doi.org/10.1016/j.ecoinf.2023.102352, 2023.

Thorndahl, S., Beven, K. J., Jensen, J. B., and Schaarup-Jensen, K.: Event based uncertainty assessment in urban drainage modelling, applying the GLUE methodology, Journal of Hydrology, 357, 421–437, https://doi.org/10.1016/j.jhydrol.2008.05.027, 2008.

Toum, E.: HBV.IANIGLA: Modular Hydrological Model, , 0.2.6, https://doi.org/10.32614/CRAN.package.HBV.IANIGLA, 2019.

Trotter, L., Saft, M., Peel, M. C., and Fowler, K. J. A.: Symptoms of Performance Degradation During Multi-Annual Drought: A Large-Sample, Multi-Model Study, Water Resources Research, 59, e2021WR031845, https://doi.org/10.1029/2021WR031845, 2023.

Wu, S., Yan, M., and Zhao, J.: Efficacy of evapotranspiration products on improving streamflow prediction in ungauged catchments under various hydrological models and climates, Journal of Hydrology, 677, 135857, https://doi.org/10.1016/j.jhydrol.2026.135857, 2026a.

Wu, S., Yang, Y., Li, C., Yang, W., and Zhao, J.: Evaluating Evapotranspiration Simulation Performance in 30 Conceptual Hydrological Models: Insights Into ET Representation Across Diverse Climates, Water Resources Research, 62, e2025WR040017, https://doi.org/10.1029/2025WR040017, 2026b.

Xu, T., Xiao, S., Reuschen, S., Wildt, N., Hendricks Franssen, H.-J., and Nowak, W.: Towards a community-wide effort for benchmarking in subsurface hydrological inversion: benchmarking cases, high-fidelity reference solutions, procedure, and first comparison, Hydrol. Earth Syst. Sci., 28, 5375–5400, https://doi.org/10.5194/hess-28-5375-2024, 2024.

Yonaba, R., Biaou, A. C., Koïta, M., Tazen, F., Mounirou, L. A., Zouré, C. O., Queloz, P., Karambiri, H., and Yacouba, H.: A dynamic land use/land cover input helps in picturing the Sahelian paradox: Assessing variability and attribution of changes in surface runoff in a Sahelian watershed, Science of The Total Environment, 757, 143792, https://doi.org/10.1016/j.scitotenv.2020.143792, 2021a.

Yonaba, R., Koïta, M., Mounirou, L. A., Tazen, F., Queloz, P., Biaou, A. C., Niang, D., Zouré, C., Karambiri, H., and Yacouba, H.: Spatial and transient modelling of land use/land cover (LULC) dynamics in a Sahelian landscape under semi-arid climate in northern Burkina Faso, Land Use Policy, 103, 105305, https://doi.org/10.1016/j.landusepol.2021.105305, 2021b.

Yonaba, R., Tazen, F., Cissé, M., Mounirou, L. A., Belemtougri, A., Ouedraogo, V. A., Koïta, M., Niang, D., Karambiri, H., and Yacouba, H.: Trends, sensitivity and estimation of daily reference evapotranspiration ET0 using limited climate data: regional focus on Burkina Faso in the West African Sahel, Theoretical and Applied Climatology, 153, 947–974, https://doi.org/10.1007/s00704-023-04507-z, 2023.

Yonaba, R., Belemtougri, A., Fowé, T., Mounirou, L. A., Nkiaka, E., Dembélé, M., Komlavi, A., Coly, S. M., Koïta, M., and Karambiri, H.: Rainfall estimation in the West African Sahel: comparison and cross-validation of top-down *vs.* bottom-up precipitation products in Burkina Faso, Geocarto International, 39, 2391956, https://doi.org/10.1080/10106049.2024.2391956, 2024.

Yonaba, R., Kiema, A., Tazen, F., Belemtougri, A., Cissé, M., Mounirou, L. A., Bodian, A., Koïta, M., and Karambiri, H.: Accuracy and interpretability of machine learning-based approaches for daily ETo estimation under semi-arid climate in the West African Sahel, Earth Sci Inform, 18, 87, https://doi.org/10.1007/s12145-024-01591-1, 2025a.

Yonaba, R., Fowé, T., Ngonga, G., Belemtougri, A., Kiema, A., Tangam, I. S., and Karambiri, H.: Hydrological evaluation of top-down and bottom-up rainfall products in West Africa: Model performance, parameter range and uncertainty propagation, Journal of Hydrology: Regional Studies, 62, 102957, https://doi.org/10.1016/j.ejrh.2025.102957, 2025b.

Yoshida, T., Hanasaki, N., Nishina, K., Boulange, J., Okada, M., and Troch, P. A.: Inference of Parameters for a Global Hydrological Model: Identifiability and Predictive Uncertainties of Climate-Based Parameters, Water Resources Research, 58, e2021WR030660, https://doi.org/10.1029/2021WR030660, 2022.

Zhang, J., Liu, Z., Li, Y., Dou, Y., Wang, M., Zhou, H., and Xu, B.: Enhancing hydrological model performance through multi-source open-data utilization in the highly managed, data-scarce basin, Journal of Hydrology, 654, 132860, https://doi.org/10.1016/j.jhydrol.2025.132860, 2025.

Zhang, Y., Chiew, F. H. S., Liu, C., Tang, Q., Xia, J., Tian, J., Kong, D., and Li, C.: Can Remotely Sensed Actual Evapotranspiration Facilitate Hydrological Prediction in Ungauged Regions Without Runoff Calibration?, Water Resources Research, 56, e2019WR026236, https://doi.org/10.1029/2019WR026236, 2020.

Zhang, Y., Huang, C., Zhang, W., Chen, J., and Wang, L.: The concept, approach, and future research of hydrological connectivity and its assessment at multiscales, Environmental Science and Pollution Research, 28, 52724–52743, https://doi.org/10.1007/s11356-021-16148-8, 2021.

Zouré, C. O., Kiema, A., Yonaba, R., and Minoungou, B.: Unravelling the Impacts of Climate Variability on Surface Runoff in the Mouhoun River Catchment (West Africa), Land, 12, 2017, https://doi.org/10.3390/land12112017, 2023.