https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Hydrological Auditing of LISFLOOD v4.1.1: Impacts of Model Setup on Water Balance Components in the Po River Basin

Francesca Moschini[1,2], Andrea Ficchí[3], and Alberto Pistocchi[1]

[1]European Commission, Joint Research Centre (JRC), Ispra, Italy
[2]Rey Juan Carlos University, Madrid, Spain
[3]Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Italy

Correspondence: Francesca Moschini (francesca.moschini@ec.europa.eu)

Abstract. In recent years, large-scale hydrological models have been increasingly used at regional and global scales to support decision making. Their realism in simulating water balance components is crucial for building trust across different use cases. Hydrological models may reproduce streamflow well but misrepresent other fluxes, due to internal fluxes compensations and equifinality. Therefore, alternative setups can benefit specific applications by improving the representation of relevant water balance components. "Hydrological auditing" of models, i.e. a thorough critical review of their realism beyond the calibration targets (usually streamflow), provides useful insights for both practical applications and process understanding. We present one such exercise in a representative European case study using a physically-based hydrological model (LISFLOOD), widely used for flood forecasting and water resources management. We evaluate LISFLOOD v4.1.1's performance in simulating streamflow, evapotranspiration, and overall water balance in the Po River Basin, a complex and highly managed basin in Northern Italy. Six alternative model setups are tested, including different soil layers depths and preferential flow representations. Results show that the model setup currently used in the European Flood Awareness System (EFAS) v.5 performs best in terms of streamflow simulation, particularly at the daily time step, but tends to underestimate evapotranspiration. In turn, this may lead to an overestimation of groundwater recharge and a poor water balance representation. The use of the Budyko framework as a diagnostic tool reveals that model setups without preferential flow better match the expected long-term water balance, but reduce daily streamflow performance. The study highlights the importance of evaluating model performance and auditing alternative parametrizations to ensure accurate simulations of water balance components, crucial for water resources management. We propose criteria to improve the calibration of the LISFLOOD model in a flexible and target-driven way, to better support water resources management in complex river basins.

1 Introduction

The importance of a sound quantification of hydrological variables for water resources management cannot be understated. With surging demand for adaptation to climate change and a water-resilient economy, hydrological models are increasingly called upon to support the analysis of scenarios and to inform decision making from local to regional scales (Kumar et al., 2025). They are essential tools to identify cost-effective, no-regret and, when possible, multifunctional measures enabling a rational and fair use of scarce resources, while supporting resilience and sustainable development goals (e.g., Granata and

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Di Nunno, 2025; Quaranta et al., 2021). Various sectoral applications, however, require an accurate representation of different hydrological variables. For example, soil moisture is key for agricultural drought monitoring and irrigation management (e.g., Zhang et al., 2025), evapotranspiration is central to vegetation and ecosystem monitoring (e.g., Fluhrer et al., 2025), groundwater storage underpins water supply planning (e.g., Abbas et al., 2025), and streamflow is essential for flood forecasting and early warning (e.g., Nearing et al., 2024), as well as hydrological drought monitoring and management (e.g., Cammalleri et al., 2017). Given the diverse application and challenges in the management of water resources, different sectors require precise estimation of specific key target variables, whose quantification should be tailored to their needs. Therefore, hydrological models should be either designed to quantify ad-hoc variables with an accuracy that aligns with the user requirements, or, in the case of process-based (semi-)distributed models, calibrated on variables other than those directly relevant for decision making, to ensure that these variables are represented in a physically consistent manner and with sufficient accuracy for the intended use case and spatial context. One prominent example of a hydrological model in use for different operational applications in the European Union, especially in large trans-boundary basins, is the open-source OS-LISFLOOD (Van Der Knijff et al., 2010). LISFLOOD is a physically-based, spatially-distributed model developed by the European Commission's Joint Research Centre (JRC) and originally used for flood forecasting at European (Matthews et al., 2025b) and global scale (Matthews et al., 2025a), but later employed also in drought monitoring (Cammalleri et al., 2015, 2017) and climate change impact analysis focusing on the appraisal of water resources management measures (Bisselink et al., 2020; De Roo et al., 2021, 2023). The model is presently calibrated on the basis of streamflow observations only, as is common practice for many complex and computationally intensive large-scale hydrological models. As the needs for applications beyond the original model scope (flood forecasting) grow, it is essential to understand the extent to which the predictions for all the variables represented by the model can be trusted.

Some hydrological modelling studies have attempted to incorporate additional variables into model validation, such as evapotranspiration and soil moisture (Orth and Seneviratne, 2015), groundwater levels (Pelletier and Andréassian, 2022) and terrestrial water storage (Jensen et al., 2025). However, challenges remain even in ensuring consistency of fluxes across spatial and temporal scales, while respecting simple mass-conservation equations (Kumar et al., 2013; Samaniego et al., 2017; Ficchì et al., 2019), and observed spatio-temporal dynamics (Rajib et al., 2018; Kraft et al., 2022).

Multi-variable calibration can improve model realism but is rarely performed so far, mostly in local or regional studies (Döll et al., 2024; Guo et al., 2024) also due to data challenges, and may introduce trade-offs (Döll et al., 2024; Pelletier and Andréassian, 2022) or parameter identifiability issues (Döll et al., 2024), especially when different data sources present different error structures. Linked to this, also the sensitivity of hydrological model performances and parameters to imperfect knowledge of water fluxes and inputs, like potential evapotranspiration (Andréassian et al., 2004) and groundwater fluxes (Gleeson et al., 2021) has been rarely studied. Despite some progress over the last few decades and increased data availability especially from satellite (Huang et al., 2025), there is still limited understanding of how model parameterizations affect the internal fluxes and processes representation in large-scale models, and systematic evaluations across multiple water fluxes and storages are needed.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Such evaluations would be essential in moving towards a diagnostic approach for model evaluation that relies on hydrological theory and process understanding to support the detection of the causes of performance limitations and the resolution of model inadequacies (Yilmaz et al., 2008). This approach remains less systematic and mature than calibration practices, partly due to a lack of diagnostic evaluation criteria and structured procedures that can trace errors to specific processes or subsystems within a model. Such limitations hinder model improvements aligned with new application needs, especially when predictions beyond streamflow are required. A systematic understanding of how internal processes and water balance components behave under different model configuration of the same model and among different models is still lacking, including in LISFLOOD. This understanding becomes essential when the model is used beyond flood forecasting and addressing such a diagnostic gap is the central motivation of the present work. This contribution examines the performance of the current LISFLOOD v4.1.1 model operational setup, as incorporated in EFAS v.5 (see https://confluence.ecmwf.int/display/CEMS/EFAS+v5.0), in terms of the overall water balance behaviour, besides the model's capability to reproduce observed streamflow. Using the Po River Basin as a test bed, we explore how the model predictions change when excluding preferential flow, a model component of LISFLOOD, and when changing the representation of soils. By comparing the EFAS v.5 model parametrization with a set of alternative parametrizations, we explore how these affect the model's ability to quantify hydrological variables relevant to decision making, namely evapotranspiration, runoff and infiltration. Based on our analysis, we suggest guiding criteria for the calibration of the model in order to improve simulations beyond the scope of flood or drought forecasting.

2 Materials and methods

2.1 The LISFLOOD model

LISFLOOD is an open-source hydrological and flood simulation model developed by the European Commission's Joint Research Centre (JRC) (Burek et al., 2013; Van Der Knijff et al., 2010). As a distributed, process-based hydrological model, it can be used to simulate the entire hydrological cycle, including rainfall-runoff processes, river routing, floodplain inundation processes, and human influence on the water system, such as water withdrawals and reservoirs. The model is currently used in two of the operational components of the Copernicus Emergency Management Service (CEMS), as it is run within the European and Global Flood Awareness Systems, i.e., EFAS [https://european-flood.emergency.copernicus.eu/] and GloFAS [https://global-flood.emergency.copernicus.eu/] for flood forecasting, and provides variables for drought monitoring in the European and Global Drought Observatories (EDO and GDO) [https://drought.emergency.copernicus.eu]. In addition, it has been used for various studies informing water-related policies, for example in evaluating nature-based solutions for water retention, as well as water savings and nutrient reduction measures on water availability and quality at the European level (Burek et al., 2012; De Roo et al., 2012). The EFAS v.5.0 setup covers the European continent at a spatial resolution of about 1' (~1.5 km) and and a temporal resolution of 6 hours. Input data can be found in the JRC data catalogue (https://data.jrc.ec.europa.eu/), and model results are distributed through the CEMS Early Warning Data Store (https://ewds.climate.copernicus.eu/). The model incorporates several modules representing all key hydrological processes, including a snowpack balance routine (based on the degree-day method), modules for interception of rainfall, evapotranspiration and water uptake by vegetation, a soil water

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

balance component describing three soil layers (from the root zone to the water table), a saturation excess mechanism for rainfall-runoff transformation and a preferential flow mechanism for aquifer recharge (bypassing the soil layers). The latter, combined with percolation from the soil, feeds an upper groundwater compartment represented as a linear reservoir, in turn feeding a lower groundwater compartment, and both contribute to streamflow with lateral flow. The model includes groundwater losses to correct for excess recharge of the lower groundwater compartment that should be in effect balanced by abstractions or transfers within aquifers. The saturation excess module is based on the conceptual Xinjiang-VIC-Arno model (Todini, 1996), while preferential flow is described with a non-linear reservoir equation as a function of soil water content. The saturation excess and the contribution of groundwater are routed through the stream network using a kinematic wave approximation of the St.Venant equations. Lakes and reservoirs are also modelled in LISFLOOD, where regulation rules can be accounted for to simulate their impact on streamflow and water balance. In the EFAS setup, the model is currently calibrated so that the outputs match observed streamflow at a number of gauging stations across Europe by optimising an objective function, which measures the goodness of fit between model simulations and observations. The chosen goodness-of-fit criterion to optimise is the modified Kling-Gupta Efficiency (KGE) (Kling et al., 2012), commonly adopted for the calibration of hydrological models (e.g., Melsen et al. (2025)). The 14 model parameters that undergo calibration are summarized in Table A1 in Appendix A.

2.2 The Po River Basin test bed

We focus our analysis on a part of the European region covered by the EFAS setup, namely the Po River Basin in Northern Italy (Figure A1, Appendix A), encompassing a total area of approximately 74 000 km^2. This basin combines a relatively limited extent (compared to the European domain) with a high variability of hydrological conditions, with elevations ranging from 4000 meters in the Alps to the sea level at the delta, and a landscape shaped by the interplay of geological processes, such as tectonic activities, erosion and deposition (Livani et al., 2023). This landscape hosts diverse habitats, including wetlands, floodplains, and riparian forests, and plays a significant role in the biodiversity of the region, supporting various species of fishes, birds, and other wildlife. Most of the region has a mild-continental climate, with annual precipitation ranging between 750 mm and 1200 mm and falling mainly in spring and autumn, and average temperature from 5 to 15 degrees Celsius (Vezzoli et al., 2015). The area is interesting not only for the complexity of the landscape, but also for human interactions with the water cycle: the basin generates 40 % of the Italian GPD (Musolino et al., 2018), while representing 23 % of Italian territory. Agriculture is the dominant sector in terms of water demand (16 Mm^3, 80 % of total demand) followed by industry (24 %), municipal use (14 %) and energy production (8 %). Historically water-rich, in the past decades the region has witnessed an increase in water demand contrasting with lower water availability, due to more severe droughts and long term fluctuations in streamflow (Montanari, 2012). The context underscores the importance of a correct representation of all phases of the water cycle and water storage components for decision support across sectors. Our simulations are performed with the EFAS v.5 setup of the LISFLOOD model, over the period 1990-2021. Within the river basin, we considered 60 streamflow gauging stations for both calibration and evaluation of the model (Figure A1, Appendix A). Each station has recorded data covering between 5.5 and 23 years of daily discharge data.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

2.3 Model representation of water balance components - soil, groundwater and actual evapotranspiration

This contribution focuses on the water balance components in the soil and aquifer compartments and on the representation of actual evapotranspiration (AET) in LISFLOOD. These water fluxes are empirically known to have a strong influence on other hydrological variables and processes, and model inadequacies in one of them potentially lead to spurious internal compensations with other fluxes (Ficchì et al., 2019; Le Moine et al., 2007) to close the water balance (Beven, 2001). The soil, groundwater and evaporation fluxes are difficult to measure, but bear a practical importance when it comes to water resources management (e.g., for municipal and agricultural uses). Hence, it is important to accurately represent these hydrological processes and select the respective parameters appropriately. In LISFLOOD, water that reaches the soil has two ways of moving through the soil matrix and its three layers (Figure 1). It can either infiltrate into the unsaturated soil zone, where evapotranspiration takes place, or drain directly to the groundwater, bypassing the unsaturated zone. The first process, based on the conceptual Xinjiang-VIC-Arno model (eq 5) (Todini, 1996), is regulated by the empirical, calibrated parameter (bInfilt), which is used to approximate the saturated fraction of a model cell. The saturated fraction is then used to calculate the amount of water that becomes direct runoff (eq 4), representing the portion of precipitation that does not infiltrate into the soil and instead flows directly into the stream network. The preferential bypass flow (ByPass) mechanism computes the amount of water that drains directly to the groundwater zone, given the empirical parameter (PowerPrefFlow) and the relative saturation of the soil at a given timestep (eq 1). The parameter PowerPrefFlow controls the rate of preferential flow, being the exponent of a power law of the relative saturation of the first two soil layers. Equations are presented here, in the same order as they are compiled in the model (see https://ec-jrc.github.io/lisflood-model/ for further details).

First, the preferential flow is calculated from the relative saturation (RelSat) of the first two soil layers, and the Available Water for Infiltration (AWI), which represent the precipitation that reaches the soil surface, as:

$$PrefFlow = (RelSat^{PowerPrefFlow}) \cdot AWI \tag{1}$$

$$AWI = AWI - PrefFlow \tag{2}$$

Where PowerPrefFlow is a calibrated model parameter. The remaining water (eq 2) is then partitioned between direct runoff(eq 4) and water seeping into the unsaturated zone as Infiltration (eq 3).

$$Infiltration = max(min(AWI, InfiltrationPot), 0.) \tag{3}$$

$$Runoff = AWI - PrefFlow - Infiltr \tag{4}$$

Where Infiltration Potential (InfiltrationPot) is calculated using the Xinanijang-Arno model formulation (Todini 1996) as:

$$InfiltrationPot = StoreMaxPervious \cdot (1.0 - SatFraction)^{PowerInfPot} \tag{5}$$

and StoreMaxPervious and PowerInfPot are equal to:

$$StoreMaxPervious = WS1/(bInfilt + 1) \tag{6}$$

$$PowerInfPot = (bInfilt + 1)/b \tag{7}$$

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

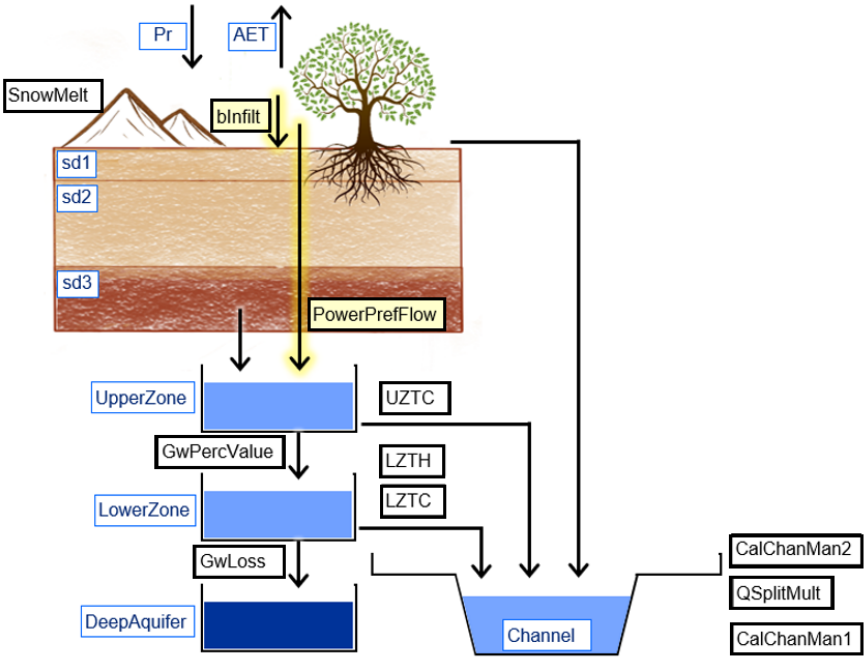

Figure 1. Schematic of the main processes of LISFLOOD model. Water fluxes are represented by arrows, with the infiltration and preferential flow processes (and respective parameters) highlighted in yellow. For visual clarity, only 11 of the 14 calibrated parameters are shown in bold black; the remaining 3 parameters, which relate to lake and reservoir dynamics, are not included in the figure but are listed in Table A1 in Appendix A), while the main processes and model compartments are reported in blue.

PowerPrefFlow and bInfilt (highligthed in yellow in Figure 1) are two of the 14 model calibrated parameters (Table A1, Appendix A), and specifically, they are among the 7 calibrated parameters that characterize the unsaturated and saturated soil zones. While a calibration of the model can elicit combinations of the parameters that reproduce observed streamflow to a comparable extent, different combinations of bInfilt and PowerPrefFlow may lead to very different simulations of the soil wetness, aquifer state and AET. As both saturation excess and ByPass flow are modelled empirically and independently, the calibration of the two parameters based only on the streamflow target does not guarantee to yield a physically consistent representation of soil water flows. This in turn influences also other model water-balance components. For example, the water content in the first 2 soil layers of the unsaturated zone (i.e., RelSat) represents the water available to plants for transpiration, which depends on soil saturation. Moreover, when the soil water content is low and cultivations lack sufficient water, the irrigation module is triggered and activates water withdrawal from surface and groundwater. As a consequence, the inaccurate representation of soil wetting and drying can translate into a over/underestimation of both actual evapotranspiration and water usage in agriculture.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Another aspect of crucial importance is the assumed soil depth. This largely determines the maximum amount of water available for soil and plants transpiration, however its elicitation remains challenging (e.g., Fan et al., 2019). Unlike in other distributed models, such as VIC, (Hamman et al., 2018), in LISFLOOD soil depth is not a calibrated parameter. LISFLOOD considers 3 soil layers: the first is the superficial soil (sd1) with a fixed depth (50 mm), while the depths of the second (sd2) and third (sd3) layers are assigned on the basis of soil properties derived from external data sources (e.g., depth to bedrock, depth to water table), as described in Section 2.4. LISFLOOD accounts for water redistribution across the soil layers, but evapotranspiration (constrained not only by energy but also by available water) occurs only in the first 2 soil layers, which represent the superficial and upper soil extending to the roots depth. Obviously, the chosen representation of soil layers depths has a potentially strong impact on model calibration, model performance and overall water balance. Finally, the model has two parameters controlling the recharge of the lower from the upper aquifer (GwPercValue), and the leakage of water from the lower to the deep aquifer (GwLoss). These add flexibility in the production of streamflow, but their values reverberate in the aquifer water balance. The purpose of this study is to explore how different parametrizations of the LISFLOOD model influence its performance in reproducing streamflow, soil water balance and AET. In particular we test the effect of (i) excluding the ByPass flow mechanism (highlighted in yellow in Figure 1) and (ii) using different soil depth configurations, since they both play a role in how water moves through the soil, hence affecting the soil water balance and the representation of AET.

2.4 Setup of numerical experiments

We design five new different model experimental setups (Table 1) and we compare the results against the current EFAS setup, considered as the benchmark. Two new configurations for soil depth 2 and soil depth 3 were designed by adopting different parametrizations. The same alternative model configurations, including the one with the original soil depths, were also calibrated excluding the ByPass mechanism, meaning that water can reach the groundwater compartment only through the unsaturated soil zone. Besides that, the input data used for all setups was identical to the one used for the benchmark model, which consist in 60 streamflow stations (used for model calibration), same static parameters (Choulga et al., 2023) and gridded time series of weather forcings from the EMO-1, European Meteorological Observations (v1) dataset (Thiemig et al., 2022a), based on observational data from over 35'000 weather stations over Europe, and covering the period 1990-2021 at 6-hourly timestep for precipitation and temperature (Thiemig et al., 2022b). The new soil configurations were calibrated following the same calibration routine of EFAS v5.0 (ECMWF, 2022).

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Table 1. Description of tested model setups, with 3 different soil depth configurations (including the benchmark, i.e., EFAS 5.0) and with-/without the preferential (ByPass) flow mechanism.

	Total soil depth = Depth to bedrock (Hengl et al., 2017)	**Total soil depth $\leq$ Water table depth (Fan et al., 2013)**	**Total soil depth $\leq$ 3 m**
Preferential flow included	Benchmark: the originally calibrated LISFLOOD used in EFAS 5.0.	BP-WT: Soil depth cannot exceed the water table depth (WTD).	BP-3M: Soil depth cannot exceed the minimum of WTD and 3 m.
Preferential flow excluded	NOBP: the benchmark setup excluding preferential flow.	NOBP-WT: Preferential flow is excluded and soil depth does not exceed WTD.	NOBP-3M: Preferential flow is excluded and soil depth does not exceed 3 m.

The three soil depth configurations (the benchmark and the two new representations) for layers 2 (sd2) and 3 (sd3) are shown in Figure 2. The benchmark soil depth configuration is the current setup (EFAS v5.0), derived from the so-called 'Absolute depth to bedrock' dataset from ISRIC (Hengl et al., 2017) and the root depth of forest and non-forest vegetation from FAO (Choulga et al., 2023). The second soil depth configuration assumes depth as the minimum between the ISRIC depth to bedrock data and the water table depth (WTD) derived from the dataset of Fan et al. (2013); in this second configuration, roots depth was also taken from another study (Fan et al., 2017), and used to compute the thickness of the soil depth of the second layer following the methodology from Burek et al. (2013). The third layer depth configuration is estimated as for the second, except in that it limits the maximum soil depth to 3 m.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

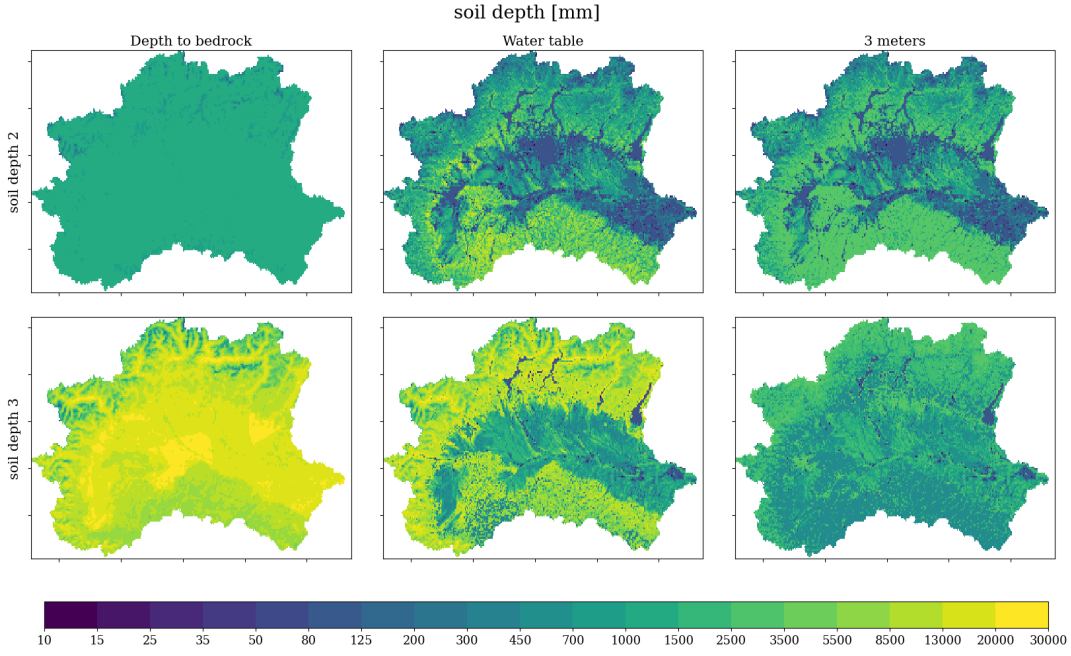

Figure 2. Soil depth (in millimeters) of soil layers 2 and 3 for the three alternative soil configurations used in our experiments.

The five new model configurations were calibrated, following the same methodology used to calibrate EFAS v5.0 (ECMWF, 2022). The calibration tool, based on the DEAP algorithm (Fortin et al., 2012), was employed and the sub-catchments draining to the 60 river gauge stations were calibrated following a nested approach, where head-catchments are calibrated first and then the simulated streamflow at their outlets (as calibrated) is the input of the downstream sub-catchment. A total of 14 parameters are calibrated in the two experiments with the active ByPass feature (BP-WT and BP-3M), as listed in Table A1, the benchmark is run with the EFAS 5.0 calibrated parameters. In contrast, in the three experiments without the ByPass mechanism, 13 parameters are calibrated, excluding the PowerPrefInfilt parameter, which specifically controls the preferential flow mechanism. Across the entire EFAS domain (Europe), the model operates at a 6-hourly timestep. This enables sub-daily calibration where sub-daily observations are available. For stations where only daily observations are available, the model output is temporarily resampled to match the daily timestep observations. For the Po Basin study area, only daily time series data were available at all 60 stations. The calibration process uses the modified Kling-Gupta Efficiency (KGE) (Kling et al., 2012) as the objective function, which is described in the following section.

2.5 Model performance criteria

2.5.1 Comparison with observed streamflow

We compare the six model setups presented above against observed streamflow at daily, monthly, and yearly time step. The comparison is mainly based on the modified Kling-Gupta Efficiency (KGE) (Kling et al., 2012), which is also used as objective

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

function for model calibration. In addition, we analyse its three components: correlation, bias ratio and variability (or spread) ratio, to characterise the performance on these three basic aspects. KGE is defined as one minus the Euclidean Distance (ED) of these components, computed as function of simulated and observed streamflows, from their ideal values, as follows:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{8}$$

where β is the bias ratio (i.e., ratio of mean simulated over mean observed flow), r is the Pearson correlation coefficient and γ is the variability ratio (i.e., ratio of the coefficients of variation). The ideal value of the KGE components r, γ and β is 1, resulting in a ED of these components equal to zero, which then maximizes the value of KGE. The KGE on untransformed streamflow is based on model residuals that are higher for high flows and consequently tend to give less prominence to errors on low flows (Santos et al., 2018; Garcia et al., 2017). While appropriate for flood applications, this objective function may not ensure a balanced model calibration when looking at low flows or at average regimes, which are crucial for drought monitoring and water resources applications.

To evaluate how different model configurations affect various flow regimes, including low flows, in addition to the KGE computed over the whole time series of flows, we evaluate the simulated flow duration curve (FDC) and key behavioural catchment functions using four signature metrics (Yilmaz et al., 2008): the percent bias in overall runoff ratio (%BiasRR), the percent bias in FDC midsegment slope (%BiasFMS), the percent bias in FDC high-segment volume (%BiasFHV), and the percent bias in FDC low-segment volume (%BiasFLV).

The %BiasRR is a signature measure of overall water balance and indicates whether the model over- or under-estimates (when positive or negative, respectively) the runoff compared to observations.

$$\%\text{BiasRR} = \frac{Q_{\text{sim}} - Q_{\text{obs}}}{Q_{\text{obs}}} \times 100 \tag{9}$$

where Q_{obs} is the observed streamflow and Q_{sim} is the simulated streamflow.

The %BiasFMS represents the percent bias of the slope of the mid-segment of the FDC, between 0.2–0.7 flow exceedance probabilities. Positive values indicate that the model overestimates the slope of the midsegment (steeper, and hence flashier than observed), while a negative %BiasFMS indicates underestimation (flatter slope than observed). Ideally, %BiasFMS should be close to zero, indicating good agreement between observed and simulated mid-segment slopes. The FDC midsegment slope can be seen as a signature of vertical redistribution of soil moisture, hence of soil storage capacity, with the importance of overland flow being larger for steeper slopes, while a slower and more sustained groundwater flow response is associated to flatter midsegment slopes.

$$\%\text{BiasFMS} = \frac{[\log(Q_{\text{sim},m1}) - \log(Q_{\text{sim},m2})] - [\log(Q_{\text{obs},m1}) - \log(Q_{\text{obs},m2})]}{[\log(Q_{\text{obs},m1}) - \log(Q_{\text{obs},m2})]} \times 100 \tag{10}$$

$m1$ and $m2$ refer to the 0.2 and 0.7 flow exceedance probabilities.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

The percentage bias in high (%BiasFHV, 0–0.02 flow exceedance probabilities) and low (%BiasFLV, 0.7–1.0 flow exceedance probabilities) flow segment volumes are used to assess the goodness of fit of the extreme high and low portions of the FDC, respectively. Positive values indicate an overestimation of high/low flows by the model. The ideal value is zero for both, which indicates that the model is perfectly able to reproduce high/low flows. As hydrological signatures, the %BiasFLV indicates the goodness of fit in long-term baseflow response (as the total volume of the low-flow segment can be seen as an index of baseflow response), while the %BiasFHV is associated to the level of both vertical and temporal redistribution of water.

$$\%\text{BiasFHV} = \frac{\sum_{h=1}^{H}(Q_{\text{sim},h} - Q_{\text{obs},h})}{\sum_{h=1}^{H} Q_{\text{obs},h}} \times 100 \tag{11}$$

$$\%\text{BiasFLV} = \frac{\sum_{l=1}^{L}(Q_{\text{sim},l} - Q_{\text{obs},l})}{\sum_{l=1}^{L} Q_{\text{obs},l}} \times 100 \tag{12}$$

where h = 1, 2,...H are the flow indices for flows within the respective exceedance probabilities intervals.

Results are considered satisfactory when bias values are within the ± 30 % range and unsatisfacory is greater than ± 30 % as summarized by Cislaghi et al. (2020) from previous literature (Herbst et al., 2009; Mendoza et al., 2015; Pfannerstill et al., 2014).

2.5.2 Compliance with the Budyko functional relationship

Besides comparison of model simulations against streamflow observations, it has been proposed that models should be also verified against empirical functional relationships that are commonly observed in hydrological data (Gnann et al., 2023; Yilmaz et al., 2008), among which the well-known Budyko model (Budyko, 1974) is probably the most commonly used. Sometimes referred to as the Turc-Budyko non dimensional graph (Coron et al., 2015; Turc, 1954), the Budyko model describes the long-term annual average water balance of a catchment, using a simple empirical equation that relates the long-term annual evaporative index, i.e., long-term mean annual actual evapotranspiration (AET) divided by long-term mean annual precipitation (Pr), to the aridity index, which is equal to mean annual potential evapotranspiration (PET) divided by mean annual Pr. This relationship can be interpreted, hydrologically, as arising from the competition between energy and water availability (Chen and Sivapalan, 2020). Although many alternative forms of the relationship have been proposed in the literature (Andréassian and Perrin, 2012; Chen and Sivapalan, 2020), here we refer to the original, non-parametric equation given in Budyko (1974):

$$\Psi = \phi[tanh(\phi^{-1}) \cdot (1 - e^{-\phi})]^{0.5} \tag{13}$$

where Ψ is the evaporative index AET/Pr and ϕ is the aridity index PET/Pr. A key assumption is that variations in catchment storage are insignificant in the long term and that the catchment remains a closed water system, unaffected by anthropogenic influences and not exchanging water with other catchments. Since the LISFLOOD model operates as a column model without lateral flow between cells, each grid cell can be considered a small, isolated catchment. Under these assumptions,

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

we expect that, over long timescales, the behaviour of an individual cell will conform to the Budyko curve. Grid cells affected by substantial anthropogenic influences were excluded, by removing those with more than 80 % irrigated crops, or 80 % sealed surface, and with 100 % open water and rice cultivation (see AppendixA), since these characteristics would cause deviations from the assumptions under which the Budyko model is valid.

For each pixel in the modelled domain, the aridity index, PET/Pr, and the evaporative index, AET/Pr, have been calculated for each experiment and compared against the Budyko curve. Under such conditions, the model allows to detect the mean "climatological" AET, enabling the estimation of average annual water surplus from the catchment, and consequently of the annual streamflow volume. Indeed, any surplus of precipitation with respect to AET would practically coincide with the streamflow volume (Q), from the closure of the water balance (Q=P-AET), in absence of human withdrawals and of natural inter-catchment groundwater flows (e.g. (Ballarin et al., 2022)). The framework has been extensively used to identify catchments undergoing shifts in water balance due to climate change (Senbeta et al., 2023), impact of land use change on runoff generation (Gnann et al., 2023; Jaramillo and Destouni, 2014). Given its simplicity, the Budyko framework offers a tool to evaluate (Gnann et al., 2023; Koppa et al., 2021) and/or calibrate (Greve et al., 2020) complex hydrological models. In previous tests at the European scale (Pistocchi et al., 2019) and, more specifically, in the Danube catchment (Pistocchi et al., 2015), it has been shown that the Budyko relationship predicts observed annual average streamflow volumes quite accurately; for the Danube, the comparison yields a correlation of 0.89, a relative mean error of +8 %, and a relative standard deviation error of +9 %. These findings have been confirmed by a more recent analysis, showing that the Budyko relationship generally predicts annual streamflow volumes well in line with observations in several European catchments (Pistocchi et al., 2024). The deviation of simulated AET and runoff from the predictions of the Budyko equation highlights situations where AET or runoff differ significantly from what one would expect based on the climatological characteristics of the catchment. We measure the relative difference of AET/Pr and Q/Pr from the values predicted with the above Budyko equation, that we call Budyko Distance (BD), for each grid cell of the modelled study area using the formula:

$$BD_{AET} = \frac{AET_{sim} - AET_{budyko}}{AET_{budyko}} \tag{14}$$

$$BD_{Q} = \frac{Q_{sim} - Q_{budyko}}{Q_{budyko}} \tag{15}$$

where AET_{budyko} \$Q_{budyko}$ is the long-term AET\calculated using the Budyko equation (13) and AET_{sim} \$Q_{sim}$ is the long-term AET simulated by the model. The closer BD is to zero, the more consistent is the model long-term partitioning of Pr between Q and AET with the expected Budyko relationship. In practice, we consider results acceptable when BD is between -0.2 and 0.2 ($\pm$ 20 %), in line with previous studies (Li et al., 2014; Gentine et al., 2012), where a general acceptance was set to a 10 % difference from AET_{budyko}, but larger deviations could occur for other physical reasons, such as vegetation and land-use changes, (Gunkel and Lange, 2017; Jaramillo and Destouni, 2014) as well as snow dynamics and snow fraction changes (Zaerpour et al., 2024).

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

2.5.3 Water balance closure

For a hydrological model, we expect the water balance to close, meaning that in the long term the cumulative precipitation should equal the sum of cumulative actual evapotranspiration, cumulative streamflow volume, and cumulative net groundwater exchanges due to sub-surface connectivity. LISFLOOD, being a column model, does not simulate does not simulate inter-catchment groundwater inflows, but ground water losses instead. In closed topographic catchments, groundwater losses are often introduced to ensure the water balance closure and compensate for excessive groundwater recharge. A large contribution of groundwater losses indicates either that the catchment leaks a groundwater volume to neighbouring catchments or that the water balance calculation is unable to explain where the corresponding volume should be allocated, thus adjusting it with large losses. For the Po Basin, however, such gains are assumed to be negligible because the dominant groundwater flow paths and aquifer boundaries lie within the surface-water catchment (Beretta et al., 2025). Based on these considerations we compare the performance of different model setups also in terms of how they break down precipitation in the various components of the water balance. This holds insofar as we neglect the changes in the internal storage of the system (snow, lakes and reservoirs, soil and aquifer water content), which is an acceptable assumption over an extended multi-year period.

3 Results

3.1 Streamflow

Across the 60 calibrated sub-catchments, model performance is primarily controlled by the representation of preferential flow, with soil depth playing a secondary role (Figure 3). At the daily scale, setups including the ByPass mechanism consistently achieve higher KGE values, while those without ByPass perform worse, regardless of soil depth. The setups with the ByPass mechanism perform in the same way (yielding almost identical KGE values), irrespectively of the assumptions on soils. For the setups without ByPass, a slight improvement is observed when soil thickness is reduced (moving from NOBP to NOBP-3M), but the difference remains small compared to the gap between ByPass and no-ByPass configurations. The main cause of the KGE performance decline for the setups without ByPass is the degradation of the correlation, followed by the variability ratio, while the bias remains nearly identical across all setups and timescales (Friedman test p-value above 0.05). This suggests that the ByPass has an essential role in the temporal redistribution of water in LISFLOOD that cannot be easily compensated by other model components, while its contribution in the long-term water balance could be easily replaced. At the monthly and especially annual scales, the performances are more similar than at daily scale for all models, as shown by the increasing p-value in KGE. In this case, the setup without ByPass and with the thinnest soils (NOBP-3M) achieves the same performance as all setups with ByPass (with negligible differences), while thicker soils cause a worsening of the performance in the absence of ByPass.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

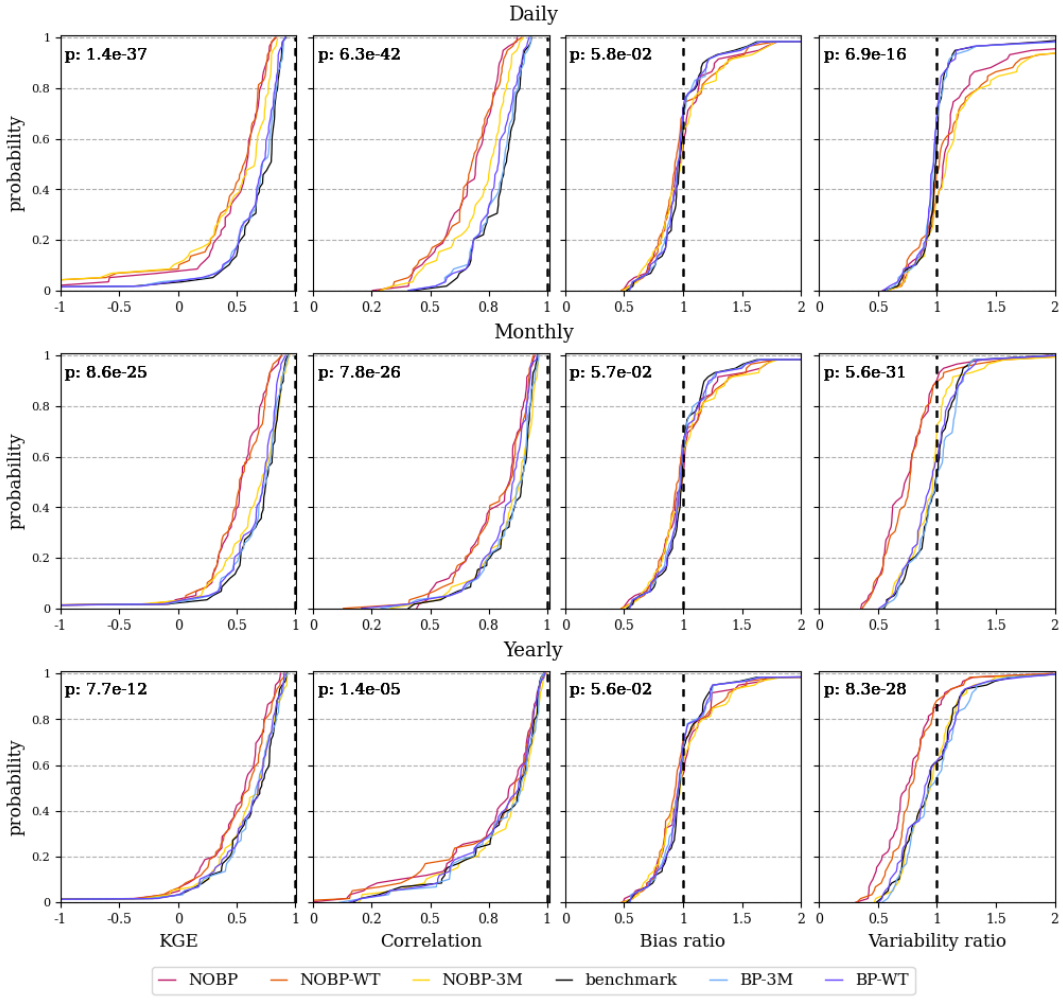

Figure 3. Empirical cumulative distribution functions (CDFs) of KGE and its three components (correlation, bias and variability ratio) across all the 60 calibrated sub-catchments for the six alternative LISFLOOD setups (including the benchmark), evaluated at daily (top row), monthly (middle) and yearly (bottom) time scale. In the top left corner of each plot, the minimum p-value for all pairs of the Friedman test (FT) is shown: $p > 0.05$ indicates no significant differences among the six LISFLOOD setups, while $p < 0.05$ indicates significant differences.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

The spatial patterns of overall daily performance confirm the consistent advantage of including the ByPass mechanism, as setups with ByPass achieve higher KGE values across the basin (Figure B1 in Appendix B). The largest difference in performance is seen in the western part of the basin and in the floodplain, downstream (southern) catchments. In contrast, sub-catchments in the northeastern part of the basin exhibit poorer performance than the basin average, regardless of the presence or absence of the ByPass. A larger number of catchments (up to 3 additional cases) have a KGE below -0.41,meaning that the model is worse than a simple mean flow benchmark (Knoben et al., 2019) , when the ByPass is excluded; furthermore, up to 3 more basins per configuration get to perform worse than a simple mean flow benchmark without ByPass. An analysis of the single components of KGE (Figures B2, B3, B4) reveals that the correlation degrades across the whole catchment in the experiments without ByPass; the variability ratio generally worsens in all experiments, especially in catchments in the north part of the basin; the bias ratio does not show major differences across the basin.

The most frequently best-performing setups are BP-3M and the benchmark, dominating respectively in the western and southern parts of the basin (see Figure B5 in Appendix B). The setups with ByPass enabled perform best in almost all sub-catchments, with only four Alpine sub-catchments in the north-east showing negligible differences. Even if each of the different soil configurations with active ByPass leads to the best performance in several cases, most catchments exhibit negligible differences in performance by changing only the soil depth. On the other hand, without ByPass the three soil configurations show substantial differences, with a general better performance of NOBP-3M (thinnest sd3 soil), especially in the floodplain, whereas in the Alpine sub-catchments the differences are negligible.

3.2 Flow Duration Curve

Flow duration curve (FDC) diagnostics reveal clear differences between setups, with the representation of preferential flow strongly shaping model behaviour, especially across average and high-flow regimes as revealed by %BiasFMS and %BiasFHV (Figure 4).

Overall water balance bias (%BiasRR) is comparable across all the experiments, with slightly better performance when the ByPass is enabled. In contrast, the mid-segment of the FDC (%BiasFMS) shows the most marked differences: in all experiments without the ByPass, more than 60 % of the catchments exhibit negative values, indicating overly sustained/slow-varying flows compared to observations, while this behaviour is much less frequent with ByPass (below 40 %); this problem is most pronounced in NOBP and NOBP-WT, where only ≈10 % of the catchments fall within the ±30 % acceptability range. The best performing experiment without the ByPass is the NPBP-3M with 40 % of catchments within the acceptability range. The experiments with ByPass show broadly similar behaviour, with more than 60 % of simulations being more flashy than observations and 50 % of catchments within the acceptability range. High-flow performance (%BiasFHV) is also generally better in the experiments with ByPass, as they show ≈85 % of the stations within the acceptability range, while these drop substantially without ByPass. Among the no-ByPass experiments, NOBP performs best, followed by the NOBP-WT and the NOBO-3M.

Low flows (%BiasFLV) are more problematic across all experiments: only ≈40 % and ≈55 % of catchments fall within the acceptability range (±30 %), in the experiments without and with ByPass respectively. In the no-ByPass experiments, the

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

model tends to overestimate low flows more frequently than with ByPass. In terms of %BiasFLV, the BP-3M experiments is the best performing among the experiments with the ByPass active, with more than 50 % of the catchments within the acceptability range. The benchmark and the BP-WT experiments tend to have more catchments where low flows are underestimated.

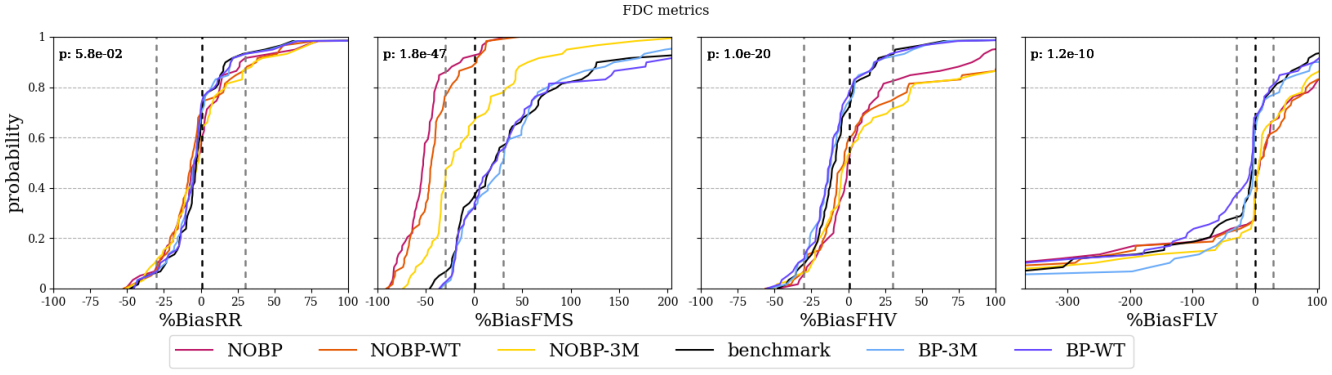

Figure 4. Flow duration curve (FDC) signature metrics (%BiasRR, %BiasFMS, %BiasFHV, %BiasFLV); the black dashed vertical line indicates the optimal values, while the grey dashed lines define the acceptability range at ±30 %.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

3.3 Calibrated parameters

The frequency distributions of the 14 calibrated parameters across the 6 experiments highlight how parameter sensitivity depends on the model setup (Figure 5).

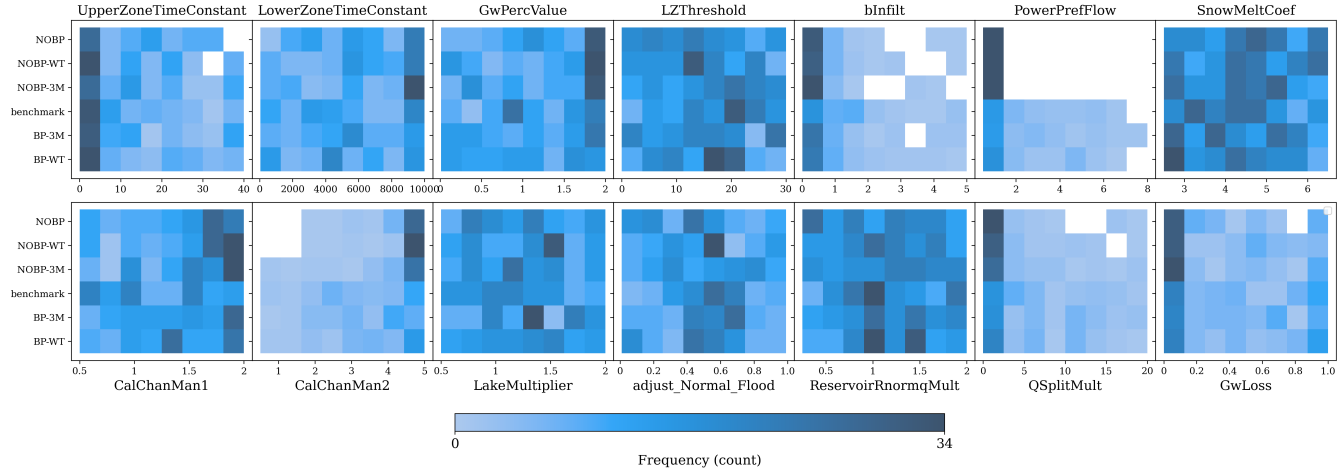

Figure 5. Frequency distributions of the 14 calibrated parameters values across the 60 sub-catchments for each of the six experiments. Colour shades represent the frequency of occurrence of a parameter within the values on the horizontal axes. Darker shades indicate more frequent parameter values. Patterns of vertical bands suggest stability across experiments, while shifts in distribution reflect sensitivity to model setup.

All the experiments including the ByPass do not show any significant difference in the frequency distribution of the calibrated parameters as identified by the Kolmogorov-Smirnov test (KST) with significance level 5 %, suggesting that changes in soil depth do not cause substantial changes in other parameters and do not lead to systematic internal flux compensations. In contrast, in the experiments without ByPass, there are significant changes (KST at 5 % significance) in the distribution of several parameters (CalChanMan2, QsplitMult, Gwloss, bInfilt, and GwPerc), pointing to more systematic compensations in the absence of preferential flow. These parameters affect the speed of propagation of floods, the losses from the lower aquifer, the saturation excess, and the percolation from the upper to the lower aquifer; the distribution of these calibrated parameters for each catchment are showed in Appendix B (Figures B6-B10). Compared to the ByPass setups, the changes in routing parameters (CalChanMan2 and QsplitMult) in the experiments without the ByPass suggest slower channel routing (higher value of CalChanMan2) and greater diversion to floodplains or to the overbank part of the channel (lower QsplitMult). Soil-groundwater parameters adjust by clearly compensating the lack of the ByPass component: bInfilt shows a significant decrease, indicating that water infiltrates more in the unsaturated zone; GwPerc is higher implying that water moves from the upper to the lower groundwater zone at a higher rate; GwLoss tends to be lower meaning that there is less need to compensate for excess water in the deep aquifer that cannot be explained by other physical processes. Overall, excluding the ByPass forces compensatory parameter adjustments that redistribute water through alternative pathways, replacing the aquifer recharge otherwise captured by the preferential flow process.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

3.4 Compliance with the Budyko functional relationship

3.4.1 Evapotranspiration

The setup with no ByPass and original soil configuration (NOBP) shows the pattern with the highest overall proximity to the Budyko curve (see Figure 6; mean distance: 0.081), followed by other setups without ByPass (NOBP-WT and NOBP-3M). Among the experiments with the ByPass, the benchmark is the one performing best (mean distance from the Budyko curve equal to 0.148). The other setups with ByPass (BP-WT and BP-3M) approach the Budyko curve when AET/Pr <0.5 and PET/Pr <0.4, while at higher PET/Pr levels there are many more cases farther from the Budyko curve. The same trend is visible, but weaker, with the other setups without ByPass (NOBP-WT and NOBP-3M). This indicates that when the soil depth is the same, excluding the ByPass tends to improve the compliance with a Budyko functional relationship for evapotranspiration. The spatial distribution of the distance from the expected Budyko evapotranspiration clearly highlights a "patchy" pattern (Figure B11 in Appendix B), especially in the ByPass experiments, which resembles the distribution of the bInfilt (Figure B6) and PowerPref (Figure B7) parameters. In general, for most of the basin the modelled AET is lower compared to the AET calculated using Budyko, highlighting a widespread non-compliance, especially for setups with the ByPass. Budyko non-compliance is also particularly visible where soil depth of the second layer is thin, as for example east of the floodplain, where soil depth is below 50 cm for NO/BP-WTD (Figure 2), and the actual evaporative index is half as expected by Budyko.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

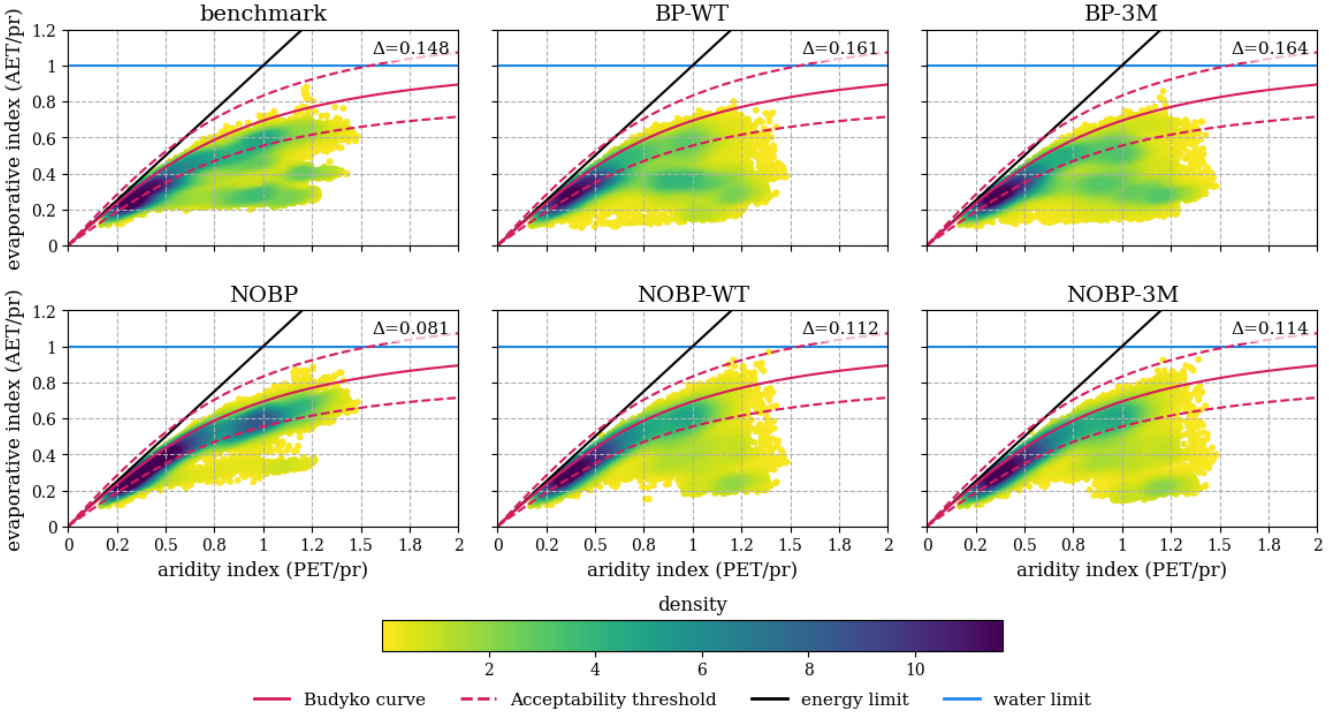

Figure 6. Density plot of Aridity Index vs Evaporation Index of each pixel of the masked Po River Basin domain; the magenta curve represents the Budyko relationship (described in equation 13) and the dotted magenta curve the acceptability range. On the top right corner of each plot the average distance (adimensional) of each pixel from the Budyko curve is indicated. The points falling between the dotted and continuos magenta lines are considered Budyko compliant. The blue line represents the water limit, where AET<P, and the green line represents the energy limit, where AET $\leq$ PET. Those two limits cannot be exceeded in systems that comply with the Budyko hypothesis.

3.4.2 Runoff

In a similar way we compare total runoff from LISFLOOD and the Budyko relationship (Figure 7). Also in this case the original setup excluding the ByPass (NOBP) shows the highest compliance with the Budyko framework, with the highest density of points close to the Budyko curve (mean distance equal to 0.122). The other experiments perform in a similar way, with a higher density of pixels close to the Budyko curve in the most energy limited part of the plot, and spreading out towards the water limited side of the plot. The setups with active ByPass have more of a scattered behaviour than the no-ByPass experiments, with the mean distance from Budyko being higher in all cases. It is possible to distinguish high density concentration of points close and parallel to the water limit in all experiments but in particular in the benchmark, denoting that there are grid-cells where runoff is significantly higher than what suggested by the Budyko framework.

In all experiments, many grid-cells cross the energy limit (green line); assuming that PET and precipitation data is correct, runoff ought to be above the energy limit threshold, causes of this behaviour are linked to a overestimation of Precipitation, an over estimation of Q, and/or an underestimation PET. Otherwise this behaviour occurs mostly in leaky catchments (Andréassian

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

and Perrin, 2012), so where water contributes to the replenish of the aquifer. Pixels that cross the energy limit and where the runoff is lower than the expected Budyko runoff are the red/negative values in Figure B12 in Appendix B. We can observe than in most of those catchments, the Gwloss parameters (Figure B8) is close to 1, flagging the tendency of the model in storing water in the deep aquifer, at the same time in some west catchments, the bias (Figure B3) remains positive which could lead to an hypothesis of overestimation in precipitation. When the bias is negative, the hypothesis is that the model tends to move water to the deep aquifer in order to achieve to maximize KGE.

On the other hand positive values in Figure B12 mean that the runoff is higher than the one expected by the Budyko equation (points that are above the red line in Figure 7). Generally most of the catchments in the floodplain tend to show this behaviour, with many of them showing a negative bias (Figure B3), which could be due to an underestimation of precipitation and the model tendency to transform precipitation in runoff only in order to match observed discharge and/or a compensation of low upstream discharge entering the catchment.

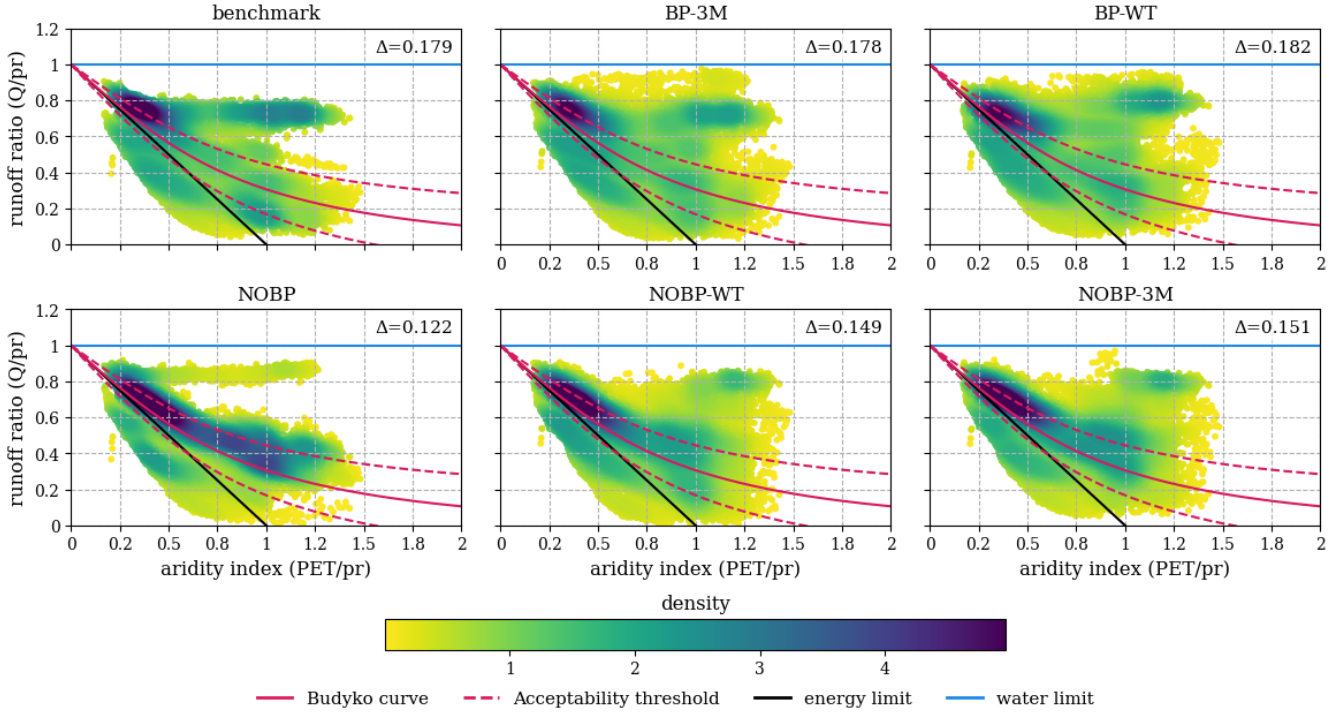

Figure 7. Density plot of aridity index vs runoff/P. The blue line represent the water limit, where runoff<precipitation, the green line is the energy limit, derived from the relation PET $\geq$ AET, hence runoff/Pr $\geq$ 1-PET/Pr. On the top right corner of each plot the average distance (adimensional) of each pixel from the Budyko curve is indicated.

The setup with no ByPass and original soil depth (NOBP) show smaller discrepancies throughout the basin. This is particularly clear from the CDFs (Figure 8) as the NOBP experiment is the one with the narrowest dispersion around the ideal value (BP=0). The second best performing experiments are the remaining 2 without the ByPass (NOBP-WT and NOBP-3M) which

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

show almost identical distributions of BP. The Bypass experiments have more than 60 % analysed area exceeding 20 % (0.2 in Figure B12) of the acceptable range. Runoff is compliant in less than 60 % of the pixels, with the experiments without ByPass performing slightly better compared to the ones with the ByPass.

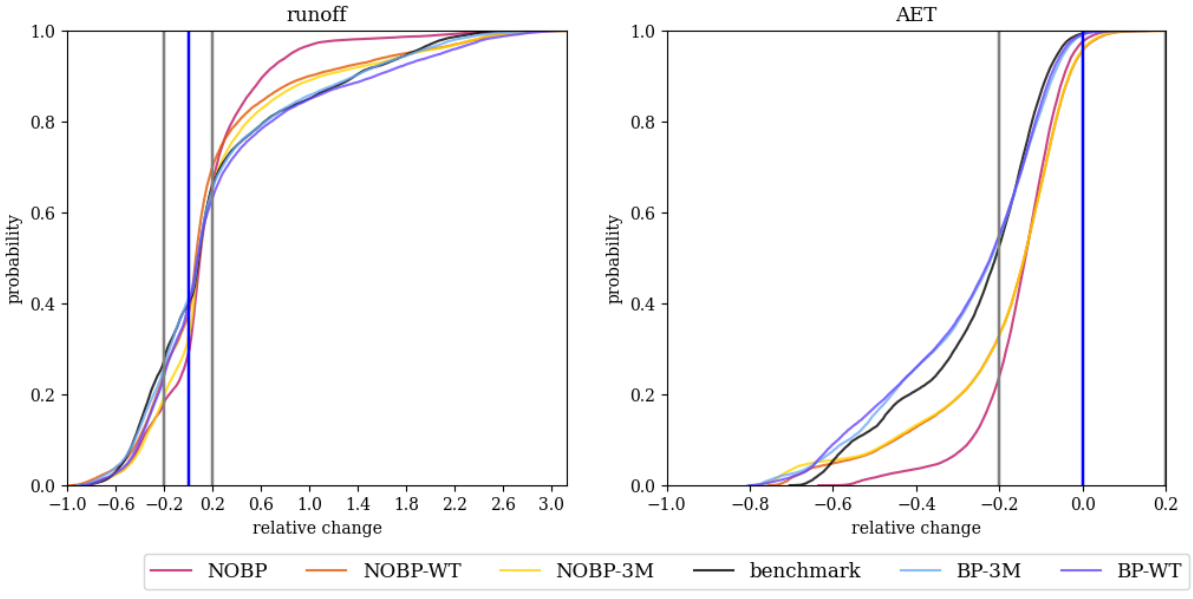

Figure 8. Cumulative Distribution Function (CDF) of the Budyko Distance (BD) calculated for each pixel of the Po river basin for total runoff (left), and actual evapotranspiration (right). The ideal value of null BD (solid blue vertical line) corresponds to a perfectly consistent model long-term partitioning of Pr between Q and AET with the expected Budyko relationship, while results of BD between -0.2 and 0.2 are considered acceptable (grey vertical lines). A few negative and positive outlier values have been removed.

3.4.3 Parameters interplay

The distributions of the Budyko Distance of AET (e.g., Figure B11) suggest an intuitive relationship between soil depth, AET and calibrated parameters bInfilt and PowerPrefFlow (Figure B6 and B7). Indeed, significant (Pearson's) correlation values are found between BD and several analysed parameters (Figure 9): bInfilt, PowerprefFlow, average sd3 and average sd2 per catchment, versus the average relative difference from Budyko AET for each catchment. Only catchments with over 30 % valid pixels (showed in Figure A2, Appendix A) were included in the analysis.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

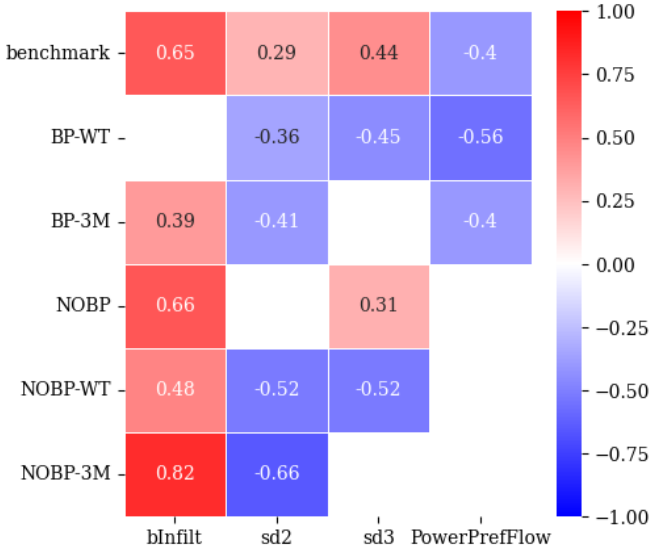

Figure 9. Correlation between Budyko Distance and parameter value (for parameters bInfilt, sd2, sd3, PowerPrefFlow). Cells are coloured only when the correlation is significant ($p<0.05$).

The bInfilt parameter (exponent for infiltration capacity of the soil) shows consistent positive correlation with BD across all experiments, but BP-WT, indicating that as bInfilt increases, the difference between modelled AET and Budyko AET also increases. This suggests that infiltration capacity plays a key role in shaping the deviation from the Budyko curve.

The sd2 parameter (depth of the second soil layer) shows a weak-moderate negative correlation in all experiments except NOBP, with the strongest inverse relationships in experiments without the ByPass mechanism. This implies that shallower soil layers (lower sd2 values) are associated with reduced AET, highlighting the importance of soil depth in AET. At the same time the lack of significance in the correlation between sd2 and NOBP, shows that without the ByPass mechanism and with a thick sd2 the BD is detangled from the sd2 value.

For sd3 (depth of the third soil layer), a weak positive correlation appears in the NOBP and benchmark experiments—where sd3 values tend to be higher (Figure 2). In contrast, experiments with thinner sd3 layers exhibit moderate to weak negative correlations. These findings suggest that, although sd3 is not explicitly used in the AET calculation in LISFLOOD, its magnitude appears to influence AET outcomes.

The PowerPrefFlow parameter demonstrates a weak to moderate negative correlation. Notably, in the BP-WT experiment, PowerPrefFlow shows a stronger inverse correlation than bInfilt, suggesting that the ByPass mechanism does influence the AET dynamics, more than sd2 parameter.

3.5 Water Balance Closure

Figure 10 shows the sum of AET, total runoff and the groundwater loss expressed in mm/year. Observed precipitation, which is equivalent in all experiments, is slightly lower than the sum of the 3 modelled components, which means that in all experiments

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

there is more water leaving the system than water entering it. The sum of the three components (AET, Q and groundwater loss) is similar in all experiments, however the actual evapotranspiration is higher in the experiments without the ByPass (with an average increase of around 15 % with respect to corresponding experiments with ByPass). On the other hand the groundwater loss is higher in the experiments with the ByPass (with an average increase of more than 20 %). Across all experiments, the maximum AET (NOBP) is 437 mm/year, which lies within the Budyko acceptability range, and is ≈18 % higher than the minimum one of 370 mm/year (BP-WT), which falls well below the acceptability range. The highest deep aquifer volume is found in the experiments with the ByPass, where the water moving to the deep aquifer is between 18 % and 34 % higher then in the experiments without the ByPass.

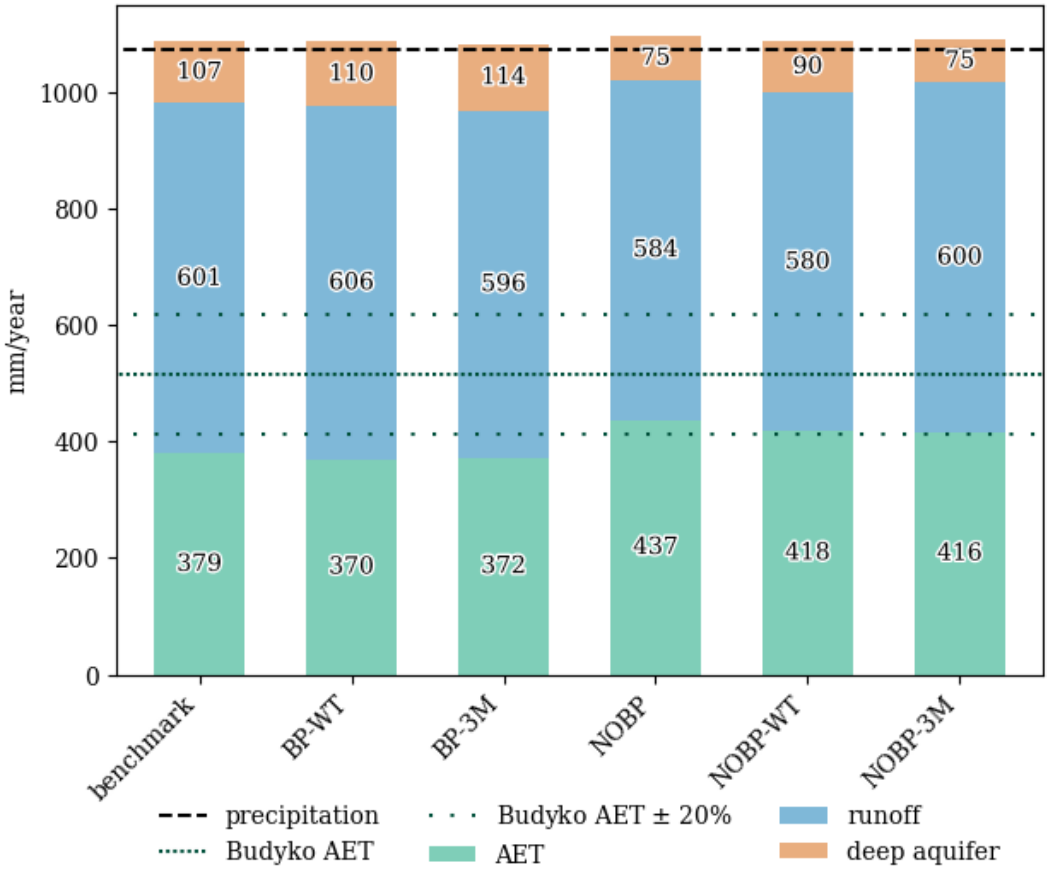

Figure 10. Water balance components (in mm/year) for all experiments, with bars showing the partitioning of water into runoff, AET and GwLoss (deep acquifer recharge). The horizontal black dashed line represents the observed precipitation, while the dark green densely dotted line represents Budyko-estimated AET (from equation 13), and the dark green loosely dotted lines represent the acceptability range of ± 20 %.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

4 Discussion

The calibration of six alternative setups of LISFLOOD has highlighted that the original model setup, with the ByPass mechanism enabled, remains the best performing one in terms of overall streamflow accuracy (KGE and its components). Among the ByPass-enabled setups (benchmark, BP-WT, BP-3M) performance differences are minimal, regardless to the assumptions made on soil depth. The ByPass reduces the influence of the soil depth, compared to the experiments without the Bypass (NOBP, NOBP-WT, NOBP-3M), by conveying water directly to the upper groundwater zone of the model. This mechanism can prevent water from infiltrating in the unsaturated zone, causing lower evapotranspiration (AET) and influencing the accumulation of water in the deep aquifer (Figure 1). We can argue that the ByPass shifts the water balance towards infiltration in the deep aquifer at the cost of evapotranspiration. As this may cause excessive return flow from the lower zone to the streams, the model calibration tends to elicit a higher percolation in the deep aquifer (GwLoss parameter), meaning that larger volumes of water are taken out of the system (Figure 10). While useful to improve general-purpose streamflow accuracy (e.g., KGE), the spurious increase in groundwater losses is purely empirical and driven solely by streamflow observations (via calibration), flagging a surrender by the model in quantifying certain shares of the water balance in a physically-consistent way. Conversely, model setups excluding the process of ByPass do not offer sufficient flexibility to achieve the same performance levels (e.g., KGE). However, these setups show a similar ability to the ByPass experiments in reproducing average runoff (%BiasRR in Figure 4) and a higher compliance with the Budyko functional relationship, which we may argue is equivalent to a more realistic response in terms of the overall water balance of the river basin; this is a case of producing the 'right results for the right reasons' (Beven, 2018), where model fluxes and states better align with theoretical expectations (as observations are not available for some variables) which might lead to more robust performance under non-stationary or changed conditions (Beven et al., 2022). When we compare streamflow, while the individual high-flow events (e.g., at daily resolution) are better captured by the benchmark, the monthly and annual flows are predicted by the model with and without preferential flow in a comparable way, particularly in the experiment with the shallowest soil (NOBP-3M). For low-flows, both types of setups show comparable performance, indicating that the different model configurations tested are not able to improve the reproduction of low flows. We can argue that this result can be a combined effect of the high model flexibility together with the objective function used for calibration (KGE), which tends to favour accuracy on high flows rather than low flows or other regimes (Mizukami et al., 2019; Gauch et al., 2023).

The ByPass mechanism represents the well-known hydrological process known as preferential flow, which refers to fast water movement through the soil matrix (Clothier et al., 2008), travelling along specific pathways deep into the subsurface, often below the root zone. Recent studies (e.g., Kocian and Mohanty, 2024) have also highlighted how seasonality and soil properties influence the occurrence of preferential flow events.

Given its localized and heterogeneous nature, it becomes challenging to model preferential flow, especially in large scale semi-distributed models, which rely on averaged parameters over large areas. In the case of LISFLOOD, the PowerPrefFlow parameter (i.e., exponent in the preferential flow function) is homogeneously applied in each calibrated sub-catchment, removing localized and temporal effects, but also contributing to dry soil and low AET together with other model parameters. Our

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

505 analysis (Figure 9) indicates that there is an interplay between PowerPrefFlow, saturation excess (bInfilt), and soil depths for the estimation of AET. Soil depth plays a more prominent role in the experiments without ByPass and for sd2 values thinner than 1 meter (as in WT and 3M experiments); on the other hand, in the ByPass experiments, the PowerPref flow becomes more important than soil depth in reducing AET from the expected Budyko value. Moreover, the stream network parameters controlling water transfer speed and correlation are less important in the setups including the ByPass (Figure 5), because the 510 PowerPrefFlow parameter strongly controls the quantity and timing of fast/subsurface runoff, influencing the flow timing more than the parameters that describe channel properties. These findings reinforce our argument that the ByPass and the Xinjiang-VIC-Arno components, as currently parametrized, can uncontrollably lead to diminish the hydrological realism of the water balance, while adding flexibility and improving accuracy on high-flow events.

The moderate-high correlation between saturation excess (bInfilt) and the relative difference from the Budyko AET (Figure 515 9), indicates that limiting the upper bound of saturation excess (bInfilt, currently between 0.1-5) could improve the representation of AET by forcing water to infiltrate in the vadose zone and hence make it available to plants transpiration. When calibrating with a lower saturation excess (bInfilt), the PrefPowerFlow would have to automatically adjust to find the best combination of parameters leading to the best performance. However, this parameter compensation process should be carefully evaluated, since some performance metrics (e.g., KGE) may significantly deteriorate under these conditions, highlighting trade-520 offs between overall performance metrics and internal consistency of water balance components, as demonstrated in this study. Adding constraints on saturation-excess parameters (bInfilt) would make the role of soil depth more prominent, as the AET and KGE results show in the experiment without the ByPass. Shallower soil depth leads to lower AET and a minimum depth of 1 metre seems to ensure a correct AET estimation if saturation excess (bInfilt) and PowerPrefFlow are correctly parametrized. That is supported by the AET performance in the NOBP and the benchmark that show a good Budyko compliance, unless 525 when saturation excess (bInfilt) is above 4 (Figure 6).

Changing the soil depth has shown to help improve the KGE performance in the NOBP-3M setup, which excluded the ByPass. In particular, the third soil layer of the model is assumed to not contribute to AET, and works only as a buffer between the unsaturated zone and the upper groundwater zone. In the experiments without ByPass, a deeper soil delays the transfer of water and contributes to a lower performance. Shallower soils allow water to move faster to the upper groundwater zone, 530 from where it contributes to subsurface runoff. These findings suggest that a way to enhance Budyko compliance without compromising streamflow performance (KGE) could be by constraining saturation excess (bInfilt) and adjusting the third soil depth. While including sd3 in the calibration process could further improve model results, it also increases the risk of over-parametrization, potentially leading to unexpected or unreliable outcomes. Careful consideration is therefore needed when expanding and/or constraining the parameter set and/or modifying the preferential flow process. This study highlights the need 535 of a deeper investigation on how the LISFLOOD model simulates the vadose zone, assessing the which processes should be constrained or modified to ensure that the model achieves the right response (streamflow) for the right reasons (model states and fluxes are internally consistent) (Beven and Cloke, 2012; Wagener and Pianosi, 2019). Sensitivity and uncertainty analysis have been carried out in the past for LISFLOOD (Zajac et al., 2017; Bisselink et al., 2016; Feyen et al., 2008), but their focus was solely on streamflow simulations. Applying Global Sensitivity Analysis (GSA), recognized as a valid and

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

powerful tool for guiding hydrological model development and quantifying the influence of modelling choices (Wagener and Pianosi, 2019), would therefore provide essential insight into how different parameters influence various processes and their interactions. However, to achieve robust and reliable sensitivity estimates, GSA typically requires a large number of model runs (Wagener and Pianosi, 2019). This computational demand presents a significant challenge for semi-distributed hydrological models, which are typically computationally intensive due to their spatial detail and complexity. To address these limitations, combining GSA with empirical relationships, such as the Budyko framework, and incorporating spatially distributed Earth Observation (EO) products like soil moisture and evapotranspiration (ET) data offers a promising path forward, the Po basin itself in an excellent example of EO-based reconstructed water cycle ((Brocca et al., 2024). This approach can help better constrain and represent AET and other hydrological processes, even at the grid-cell level, hence improving realism while minimizing adverse impacts on streamflow simulations.

Another aspect to consider is the general design of the LISFLOOD model, which implicitly incorporates human influences by using non-naturalized streamflow time series for calibration. This approach, while necessary due to the large number of calibrated stations (1903) across Europe (ECMWF, 2022) and the diversity of data sources, may introduce biases when human influences are not fully accounted for (Terrier et al., 2021). Given the general lack of knowledge of water management practices, including inter-basin transfers-information that is often unavailable, improving the explicit modelling of these influences is an open challenge. This knowledge gap, combined with the model's calibration routine, can lead to systematic errors affecting the water balance components. For instance, in basins where water is artificially transferred from other catchments, or the flow is heavily regulated, the model may underestimate AET. This occurs because additional inflows increase streamflow, which the model may balance by reducing AET or soil water storage, thus closing the water balance at the expense of accurately representing other hydrological processes. A further concern arises from the absence of explicit groundwater-transfer processes in LISFLOOD combined with the cascading calibration approach employed for the calibration of EFAS and GLOFAS. In catchments where groundwater moves ("leaky catchments") to neighbouring basins, the model cannot represent this transfer. During calibration, the water that is actually moved via groundwater may instead be assigned to the deep aquifer storage (gwloss). Consequently, the leaky catchment may show lower simulated runoff and/or AET (potentially below the theoretical optimum such as a Budyko curve metric), while the receiving catchment might show an overestimation — or at least a mismatch — of runoff or AET.(Andréassian and Perrin, 2012)

A similar issue arises when precipitation is over- or under-estimated: if precipitation is underestimated, the model may achieve a good streamflow fit by lowering AET, if precipitation is overestimated, the model may compensate by increasing deep aquifer storage (gwloss).

In the context of LISFLOOD calibration, and more generally in semi-distributed models, the Budyko framework proves to be an effective tool for model diagnostic, for a better interpretation of model results, and verification of simulated AET. As also suggested by Gnann et al. (2023), this verification is necessary for a reliable use of model outputs that are not directly calibrated. The Budyko framework, and the BD, could be employed to constrain the parameter space prior calibration, ensuring that only physically meaningful values are considered. Alternatively, it can be integrated as a metric during calibration or used in post-calibration evaluation to guide parameter selection/tuning, reducing equifinality, and identify combinations that

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

underestimate AET. This path aligns with the "expert-based model evaluation" approach proposed by Gleeson et al. (2021), who advocate for large-scale model behaviour testing against expected system dynamics derived from hydrological theory and expert expectations (with or without direct observations). When combined with streamflow observations, the Budyko-based analysis may also serve as a valuable diagnostic for detecting errors in model input data, such as meteorological forcing, thereby enhancing model robustness and credibility.

However, care must be taken when applying the Budyko relationship as an objective function in a multi- or single-objective calibration, particularly in catchments heavily influenced by human activities or that underwent significant land-use changes that are not accounted for in the model. In such cases, the theoretically expected partitioning between runoff and AET may not hold, and enforcing Budyko constraints could lead to suboptimal or unexpected results.

5 Conclusions

We have examined how the LISFLOOD hydrological model reproduces the water balance components across six setups on the representative case-study of the Po river basin in Northern Italy. We have shown that the setup with the preferential flow (ByPass) mechanism currently implemented in EFAS v.5 outperforms the tested alternative setups without the ByPass in terms of streamflow accuracy (based on the KGE and its components). However, this owes to a large extent to the inclusion of a preferential flow process, which can cause underestimation of infiltration in soils and of actual evapotranspiration, as well as a consequent overestimation of recharge to aquifers. To compensate, the model relies on an empirical parameter to adjust groundwater losses, which retains limited physical meaning and may limit the validity of the model when addressing water resources management questions beyond streamflow (e.g. aquifer recharge, estimation of irrigation requirements and availability of groundwater resources). The deviations between the modelled water balance components and Budyko-based predictions are particularly apparent when the second soil layer is shallow (< 1 meter) and for low values and high values of PowerPrefFlow and bInfilt respectively.

Further research is needed to understand how to improve the model and/or constrain parameters to keep a good performance in terms of KGE and an acceptable representation of AET and groundwater components. Experiments excluding the preferential flow show a higher consistency with the Budyko framework, but lower accuracy (KGE) against observed streamflow at daily time step. However, aggregated streamflow volumes at monthly and yearly scale, as well as low flows at daily step are predicted with accuracy comparable with the setups with preferential flow. While a performance metric targeting the goodness-of-fit of daily observed and simulated streamflows (e.g., KGE) is key when considering flood applications, a plausible representation of soil moisture, evapotranspiration and groundwater dynamics is also essential for water resources management and climate adaptation. For LISFLOOD, we argue that using traditional performance metrics (e.g., KGE) for a model setup with preferential flow can be preserved by appropriately constraining other model parameters, without deviating excessively from the Budyko model. More generally, this study highlights the importance of integrating diagnostic tools like the Budyko framework into the evaluation and possibly the calibration of large-scale hydrological models, to ensure that high model performance is achieved not only for the right streamflow results, but also for the right reasons.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Code and data availability.

- The LISFLOOD model v4.1.1 is available at: https://github.com/ec-jrc/LISFLOOD-code/tree/v4.3.1 (last access: 15 January 2026) (https://doi.org/10.5281/zenodo.15430199) (Moschini et al. (2026))
- The model calibration tool can be found at: https://github.com/ec-jrc/LISFLOOD-calibration/tree/1.1.0 (last access: 15 January 2026) (https://doi.org/10.5281/zenodo.15430199) (Moschini et al. (2026))
- The LISFLOOD parameter maps are available at: https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/CEMS-EFAS/LISFLOOD_static_and_parameter_maps_for_EFAS/ (last access: 15 January 2026) PID: http://data.europa.eu/89h/f572c443-7466-4adf-87aa-c0847a169f23
- The meteorological forcing datasets are available at: https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/CEMS-EFAS/meteorological_forcings/. (last access: 15 January 2026) DOI: 10.2905/0BD84BE4-CEC8-4180-97A6-8B3ADAAC4D26
- The soil depth maps, data processing scripts, and the code of the LISFLOOD model, calibration tool, and data analysis used in this study are available at: https://github.com/r3dmos/LISFLOOD_Budyko, (https://doi.org/10.5281/zenodo.15430199) (Moschini et al. (2026)).

Appendix A: Calibrated parameters and study area

Study area (Figure A1) and model cells used in the Budyko analysis (Figure A2)

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

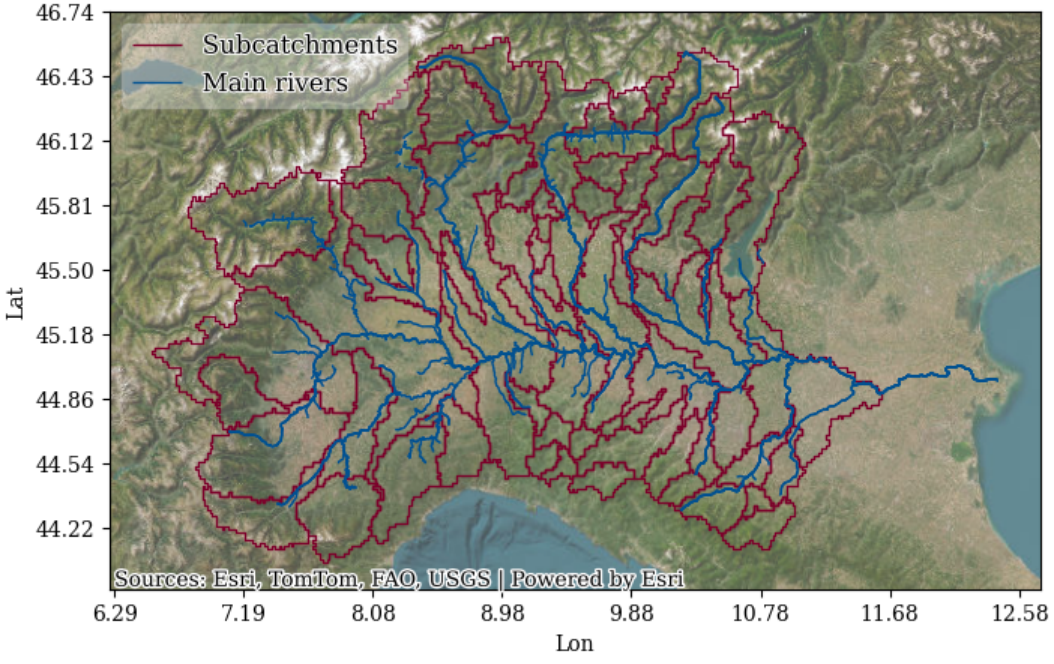

Figure A1. Study area, the Po river basin, with the 60 calibrated sub-catchments (boundaries in dark red) and the main rivers (dark blue lines). Satellite background was retried from Esri. (n.d.). World Imagery. https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

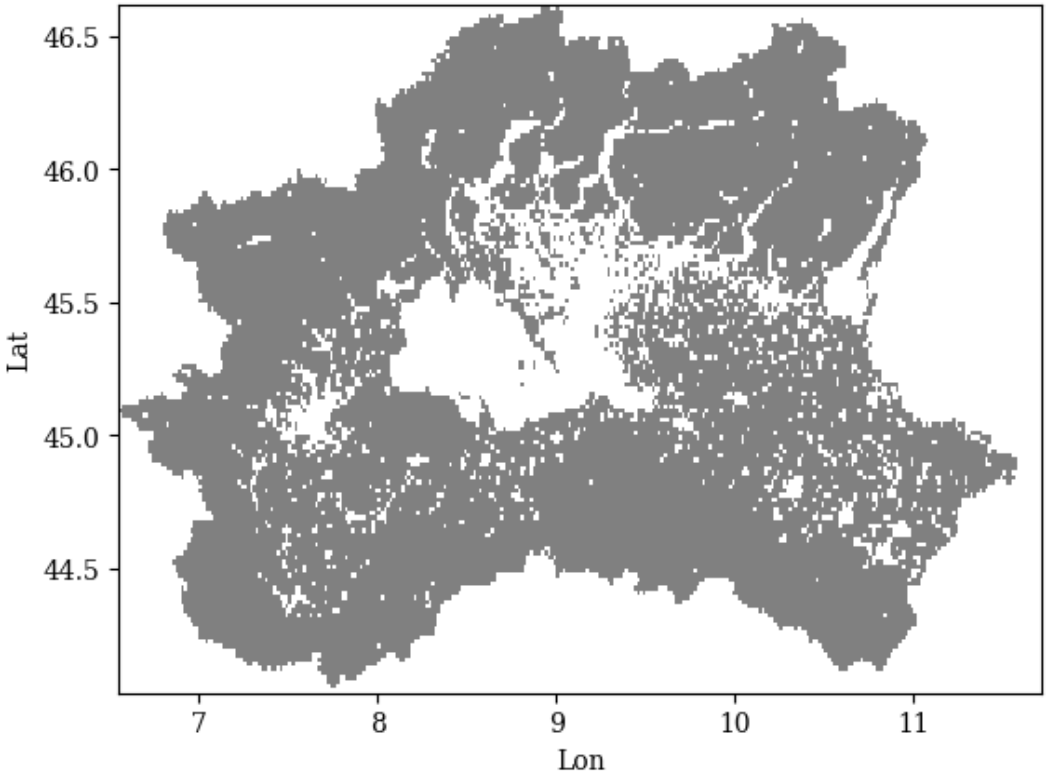

Figure A2. Model domain, showing cells included in the AET/Budyko analysis and the water balance diagnostics (Figure 10) in grey (the white area was excluded due to land cover criteria detailed in Section 2.5.2).

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Table A1. LISFLOOD model parameters subject to calibration (ECMWF, 2022)

Parameter name	Description	Min	Max	Default
bInfilt	Exponent in Xinanjiang equation for infiltration capacity of the soil [-]	0.01	5	0.5
PowerPrefFlow	Exponent in the empirical function describing the preferential flow (i.e. flow that bypasses the soil matrix and drains directly to the groundwater) [-]	0.5	8	4
SnowMeltCoef	Snow melt rate in degree day model equation [mm/(C day)]	2.5	6.5	4
UpperZoneTimeConstant	Time constant for upper groundwater zone [days]	0.01	40	10
GwPercValue	Maximum percolation rate from upper to lower groundwater zone [mm/day]	0.01	2	0.8
LowerZoneTimeConstant	Time constant for lower groundwater zone [days]	40	500	100
LZThreshold	Threshold to stop outflow from lower groundwater zone to the channel [mm]	0	30	10
GwLoss	Maximum loss rate out of lower groundwater zone expressed as a fraction of lower zone outflow [-]	0	1	0
QSplitMult	Multiplier to adjust discharge triggering floodplains flow [-]	0	20	2
CalChanMan1	Multiplier for channel Manning's coefficient n for riverbed [-]	0.5	2	1
CalChanMan2	Multiplier for channel Manning's coefficient n for floodplains [-]	0.5	5	1
adjustNormalFlood	Multiplier to adjust reservoir normal filling (balance between lower and upper limit of reservoir filling) [-]	0.01	0.99	0.8
ReservoirRnormqMult	Multiplier to adjust normal reservoir outflow [-]	0.25	2	1
LakeMultiplier	Multiplier to adjust lake outflow [-]	0.5	2	1

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Appendix B: Results

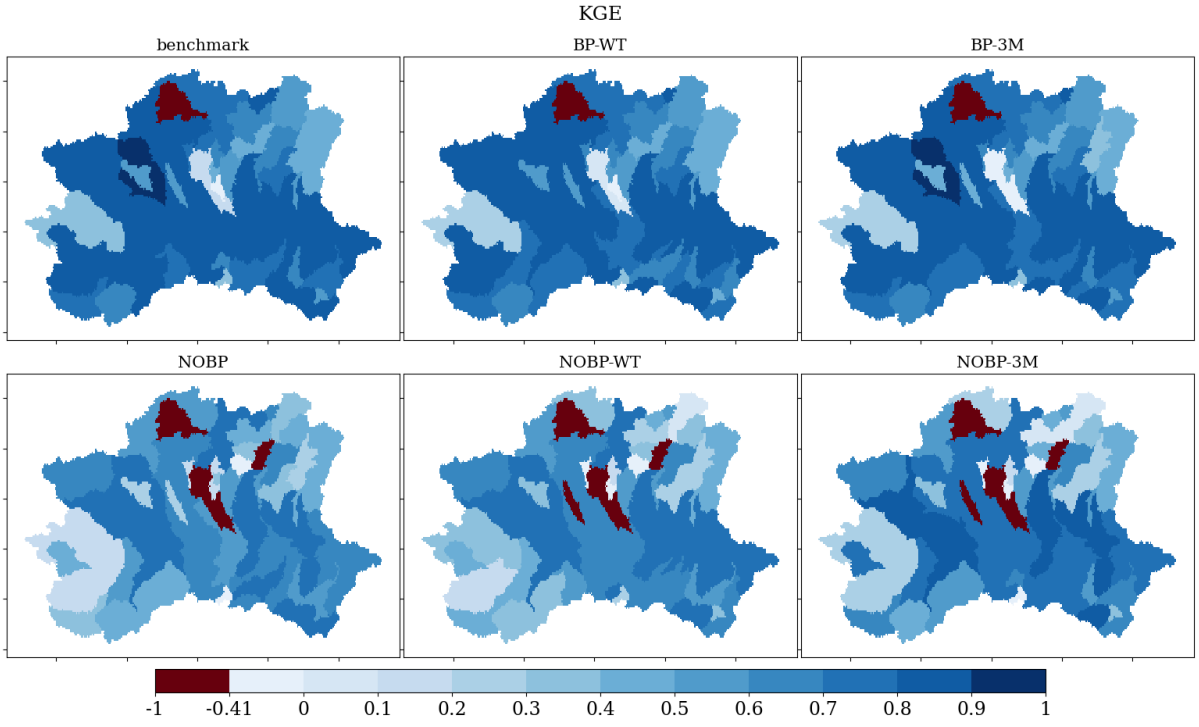

Figure B1. KGE for all the calibrated sub-catchments in all the 6 experiments (the optimal value for KGE is 1).

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

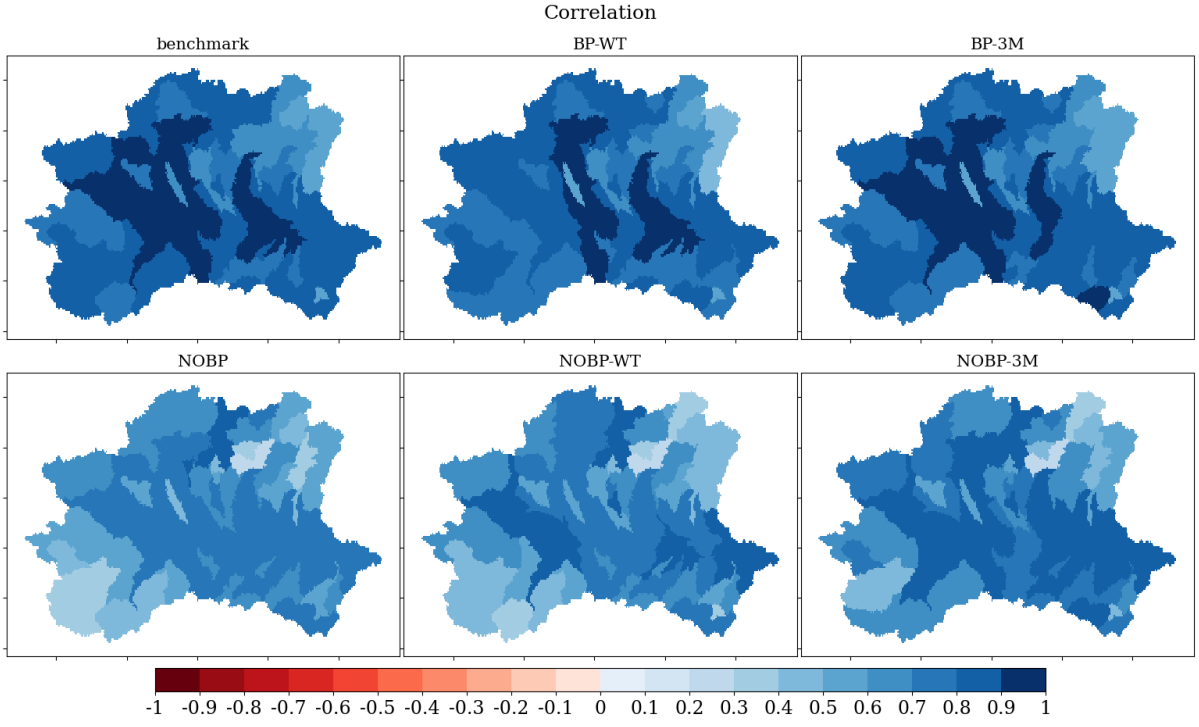

Figure B2. Correlation for all the calibrated sub-catchments in all the 6 experiments (optimal value is 1).

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

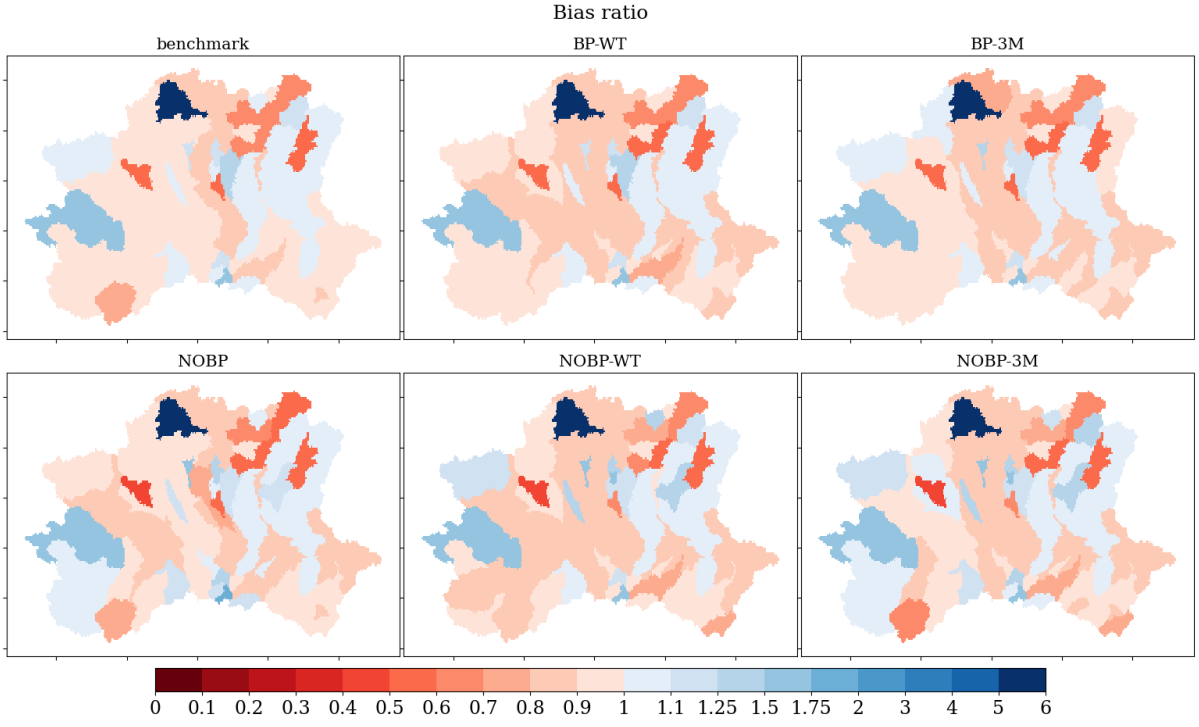

Figure B3. Bias ratio for all the calibrated sub-catchments in all the 6 experiments (optimal value is 1).

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

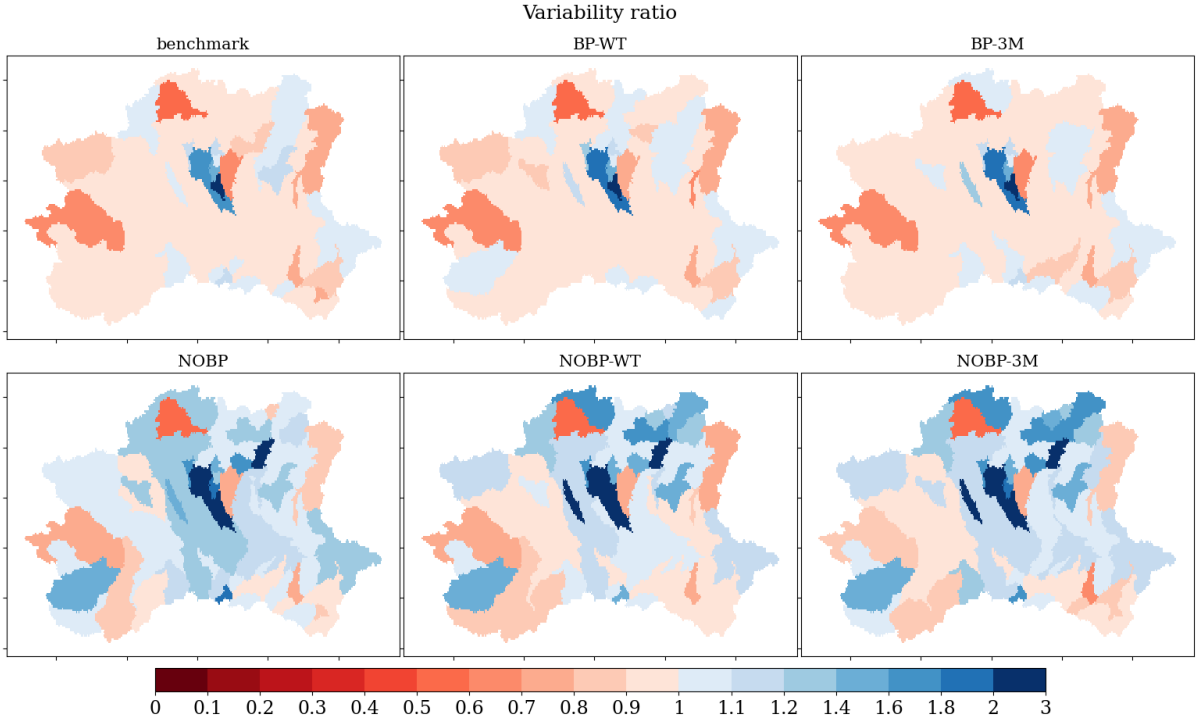

Figure B4. Variability ratio for all the calibrated sub-catchments in all the 6 experiments (optimal values is 1).

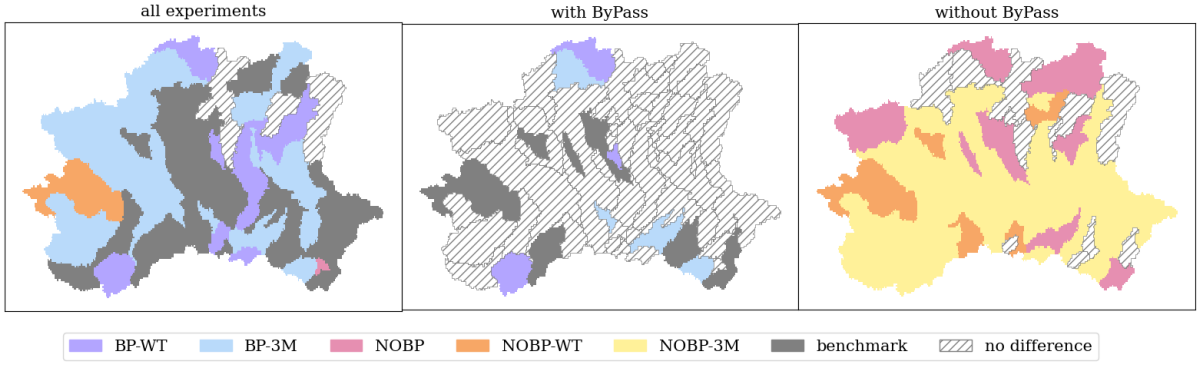

Figure B5. Best performing setup in terms of KGE (at daily scale) across sub-catchments, considering all the experiments (left panel), the subset of experiments with ByPass (middle), and without ByPass (right). When the difference of KGE among all the experiments is below 0.05, the performance of all setups is comparable (we define it as "no difference") and the sub-catchment area is filled with oblique grey lines.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

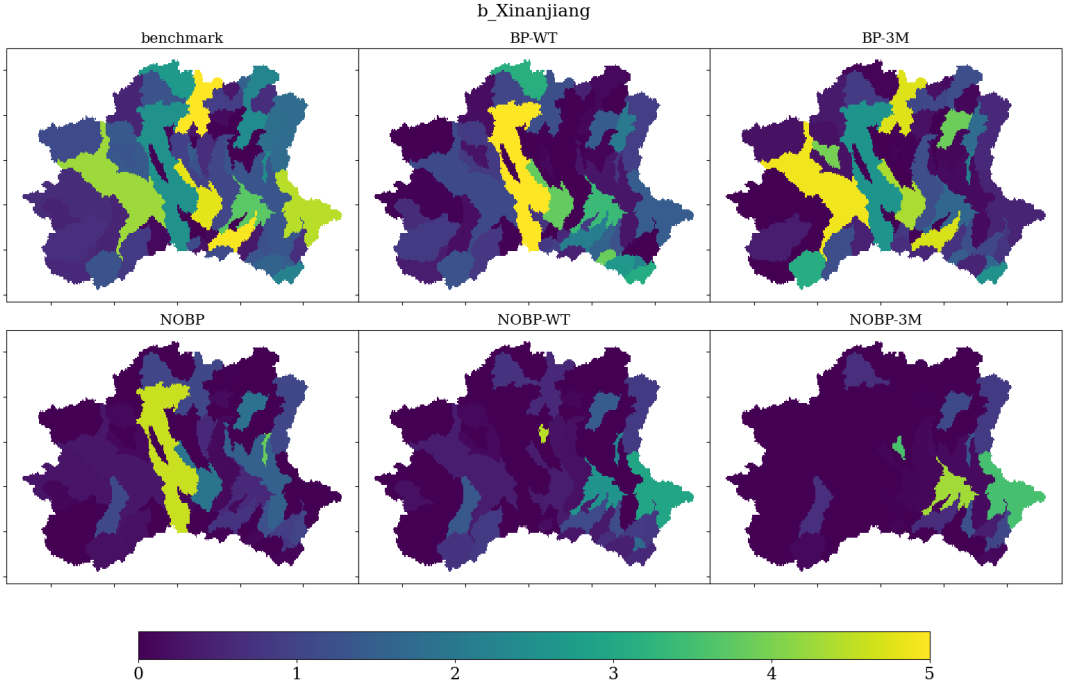

Figure B6. Calibrated parameter bInfilt for each experiment setup

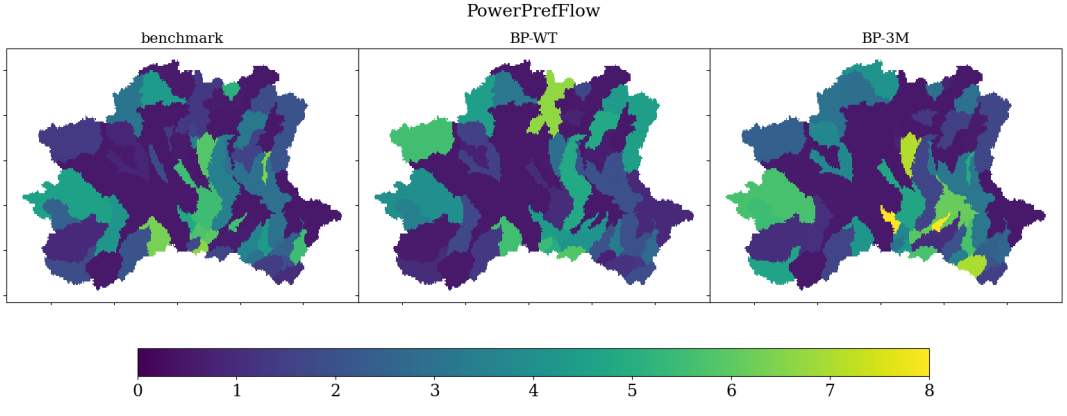

Figure B7. Calibrated parameter PowerPrefFlow for each experiment setup. The parameter was not used for the experiments without ByPass

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

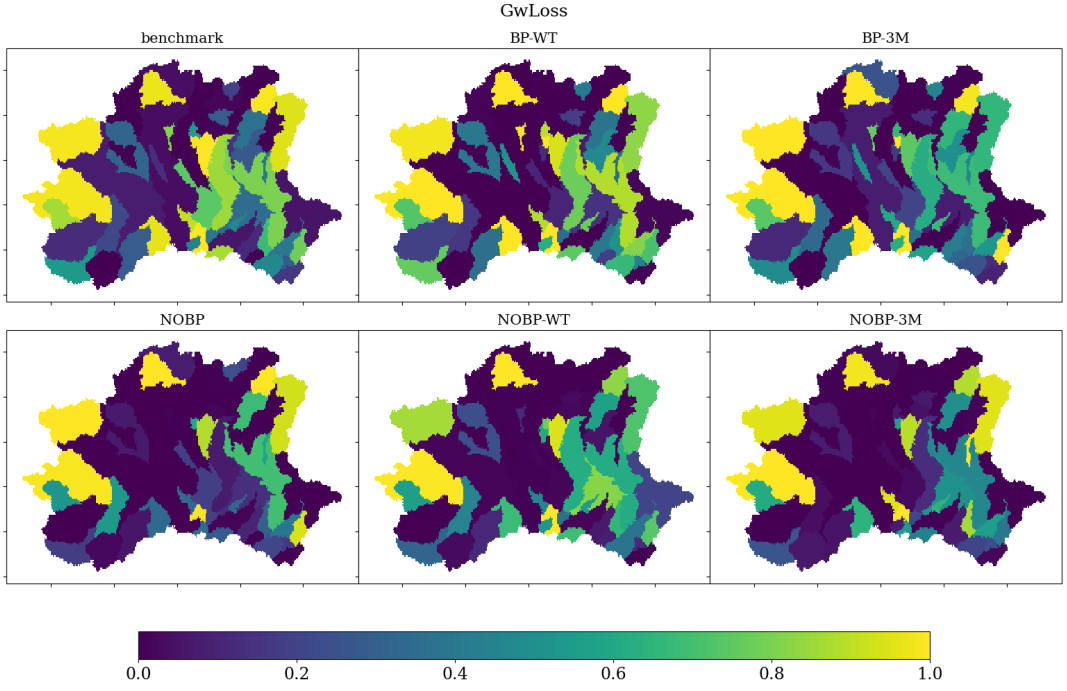

Figure B8. Calibrated parameter GwLoss for each experiment setup

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

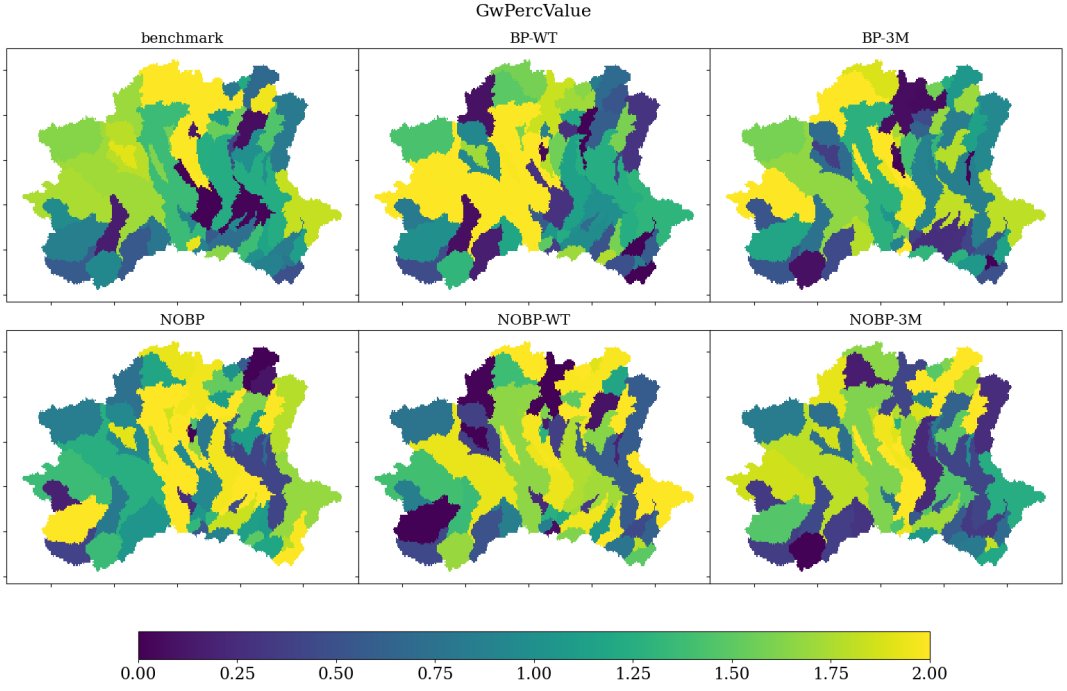

Figure B9. Calibrated parameter GwPerc for each experiment setup

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

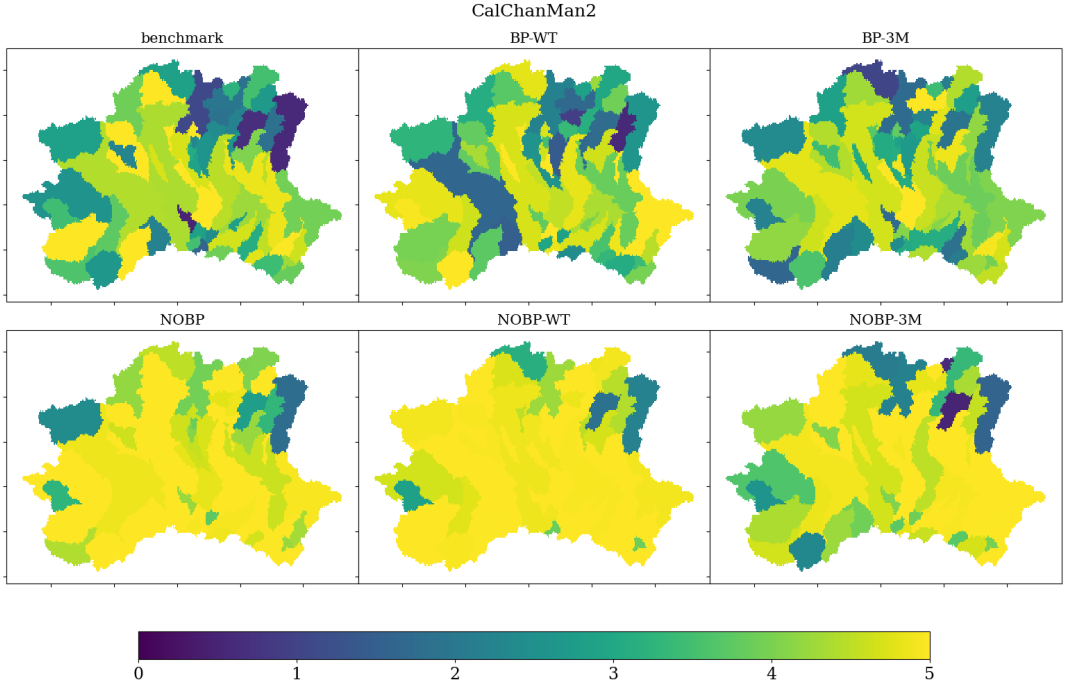

Figure B10. Calibrated parameter CalChanMan2 for each experiment setup

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

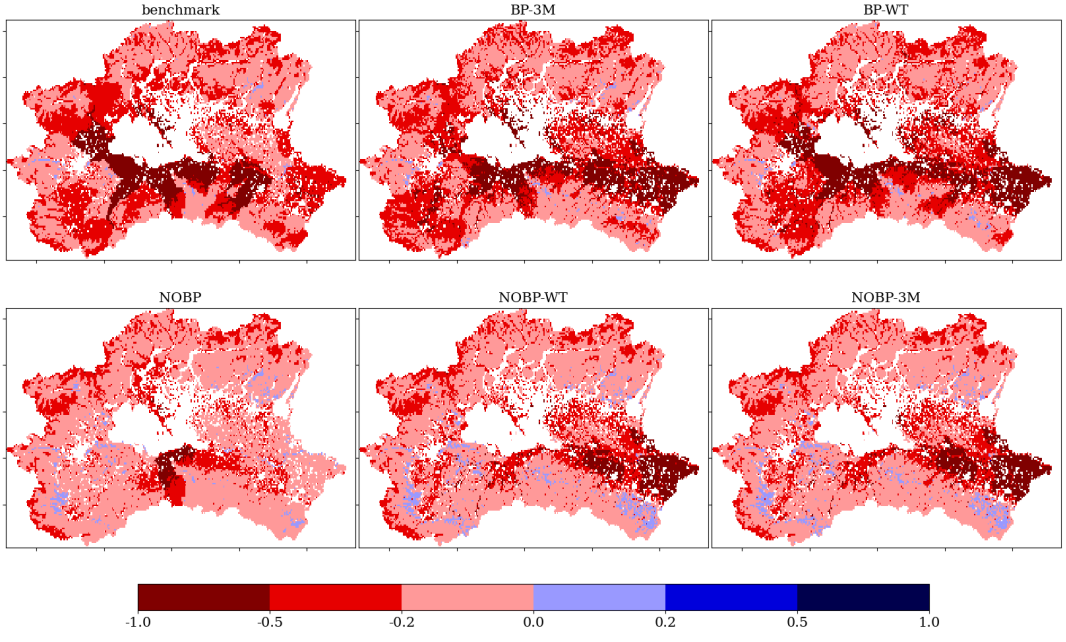

Figure B11. Relative difference of modelled annual average AET from the expected Budyko AET per pixel. Negative values (represented in red) mean that the modelled AET is lower compared to AET calculated using Budyko, while positive values (blue) mean that AET is higher compared to Budyko.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

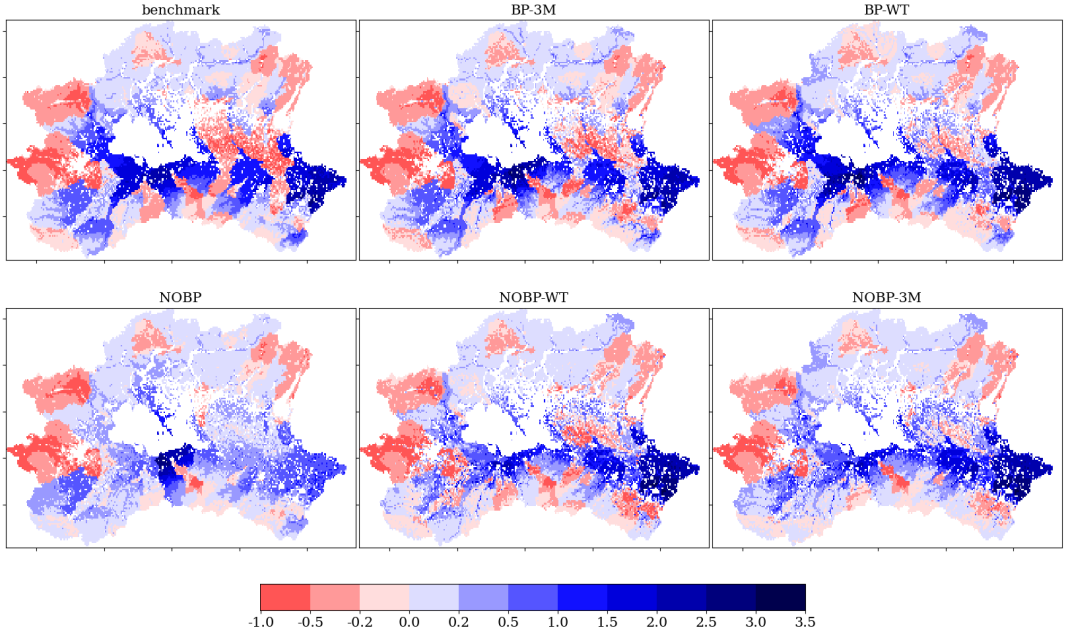

Figure B12. Relative difference of modelled annual average runoff (P-AET) from the expected Budyko runoff per pixel. Negative values (represented in red) mean that the modelled runoff is lower compared to runoff calculated using Budyko. Positive values (represented in blue) mean that the modelled runoff is higher compared to the runoff calculated using the Budyko equation.

Author contributions. FM and AP designed and conceptualized the study. FM run the experiments and analysed the results. FM wrote the paper and integrated feedback and additional analysis suggested by AP and AF. All authors significantly contributed to the realization of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We would like to thank the CEMS Hydrological Data Collection Centre for providing the historical streamflow data.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

References

Abbas, S. A., Bailey, R. T., White, J. T., Arnold, J. G., and White, M. J.: Estimation of groundwater storage loss using surface–subsurface hydrologic modeling in an irrigated agricultural region, Scientific Reports, 15, 8350, https://doi.org/10.1038/s41598-025-92987-6, 2025.

Andréassian, V. and Perrin, C.: On the ambiguous interpretation of the Turc-Budyko nondimensional graph, Water Resources Research, 48, https://doi.org/10.1029/2012WR012532, publisher: John Wiley & Sons, Ltd, 2012.

Andréassian, V., Perrin, C., and Michel, C.: Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models, J. Hydrol. (Amst.), 286, 19–35, 2004.

Ballarin, A. S., Oliveira, P. T. S., Marchezepe, B. K., Godoi, R. F., Campos, A. M., Campos, F. S., Almagro, A., and Meira Neto, A. A.: The impact of an open water balance assumption on understanding the factors controlling the long-term streamflow components, Water Resour. Res., 58, 2022.

Beretta, G. P., La Vigna, F., Camera, C. A. S., Gafà, R. M., Citrini, A., Martarelli, L., Monti, G. M., Roma, M., Silvi, A., Vitale, V., Fiori, C., Masetti, M., Pascarella, F., and Proietti, R.: Carta Idrogeologica d'Italia alla scala 1:500.000 – 4 Fogli, 2025.

Beven, K.: On hypothesis testing in hydrology, Hydrological Processes - HYDROL PROCESS, 15, 1655–1657, https://doi.org/10.1002/hyp.436, 2001.

Beven, K., Lane, S., Page, T., Kretzschmar, A., Hankin, B., Smith, P., and Chappell, N.: On (in)validating environmental models. 2. Implementation of a Turing-like test to modelling hydrological processes, Hydrological Processes, 36, e14 703, https://doi.org/10.1002/hyp.14703, publisher: John Wiley & Sons, Ltd, 2022.

Beven, K. J.: On hypothesis testing in hydrology: Why falsification of models is still a really good idea, WIREs Water, 5, e1278, https://doi.org/https://doi.org/10.1002/wat2.1278, _eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wat2.1278, 2018.

Beven, K. J. and Cloke, H. L.: Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et al, Water Resour. Res., 48, 2012.

Bisselink, B., Zambrano-Bigiarini, M., Burek, P., and de Roo, A.: Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions, J. Hydrol. Reg. Stud., 8, 112–129, 2016.

Bisselink, B., Bernhard, J., Gelati, E., Adamovic, M., Guenther, S., Mentaschi, L., Feyen, L., and de Roo, A.: Climate change and Europe's water resources, Tech. Rep. EUR 29951 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-10398-1, https://doi.org/10.2760/15553, jRC118586, 2020.

Brocca, L., Barbetta, S., Camici, S., Ciabatta, L., Dari, J., Filippucci, P., Massari, C., Modanesi, S., Tarpanelli, A., Bonaccorsi, B., et al.: A Digital Twin of the terrestrial water cycle: a glimpse into the future through high-resolution Earth observations, Frontiers in Science, 1, 1190 191, 2024.

Budyko, M.: Climate and Life, 1974.

Burek, P., Mubareka, S., Rojas Mujica, R., De Roo, A., Bianchi, A., Baranzelli, C., Lavalle, C., and Vandecasteele, I.: Evaluation of the effectiveness of Natural Water Retention Measures - Support to the EU Blueprint to Safeguard Europe's Waters, Scientific analysis or review, Policy assessment, Anticipation and foresight LB-NA-25551-EN-N, Joint Research Centre, Luxembourg (Luxembourg), ISBN 978-92-79-27021-5, ISSN 1831-9424, https://doi.org/10.2788/5528, 2012.

Burek, P., van der Knijff, J., and Roo, A. D.: LISFLOOD. Distributed Water Balance and Flood Simulation Model, https://doi.org/10.2788/24719, 2013.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Cammalleri, C., Micale, F., and Vogt, J.: On the value of combining different modelled soil moisture products for European drought monitoring, Journal of Hydrology, 525, 547–558, 2015.

Cammalleri, C., Vogt, J., and Salamon, P.: Development of an operational low-flow index for hydrological drought monitoring over Europe, Hydrological Sciences Journal, 62, 346–358, 2017.

Chen, X. and Sivapalan, M.: Hydrological Basis of the Budyko Curve: Data-Guided Exploration of the Mediating Role of Soil Moisture, Water Resources Research, 56, e2020WR028 221, https://doi.org/10.1029/2020WR028221, publisher: John Wiley & Sons, Ltd, 2020.

Choulga, M., Moschini, F., Mazzetti, C., Grimaldi, S., Disperati, J., Beck, H., Salamon, P., and Prudhomme, C.: Technical Note: Surface Fields for Global Environmental Modelling, EGUsphere, https://doi.org/10.5194/egusphere-2023-1306, 2023.

Cislaghi, A., Masseroni, D., Massari, C., Camici, S., and Brocca, L.: Combining a rainfall–runoff model and a regionalization approach for flood and water resource assessment in the western Po Valley, Italy, Hydrological Sciences Journal, 65, 348–370, https://doi.org/10.1080/02626667.2019.1690656, 2020.

Clothier, B., Green, S., and Deurer, M.: Preferential flow and transport in soil: progress and prognosis, European Journal of Soil Science, 59, 2–13, 2008.

Coron, L., Andréassian, V., Perrin, C., and Moine, N. L.: Graphical tools based on Turc-Budyko plots to detect changes in catchment behaviour, Hydrological Sciences Journal, 60, 1394–1407, https://doi.org/10.1080/02626667.2014.964245, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626667.2014.964245, 2015.

De Roo, A., Burek, P., Gentile, A., Udias, A., Bouraoui, F., Aloe, A., Bianchi, A., La Notte, A., Kuik, O., Elorza Tenreiro, J., Vandecasteele, I., Mubareka, S., Baranzelli, C., Van Der Perk, M., Lavalle, C., and Bidoglio, G.: A multi-criteria optimisation of scenarios for the protection of water resources in Europe: Support to the EU Blueprint to Safeguard Europe's Waters, Scientific report LB-NA-25552-EN-N, Joint Research Centre, Luxembourg (Luxembourg), ISBN 978-92-79-27025-3, ISSN 1831-9424, https://doi.org/10.2788/55540, 2012.

De Roo, A., Trichakis, I., Bisselink, B., Gelati, E., Pistocchi, A., and Gawlik, B.: The water-Energy-Food-ecosystem nexus in the Mediterranean: Current issues and future challenges, Front. Clim., 3, 2021.

De Roo, A., Bisselink, B., and Trichakis, I.: Water-Energy-Food-Ecosystems pathways towards reducing water scarcity in Europe, Tech. Rep. EUR 31680 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-68-08067-2, https://doi.org/10.2760/478498, jRC133439, 2023.

Döll, P., Hasan, H. M. M., Schulze, K., Gerdener, H., Börger, L., Shadkam, S., Ackermann, S., Hosseini-Moghari, S.-M., Müller Schmied, H., Güntner, A., and Kusche, J.: Leveraging multi-variable observations to reduce and quantify the output uncertainty of a global hydrological model: evaluation of three ensemble-based approaches for the Mississippi River basin, Hydrol. Earth Syst. Sci., 28, 2259–2295, 2024.

ECMWF: EFAS v5.0 - Calibration Methodology and Data, https://confluence.ecmwf.int/display/CEMS/EFAS+v5.0+-+Calibration+Methodology+and+Data, accessed May 25, 2025, 2022.

Fan, Y., Li, H., and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth, Science, 339, 940–43, 2013.

Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., and Otero-Casal, C.: Hydrologic Regulation of Plant Rooting Depth, Proceedings of the National Academy of Sciences of the United States of America, 114, 10 572–77, 2017.

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., Brooks, P. D., et al.: Hillslope Hydrology in Global Change Research and Earth System Modeling, Water Resources Research, 55, 1737–72, 2019.

Feyen, L., Kalas, M., and Vrugt, J. A.: Semi-distributed parameter optimization and uncertainty assessment for large-scale streamflow simulation using global optimization / Optimisation de paramètres semi-distribués et évaluation de l'incertitude pour la simulation de débits à grande échelle par l'utilisation d'une optimisation globale, Hydrol. Sci. J., 53, 293–308, 2008.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Ficchì, A., Perrin, C., and Andréassian, V.: Hydrological modelling at multiple sub-daily time steps: Model improvement via flux-matching, J. Hydrol. (Amst.), 575, 1308–1327, 2019.

Fluhrer, A., Baur, M. J., Piles, M., Bayat, B., Rahmati, M., Chaparro, D., Dubois, C., Hellwig, F. M., Montzka, C., Kübert, A., Mueller, M. M., Augscheller, I., Jonard, F., Schellenberg, K., and Jagdhuber, T.: Assessing evapotranspiration dynamics across central Europe in the context of land–atmosphere drivers, Biogeosciences, 22, 3721–3746, https://doi.org/10.5194/bg-22-3721-2025, 2025.

Fortin, F.-A., Rainville, F.-M. D., Gardner, M.-A., Parizeau, M., and Gagné, C.: DEAP: Evolutionary Algorithms Made Easy, The Journal of Machine Learning Research, 13, 2171–2175, 2012.

Garcia, F., Folton, N., and Oudin, L.: Which Objective Function to Calibrate Rainfall–runoff Models for Low-Flow Index Simulations?, Hydrological Sciences Journal, https://doi.org/10.1080/02626667.2017.1308511, 2017.

Gauch, M., Kratzert, F., Gilon, O., Gupta, H., Mai, J., Nearing, G., Tolson, B., Hochreiter, S., and Klotz, D.: In Defense of Metrics: Metrics Sufficiently Encode Typical Human Preferences Regarding Hydrological Model Performance, Water Resources Research, 59, e2022WR033 918, https://doi.org/10.1029/2022WR033918, publisher: John Wiley & Sons, Ltd, 2023.

Gentine, P., D'Odorico, P., Lintner, B. R., Sivandran, G., and Salvucci, G.: Interdependence of climate, soil, and vegetation as constrained by the Budyko curve, Geophysical Research Letters, 39, L19 404, https://doi.org/10.1029/2012GL053492, 2012.

Gleeson, T., Wagener, T., Döll, P., Zipper, S. C., West, C., Wada, Y., Taylor, R., Scanlon, B., Rosolem, R., Rahman, S., Oshinlaja, N., Maxwell, R., Lo, M.-H., Kim, H., Hill, M., Hartmann, A., Fogg, G., Famiglietti, J. S., Ducharne, A., de Graaf, I., Cuthbert, M., Condon, L., Bresciani, E., and Bierkens, M. F. P.: GMD perspective: The quest to improve the evaluation of groundwater representation in continental- to global-scale models, Geosci. Model Dev., 14, 7545–7571, 2021.

Gnann, S., Reinecke, R., Stein, L., Wada, Y., Thiery, W., Müller Schmied, H., Satoh, Y., et al.: Functional Relationships Reveal Differences in the Water Cycle Representation of Global Water Models, Nature Water, 1, 1079–90, 2023.

Granata, F. and Di Nunno, F.: Pathways for Hydrological Resilience: Strategies for Adaptation in a Changing Climate, Earth Systems and Environment, https://doi.org/10.1007/s41748-024-00567-x, 2025.

Greve, P., Burek, P., and Wada, Y.: Using the Budyko Framework for Calibrating a Global Hydrological Model, Water Resources Research, 56, https://doi.org/10.1029/2019wr026280, 2020.

Gunkel, A. and Lange, J.: Water scarcity, data scarcity and the Budyko curve—An application in the Lower Jordan River Basin, Journal of Hydrology: Regional Studies, 12, 136–149, https://doi.org/https://doi.org/10.1016/j.ejrh.2017.04.004, 2017.

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, J. Hydrol. Reg. Stud., 55, 101 944, 2024.

Hamman, J. J., Nijssen, B., Bohn, T. J., Gergel, D. R., and Mao, Y.: The Variable Infiltration Capacity model version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility, Geoscientific Model Development, 11, 3481–3496, 2018.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., et al.: SoilGrids250m: Global gridded soil information based on machine learning, PLoS one, 12, e0169 748, 2017.

Herbst, M., Gupta, H. V., and Casper, M. C.: Mapping model behaviour using Self-Organizing Maps, Hydrology and Earth System Sciences, 13, 395–409, https://doi.org/10.5194/hess-13-395-2009, 2009.

Huang, J., Sehgal, V., Alvarez, L. V., Brocca, L., Cai, S., Cheng, R., Cheng, X., Du, J., El Masri, B., Endsley, K. A., Fang, Y., Hu, J., Jampani, M., Kibria, M. G., Koren, G., Li, L., Liu, L., Mao, J., Moreno, H. A., Rigden, A., Shi, M., Shi, X., Wang, Y., Zhang, X., and Fisher, J. B.:

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Remotely sensed high-resolution soil moisture and evapotranspiration: Bridging the gap between science and society, Water Resour. Res., 61, 2025.

Jaramillo, F. and Destouni, G.: Developing water change spectra and distinguishing change drivers worldwide, Geophys. Res. Lett., 41, 8377–8386, 2014.

Jensen, L., Dill, R., Balidakis, K., Grimaldi, S., Salamon, P., and Dobslaw, H.: Global 0.05° water storage simulations with the OS LISFLOOD hydrological model for geodetic applications, Geophys. J. Int., 241, 1840–1852, 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, 264–77, 2012.

Knoben, W. J. M., Freer, J. E., and Woods, R.: Technical note: Inherent benchmark ? Comparing Nash-Sutcliffe Kling-Gupta efficiency scores, Hydrol. Earth Syst. Sci, 23, 4323–4433, 2019.

Kocian, L. and Mohanty, B. P.: Characterizing large-scale preferential flow across Continental United States, Vadose Zone Journal, 23, e20 316, 2024.

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, e2020WR028 658, 2021.

Kraft, B., Jung, M., Körner, M., Koirala, S., and Reichstein, M.: Towards hybrid modeling of the global hydrological cycle, Hydrol. Earth Syst. Sci., 26, 1579–1614, 2022.

Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resour. Res., 49, 360–379, 2013.

Kumar, V., Jahangeer, J., Singh, R., and Dikshit, P. K. S.: Editorial: Advancement in hydrological modeling and water resources management for achieving Sustainable Development Goals (SDGs), Frontiers in Water, Volume 7 - 2025, https://doi.org/10.3389/frwa.2025.1599795, 2025.

Le Moine, N., Andréassian, V., Perrin, C., and Michel, C.: How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments, Water Resour. Res., 43, 2007.

Li, H.-Y., Sivapalan, M., Tian, F., and Harman, C.: Functional approach to exploring climatic and landscape controls of runoff generation: 1. Behavioral constraints on runoff volume, Water Resources Research, 50, 9300–9322, https://doi.org/10.1002/2014WR016307, 2014.

Livani, M., Petracchini, L., Benetatos, C., Marzano, F., Billi, A., Carminati, E., Doglioni, C., et al.: Subsurface Geological and Geophysical Data from the Po Plain and the Northern Adriatic Sea (north Italy), Earth System Science Data, https://doi.org/10.5194/essd-15-4261-2023, 2023.

Matthews, G., Baugh, C., Barnard, C., Carton De Wiart, C., Colonese, J., Grimaldi, S., Ham, D., Hansford, E., Harrigan, S., Heiselberg, S., Hooker, H., Hossain, S., Mazzetti, C., Milano, L., Moschini, F., O'Regan, K., Pappenberger, F., Pfister, D., Rajbhandari, R. M., Salamon, P., Ramos, A., Shelton, K., Stephens, E., Tasev, D., Turner, M., van den Homberg, M., Wittig, J., Zsótér, E., and Prudhomme, C.: Chapter 15 - On the operational implementation of the Global Flood Awareness System (GloFAS), in: Flood Forecasting (Second Edition), edited by Adams, T. E., Gangodagamage, C., and Pagano, T. C., pp. 299–350, Academic Press, second edition edn., ISBN 978-0-443-14009-9, https://doi.org/https://doi.org/10.1016/B978-0-443-14009-9.00014-6, 2025a.

Matthews, G., Baugh, C., Barnard, C., De Wiart, C. C., Colonese, J., Decremer, D., Grimaldi, S., Hansford, E., Mazzetti, C., O'Regan, K., Pappenberger, F., Ramos, A., Salamon, P., Tasev, D., and Prudhomme, C.: Chapter 14 - On the operational implementation of the European Flood Awareness System (EFAS), in: Flood Forecasting (Second Edition), edited by Adams, T. E., Gangodagamage, C., and Pagano, T. C.,

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

pp. 251–298, Academic Press, second edition edn., ISBN 978-0-443-14009-9, https://doi.org/https://doi.org/10.1016/B978-0-443-14009-9.00005-5, 2025b.

Melsen, L. A., Puy, A., Torfs, P. J. J. F., and Saltelli, A.: The rise of the Nash-Sutcliffe efficiency in hydrology, Hydrol. Sci. J., 70, 1248–1259, 2025.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., and Gupta, H.: Are we unnecessarily constraining the agility of complex process-based models?, Water Resources Research, 51, 716–728, 2015.

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., and Kumar, R.: On the choice of calibration metrics for "high-flow" estimation using hydrologic models, Hydrology and Earth System Sciences, 23, 2601–2614, https://doi.org/10.5194/hess-23-2601-2019, 2019.

Montanari, A.: Hydrology of the Po River: looking for changing patterns in river discharge, Hydrology and Earth System Sciences (HESS), 16, 3739–3747, https://doi.org/10.5194/hess-16-3739-2012, 2012.

Moschini, F., Ficchì, A., and Pistocchi, A.: Hydrological Auditing of LISFLOOD: Impacts of Model Setup on Water Balance Components in the Po River Basin - supporting data and software, https://doi.org/10.5281/zenodo.15430199, 2026.

Musolino, D., Vezzani, C., and Massarutto, A.: Drought Management in the Po River Basin, Italy, in: Drought, Wiley, https://doi.org/10.1002/9781119017073.ch11, 2018.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T. Y., Weitzner, D., and Matias, Y.: Global prediction of extreme floods in ungauged watersheds, Nature, 627, 559–563, https://doi.org/10.1038/s41586-024-07145-1, 2024.

Orth, R. and Seneviratne, S. I.: Introduction of a simple-model-based land surface dataset for Europe, Environ. Res. Lett., 10, 044 012, 2015.

Pelletier, A. and Andréassian, V.: On constraining a lumped hydrological model with both piezometry and streamflow: results of a large sample evaluation, Hydrol. Earth Syst. Sci., 26, 2733–2758, 2022.

Pfannerstill, M., Guse, B., and Fohrer, N.: Smart low flow signature metrics for an improved overall performance evaluation of hydrological models, Journal of Hydrology, 510, 447–458, https://doi.org/https://doi.org/10.1016/j.jhydrol.2013.12.044, 2014.

Pistocchi, A., Gelati, E., Beck, H., Lavalle, C., Bisselink, B., and Feher, J.: Climate Change and the Danube Region: Implications for Energy Production and Environmental Impact, Tech. Rep. LB-NA-27700-EN-N, Joint Research Centre (European Commission), ISBN 978-92-79-54582-5, ISSN 1831-9424, https://doi.org/10.2788/375680, 2015.

Pistocchi, A., Dorati, C., Aloe, A., Ginebreda, A., and Marcé, R.: River pollution by priority chemical substances under the Water Framework Directive: A provisional pan-European assessment, Science of The Total Environment, 662, 434–445, https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.12.354, 2019.

Pistocchi, A., Bisselink, B., Moschini, F., Quarante, E., Trichakis, Y., Bouraoui, F., Grizzetti, B., Hidalgo González, I., Udias, A., and Zal, N.: Human freshwater appropriation in Europe: a preliminary assessment of current conditions, knowledge gaps and resilience prospect, Tech. rep., Publications Office of the European Union, Luxembourg, jRC141278, 2024.

Quaranta, E., Dorati, C., and Pistocchi, A.: Water, energy and climate benefits of urban greening throughout Europe under different climatic scenarios, Scientific Reports, 11, 12 163, https://doi.org/10.1038/s41598-021-88141-7, 2021.

Rajib, A., Merwade, V., and Yu, Z.: Rationale and efficacy of assimilating remotely sensed potential evapotranspiration for reduced uncertainty of hydrologic models, Water Resour. Res., 54, 4615–4637, 2018.

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., Eisner, S., Müller Schmied, H., Sutanudjaja, E. H., Warrach-Sagi, K., and Attinger, S.: Toward seamless hydrologic predictions across spatial scales, Hydrol. Earth Syst. Sci., 21, 4323–4346, 2017.

https://doi.org/10.5194/egusphere-2026-423
Preprint. Discussion started: 12 February 2026

Santos, L., Thirel, G., and Perrin, C.: Technical Note: Pitfalls in Using Log-Transformed Flows within the KGE Criterion, Hydrology and Earth System Sciences, 22, 4583–91, 2018.

Senbeta, T. B., Karamuz, E., Kochanek, K., Napiórkowski, J. J., and Romanowicz, R. J.: Budyko-Based Approach for Modelling Water Balance Dynamics Considering Environmental Change Drivers in the Vistula River Basin, Poland, Hydrological Sciences Journal, 68, 655–69, 2023.

Terrier, M., Perrin, C., De Lavenne, A., Andréassian, V., Lerat, J., and Vaze, J.: Streamflow naturalization methods: a review, Hydrological Sciences Journal, 66, 12–36, 2021.

Thiemig, V., Gomes, G. N., Skøien, J. O., Ziese, M., Rauthe-Schöch, A., Rustemeier, E., Rehfeldt, K., Walawender, J. P., Kolbe, C., Pichon, D., Schweim, C., and Salamon, P.: EMO-5: a high-resolution multi-variable gridded meteorological dataset for Europe, Earth Syst. Sci. Data, 14, 3249–3272, 2022a.

Thiemig, V., Gomes, G. N., Skøien, J. O., Ziese, M., Rauthe-Schöch, A., Rustemeier, E., Rehfeldt, K., et al.: EMO-5: A High-Resolution Multi-Variable Gridded Meteorological Dataset for Europe, Earth System Science Data, 14, 3249–72, 2022b.

Todini, E.: The ARNO Rainfall–runoff Model, Journal of Hydrology, 175, 339–82, 1996.

Turc, L.: Le bilan d'eau des sols: Relation entre la precipitations, l'evaporation et l'ecoulement, Annales Agronomiques, Serie A, 5, 491–495, https://cir.nii.ac.jp/crid/1370004233293014026, 1954.

Van Der Knijff, J., Younis, J., and De Roo, A.: LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, International Journal of Geographical Information Science, 24, 189–212, 2010.

Vezzoli, R., Mercogliano, P., Pecora, S., Zollo, A., and Cacciamani, C.: Hydrological simulation of Po River (North Italy) discharge under climate change scenarios using the RCM COSMO-CLM, Science of The Total Environment, 521-522, 346–358, https://doi.org/https://doi.org/10.1016/j.scitotenv.2015.03.096, 2015.

Wagener, T. and Pianosi, F.: What has Global Sensitivity Analysis ever done for us? A systematic review to support scientific advancement and to inform policy-making in earth system modelling, Earth Sci. Rev., 194, 1–18, 2019.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res., 44, 2008.

Zaerpour, M., Hatami, S., Ballarin, A. S., Knoben, W. J. M., Papalexiou, S. M., Pietroniro, A., and Clark, M. P.: Impacts of agriculture and snow dynamics on catchment water balance in the U.S. and Great Britain, Commun Earth Environ, 5, 2024.

Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F. A., and Beck, H.: The impact of lake and reservoir parameterization on global streamflow simulation, J. Hydrol. (Amst.), 548, 552–568, 2017.

Zhang, W., Wu, Y., Guo, H., Li, W., An, J., and Wang, S.: Temporal and spatial response of agricultural drought to meteorological drought in inner Mongolia plateau inland river basin, Scientific Reports, 15, 28 225, https://doi.org/10.1038/s41598-025-14236-0, 2025.