https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

The water balance quagmire: a zoom *out* to close *in* on climate forcing biases

Matteo Rosales[1,2], Fanny J. Sarrazin[1], Frédéric Hendrickx[2], and Vazken Andréassian[1]

[1]Université Paris-Saclay, INRAE, HYCAR, Antony, France
[2]Électricité de France, R&D, Chatou, France

Correspondence: Matteo Rosales (matteo.rosales@inrae.fr)

Abstract. Non closure of the interannual water balance is a major issue in hydrological modelling. Because the potential causes (e.g. forcing biases, groundwater exchanges, anthropogenic influences, streamflow uncertainties) are so inextricably intertwined, many modelers resort to the calibration of 'sweep' parameters, leading to *ad hoc* corrections of water balance components to improve performance metrics. Such practices can be beneficial when applied 'for the right reasons', although this is seldom possible to verify. In this paper, we look at the untangling of water balance closure through the lens of scale: we hypothesize that the spatial variability of the water balance anomaly signal can be used to separate the different causes of non-closure and, in particular, to discriminate between those which have local determinants (such as groundwater leaks) and those involving non-local or systematic drivers (such as forcing biases). To test the idea, we compute long-term water balance anomalies for 3641 near-natural gauged catchments across Europe, using the CERRA-Land climate reanalyses (1984-2024). We then apply Thin Plate Splines to extract the *mesoscale* imbalance, a continuous field of low-frequency variability. We show that the extracted signal is most likely driven by the forcing biases of the CERRA-Land reanalysis. The approach is a step forward towards extracting the contribution of climate forcing biases across gauged catchments and beyond. The hypothesis and the associated methodology provide a tool to contain the uncertainties of 'sweeping', paving the way for hydrologically-grounded mesoscale corrections of the forcing variables in large-scale distributed hydrological models.

1 Introduction

1.1 On closing the long-term water balance

The interannual water balance (IWB) closure equation writes as follows:

$$P - E_A + \Delta S + I_{GF} + I_H = Q \tag{1}$$

where P, E_A and Q refer to the interannual means of total annual precipitation, actual evaporation and specific discharge at the catchment outlet; ΔS is the mean annual change in storage; I_{GF} broadly refers to intercatchment groundwater flows (Le Moine et al., 2007) and I_H to human influences. Each of these components is defined in $mm\ yr^{-1}$. Whilst P, E_A and Q are positive quantities, both ΔS, I_{GF} and I_H may take either positive (gain) or negative (loss) values.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

Any error in the estimation of each of these fluxes entails water balance non-closure. We define the IWB non-closure anomaly Δ_{IWB} as follows:

$$\Delta_{IWB} = Q - (P - E_A + \Delta S + I_{GF} + I_H) \quad (2)$$

The IWB is only a special case for the more general closure problem applied to any spatial or temporal scale, which Beven (2006) described as the Holy Grail of scientific hydrology. Yet, because the complexity of the water balance decreases as the timestep increases (Coutagne and de Martonne, 1934), addressing the water balance closure at the *interannual timestep* allows to filter out those transient phenomena (e.g. annual aquifer fluctuations) which make the parent problem (searching for the Holy Grail) considerably more challenging.

Nevertheless, IWB non-closure is still a ubiquitous issue. Measuring or modelling even the long-term average fluxes across catchment boundaries is arduous, as observations of the hydrological processes are often too scarce and too uncertain. In spite of technological developments, some sources of water balance non-closure have been discussed for over a century, such as the near-impossible estimation of IGF, unaccounted human influences, long-term storage fluctuations, biased climate forcings, streamflow measurement errors (Lugeon, 1928; Penman, 1950; Flerchinger and Cooley, 2000; Rakovec et al., 2016; Safeeq et al., 2021; Luo et al., 2025).

1.2 Sweep parameters and the risk of heedless water balance closure

Why does it matter so much that we achieve long-term water balance closure over the historical period? Given the widespread deployment of hydrological models for climate change impact studies (Beniston, 2012; Vezzoli et al., 2015; Bisselink et al., 2018; Sauquet et al., 2024; Marx et al., 2018), our inability to close the IWB raises concerns about their suitability for temporal extrapolation, as IWB closure is a fundamental prerequisite of model calibration. Popular performance metrics in hydrology often include - implicitly or explicitly - some measure of the average bias - namely, the average mismatch between observed and simulated discharge (Gupta et al., 2009). When computed for long timeseries, bias can also be interpreted as a measure of the IWB closure, as shown through the following identity:

$$\begin{aligned} &Q_{obs} - Q_{sim} \\ &= Q_{obs} - (P - E_A + \Delta S + I_{GF} + I_H) \\ &= \Delta_{IWB} \end{aligned} \quad (3)$$

Therefore, many hydrological models include some lever to force IWB closure as a means of increasing overall model performance. Generally, modellers rely on calibration to tune the average partition of rainfall between discharge and evaporation. Sometimes, models also include additional parameters which are specifically designed to correct (or account for) the causes of long-term non-closure. Such parameters may come in the form of additive or multiplicative corrections of the forcing variables (Le Moine et al., 2007; Lussana et al., 2018), the simulated streamflow (Clerc-Schwarzenbach et al., 2026) or the catchment surface area (Rouhier, 2018), but also in the form of the outright 'removal' or 'creation' of water at some stage of the hydrological process (Perrin et al., 2003; Samaniego et al., 2010) to account for I_{GF}. Because modellers cannot address all the

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

causes for non-closure at once, only one parameter is often used to freely adjust the water balance and to compensate for all non-closure causes at once, which equates to sweeping all the hydrological unknowns under the rug of a single parameter, hence the notion of 'sweeping' (Clerc-Schwarzenbach et al., 2026).

Regrettably, 'sweep' parameters are often calibrated alongside other parameters, thereby affecting the entire model (Andréassian et al., 2001; Oudin et al., 2006; Kavetski et al., 2006b; Vrugt et al., 2008). Consequently, indiscriminate sweeping can lead to unrobust parameter selection, which undermines a model's potential for temporal and spatial (regionalisation) extrapolation (Kavetski et al., 2006a; Coron et al., 2012; Clerc-Schwarzenbach et al., 2026). Moreover, one may argue that the unrestricted calibration of sweep parameters hinders the progress of hydrological science, because it curtails our ability to falsify inadequate model structures. By design, these parameters are adjusted to compensate for any cause of non-closure. In doing so, they can obscure the physical meaning of model performance. If all models are allowed to enforce IWB closure through additional calibrated degrees of freedom, they might reproduce average streamflow equally well, regardless of whether the underlying process representation is realistic. Consequently, 'sweeping' prevents us from using IWB non-closure as a diagnostic criterion to identify and reject the poorest - or "non-behavioural" (Beven, 2001) - models. The apparent convergence of model performance would therefore come at the expense of model discrimination and physical interpretability.

1.3 Making sure we sweep the 'right' fluxes for the 'right' reasons

In spite of the risks, and as Clerc-Schwarzenbach et al. (2026) demonstrated, 'sweep' parameters can still provide a robust and sensible tool, especially when one succeeds in making them 'right for the right reasons'. Targeted and informed corrections of the forcing biases can help constrain our modelling tools and their structural uncertainties (Renard et al., 2011). How, then, can we make sure to only sweep what must be? In other words, how can we identify and quantify IWB non-closure causes among the long list of likely candidates?

The answer to this question changes depending on the spatial scale and the modelling paradigm. Closing the IWB for a lumped model of a small experimental watershed with comprehensive monitoring data, raises different issues compared to IWB closure of distributed models over entire continents. In the former case, lower spatial heterogeneity (Kane and Yang, 2004) entails higher confidence in the observed fluxes (discharge, precipitation and evaporation) and the storage terms, thus making it possible to better unwind the multiple non-closure causes and to identify precisely which ones might prevail (e.g. deep percolation (Chauvin et al., 2011) or permafrost melt (Zhuravin, 2004)).

In contrast, large-scale distributed models are commonly calibrated against extensive datasets comprising several hundred to several thousand gauged catchments (Troy et al., 2008; Samaniego et al., 2017). Data scarcity and heterogeneity severely limit the ability to disentangle the multiple sources of water balance non-closure, and thus to calibrate the value of sweep parameters. More importantly, the deployment of distributed models hinges on the ability to regionalize sweep parameters in ungauged areas and across computational units. Because water balance non-closure is often catchment-specific and reflects local processes and local data limitations, these parameters tend to lack transferability in space (Degenne et al., 2026). As a result, the closure of the IWB is rarely satisfied over the ungauged domain, at the level of individual grid cells or computational units in large-scale distributed models. This limitation is reflected in model evaluation studies. For example, several pan-European

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

hydrological models - including VIC-WUR, E-HYPE and EFAS-v5/LISFLOOD - exhibit substantial streamflow biases, with reported deviations reaching up to 50% of observed discharge in some regions (Copernicus Climate Change Service, 2021; ECMWF, 2024). Samaniego et al. (2019, 2017) likewise compared the performances of distributed hydrological models, respectively across the European continent (mHM, VIC, PCRGLOBWB and Noah-MP) and the Rhine basin at Lobith (mHM, PCR-GLOBWB and WaterGAP), revealing similar issues with average biases.

Thereon, the question remains of how to make sure to only sweep what must be, whenever we deal with ungauged basins, especially when the target domain is so big that we cannot afford a thorough investigation of non-closure causes. The best way to answer this question is actually to rephrase it: to choose what to sweep means to choose *where* to adjust *which* variable of the IWB closure. As explained below, rephrasing the IWB closure problem is also hinting to the solution, because asking "*where*" sheds unexpected light on the answer to the "*which*" question.

1.4 Zooming-out to unwind water balance non-closure?

By examining the *where* in relation to the *which*, we do not mean to assess non-closure causes across thousands of mesh units, a task which we previously deemed intractable. Rather, we mean to look at the spatial patterns of IWB non-closure across the gauged domain. We make the assumption that each non-closure cause has a distinct spatial variability, which means we can run spatial analyses of the variability to deconvolve the non-closure problem.

Specifically, precipitation products may carry systematic and spatially coherent biases across mountain ranges or latitude bands (e.g. wind-induced snowfall undercatch (MeteoSwiss, 2013; Prein and Gobiet, 2017; Lussana et al., 2018; Hiebl and Frei; Adam et al., 2006)). Consequently, large-sample studies have pointed to forcing biases as likely explanations for spatially coherent IWB signals, especially over mountains and snowier latitudes. Wang et al. (2014) made use of 370 small to medium-sized gauged basins to study water balance closure across Canada, revealing the paramount role of precipitation biases along the Canadian Cordillera, but also the likely impact of snow undercatch in Northern Canada. Rakovec et al. (2016) (section 3.(e), Figure 9) addressed water-imbalance across 179 basins in Europe; they found that water balance non-closure was generally independent of physiographic catchment properties, suggesting the paramount role of forcing biases. Several other water balance studies have confirmed the prevalence of precipition and potential evaporation biases at continental (Landerer et al., 2010; Troy et al., 2011; Zheng et al., 2025) and global scales (Sahoo et al., 2011; Zhang et al., 2018; Lehmann et al., 2022; Tan et al., 2022; Luo et al., 2025).

By contrast, Coxon et al. (2015), examined streamflow measurement uncertainties across 500 UK gauging stations, and showed "no clear spatial patterns" in the distribution of uncertainties, whose mean value was 0. In terms of IGF, regional groundwater transfers have been documented (Schaller and Fan, 2009); however, their influence on water balance declines as catchment size increases (Fan, 2019), implying that water exports and imports might balance out beyond a certain scale. Consequently, we build on the variability assumption to make the following hypothesis:

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

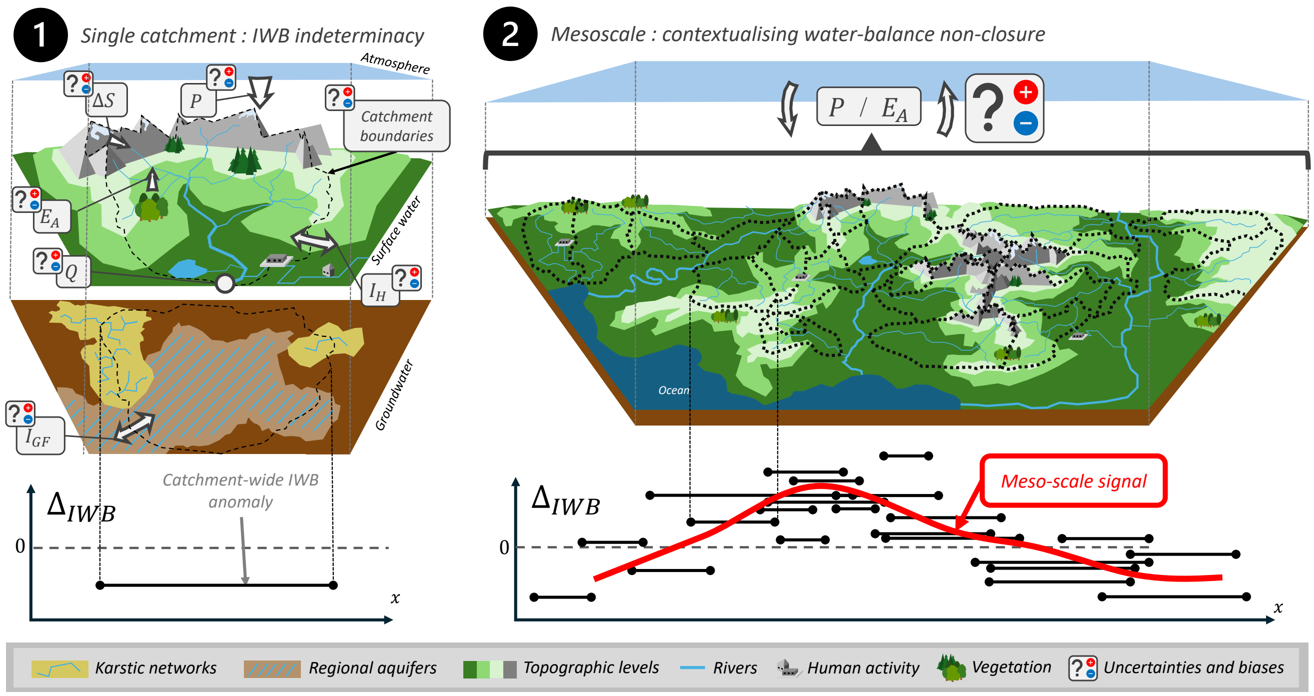

Figure 1. Conceptual comparison of the water balance closure problem at the catchment scale and the mesoscale. Panel 1 describes the many sources of non-closure at the local scale (see Eq. 1 for notations), hindering our ability to 'sweep' for the right reasons. Panel 2 shows the potential gain from taking a step back to look at water imbalance within a catchment's geographical context. The two graphs below show the corresponding water balance anomaly, plotted along some spatial axis with coordinate x; because the anomaly is defined at the catchment scale, we choose a segment representation to highlight the spatial extent of each catchment. On the bottom right, the red line gives a rough idea of what the low-frequency mesoscale signal may look like, set against the catchment-wide anomalies in black.

$\mathcal{H}$-0 *There exists a spatial scale, which we name the mesoscale, at which the spatial variability of the IWB anomaly signal can only be ascribed to the climate forcing biases. Correspondingly, the mesoscale is also the scale at which all other IWB non-closure causes are filtered out, because their spatial variability is believed to be higher.*

Figure 1 provides a visual interpretation of $\mathcal{H}$-0. The Figure shows that a good way of thinking about catchment-scale water balance consists in trying to put the catchment back in its mesoscale geographical context, because the context could tell us what we should or should not be sweeping. While

1.5 Scope of the paper

This paper aims to investigate $\mathcal{H}$-0. Specifically, we wish to address the following questions:

1.5.A Is there a spatially coherent low-frequency component in the water balance non-closure signal? How do we define this component, and how do we quantify it?

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

1.5.B Are the forcing biases the chief drivers of these low-frequency fluctuations? In other words, is the low-frequency signal a good match for the mesoscale imbalance as defined in $\mathcal{H}$-0?

1.5.C What new insights can the mesoscale hypothesis bring to the regionalisation of sweep parameters?

.

To that end, we choose to focus on the European continent, and we introduce ENTWASH, a high-quality network of low-influence streamflow gauges, selected from a number of national and international sources. The climate forcing variables are given by the CERRA and CERRA-Land reanalyses (Verrelle et al., 2022; Ridal et al., 2024). The water balance is then computed with a Budyko-style interannual model (Andréassian et al., 2016). Thin Plate Splines (Hutchinson and Gessler, 1994) have been chosen for the extraction of the mesoscale component.

2 Materials and methods

2.1 Climate forcing data

Having chosen to focus on the European continent, we test our hypothesis on the CERRA and CERRA-Land reanalyses (Verrelle et al., 2022; Ridal et al., 2024) to compute all relevant forcing variables. CERRA is a high resolution ($5\ km \times 5\ km$) European dataset containing atmospheric and surface variables. It runs HARMONIE-ALADIN, a numerical system for weather prediction and assimilation. CERRA-Land is a refinement of CERRA with a two-dimensional total precipitation analysis system (MESCAN) at the daily timestep, drawing from a pan-European network of rain gauges.

In addition to daily total precipitation, we retrieved five three-hourly variables from CERRA-Land to compute potential evaporation at the daily timestep: $2\ m$ temperature (T_{2m}), $2\ m$ relative humidity (h_r), $10\ m$ wind speed (u_{10}), surface net solar radiation (R_s) (which accounts for surface albedo) and surface net thermal radiation (R_{th}). Reference potential evaporation E_P is computed according to the FAO Penman-Monteith formulation (Allen et al., 1998).

2.2 Streamflow data: ENTWASH

The degree of human alteration of the observed streamflow may affect our ability to assess $\mathcal{H}$-0. Human-influences may lead to widespread water balance non-closure, especially in relation to water consumption due to irrigation practices over large areas. Additionnally, when the density of the gauged network is low, local anthropogenic influences will have a stronger impact on our ability to extract a mesoscale signal as defined in $\mathcal{H}$-0. Therefore, filtering out the most influenced gauges should be beneficial to our analyses of the IWB. Thus, we assemble a specific streamflow dataset of more than 4000 gauges across Europe (Figure 2), which we call ENTWASH (European Near-natural Timeseries for Water balance Analysis Studies in Hydrology). ENTWASH draws from multiple sources, which are listed in table 1. The final selection ensures that all ENTWASH catchments have been subject to some form of influence evaluation, albeit through evaluation frameworks that vary across providers (see appendix A for details). We used the MERIT (Yamazaki et al., 2017) digital elevation model (DEM) alongside the Joint Research Center river network (Vogt et al., 2007) to delineate the catchments of the ENTWASH dataset.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

Dataset source	Country / Region	Type of influence[1]	N	Δt	$\overline{\mathcal{A}}$	$\mathcal{A}_{\Updownarrow}$
International large sample datasets:						
ROBIN (Turner et al., 2025)	World	Multi-criteria	863	D	384	135
GRDC-CSS (GRDC, 2026)	Europe	Multi-criteria	22	D	6563	1174
LamaH-CE (Klingler et al., 2021)	Central Europe	Multi-criteria	400	D	260	101
National large-sample datasets:						
BULL (Senent-Aparicio et al., 2024)	Spain	Multi-criteria	131	D	546	371
CAMELS-CH (Höge et al., 2023)	Switzerland	Multi-criteria	13	D	194	
CAMELS-CZ (CHMI et al., 2025)	Czechia	Upstream storage	134	D	195	73
CAMELS-DE (Loritz et al., 2024)	Germany	Upstream storage	1265	D	291	132
CAMELS-DK (Liu et al., 2025)	Denmark	Multi-criteria	222	D	139	63
CAMELS-FR (Delaigue et al., 2025)	France	Upstream storage	489	D	765	191
CAMELS-FI (Seppä et al., 2025)	Finland	Multi-criteria	122	D	807	453
CAMELS-LUX (Nijzink et al., 2025)	Luxembourg	Multi-criteria	26	D	135	57
EStreams (IT) (do Nascimento et al.)	Europe (Italy)	Upstream storage	99	W	388	207
Explore2 (Strohmenger et al., 2024)	France	Upstream storage	150	D	2390	308
LamaH-Ice (Helgason and Nijssen, 2024)	Iceland	Multi-criteria	25	D	457	237
Timeseries from national services or private organisations:						
EDF[4] - Personal contact[2]	France	Naturalised	26	D	5002	6598
SHMÙ - Personal contact	Slovakia	Local expertise	17	M	50	48
NIMH - Personal contact	Bulgaria	Local expertise	1	M	924	924
RHMZ Srbije - Online archive	Serbia	Local expertise	9	D	217	159
IMGW-PIB[5] - Online archive	Poland	Multi-criteria	127	D	760	553
HMC-RA - Personal Contact	Armenia	Local expertise	2	M	1970	1970
Timeseries from published regional studies[3]:						
(Radevski et al., 2018)	North Macedonia	Local expertise	12	M	98	67
(de Lavenne et al., 2022) (SMHI)	Sweden	Local expertise	122	D	731	205
(Gnjato et al., 2023) (RHMZ-RS)	Bosnia-Herze.	Local expertise	4	M	5524	5849

Table 1. Streamflow timeseries included in ENTWASH. Columns N, Δt, $\overline{\mathcal{A}}$ and $\mathcal{A}_m$, respectively: the number of streamflow timeseries, the time resolution (D: daily, W: weekly, M: monthly), the average surface area (km^2, global mean : 564) and the median surface area (km^2, global mean : 147). [1]: the "Type of influence" column refers to the criteria that were used to evaluate human alterations; for more detailed explanations, please refer to tables A1 , A2 and A3. [2]: Wherever the "Dataset source" column reads "Personal contact", streamflow data was retrieved through private requests; see Table B1 in the Appendix. [3]: Timeseries from published regional studies were shared by the original authors. [4]: the EDF gauges come from the Explore2 project (Strohmenger et al., 2024), but were not made publically available. [5]: the selection of Polish gauges was based on ROBIN, but the data was retrieved from IMGW, in accordance with the country's data policy.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

Because we focused on near-natural rivers, table 1 shows that the average surface area of the selected catchments is rather small, falling below the spatial resolution of the CERRA forcing, which comes at 25 km^2 (see Figure 2 (b)). For the present application, we claim that catchment size is less important than the density and spatial coverage of the ENTWASH dataset within a region. As illustrated in the right-hand panel of Figure 1, the objective is to derive a low-frequency signal constrained by multiple catchment-scale IWB anomalies. A dense network of smaller catchments still provides stronger spatial constraints and improves the robust characterization of regional-scale variability.

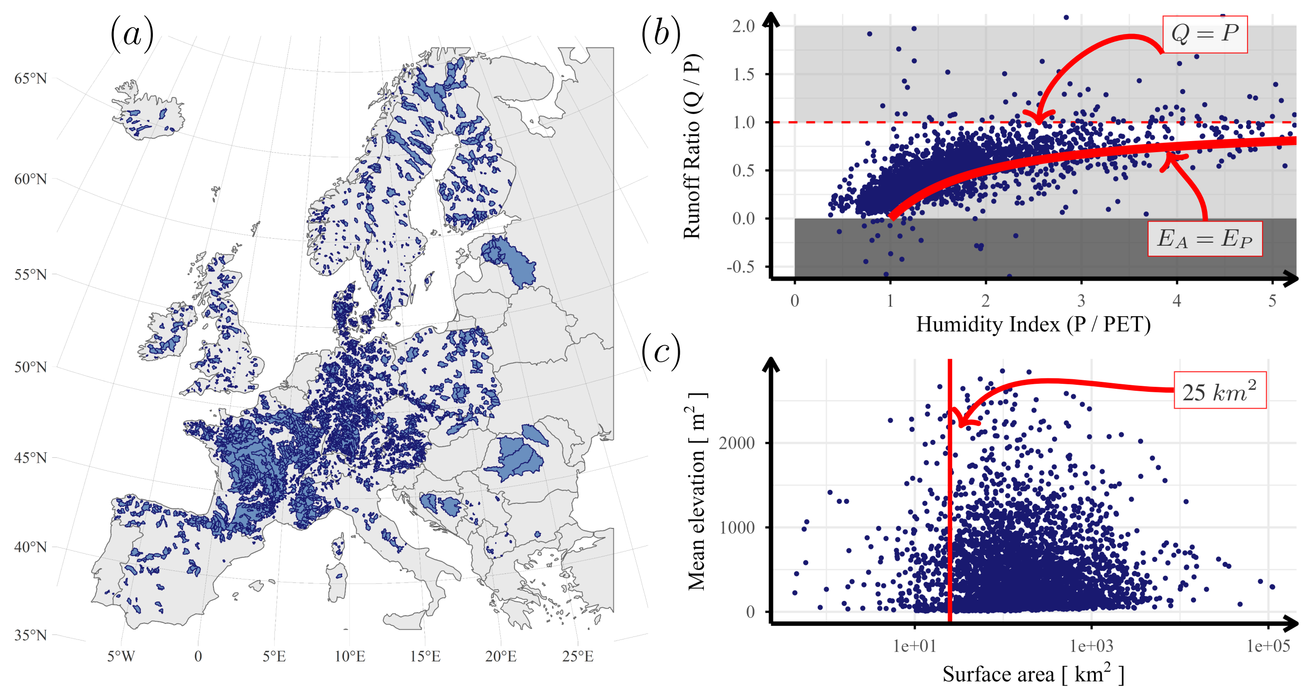

Figure 2. Overview of the ENTWASH product (in dark blue) after unnesting, mapped in 3 different ways: gauged catchment boundaries across the continent (a); associated Budyko diagram (b); catchment mean elevation against surface area (c). Negative values are due to the unnesting procedure.

2.3 Interannual (long-term) streamflow model

Closing the IWB requires an estimation of actual evaporation. Here, we take advantage of the widely adopted *Budyko hypothesis* (Budyko, 1948), stating that the long-term average of the actual evaporation is, to a first approximation, a function of the average total precipitation P and the aridity index E_P/P (E_P is the potential evaporation, see section 2.1). The idea is conveyed by Budydko-style graph as shown in panel (b) of Figure 2, supporting the hypothesis of an overarching relation between

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

the runoff coefficient and the humidity index P/E_P. More specifically, we choose to use the Turc-Mezentsev formulation (Andréassian and Sari, 2019), so that the long-term water balance anomaly writes:

$$\begin{aligned}\Delta_{IWB} = Q_{obs} - P + \left(P^{-n} + E_P^{-n}\right)^{-1/n} \\ - \Delta S_{glac} - \cancel{I_{GF}} - \cancel{I_H}\end{aligned} \tag{4}$$

where n is a shape parameter, whose value may change to account for local differences in the partition between evaporation and runoff. Because the ENTWASH dataset only comprises near-natural catchments, we expect I_H to be negligible. For the change in storage, we focus on the contribution of glacial melt ΔS_{glac}. Under global warming, glacial melt is a long-term change in storage (Dussaillant et al., 2025), which results in a net (generally positive) contribution to river discharge. In Iceland, Scandinavia and the Alps, glacial melt is widespread and we might expect it to interfere with the mesoscale signal. Dussaillant et al. (2025) provided estimates of average annual glacier mass change over the period 1976-2024. Glaciated fractions for each ENTWASH catchment were derived from the GLIMS Glacier Database (GLIMS and NSIDC, 2005; Raup et al., 2007). The combination of melt rates and glaciated areas enabled us to compute ΔS_{glac}. Finally, in the absence of a reliable estimate of I_{GF}, its contribution feeds the total imbalance Δ_{IWB}.

The shape parameter n of the Turc-Mezentsev model is regionalised following the methodology of de Lavenne and Andréassian (2018), who proposed to link the shape parameter to an index of synchronicity between precipitation and potential evaporation. The idea stems from an earlier comparison of several regionalisation methods for the shape parameter (Lebecherel et al., 2013), highlighting the need to incorporate synchronicity. Here, we adopt the formula of the synchronicity index of Andréassian et al. (2026).

2.4 Catchment unnesting and final selection

Nested catchments are not adequate for studying the spatial variability of IWB, as non-closure sources may change from upstream to downstream. Therefore, we carry out catchment unnesting. The procedure consists in deducting all the upstream inflows from the outflow of the downstream basin.

Unnesting is only meaningful if the upstream and downstream streamflow data are all available on the same period, which is not always the case in the ENTWASH dataset. To address this issue, we rescale the average annual discharge of nested catchments to a common reference period, based on the change in the Turc-Mezentsev streamflow (Q_{sim}, section 2.3) for a given change in aridity. A more detailed explanation is given in the appendix D. After applying the unnesting procedure, the resulting values of Q are found to be negative for a subset of catchments, leading to their exclusion from the analysis.

Annual discharge can only be computed (from daily, weekly or monthly timeseries, see table 1) for those years without a single missing datapoint, whereas the interannual average is computed as the average of all the valid years. Only those catchments with more than 5 valid years were kept to run the mesoscale analyses. The unnested ENTWASH dataset comprises 3614 catchments. After applying the unnesting procedure, the resulting values of Q are found to be negative for a subset of catchments, leading to their exclusion from the analysis. Likewise, those catchments which exhibited a ratio $(Q - \Delta_{glac})/P$

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

greater than 2 were left out, as a preemptive filter on the most non-conservative catchments (only a $> 50\%$ negative bias on P would lead to $(Q - \Delta_{glac})/P > 2$). The final dataset for mesoscale analyses comprised 3567 catchments.

2.5 Extraction of the mesoscale water balance anomaly

As previously stated, the first aim of this paper consists in evaluating hypothesis $\mathcal{H}$-0. To reach that goal, we need to extract the so-called mesoscale signal, namely the part of the IWB anomaly signal whose variability can be ascribed solely to the forcing biases. The mesoscale signal is thereby understood as a continuous field across the study domain, such that our definition of signal extraction comprises both the interpolation of locally computed water balance anomalies and the extraction of a coherent pattern from the signal. To that end, we made of use Thin Plate Splines (TPS), a popular solution to interpolate rain gauges (Hutchinson, 1995; Cornes et al., 2018). The appeal of these methods is their ability to provide *explicit and intuitive control* over the smoothness of the resulting interpolation.

A Thin Plate Spline is a piecewise polynomial which minimises the sum of two metrics: the total error (distance to observations) and the polynomial's total curvature. The trade-off between the curvature and the error is determined by a smoothing parameter λ_{TPS}. The smaller λ_{TPS} is, the more the interpolator sticks to the observation. Formally, if f is a one-dimensional spline, y_i the observations, x_i the associated coordinates and $\mathcal{S}$ the domain, the spline should minimise the following quantity:

$$\sum_i (f(x_i) - y_i)^2 + \lambda_{TPS} \int_S f^{(2)}(x)^2 dx \quad (5)$$

where $f^{(2)}$ is the $2-nd$ order derivative of f. Although higher-order splines (polyharmonic splines) can be used, their primary effect is to alter the smoothness of the interpolant. In this work, we choose to retain the classical thin plate formulation (second order), and regulate smoothness through λ_{TPS} alone. Moreover, we use a three-dimensional spline, allowing elevation to be incorporated alongside horizontal coordinates among the interpolator's inputs, a useful property as we want to capture the effect of climate forcing biases. Prior to running the interpolation, we also checked the value Q/P for all unnested catchments, and we left out the most extreme cases ($Q/P > 10$).

To investigate question 1.5.A of Section 1.5, we opted to analyse the interpolated signal across a range of values for the smoothing parameter λ_{TPS} (from 10^{-4} to 10^{1}, on a logarithmic scale). The `Tps()` R function (from the `fields` R library) was used to perform the interpolation.

The interpolated signal is denoted by Δ_{IWB}^{MS} (MS referring to the mesoscale). We also define the residual signal over the gauged network:

$$r\Delta_{IWB} = \Delta_{IWB} - \overline{\Delta_{IWB}^{MS}}^{c} \quad (6)$$

where $\overline{X}^{c}$ is the catchment-averaged value of variable X. The residual signal is therefore the remaining anomaly after the deduction of the interpolated signal. The residual is meaningful when compared to the original anomaly: if it is less than Δ_{IWB}, it means that the deduction of the interpolated signal improves non-closure; if it is greater, the deduction strengthens non-closure, which may reveal compensation between low-frequency variability and high frequency variability.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

2.6 Protocol for investigating the relevance of the $\mathcal{H}$-0 hypothesis

Directly testing $\mathcal{H}$-0 would require knowledge of the very biases we aim to infer. Thus, the best tool we have to evaluate $\mathcal{H}$-0 is to check if the mesoscale signal provides a spatial distribution of the IWB anomaly that matches our perception of the dominant process controls of the anomaly (Wagener et al., 2022).

CERRA-Land includes a two-dimensional total precipitation analysis through the MESCAN system. This means that the CERRA-Land gridded product may inherit some of the observational biases which usually affect precipitation measurements (Abbas et al., 2026). As discussed in the introduction, precipitation biases from observations are often linked to wind and snow undercatch. At higher elevations, further complications stem from sampling biases and the lack of representativeness (Prein and Gobiet, 2017; Dura et al., 2024; Maier et al., 2026). In practice, we expect that the greater the amount of snowfall or the greater the elevation, the greater the bias. Therefore, if the mesoscale biases of CERRA-Land are driven by the climate forcings alone, we expect orographic relief and the fraction of solid precipitation to be positively correlated with IWB anomalies. The correlations are investigated in sections 3.3 and 3.4, using the CERRA snowfall water equivalent to compute the average solid fraction. In terms of potential evaporation biases, mesoscale spatial patterns of water balance non-closure are less-predictible, due to the many climatic variables and the modelling assumptions required for computing E_P (Allen et al., 1998). Hence, in the present paper, we mostly focus on elevation and solid precipitation to test the consistency of $\mathcal{H}$-0.

As a complement, we also examine the value of incorporating interannual glacial melt in Eq. (4). Long-term glacial storage change is a good example of another source of potential IWB non-closure in hydrological models, whose spatial variability could well interfere with that of forcing biases (see section 2.3).

3 Results

3.1 Evidence for widespread coherent patterns of non-closure

Figure 3, displays the interannual water balance anomaly Δ_{IWB} across the ENTWASH gauged domain (panel (a)), and the corresponding TPS extraction Δ_{IWB}^{MS} (panel (b)). The anomaly is shown in colors, from red to blue denoting positive (observed $>$ predicted) to negative (observed $<$ predicted) imbalance, with the starkest colors denoting stronger imbalance.

In spite of the relative sparsity of the gauged network, the map of Δ_{IWB} in panel (a) is already hinting to the existence of a widespread coherent signal. The prevalence of red in some regions (Northern Europe, the Pyrenees, the Cantabrian and Galician ranges, the Massif Central, the Western Alps or even southern Germany) points to an excess of discharge: the observed outflow at the catchment outlet is greater than what the water balance model predicts. Conversely, the prevalence of blue (Western France, the Baltics, the Central Appenines, much of the Eastern Alps and southern Bohemia) indicates a configuration where the observed outflow is lower than what we expect from the water budget.

Below the European map of Δ_{IWB}, we take a closer look to the Upper Rhine (panel (b)) and Northwest German lowlands (panel (c)), where the gauged density and the spatial coverage of ENTWASH is high. The close-up view confirms our intuition that catchment-scale water balance is not spatially coherent: widely different values may occur in adjacent basins, with locally

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

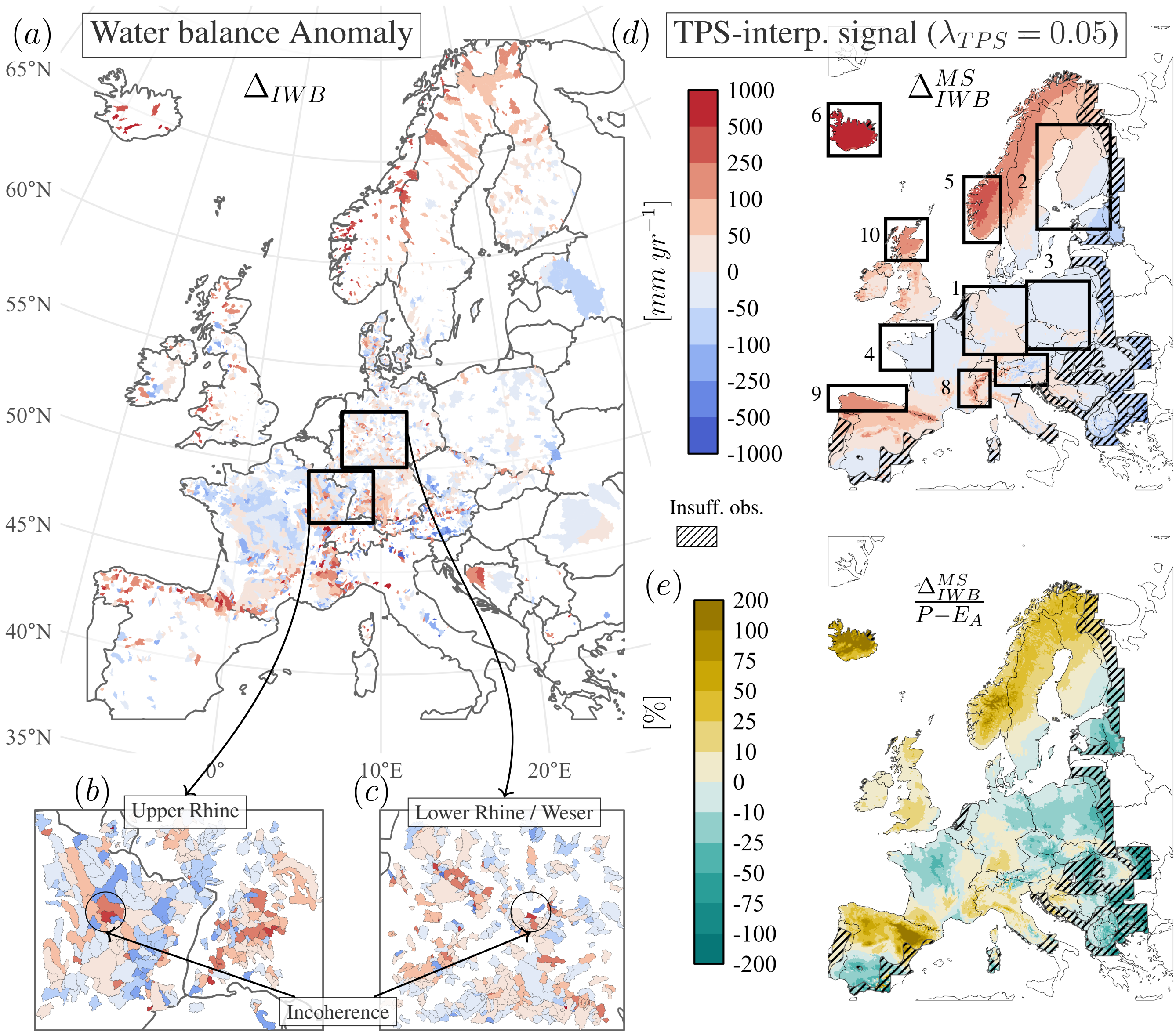

Figure 3. Evidence for widespread and spatially coherent patterns of water balance non-closure. Panel (a): IWB anomaly, plotted for all valid catchments in ENTWASH, with associated zoomed-in windows in panels (b) and (c). Panel (d): TPS-interpolated signal of IWB non-closure anomaly, in the same colormap. Panel (e): TPS-interpolated signal of IWB non-closure anomaly normalised by $P - E_a$. The choice of $\lambda_{TPS} = 0.05$ is explained in section 3.3. Hatched areas in panels (d) and (e) denote reduced confidence in the interpolated signal, reflecting the low spatial density of the ENTWASH dataset, as assessed through kernel density estimation (KDE) - the threshold was set to 1 gauge in 20000 km^2 (roughly the size of Slovenia). The numbered black boxes of panel (d) refer to Figure 5.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

stark discontinuities reaching up to almost 1000 $mm.yr^{-1}$. The two panels also highlight the value of assessing water balance non-closure of individual catchments within the surrounding geographical context: a red spot in a largely blue region (Northeast France) tells a different story than a red spot in a largely red region (Southwest Germany).

The TPS-interpolated signal of the IWB anomaly Δ_{IWB}^{MS} is shown on the right-hand side of Figure 3, for a chosen value of the smoothing parameter λ_{TPS} (see section 3.3 for more details on this choice). A visual comparison of panels (a) and (d) shows that the signal extraction seems to capture the main features of the IWB across Europe. The Northern Iberian range, the Western Alps, Southern Germany and Northern Europe all exhibit positive water imbalance, whereas negative imbalance is found in Western France, the Austrian Alps, the Northeastern European plains and the Central Appennines. The map of Δ_{IWB}^{MS} also reveals that the orography plays a key role in shaping the spatial patterns of the IWB. The starkest imbalance is found in the Pyrenees, the Alps, the mains reliefs of the British Isles and the Scandinavian range.

In panel (e), we also map the relative imbalance, namely the ratio of the interpolated signal to the budget of climatic fluxes $P - E_A$. Across much of Western Scandinavia, Iceland, the Pyrenees and northern Iberia, water balance non-closure amounts to more than 25 % of $P - E_A$; conversely, in the remaining ranges, such as the British Isles, the Massif Central, the Alps, the Carpathians and the Dinaric range, relative imbalance is weaker, rarely exceeding 25 %. Moreover, panel (e) sheds light on the crucial distinction between absolute and relative anomalies in the European lowlands. Here, in the Polish, Austrian and French lowlands, we see that the negative water balance anomaly may in fact exceed 25 % of $P - E_A$, in spite of the weaker magnitude of the absolute water balance anomalies (panel (d)). Altogether panel (e) tells us that the magnitude of these patterns is hydrologically significant across most of Europe.

The interpolated signal is less convincing in central and southern Spain. The stark and widespread positive signal of the interpolated anomaly Δ_{IWB}^{MS} (Figure 3, panel (d)) across the Iberian peninsula does not match what we observe in the original map of Δ_{IWB} (Figure 3, panel (a)), where the anomaly for most catchments outside of the Northern Iberian Ranges is either negative or weakly positive. The issue is even starker in panel (e), indicating that the water balance anomaly even exceeds 100 % of the modelled runoff, especially over the Ebre basin. The most likely explanation lies in the lower density of the ENTWASH dataset across Spain (see section 3.2 below).

Overall, Figure 3 shows that a spatially coherent water balance non-closure signal can be extracted (as long as the original dataset of gauged catchments is dense enough), and it exhibits stark spatial patterns across the European continent.

3.2 Low-frequency non-closure signal: trade-offs and sensitivity to the gauge density

The previous paragraph offered a clear illustration of the coherent patterns of IWB non-closure. Here, we analyse the results at a finer spatial resolution to understand if the TPS-interpolator is an adequate tool to capture the general properties of the anomaly signal, and to filter out the more local signals. Specifically, we compare the distributions of the IWB anomalies Δ_{IWB} and the corresponding TPS extraction Δ_{IWB}^{MS} in a few selected regions, and we analyse to what extent the latter preserves the statistical features of the former.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

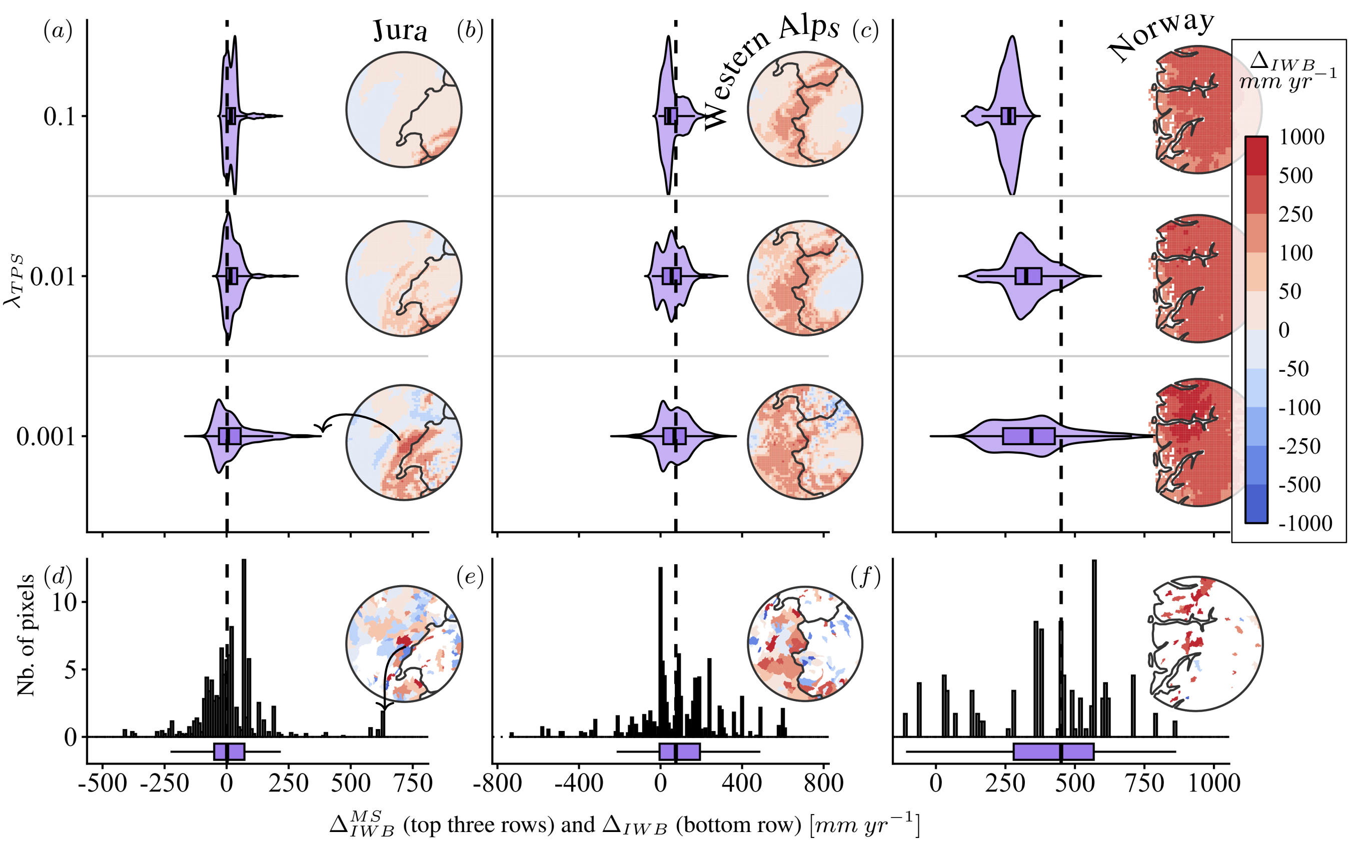

Figure 4. Three case studies to compare the regional distribution of the mesoscale signal to the original distribution of anomalies. In each panel - (a), (b) and (c) - we focus on a different region, shown by the inset maps (Jura, Western Alps and Norway). The first three rows describe the distribution of the interpolated mesoscale signal within the circular region shown by the inset map, for $\lambda_{TPS} \in \{0.1, 0.01, 0.001\}$. Boxplots (median and interquartile range (IQR); whiskers extend to 1.5 × IQR) are superimposed onto the associated violin plots. The histograms of the bottom row describe the distribution of Δ_{IWB}, also summed up by the underlying boxplots. In each panel, the median of the Δ_{IWB} distribution is highlighted by the vertical dashed lines across all rows.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

In Figure 4, we propose to focus on three different case studies across the continent. In panels (a), (b), (c) and (d), we test whether the TPS interpolation is able to remove local anomalies, depending on the choice of the smoothing parameter λ_{TPS}, and we assess the corresponding trade-off in terms of agreement between Δ_{IWB} and Δ_{IWB}^{MS} at the regional scale.

In panels (a) and (d), we center the analysis on the Loue river in the Jura mountains. The Loue is a notorious case of a karstic resurgence which is fed by losses from the neighbouring river Doubs (Charlier et al., 2014). The impact of the corresponding IGF on the IWB is visible in the circular map of Δ_{IWB} over the Jura region (panel (d)): the Loue at Chenecey-Buillon (red polygon in the middle) is in stark contrast with the Doubs at la Cluse-et-Mijoux (adjacent blue polygon south of the Loue catchment). The substantial water imbalance of the Loue catchment also features in the histogram of panel (d), in the form of a clear and isolated spike at $\Delta_{IWB} = 600\ mm\ yr^{-1}$. In panel (a), we see that the choice of the smoothing parameter λ_{TPS} is critical to erase the impact of the local anomaly from the TPS-interpolated signal. When $\lambda_{TPS} = 0.001$, the interpolated signal Δ_{IWB}^{MS} is clearly distorted by the Loue anomaly, with a long positive tail extending to $200\ mm\ yr^{-1}$, corresponding to the darker shade of red in the associated map. Conversely, when $\lambda_{TPS} \geq 0.1$, the positive tail vanishes. Aside from filtering the outliers, the interpolated signal also appears to preserve the main features of the original distribution, as the interquartile range of Δ_{IWB}^{MS} falls within the interquartile range of Δ_{IWB}.

Another valuable case study is found in panels (b) and (e), centered on the Arc at Saussaz in the Graian Alps, across the French-Italian border. In panel (e), the IWB anomaly of the French catchments looks overwhelmingly positive, while that of the Italian catchments is consistently negative. In this case, the jump in the IWB anomaly between two adjacent basins reflects a specific type of variability of the climate forcing variables linked to the leeward to windward sides (see Wang et al. (2014) for similar sign change of water balance anomaly). The TPS-analysis fails to capture the sign change, by producing an overarching weekly positive signal, especially when $\lambda_{TPS} = 0.1$. Lower values of the smoothing parameter would enable us to better capture the sign-change, although that would come at the expense of our ability to filter out local anomalies, like that of the Loue in the Jura Moutains. Additionnally, one might point out that the density and spatial coverage of ENTWASH is much higher in France than it is in Italy, thus creating a spatial imbalance which hinders the interpretability of the signal.

Taken together, the Jura and the Western Alps show that the extraction of a convincing mesoscale anomaly signal can be viewed as a compromise between filtering out those unwanted local anomalies which are not driven by the climate forcing biases (e.g. the Loue water imports from the Doubs watershed) and losing some valuable parts of the signal (e.g. the windward and leeward distinction in the Western Alps), especially in mountainous regions. Additionnally, the example of the Western Alps also indicates how the heteregeneous density of ENTWASH might detract from the quality of the interpolated signal.

The impact of poor gauge density and coverage is best depicted by panels (c) and (f), which focus on a region centered on the Raundalselvi River at Kinne in the Norwegian Mountains. The distribution of the TPS-interpolated signal is shown to be consistently lower than the distribution of Δ_{IWB}, whatever the value of the smoothing parameter. The issue is further exposed in Figure 5 which compares the regional average of the IWB non-closure anomaly over the rectangular boxes shown in panel (d) of Figure 3, to the regional average of the interpolated IWB. The regional average is meaningful because it functions as the sum of local contributions, effectively yielding *a regional water balance*. Across the Northern European lowlands (rectangles 1 to 4), the spatial average of Δ_{IWB} is close to the spatial average of Δ_{IWB}^{MS}. However, along mountain ranges (especially

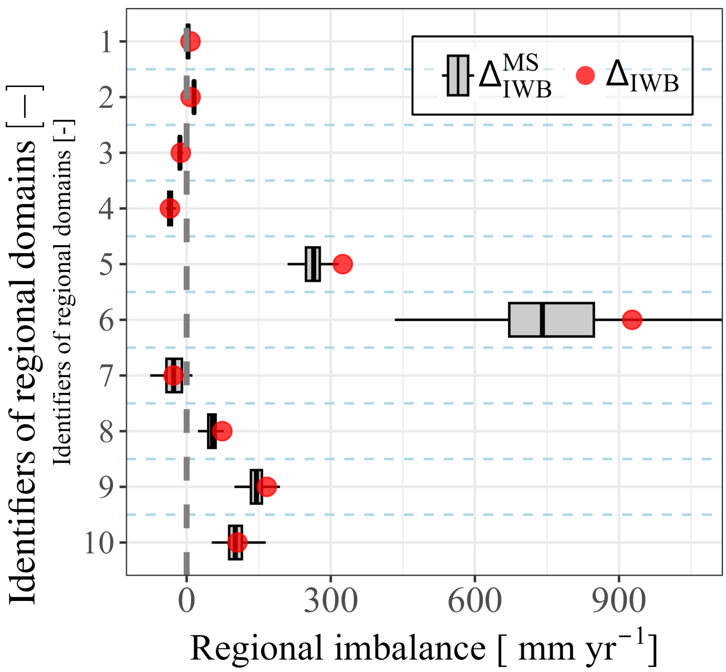

Figure 5. Regional averages of the IWB interpolated signal (gray boxplot) compared to the original IWB anomalies (red spots); the regional average is computed for each rectangular domain from panel (d) in Figure 3. The grey boxplots describe the distribution (median and interquartile range (IQR); whiskers extend to the minimum and maximum) of the regional average of the IWB interpolated signal across the λ_{TPS} range, spanning values from 10^{-4} to 0.1 at regular intervals on a logarithmic scale.

rectangles 7 and 8), the magnitude of the interpolated anomaly tends to be lower than that of the original imbalance. A possible explanation lies in the limited elevation representativeness of the ENTWASH network, which favours headwater basins located at higher elevations, while the interpolated signal extends into lowland regions. We expect that a higher density and spatial coverage of gauged catchments would reduce the discrepancy between the distribution of observed IWB anomalies and the mesoscale signal.

3.3 Forcing biases as the drivers of the mesoscale signal

In the previous sections, we shed light on a spatially coherent non-closure signal which enables us to filter out small-scale variability due to local sources of water imbalance (such as karstic transfers). Yet, it remains unclear if this variability of the IWB non-closure is driven by the biases of the CERRA-Land reanalyses, i.e. if the interpolated signal truly equates to the mesoscale signal as we defined it in $\mathcal{H}$-0. In Figure 6, we look at the correlation between Δ_{IWB}^{MS} and elevation, whereas in Figure 7, we evaluate the relation between Δ_{IWB}^{MS} and the fraction of solid precipitation outside of mountain ranges.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

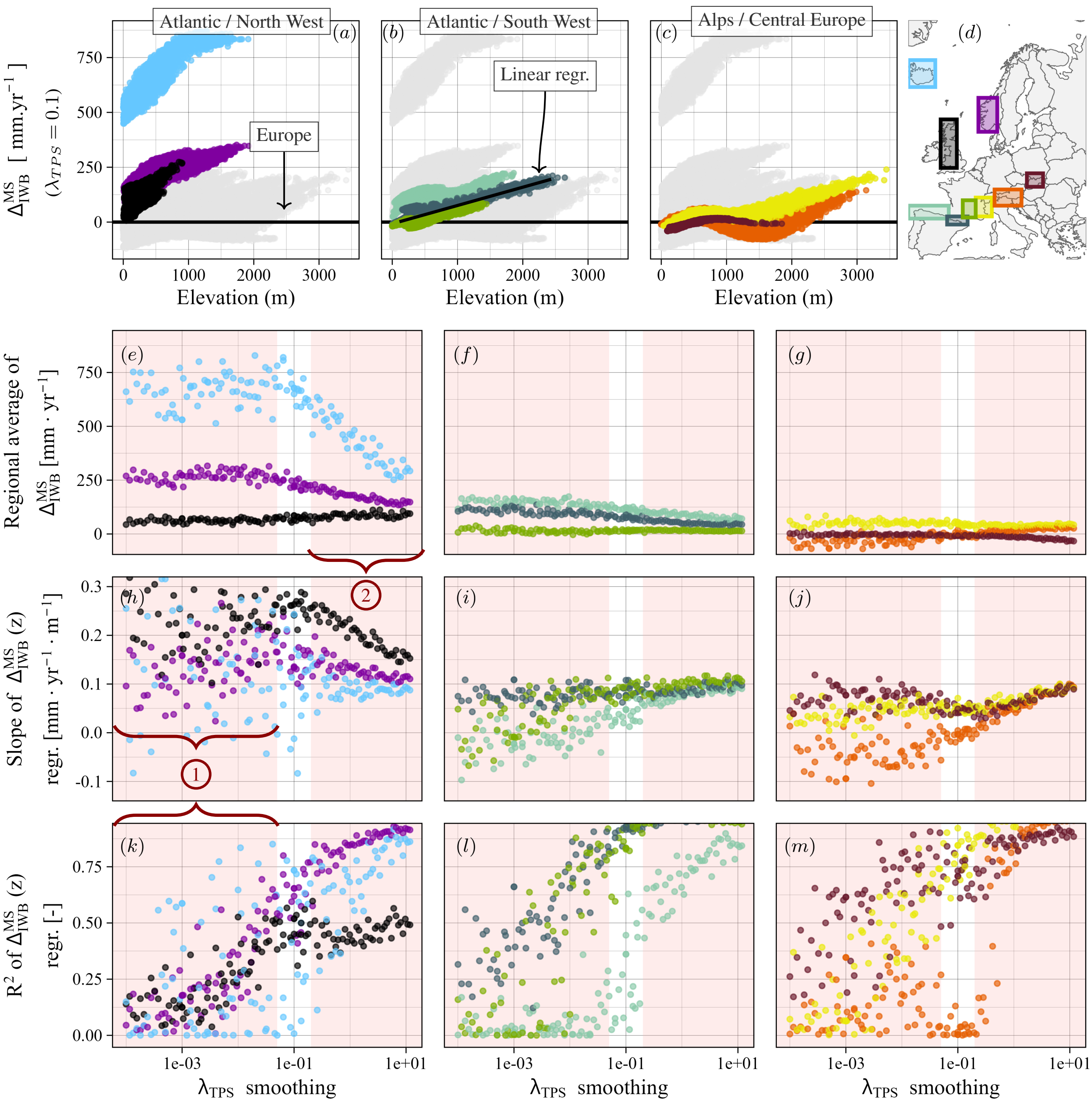

Figure 6. The role of elevation in the interpolated IWB anomaly. In panels (a) to (c), the grey scatter plot in the background refers to the complete European domain, whilst the colored subsets in the foreground refer to smaller domains, as defined in panel (d). Below, panels (e) to (g) describe the regional average of the Δ^{MS}_{IWB}, within the corresponding colored rectangles of panel (d). Likewise, panels (h) to (j) display the slope of the linear regressions of the scatter plots for each regional subset, whilst panels (k) to (m) feature the corresponding R^2 of the linear regressions.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

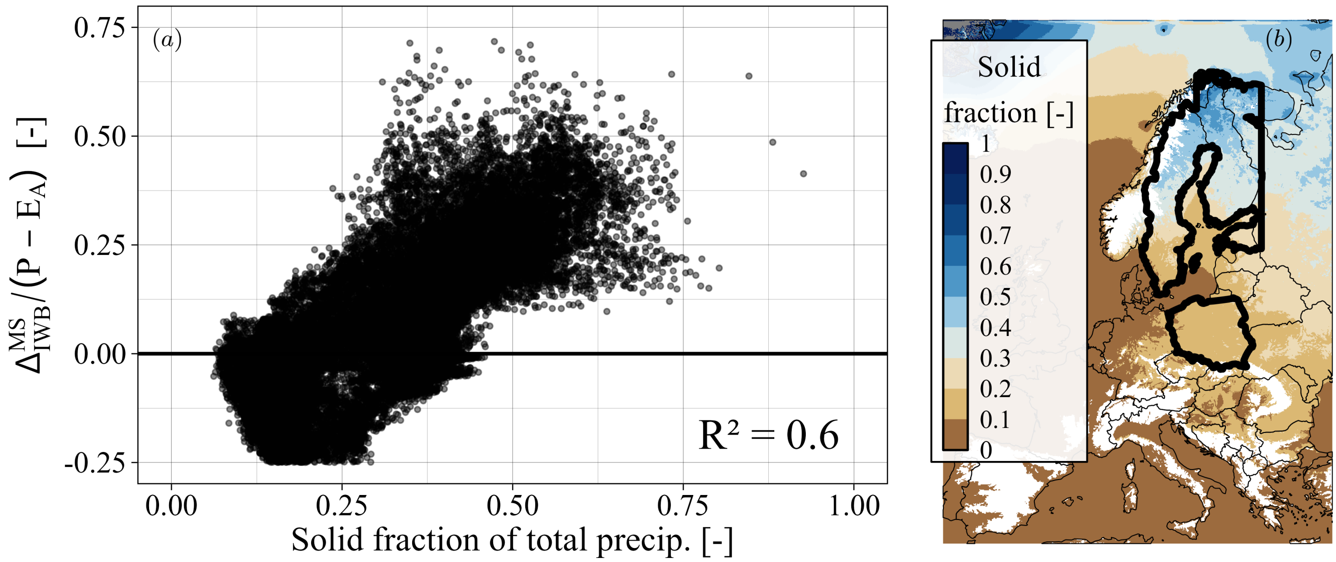

Figure 7. Correlation between the solid fraction of total precipitation (ratio of total snowfall and total precipitation) and the relative interpolated imbalance $\Delta^{MS}_{IWB}/(P-E_A)$ (with $\lambda_{TPS} = 0.05$). On the left panel, the black dots refer to the selection shown in the map of the right panel. High elevation pixels (in white on the map of panl (b)) are excluded from the analysis.

In Figure 6, we plot Δ^{MS}_{IWB} against elevation in panels (a), (b) and (c). In grey, the complete European domain is shown, whereas the colored subsets correspond to the colored boxes in panel (d). Here, the positive correlation between elevation and water imbalance is clear, although the dependance on elevation is not always linear (especially in the Alps).

Panels (h), (i) and (j) also show that the slope of the $\Delta^{MS}_{IWB}(z)$ linear regression depends on the value of the smoothing parameter λ_{TPS}. As λ_{TPS} gets smaller, the variance of the slope increases, suggesting lower confidence in the value of this vertical lapse rate. This is confirmed by panels (k), (l) and (m), which reveal that the R^2 of the linear regression between Δ^{MS}_{IWB} and elevation grows weaker as the value of the smoothing parameter decreases. This is no surprise: the smaller the roughness penalty, the better we stick to observations, causing mesoscale features such as the vertical lapse rate to become increasingly obscured by local-scale variability. Conversely, as λ_{TPS} increases, the variance of the slope decreases and the R^2 increases, pointing to a more robust estimation of the vertical lapse rate. Nonetheless, higher λ_{TPS} also leads to greater smoothing, which eventually erases the mesoscale variability: as shown in panels (h), (i) and (j), as the smoothing parameter increases, the vertical lapse rate for all regions seems to converge towards the same value, at about $0.1\ mm\ yr^{-1}\ mm\ m^{-1}$ (brace (1) in (h)). By the same token, panels (e), (f) and (g) indicate that regional averages of the interpolated signal also seem to suffer from oversmoothing when $\lambda_{TPS} > 0.2$ (brace (2) in (h)).

Panels (h) to (j) also reveal that the vertical lapse rate of Δ^{MS}_{IWB} changes depending on the mountain range: over the British Isles, the vertical gradient reaches $0.2\ mm\ yr^{-1}\ m^{-1}$ (panel (e)), set against $0.1\ mm\ yr^{-1}\ m^{-1}$ in the Pyrenees (panel (f)).

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

This means that the three-dimensional TPS interpolator is able to provide an elevation-dependent mesoscale signal of IWB non-closure, but also to offer a description of the variability of this dependance across different regions. The variability conveys the fact that orography-related sources of non-closure differ across the continent (e.g. snow and wind-induced precipitation undercatch), although a finer investigation of the drivers of this variability is out of our scope.

Hence, the meaning of Figure 6 is two-fold. On the one hand, it shows that the mesoscale signal is consistent with our expectation to find a vertical lapse rate of IWB non-closure, which is likely related to orographic biases on total precipitation. On the other hand, it provides more insights into the choice of the smoothing parameter, revealing an optimal range to extract vertical lapse rates ($\lambda_{TPS} \in [0.05, 0.2]$).

Alongside the role of elevation, Figure 7 shows that the relative interpolated imbalance $\Delta^{MS}_{IWB}/(P - E_A)$ is also positively correlated ($R^2 = 0.6$) with the solid fraction of total precipitation. Here, we focus on the northeastern European plains, as the solid fraction of precipitation is mostly negligible in the rest of Europe. Remarkably, high elevation were left out (see white ares on the map of panel (b) in figure 7), meaning that the distribution on panel (a) focuses on lowlands exclusively. In turn, this means that the observed correlation is independent of the orography.

3.4 The inclusion (or not) of glacial melt

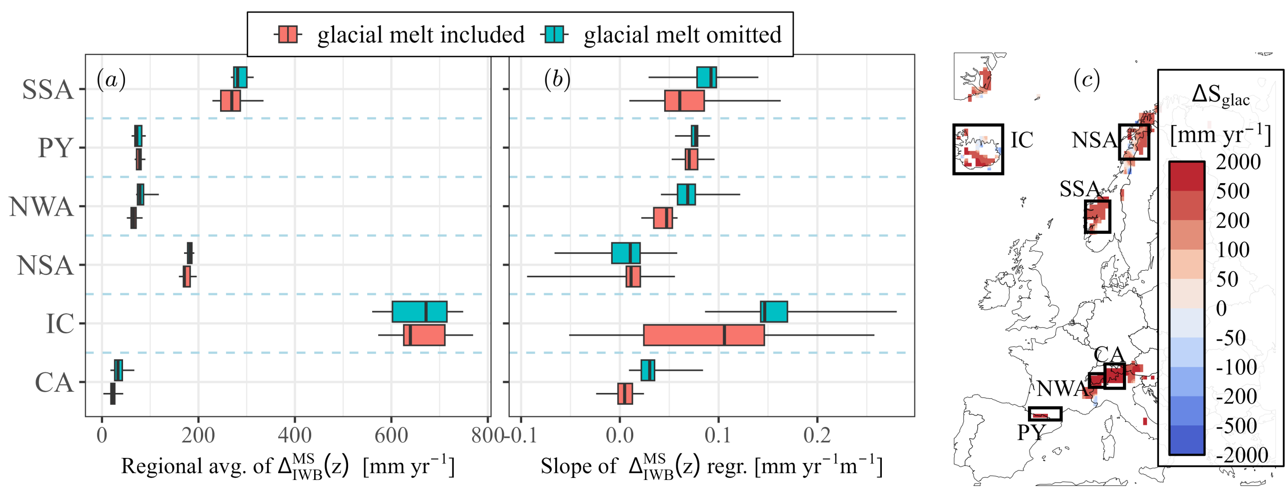

Figure 8. Comparison of the mesoscale signal when glacial melt is taken into account (red), against the signal when glacial melt is ignored (turquoise). Panel (a): regional average of the IWB interpolated signal, averaged over regional domains (SSA: South Scandinavian Alps; PY: Pyrenees; NWA: North Western Alps; NSA: North Scandinavian Alps; IC: Iceland; CA: Central Alps). Panel (b): the slope of the linear regression of the anomaly against the elevation, computed over the regional domains. The boxplots describe the distribution across a range values for the smoothing parameter λ_{TPS}, comprised between 0.05 and 0.2 (see section 3.3 and Figure 6). Panel (c): map of the interannual glacial melt as derived from Dussaillant et al. (2025).

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

The previous section aimed to highlight how the spatial pattern and vertical structure of the IWB non-closure anomaly is consistent with hypothesis $\mathcal{H}$-0. Here, we study the effect of omitting glacial melt from the IWB, to assess whether mesoscale signal extraction provides a robust framework for disentangling forcing biases from the other sources of water balance non-closure.

Figure 8 shows the result of this experiment. In panel (c), we plot the gridded average melt between 1976 and 2024, taken from Dussaillant et al. (2025); in panels (a) and (b) respectively, we then look at the regional average of the mesoscale signal, as well as the slope of the affine regression (see Figure 6) within each rectangle. In most regions, the omission of glacial melt produces larger IWB anomalies, although the effect of the omission of glacial melt is not significant. Because glaciers are mostly located at higher elevations, the vertical lapse rate is also affected by the omission, especially in Southwest Scandinavia, the Northwestern Alps and Iceland. The starker signal in Iceland is likely due both to the high percentage of glacierised areas, and to the relatively poor gauge density, which means that high levels of glacial melt over individual catchments will result in stronger mesoscale imbalance. Overall, the main takeaway is that glacial melt does not radically change the mesoscale patterns of IWB non-closure. Remarkably, in most regions, the offset due to the omission of glacial melt falls within the interquartile range of the boxplots, meaning that the uncertainty of the interpolation procedure is greater than the effect of the omission.

3.5 The residual signal and the strengthening of the imbalance

Having extracted a reasonable mesoscale signal, we can also compute the residual signal $r\Delta_{IWB}$. Figure 9 provides a general diagnostic of the residual field across the European continent. In panel (a), the residual is plotted with the same colormap as the one we used in panel (a) of Figure 3. Remarkably, the widespread patterns of positive and negative imbalance have receded, leaving us with a much less coherent map and generally weaker values. In panels (b) and (c), we compare the residual to the original signal: dark red refers to those catchments where the magnitude (absolute value) of the residual signal is greater than the magnitude of the original IWB anomaly. Although the value of the residual is less than (in absolute terms) the value of the anomaly for a majority of catchments (2155, in dark purple), the opposite is true for a substantial number of catchments (1459, in light gold). Moreover, panel (b) shows that those catchments with a higher residual are spread throughout the continent, without a clear spatial pattern.

4 Discussion

4.1 Plausibility of $\mathcal{H}$-0: precipitation biases as drivers of mesoscale imbalance

With $\mathcal{H}$-0, we hypothesised the existence of the mesoscale, a spatial scale at which the non-closure of the interranual water balance is driven by the forcing biases. The main finding of this paper is that long-term water balance closure displays widespread and spatially-coherent patterns that are positively correlated with elevation and the fraction of solid precipitation. Meanwhile, as explained in section 2.6, we expect precipitation biases to increase with elevation and with the fraction of solid precipitation.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

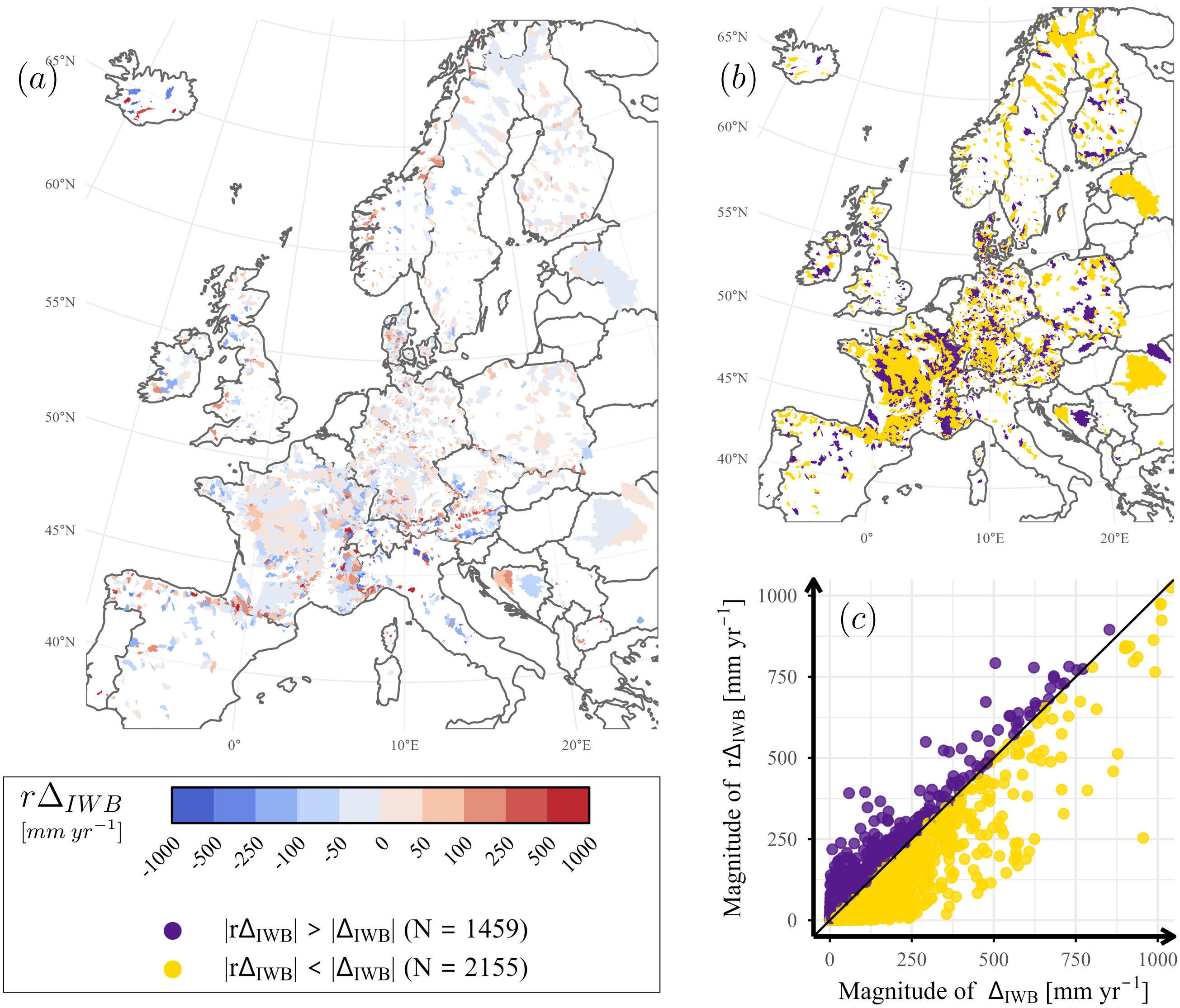

Figure 9. Comparison of the residual IWB non-closure anomaly to the raw anomaly. (a): residual signal (see Eq. (6)) across the gauged domain. (b) and (c): relative magnitude of the residual compared to the original signal. When $r\Delta_{IWB} < \Delta_{IWB}$ (in blue), the IWB non-closure anomaly after deduction of the mesoscale signal is less than its original value. When $r\Delta_{IWB} > \Delta_{IWB}$ (in red), the opposite is true.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

Thus, our findings are consistent with $\mathcal{H}$-0, because precipitation biases *can explain* much of the low-frequency variability of water imbalance. This idea is further supported by the following arguments.

Firstly, the impact of potential evaporation is expected to decrease as the latitude and the elevation increase, which means that evaporation can hardly explain the starkly positive anomalies across Scandinavia and over mountain ranges (Figure 7). Out of the forcing variables, precipitation biases yield a more consistent explanation for the observed correlations.

Secondly, the omission of glacial melt has shown that mesoscale patterns of imbalance are only weakly affected by the interannual contribution of glaciers, even in the most glaciated regions. Thus, a biased estimation of the long-term change in glacial storage is unlikely to explain the correlation of IWB anomalies with elevation.

Thirdly, our mesoscale estimate of water imbalance is consistent with typical values of wind-induced rain gauge undercatch. Over Scandinavia, where the solid fraction of precipitation accounts for up to 50 % of annual rainfall, the interpolated IWB anomaly reaches approximately 25 % of $P - E_A$ (Figures 3 and 7), in line with the range of snowfall undercatch reported by Herrnegger et al. (2015) ($10 - 50\%$ drawing from several prior studies), Prein and Gobiet (2017) ($40 - 80\%$ across Europe) and Buisán et al. (2017) ($30 - 80\%$ in Spain). Across the mountain ranges of the British Isles, where snowfall accounts for less than 10% of annual rainfall, the anomaly is closer to 10%, again consistent with typical wind-induced undercatch for liquid precipitation, which can range between $2 - 10\%$ (Herrnegger et al., 2015).

The order of magnitude of the interpolated IWB anomaly is also consistent with independent estimates of precipitation biases in observational *gridded products*. In Switzerland, MeteoSwiss (2013) reports as much as 40 % of precipitation undercatch for elevations greater than 1500 m $a.s.l$ in the RhiresD reanalysis. Maier et al. (2026) highlighted similar levels of undercatch across the Austrian Alps in the SPARTACUS reanalysis (Hiebl and Frei), using the COSERO rainfall-runoff model to close the water balance. In Norway, Lussana et al. (2018) likewise examined precipitation biases in seNorge2 with the DDD rainfall-runoff model, and found substantial underestimation of liquid precipitation (25% on average) and snow-covered area, especially over the western coast and mountains. These values are to be compared to Figure panel (e) in Figure 3, which reveals that the relative mesoscale water imbalance from CERRA-Land is locally greater than 25 % for a number of high-altitude Alpine and Scandinavian ranges (Bernese and Pennine Alps, Grisons, Vanoise and Graian Alps, Ötztal Alps and Soutwest Norway).

Each of the afore mentioned studies shows that precipitation biases could explain the magnitude and the spatial distribution of the mesoscale patterns of water imbalance. Yet, outside of moutainous and arctic regions, the plausibility of $\mathcal{H}$-0 is more challenging to prove. It should be added that while we chose to focus on the CERRA reanalysis to challenge our hypothesis on a high-quality product, the scope of this paper is not to provide a definitive assessment of the biases of CERRA. The focus remains centered on the mesoscale hypothesis and the methodology to extract it. Nonetheless, it should be mentioned that the performances of the CERRA/CERRA-Land reanalyses have been shown to be on par with national gridded products (Wood et al., 2024; Papa and Koutroulis, 2025; Dura et al., 2024; Moccia et al., 2025; Alvarez et al., 2025; Yang et al., 2026)), hence our choice to evaluate the mesoscale signal against the reported bias estimates from those reputable products.

4.2 Implications of the mesoscale analysis for the regionalisation of sweep parameters

By decoupling the low-frequency contribution of forcing biases from the other causes of non-closure, we are able to discriminate between the part of the imbalance which is intrinsic to the catchment, from that which is (hopefully...) independent. The extraction yields a plausible distributed water imbalance in the form of a continuous field over the European continent, which means we can produce *non-closure predictions in ungauged basins*. Therefore, the mesoscale analysis is best interpreted as a *regionalisation procedure* applied not to model parameters, but to the average errors on forcing inputs that drive the model.

In simpler terms, the regionalisation of the non-closure anomaly is the first step towards a streamflow-based pan-European correction of both precipitation and potential evaporation. In this respect, our work is no more than a modern take on a century-old tradition amongst hydrologists, who have long made the case for an alternative approach to precipitation and evaporation measurements - one which acknowledges that catchments are the only actual witnesses of the fluxes between the surface and the atmosphere. As forerunners, one must mention in particular Wallén (1923, 1924), who used the average discharge rate to reconstruct plausible precipitation values over the Swedish mountains, where rainfall measurements were sparse. A few decades later, Penman (1950) published a large-scale analysis of potential evaporation climatology over the British Isles, in which the water balance closure of 41 gauged catchments was used to assess the quality of catchment-scale forcing estimates, using his own evaporation formula. These foundational works paved the way for later attempts, often described as backward hydrology (Vrugt et al., 2008; Kirchner, 2009; Herrnegger et al., 2015; Hughes and Gray, 2017; Manoj et al., 2025). That being said, the distinctive argument of the present approach lies in the targeted and coupled treatment of both P and E_P as non-closure causes through the mesoscale analysis. In the future, we hope that the mesoscale hypothesis will open new paths for the simultaneous computation of correction factors for both precipitation and potential evaporation across the European continent.

More crucially, because the mesoscale signal filters out some of the complexity, it also enables us to take a step forward towards the ultimate goal of "sweeping for the right reasons" (Clerc-Schwarzenbach et al., 2026). As illustrated in Figure 1, untangling (see problem 20 in Blöschl et al., 2019) the IWB non-closure of an isolated catchment is challenging because several uncertainities and biases coincide, often leading to mutual compensation (Wang et al., 2014). This means that the IWB anomaly of a single catchment, when examined in isolation, has limited interpretive value. In Figure 9, the comparison of the residual to the original imbalance provides critical insight: for a significant number of catchments, the deduction of the mesoscale signal leads to *strengthening* the IWB non-closure anomaly. The mesoscale analysis offers a striking indication that the IWB anomaly of single catchments is in fact a **lower bound** (or a *measure "by default"*) on the errors and biases which may cause non-closure - a statement which, to the best of our knowledge, is not adressed by the modellers who attempt water balance closure.

Lastly, even the mesoscale signal may hide compensatory effects between potential evaporation and precipitation biases. At higher elevations and latitudes, we showed that precipitation biases are the preferred explanation, but partitioning the imbalance will be more challenging everywhere else.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

4.3 Blind spots of the mesoscale hypothesis

4.3.1 Should we be worried about the low-frequency variability of I_{GF}, I_H and ΔS?

Assessing the contributions of I_{GF}, I_H and ΔS to the mesoscale imbalance remains elusive. Firstly, long-term groundwater depletion over the largest aquifers (Konikow, 2015; Carlson et al., 2025) might lead to the emergence of a spatially coherent imbalance. However, Chávez García Silva et al. (2024) have shown that groundwater levels have remained stable over 1960-2020 and across 68% of wells in Portugal, Spain, France and Italy, all of them countries where we would expect high levels of groundwater depletion to sustain agriculture. Alternatively, the strong signal in Northern France (panel (e) in Figure 3), where the negative imbalance reaches as much as -25% of $P - E_A$, could be a sign of consistent water gains from the aquifers of the Paris Basin. As suggested by Schaller and Fan (2009) and Fan (2019), I_{GF} could effectively lead to spatially coherent negative IWB anomalies, whose patterns have been linked to the underlying geology. That said, the negative imbalance over Northern France (panel (e) in Figure 3) also extends across the Armorican Massif over Brittany and Normandy, whose geology is radically different from that of the Paris Basin. More importantly, panel (e) of Figure 3 shows that $Q_{obs} < Q_{sim}$ throughout the Northern European lowlands, especially in France, Czechia and around the Baltic Sea (Northeast Germany, Poland, Estonia and southeastern Sweden). At such scales, I_{GF} can hardly explain the spatial coherence of the signal, especially when the underlying geology is non-uniform. Finally, the negative imbalance in Northern France (and especially in the Paris Basin and the Beauce Region) could also stem from I_H, as irrigation practices would increase evapotranspiration at the expense of runoff. However, the impact of irrigation is inextricable from the many sources of biases regarding evaporation, which are discussed below. Overall, we believe that precipitation and evaporation biases provide the most reasonable explanation for the low-frequency patterns we showed in panel (d) of Figure 3.

4.3.2 The challenge of interpolation

As illustrated in section 3.2, the interpolation stage of our mesoscale analysis is deeply sensitive to the density and coverage (representativeness) of the gauged network. The best explanation is found by looking at Eq. (5). Decreasing the number of constraints leads to decreasing the relative contribution of the data-fitting term $\sum_i (f(x_i) - y_i)^2$, thus increasing the relative importance of the roughness penalty $\lambda_{TPS} \int_S f^{(m)}(x)^2 dx$ (integral constraint on the spline curvature). Therefore, fewer datapoints lead to more substantial deviations of the interpolated signal (regardless of the value of λ_{TPS}). It should be noted that we chose to pre-emptively exclude those catchments where $(Q - \Delta S_{glac.})/P > 2$, specifically because the impact of the catchments can be strong in the more sparsely gauged regions (e.g. Iceland and Portugal). Other interpolators might prove more robust, although we suspect that an increased gauge density remains the best solution (Hofstra et al., 2010).

Sections 3.2 and 3.3 have also revealed that the extraction of a low-frequency signal is a fine balance between oversmoothing and overfitting. Because we opted for the TPS interpolator, the compromise between smoothing and fitting is governed by the fixed roughness penalty parameter λ_{TPS}. Consequently, the same level of smoothing applies throughout the domain, implying that the characteristic length of the mesoscale signal is uniform and that the underlying causes exhibit homogeneous spatial smoothness. Yet, the definition of the mesoscale may change across different regions, since the hydrological processes whose

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

variability we want to filter out may be different from one region to the other. Future work might focus on the refinement of the present methodology with spatially varying smoothing parameters (Pintore et al., 2006; Wang et al., 2013). Moreover, classic TPS interpolation is isotropic (Wood, 2003), which means that the smoothing penalty does not depend on the direction of smoothing (i.e. λ_{TPS} is rotation-invariant). Once again, this is potentially ill-suited to our problem considering that the orography is often dominated by anisotropy (mountain ranges, peninsulas, valleys).

Interpolation issues raise questions as to the practical value of mesoscale analyses. The bottom line is the following question: is the regionalisation procedure of the IWB anomaly any good? In particular, to what extent can we safely 'downscale' the mesoscale anomaly to the catchments or mesh units of a distributed model, and what is the interpretive value of the residuel signal? Ultimately, can we use the mesoscale imbalance to improve the predictive performance of hydrological models? Addressing these important questions should be a key focus of future studies.

4.3.3 The interannual timescale and the Turc-Mezentsev model

We used long-term averages as a means of simplifying water balance analyses. This approach builds on the questionable assumption that hydrological conditions are stationary, so that interannual averages are meaningful. Koutsoyiannis (2003) analysed Northern Hemisphere temperature anomalies and showed that a "local multi-year average (5 or 25-year averages) is not stable but, rather, it exhibits significant variability". This means that the average anomaly of any given catchment is a virtual construct. For example, precipitation levels might decline while the evaporative demand increases, which also entails that systematic biases of both precipitation and evaporation might change. This amounts to questioning the relevance of sweeping at the interannual timescale and future developments may need to incorporate non-stationarity.

In turn, hydrological non-stationarity also means that the position of a catchment on a Budyko-style graph (see panel (b) in Figure 2) is not fixed, even when we look at the interannual average over several decades. In addition, the '*catchment trajectory conjecture*', stating that individual catchments will move along a Budyko-style curve under a change in aridity, is not always empirically verified (Coron et al., 2015; Reaver et al., 2022). Therefore, the Budyko class of interannual models is likely sensitive to the averaging period, which could have an impact on our ability to interpret the IWB non-closure, but also on our unnesting methodology (see supplements).

Beyond the difficult interpretation of average imbalance, the choice of the Turc-Mezentsev interannual model also raises questions. Even though the original Budyko framework has been widely deployed (Xiong and Guo, 2012; Andréassian and Perrin, 2012; Berghuijs et al., 2020), Padrón et al. (2017) has shown that "controls other than [aridity] explain at least 35% [of the variance of runoff efficiency], and up to 90% in arid climates. Likewise, Milly (1994) proposed an alternative interannual model using seven dimensionless variables, set against only one variable for Budyko-styled approaches (aridity index). Consequently, it is possible that part of the observed mesoscale anomaly is merely the consequence of model inadequacy. Further work will need to address the sensitivity of the results to that choice.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

4.3.4 Potential evaporation: the ultimate loose-end

Potential evaporation is a major source of uncertainty for mesoscale analyses, because the calibration of the interannual model is very sensitive the choice of the potential evaporation formula. A change in our estimation of E_P will change both the catchment aridity, but also the synchronicity index (see section 2.3). Thus, the position of the regionalised Turc-Mezentsev curve is E_P-dependent. Whilst the Budyko-style models are more parsimonious (parameter wise) than any other class of hydrological models, this tells us that even the calibration of such models is burdened with a lack of robustness linked to the indeterminacy of the evaporation. Unfortunately, there is no consensus pertaining to the choice of a better E_P model at continental scales and interannual timescales. Over Europe, Bouabdelli et al. (2025) shed light on the adequacy of the FAO Penman-Monteith formula for hydrological studies - although deep uncertainties remain, mostly due to the lack of observational constraints.

Additionnally, the impact of vegetation was not included in the present study, so that the absence of vegetation becomes a mere bias on potential evaporation. Vegetation could be one reason for the negative imbalance across the Northern European Plains and extending from France to Estonia, as the lack of vegetation will possibly translate into a systematic underestimation of actual evaporation. Our treatment of the lack of vegetation as a bias recognises the strong uncertainties in the representation of the vegetation properties (as discussed in Sarrazin et al., 2018). Moreover, the inclusion of vegetation in Budyko-style interannual models has been a topic of unresolved discussions. For instance, Zhang et al. (2001) proposed a dedicated formula accounting for 'plant-available water capacity', while Li et al. (2013) and Zhang et al. (2016) focused on the possibility of connecting the Turc-Mezentsev shape parameter n to vegetation indices. Nonetheless, Reaver et al. (2022) highlighted that *"the attribution of physical meaning to the specific functional forms of [parametric Budyko] curves [...] is an unsupported conjecture"*. Padrón et al. (2017) also found that vegetation-related indices *"play a minor and often nonsignificant role in the long-term partition between precipitation and potential evaporation"*. Altogether, the absence of a widely accepted consensus as to the inclusion of vegetation comes down to the interpretation of the Budyko framework, stating that runoff production is chiefly governed by the equilibrium of the atmospheric water supply and the atmospheric energy supply (equating to the atmospheric water demand). Because vegetation lies at the interface between the catchment and the atmosphere, the question of its inclusion is in fact the question of whether vegetation should be treated as a catchment property (which determines the partition of runoff and evaporation like any other catchment property) or a property of the atmospheric energy supply.

5 Conclusions

In this paper, we conjectured the existence of a mesoscale, a spatial scale at which water balance non-closure is primarily determined by the forcing biases. The main takeaway of our analyses is the existence of coherent low-frequency patterns of water balance non-closure, which are consistent with our initial hypothesis. Our confidence in the signal has been shown to be conditional to the representativeness of the gauged network, pointing to the value of large sample hydrology.

Mesoscale imbalance brings valuable insights for hydrological modelling, by showing that spatial variability can be a relevant tool to disentangle the many causes of water balance non-closure. While we agree with Kuczera et al. (2010) that water balance non-closure cannot be solved by looking at rainfall-runoff records alone, the mesoscale framework suggests that those

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

records still hold unexploited value to decompose the problem. The implications are two-fold. Firstly, future work might focus on the partioning of the mesoscale imbalance to achieve *targeted* and *distributed* sweeping of potential evaporation and precipitation biases for distributed models. Secondly, mesoscale imbalance also sheds new light on catchment-scale imbalance, through the residual signal: the strenghtening of non-closure after the deduction of the mesoscale signal is hinting to the compensatory effects which curtail the interpretive value of the IWB anomaly. It offers proof that the anomaly is at best a *lower bound* on the errors and biases which affect the many terms of the closure equation. Studying water imbalance through the lens of spatial variability reveals how often we might be dealing with good performances for the wrong reasons, due to unconstrained sweeping.

In many ways, the challenge of water balance non-closure is analogous to the ancient trial of the Lernean Hydra, the fearsome creature that sprouted a pair of new heads for each one chopped off with a sword. The infamous monster could only be defeated by burning the stumps. The myth is a figurative reminder of what we should strive to avoid: complex problems (i.e. defeating the hydrological monster of water balance non closure) cannot always be addressed with all-purpose generic solutions (i.e. chopping off water imbalance with heedless sweeping), whose apparent advantages frequently sow the seeds of inextricable problems. This paper aimed to show that the analysis of spatial variability could be the fire that helps holding back (but not defeating !) the endless return of non-closure issues.

Data availability. Streamflow data were obtained from the databases listed in Table 1. Several timeseries in the ENTWASH collection are not openly accessible, as they were kindly provided upon request for the specific purpose of writing this paper; table B1 provides a complete overview for data retrieval. The final list of station identifiers, as well as the years we chose for computing water-balance closure, can be found in the supplements. The MERIT DEM was retrieved under the CC BY-NC license, and the JRC CCM river network was retrieved under non-commercial use restrictions (©European Commission – JRC, 2007). Both the GLIMS dataset and the glacier melt datset from Dussaillant et al. (2025) are distributed under open access licenses with appropriate attribution. Finally, the country delineation for the European maps was retrieved from the `rnaturalearth` library, whose data stem from Natural Earth raster and vector data, belonging to the public domain.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

Appendix A: Detailed description of ENTWASH selection criteria based on anthopogenic influence

Dataset	Brief description of gauge selection criteria
ROBIN	All **semi-natural catchments**, with two-levels of naturalness (Turner et al., 2025). Level 1: Largely free from human disturbances such as urbanisation ($\leq$ 10% of the catchment), river engineering and water abstractions. Modest net impact of all influences on low flows and high flows and any impact stable over time. No known major changes in land use likely to impact streamflow regime. Level 2.: Fairly free from human disturbances such as urbanisation ($\leq$ 20% of the catchment), river engineering and water abstractions. Modest net impact of all influences on monthly and annual flows and any impact stable over time. For ENTWASH, we kept both levels.
ROBIN - Norway	Norwegian Reference Hydrometric Network. Selection criteria (Fleig et al., 2013) include: (a) Degree of river basin development; (b) regulations; (c) diversions (d) water use and abstraction for drinking water.
ROBIN - Ireland	Irish Reference Hydrometric Network (O'Connor et al., 2022). Selection criteria: (a) station performance at high, mean and low flows; (b) abstractions, arterial drainage and other anthropogenic influences; (c) exploratory data analysis via station hydrographs, histograms and missing data plots; (d) questionnaire developed for hydrometric experts that further queried hydrometric performance at each gauge and sought feedback on inclusion in the final network.
ROBIN - UK	UK Reference Hydrometric Network. Selection criteria (Harrigan et al., 2017): (a) detailed station metadata including water abstractions and discharges; (b) visual screening; (c) consultation with those responsible for maintenance and collection of hydrometric data.
ROBIN - Germany	Selection criteria: (a) percentage of artificial surface; (b) number of dams
ROBIN - France	French Reference Hydrometric Network. Gauge selection based on: (a) Banque Hydro (SCV, 2026) metadata on quality of low, mean and high flows, local influences; (b) visual screening; (c) automatic detection of suspiciousbreaks; (d) discussion with local providers to evaluate likely sources of alteration.
ROBIN - Spain	Selection criteria include: (a) dams and water management structures; (b) abstractions for agriculture and irrigation.
ROBIN - Portugal	No details provided by the authors.
ROBIN - Poland	Refinement on the selection of semi-natural catchments from Venegas-Cordero et al. (2022). New criteria include: (a) soil sealing data from Copernicus Imperviousness layer; (b) data on surface and groundwater withdrawals and discharges; (c) additional data on dams and hydrotechnical structures; (d) volumes of storage reservoirs. Final selection: 139 gauges in ROBIN Level 1 and 6 additional gauges in Level 2.
ROBIN - Finland	Based on a personal communication with Uusikivi Jari, stations were selected on the basis of: (a) dams, reservoirs; (b) abstractions ; (c) diversions of the water flow; (d) stability of land use has remained constant for the whole observation period; (e) expert appreciation of potential large changes in the drainage of the forests of agriculture land.

Table A1. Description of the naturality criteria of datasets which were included in ENTWASH - Part 1. In the case of ROBIN, we describe the common guidelines as well as the detailed approach of each participating country.

The selection of the ENTWASH catchments relied on several different criteria, depending on the available information or on the choices of the providers themselves. Tables A1, A2 and A3 provide a comprehensive account of such criteria.

Some datasets provided explicit flags for human influences (LamaH-CE, LamaH-Ice, ROBIN and HydCHeck), which aggregate multiple quantitative criteria, encompassing dams and regulation, but also irrigation, drinking water or industrial abstractions. The flags may take the form of an explicit ranking from low to high levels of human alteration. Only those catchments with 'low' or no alteration were included in the ENTWASH collection.

Sometimes, a simple threshold (10 mm) on the water height equivalent of total upstream storage (Delaigue et al., 2025; Strohmenger et al., 2024) was used, in addition to the visual inspection of timeseries to check for the most obvious issues

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

Dataset	Brief description of gauge selection criteria
GRDC-CSS	In ENTWASH, only for Estonia, Romania and Slovakia. Selection criteria: (a) quality, acuracy and length of timeseries; (b) degree of basin development; (c) stability of anthropogenic alterations; (d) no significant regulation or diversions: total extractions or diversions should not exceed 5% of the mean annual flow.
LamaH-CE	Four degrees of anthropogenic impact, ten impact categories, including: (a) reservoirs, lakes and channels; (b) water abstractions. For ENTWASH, catchments falling within the lowest two levels of impact were selected.
LamaH-Ice	Four degrees of anthropogenic impact, six impact categories, including: (a) upstream hydropower reservoir; (b) subsurface transboundary flow; (c) withdrawal for irrigation and revegetation; (d) natural or artificial diversion or uncertainties in watershed surface. For ENTWASH, we kept the lowest two levels (low or no influence).
CAMELS-CH	We filtered the dataset based on the BAFU flag system drawing from the HydCHeck methodology (WSL, 2024). Seven alteration sources (e.g. groundwater, agriculture, dams, etc.), whose impact on the catchment is measured with a specific metric ranging between 0 and 1. After evaluating all the alteration sources, they are incorporated into an aggregated metric on five aspects of the hydrological signature (low flow, high flow, etc.). The aggregated metric is a weighted sum of the different metrics of all seven alteration sources, that have different repercussions on the five aspects of the hydrological signature. The aggregated metric is then used to assign a label of alteration level (none, low, moderate and strong). For ENTWASH, we kept all catchments with low or no alteration on all aspects of the signature.
CAMELS-DK	Selection based on hydrological model calibration and validation (Liu et al., 2025). Jun Liu personal communication: "largely free of human influence (no reservoirs, pumping stations, weirs, or surface-water extractions, etc.). However, some control structures do exist, for example at certain lake outlets, and groundwater extraction does impact streamflow/baseflow across large parts of Denmark". We carried out additional visual screening for the final selection.
CAMELS-FI	Human influence attributes include: (a) population count and density; (b) dam and reservoirs count and capacity; (c) regulation level. For ENTWASH, we only selected basins with the lowest two levels of regulation (minor and unregulated).
CAMELS-LUX	Selection criteria include: (a) dams and derivation; (b) industrial influences; (c) urban influences.
BULL	Personal communication with Javier Senent-Aparicio: selection based on hydrological model validation, and comparison with the information provided by ROEA (Red Oficial de Estaciones de Aforos) that assesses whether basins are altered based on the presence of upstream reservoirs. BULL selection differs from ROEA classification of natural catchments; among others, the BULL methodology leads to the exclusion of the most heavily karstic catchments.

Table A2. Description of the naturality criteria of datasets which were included in ENTWASH - Part 2.

(see table A2 and A3). For CAMELS-CZ, the upstream storage criterion was refined with an additional threshold on average elevation (500m). In the case of Italy, where a nation-wide open-access streamflow dataset is anyhow missing, we relied on EStreams, coupled with the Italian Large Dams dataset (see table 1). As to the datasets under the "Local Expertise" flag, timeseries were chosen by local experts and hydrologists, based on their knowledge of basin characteristics.

The retrieval of natural timeseries was more challenging in the most arid countries, where human activities (especially irrigation) have deep-reaching consequences onto the regional hydrology. In Spain, the BULL dataset made use of the SIMPA hydrological model to identify the least affected catchments. Finally, timeseries from the French electricity provider (EDF) were naturalised, by computing the reservoir water balance (Terrier et al., 2021) to reconstruct daily streamflow values.

In several countries, multiple data sources overlapped, especially in France, Germany, Switzerland and Austria. To prevent all risks of duplicate streamflow gauges, we computed all the gauge-to-gauge distances; when the distance was less than 500 m, we assumed the two gauges to be identical, and we applied the following rule. Sometimes, the reported gauge identifiers

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

Dataset	Brief description of gauge selection criteria.
Upstream storage and/or visual screening	
CAMELS-CZ	Selection criteria: (a) headwater catchments (average elevation above 500 $m\ a.s.l.$); (b) simple threshold (10 mm) on the water height equivalent of total upstream storage.
CAMELS-DE	Selection criteria: (a) visual screening of individual hydrographs.
CAMELS-FR	Selection criteria: (a) simple threshold (10 mm) on the water height equivalent of total upstream storage, and (b) visual screening of individual hydrographs.
EStreams (IT)	Selection criteria: (a) EStreams quality gauge flag, levels "A" or "B"; (b) simple threshold (10 mm) on the water height equivalent of total upstream storage, based on the Italian Large Dam dataset (Evangelista et al., 2025).
Explore2	Selection criteria: (a) simple threshold (10 mm) on the water height equivalent of total upstream storage, and (b) visual screening of individual hydrographs.
Sweden	Selection based on de Lavenne et al. (2022): "stations having less than 5% of their area regulated are used [...]. The degree of regulation is the percentage of the volume of the mean annual flow that can be stored in reservoirs located upstream of the gauged catchment outlet. We carried out additional visual screening for the final selection.
Naturalised timeseries	
EDF	The timeseries provided by the French electricity provider EDF have been naturalised for the Explore2 project (see line above), by computing the reservoir water balance (Terrier et al., 2021) to reconstruct daily streamflow values.
Local expertise	
North Macedonia	Selection by Radevski et al. (2018), aiming for "*natural or near natural regime without significant human activities, canals, dams, ameliorative or water supply systems on the stream, and there are also waste water treatment plants that affect water regime in time, land degradations, deforestations, infrastructures etc.*"
Bosnia Herzegovina	Selection by Gnjato et al. (2023) of 5 catchments "*not affected by major anthropogenic impacts*", especially dams.
SHMÙ	Selection made by Katarína Jeneiová, based on the station metadata provided by SHMÙ.
NIMH	Selection made by Eram Artinyan, based on upstream dams and derivation.
RHMZ Srbije	Selection made by Slavimir Stefanovic: "stations that are located in head parts of the basins and for which the conditions of natural flow apply". We retrieved daily timeseries from the RHMSS online archive, and digitised them manually for the purpose of this study.
HMC-RA	Selection made by Amalya Misakyan, focusing on regulation.

Table A3. Description of the naturality criteria of those datasets which were included in ENTWASH - Part 3.

were not standard, so that identical gauges still had different IDs from one dataset to another. In Germany, the correspondance between ROBIN and CAMELS-DE draws from a personal communication with Sebastian Gnann.

Appendix B: Data retrieval overview

610
Appendix C: Catchment delineation

We used the MERIT (Yamazaki et al., 2017) digital elevation model (DEM) to delineate the catchments of the ENTWASH dataset. The DEM was burnt ($-5\ m$) with the help of the rasterised JRC CCM river network (Vogt et al., 2007), and filled (`whitebox` R package) to avoid endorheic configurations. We snapped gauge coordinates on the DEM river network. The

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

Dataset	Access condition
ROBIN	Download URL : https://https://catalogue.ceh.ac.uk/documents/3b077711-f183-42f1-bac6-c892922c81f4
GRDC-CSS	Station catalogue : https://grdc.bafg.de/data/project_datasets/climate_sensitive_stations/ Donwload portal : https://portal.grdc.bafg.de/applications/public.html?publicuser=PublicUser#dataDownload/Subregions
LamaH-CE	Donwload URL : https://zenodo.org/records/5153305
BULL	Download URL :https://doi.org/10.5281/zenodo.10605646
CAMELS-CH	Download URL : https://doi.org/10.5281/zenodo.7784632
CAMELS-CZ	Data sent before the official publication by Michal Jeníček and Ondřej Ledvinka at Charles University, Prague, Czech Republic.
CAMELS-DE	Donwload URL : https://doi.org/10.5281/zenodo.13837553
CAMELS-DK	Dowload URL : https://dataverse.geus.dk/dataset.xhtml?persistentId=doi:10.22008/FK2/AZXSYP
CAMELS-FR	Dowload URL : https://doi.org/10.57745/WH7FJR
CAMELS-FI	Dowload URL : https://doi.org/10.5281/zenodo.15853357
CAMELS-LUX	Dowload URL : https://doi.org/10.5281/zenodo.13846619
EStreams	Dowload URL : https://zenodo.org/records/10733142
Explore2	Dowload URL : https://doi.org/10.57745/MXGCX5
LamaH-Ice	Dowload URL : https://doi.org/10.4211/hs.86117a5f36cc4b7c90a5d54e18161c91
EDF	Courtesy of Electricité de France. Please contact the authors for details.
SHMÙ	Private agreement with SHMÙ. Contact : Katarína Jeneiová from the Slovak Hydrometeorological Institute, Bratislava (SHMÙ), Slovakia.
NIMH	Courtesy of Eram Artinyan at the National Institute of Meteorology and Hydrology, Bulgaria.
RHMZ Srbije	Online archive (.pdf) URL : https://www.hidmet.gov.rs/latin/hidrologija/povrsinske_godisnjaci.php. Contact for selection of gauges : Samir Ćatović and Slavimir Stevanovic, Assistant Director at RHMZ Srbije, Serbia.
IMGW-PIB	Online archive, access with `climate` R package. Contact : Tomasz Berezowski, professor at Gdańsk University of Technology and Mikołaj Piniewski at Warsaw University of Life Sciences, Warsaw (SGGW), Poland.
HMC-RA	Courtesy of Amalya Misakyan, head of Hydrology Service at the Hydrometeorology and Monitoring Center, Republic of Armenia.
North Macedonia	Courtesy of Ivan Radevski, Professor at "Ss Cyril and Methodius" University in Skopje.
Sweden	See de Lavenne et al. (2022)
Bosnia-Herzegovina	Courtesy of Slodoban Gnjato, Professor at the University of Banja Luka.

Table B1. Access conditions for all datasets included in ENTWASH

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

snapping algorithm works as a straightforward search for the pixel which minimises the area mismatch. Firstly, it takes the initial coordinates of a gauge, and compares the provider's catchment area to the DEM upstream area at the gauge location; if the relative difference is less than 1%, the algorithm stops. If not, the neighbouring pixels around the original location are tested, to find which one minimises the area mismatch. As long as the mismatch is greater than 1%, we increase the size of the neighbourhood, until the distance from the original location reaches $2\ km$. Beyond $2\ km$, the algorithm stops as soon as the mismatch does not improve from one iteration to the next or when a distance of $4\ km$ is reached. Having run the snapping method, any gauge with a final area mismatch above 10% was rejected.

Appendix D: Catchment unnesting

Here, we describe the unnesting procedure, which consists in deducting all the upstream inflows from the outflow of the downstream basin.

For each catchment we defined a vector of yearly validity:

$$\mathcal{V} = \{V_j\}_{j\in\{1984,\ldots,2025\}}, \quad \text{where } V_j \in \{0,1\} \tag{D1}$$

$V_j = 1$ if the annual total discharge for year j has a valid value, and $V_j = 1$ if not. For each pair of upstream and downstream catchments, we computed the number of years where both catchments have a valid annual discharge value:

$$\mathcal{V}^{up} \cap \mathcal{V}^{down} = \sum_{j=1984}^{2025} \min(V_j^{up}, V_j^{down}) \tag{D2}$$

Unnesting was considered appropriate whenever the upstream and downstream catchments shared 15 valid years:

$$\mathcal{V}^{up} \cap \mathcal{V}^{down} \geq 15 \tag{D3}$$

Otherwise, one of the two gauges had to be left out, in which case the larger downstream catchment is always favoured, except if its own number of valid years is less than 15. The main issue emerges when multiple nested catchments share the same downstream catchment. In that case, the condition of Eq. (D3) may not be met for all the upstream gauges. If the downstream catchment has more than 15 valid years between 1984 and 2025, we take out all the upstream gauges where D3 is not verified (see Figure D1, top); if not, the downstream gauge is taken out (see Figure D1, bottom).

When a pair of nested catchments are deemed fit for unnesting, we rescale the average discharge of either catchments to the 1985-2024 reference period. For a given catchment with a vector of yearly validity $\mathcal{V}$, the rescaling is given by:

$$Q_{rescaled} = \frac{Q_{ref}^{IM}}{Q_{\mathcal{V}}^{IM}} \times Q_{\mathcal{V}}^{obs} \tag{D4}$$

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

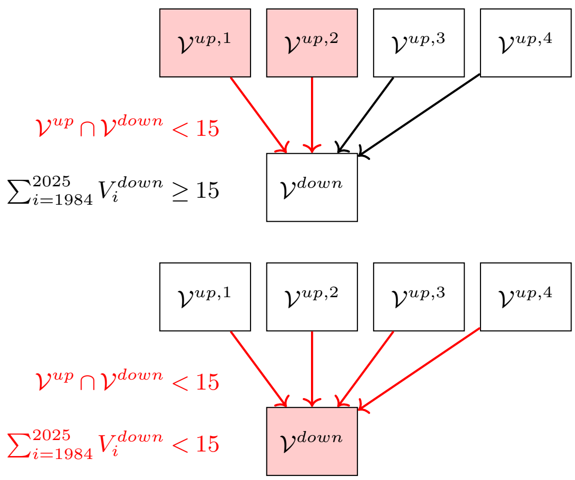

Figure D1. Unnesting strategy: red boxes are the gauges we take out, white boxes those that we keep. On top: the downstream catchment shares 15 valid years with only two upstream catchments, which leads to leaving out $\mathcal{V}^{up,1}$ and $\mathcal{V}^{up,2}$. Below: the downstream catchment shares less than 15 valid years with all upstream catchments, but it has less than 15 valid years on its own. In this case, we leave out the downstream basin.

where $Q_{\mathcal{V}}^{obs}$ is the interannual average of the observed discharge across all valid years; $Q_{\mathcal{V}}^{IM}$ is the interannual average of the modelled discharge across all valid years; IM stands for the interannual model (here, the Turc-Mezentsev model), which we described in section 2.3; Q_{ref}^{IM} is the interannual average of the modelled discharge across the complete reference period, 1984-2025. After rescaling, we could finally proceed with unnesting:

$$Q_{down}^{unnested} = Q_{rescaled}^{down} - \sum_j Q_{rescaled,j}^{up} \tag{D5}$$

Author contributions. **MR**: Conceptualisation, Methodology, Software, Formal Analysis, Investigation, Data Curation, Visualisation, Writing - Original Draft. **FJS**: Methodology, Supervision, Data curation, Writing (review and editing) **FH**: Methodology, Supervision, Funding Aquisition, Writing (review and editing) **VA**: Methodology, Data Curation, Supervision, Writing (review and editing)

Competing interests. One of the authors (FJS) is a member of the editorial board of Hydrology and Earth System Science. MR and FH are employed by Electicité de France (EDF), which partially funded this research through the CIFRE program managed by the French association for Research and technology (ANRT). The contact author has declared that none of the remaining authors has any competing interests.

Acknowledgements. **For providing data, we thank:** Amalya Misakyan at the Armenian State Hydrometeorological and Monitoring Service; Michal Jeníček and Ondřej Ledvinka at Charles University, Prague, Czech Republic; the Slovak Hydrometeorological Institute and especially Katarína Jeneiová, SHMU, Slovakia; Ivan Radevski at Cyril and Methodius University, North Macedonia; Slodoban Gnjato at University of Banja Luka, Bosnia Herzegovina; Judith Nijzink at Luxembourg Institute of Science and Technology, Luxembourg; Eram Artinyan at NIMH, Bulgaria; Caroline Kan Stettler at BAFU, Switzerland.

For the answers and clarification they provided, we thank: Sayali Pawar and Steve Turner at UKCEH, UK; Thomas Berezowski at Gdańsk University of Technology and Mikołaj Piniewski at Warsaw University of Life Sciences, Poland; Demetris Koutsoyiannis at National Technical University of Athens, Greece; Uusikivi Jari from SYKE and Iiro Seppä at Turku University, Finland; Tatjana Vujnović at DHMZ, Croatia; Mathew Herrnegger at BOKU University and Harald Kling at AFRY, Austria; Jun Liu at Geological Survey of Denmark and Greenland, Denmark; Conor Murphy at Maynooth University, Ireland; Slavimir Stevanovic and Samir Ćatović at RHMSS, Serbia; René Capell at SMHI, Sweden; Andréa Massad at the Icelandic Met Office, Iceland; Marius-Victor Bîrsan at the Romanian Academy, Romania; Luis Mediero at Universidad Politécnica de Madrid, Javier Senent Aparicio at CIDE-CSIC and José Martínez-Fernández at Universidad de Salamanca, Spain; Anja Iselin Pedersen and Anne Kristina Fleig at NVE, Norway; Petra-Schmocker Fackel at BAFU, and Bettina Schaefli at University of Bern, Switzerland; Paola Mazzoglio and Alberto Viglione at the Politecnico di Torino, and Alessia Cecchi at SIR Toscana, Italy; Sebastian Gnann at Freiburg University, Simon Mischel at GRDC, and Patricio Yeste at Potsdam University, Germany; Pierre Brigode at Rennes University and Benjamin Renard at INRAE, France.

We also thank our colleagues at INRAE, Antony, France, for the careful selection of the CAMELS gauges, which facilitated the construction of ENTWASH: Guilherme Mendoza Guimarães, Antoine Degenne, Gaëlle Tallec and François Bourgin. Special thanks to Paul Zarpas and Charles Perrin for reading the manuscript. Likewise, we thank our colleagues Emmanuel Paquet, Céline Monteil, Matthieu Le Lay, Léonard Santos and Camille Brun at EDF, for the rich and engaging discussions about the water balance 'quagmire'.

Financial support. This research has been supported by the French Association for Research and Technology (ANRT - grant no. 2024/1559), in partnership with Electricité de France (EDF).

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

References

Abbas, A., Yang, Y., Pan, M., Tramblay, Y., Shen, C., Ji, H., Gebrechorkos, S. H., Pappenberger, F., Pyo, J., Feng, D., Huffman, G., Nguyen, P., Massari, C., Brocca, L., Tan, J., and Beck, H. E.: Comprehensive Global Assessment of 24 Gridded Precipitation Datasets Across 18 428 Catchments Using Hydrological Modeling, Hydrology and Earth System Sciences, 30, 3399–3423, https://doi.org/10.5194/hess-30-3399-2026, 2026.

Adam, J. C., Clark, E. A., Lettenmaier, D. P., and Wood, E. F.: Correction of Global Precipitation Products for Orographic Effects, Journal of Climate, 19, 15 – 38, https://doi.org/10.1175/JCLI3604.1, 2006.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements, vol. 56 of *FAO Irrigation and Drainage Paper*, Food and Agriculture Organization of the United Nations (FAO), Rome, ISBN 978-92-5-104219-9, 1998.

Alvarez, I., Pereira, H., Lorenzo, M., Picado, A., and Dias, J.: Assessing the performance of the CERRA dataset in reproducing precipitation and extreme precipitation events over the Iberian Peninsula, Journal of Hydrology: Regional Studies, 60, 102 517, https://doi.org/https://doi.org/10.1016/j.ejrh.2025.102517, 2025.

Andréassian, V. and Sari, T.: Technical Note: On the puzzling similarity of two water balance formulas – Turc–Mezentsev vs. Tixeront–Fu, Hydrology and Earth System Sciences, 23, 2339–2350, https://doi.org/10.5194/hess-23-2339-2019, 2019.

Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., and Lavabre, J.: Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models, Journal of Hydrology, 250, 206–223, https://doi.org/https://doi.org/10.1016/S0022-1694(01)00437-1, 2001.

Andréassian, V., Mander, Ü., and Pae, T.: The Budyko hypothesis before Budyko: The hydrological legacy of Evald Oldekop, Journal of Hydrology, 535, 386–391, https://doi.org/https://doi.org/10.1016/j.jhydrol.2016.02.002, 2016.

Andréassian, V., Guimarães, G. M., Lerat, J., and de Lavenne, A.: Streamflow elasticity as a function of aridity, Hydrology and Earth System Sciences, 30, 1865–1876, https://doi.org/10.5194/hess-30-1865-2026, 2026.

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

Beniston, M.: Impacts of climatic change on water and associated economic activities in the Swiss Alps, Journal of Hydrology, 412–413, 291–296, https://doi.org/10.1016/j.jhydrol.2010.06.046, hydrology Conference 2010, 2012.

Berghuijs, W. R., Gnann, S. J., and Woods, R. A.: Unanswered questions on the Budyko framework, Hydrological Processes, 34, 5699–5703, https://doi.org/https://doi.org/10.1002/hyp.13958, 2020.

Beven, K.: How far can we go in distributed hydrological modelling?, Hydrology and Earth System Sciences, 5, 1–12, https://doi.org/10.5194/hess-5-1-2001, 2001.

Beven, K.: Searching for the Holy Grail of scientific hydrology: $Q_t = (S, R, \Delta t)A$ as closure, Hydrology and Earth System Sciences, 10, 609–618, https://doi.org/10.5194/hess-10-609-2006, 2006.

Bisselink, B., Bernhard, J., Gelati, E., Adamovic, M., Guenther, S., Mentaschi, L., and De Roo, A.: Future projections of water scarcity in the Danube River Basin due to land use, water demand and climate change, 2018.

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S. T., Amin, A., Andréassian, V., Arheimer, B., Aryal, S. K., Baker, V., Bardsley, E., Barendrecht, M. H., Bartosova, A., Batelaan, O., Berghuijs, W. R., Beven, K., Blume, T., Bogaard, T., de Amorim, P. B., Böttcher, M. E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Y., Chen, Y., Chifflard, P., Claps, P., Clark, M. P., Collins, A. L., Croke, B., Dathe, A., David, P. C., de Barros, F. P. J., de Rooij, G., Baldassarre, G. D., Driscoll, J. M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W. H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Bevacqua, A. G., González-Dugo, M. P., Grimaldi, S., Gupta, A. B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T. H., Inam, A., Innocente, C., Istanbulluoglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnová, S., Kolechkina, A., Krause, S., Kreamer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M. L. R., Lindquist, E., Link, T., Liu, J., Loucks, D. P., Luce, C., Mahé, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Misstear, B. D., Montanari, A., Müller-Thomy, H., Nabizadeh, A., Nardi, F., Neale, C., Nesterova, N., Nurtaev, B., Odongo, V. O., Panda, S., Pande, S., Pang, Z., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M. J., Post, D., Sierra, C. P., Ramos, M.-H., Renner, M., Reynolds, J. E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D. E., Rosso, R., Roy, T., Sá, J. H., Salvadori, G., Sandells, M., Schaefli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R. C., Skaugen, T., Smith, H., Spiessl, S. M., Stein, L., Steinsland, I., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., van Beek, R., van der Ent, R. J., van der Ploeg, M., Loon, A. F. V., van Meerveld, I., van Nooijen, R., van Oel, P. R., Vidal, J.-P., von Freyberg, J., Vorogushyn, S., Wachniew, P., Wade, A. J., Ward, P., Westerberg, I. K., White, C., Wood, E. F., Woods, R., Xu, Z., Yilmaz, K. K., and Zhang, Y.: Twenty-three unsolved problems in hydrology (UPH) – a community perspective, Hydrological Sciences Journal, 64, 1141–1158, https://doi.org/10.1080/02626667.2019.1620507, 2019.

Bouabdelli, S., Massari, C., Morlot, M., Castelli, M., and Formetta, G.: High resolution (1 km-daily) potential evapotranspiration dataset over Europe and the Mediterranean region, Journal of Hydrology, 662, 133 839, https://doi.org/https://doi.org/10.1016/j.jhydrol.2025.133839, 2025.

Budyko, M. I.: Evaporation under natural conditions, Gidrometeorologicheskoe Izdatel'stvo, Tranlated from Russian by the Israel program for Scientific Translations in 1963, 1948.

Buisán, S. T., Earle, M. E., Collado, J. L., Kochendorfer, J., Alastrué, J., Wolff, M., Smith, C. D., and López-Moreno, J. I.: Assessment of snowfall accumulation underestimation by tipping bucket gauges in the Spanish operational network, Atmospheric Measurement Techniques, 10, 1079–1091, https://doi.org/10.5194/amt-10-1079-2017, 2017.

Carlson, G., Massari, C., Rotiroti, M., Bonomi, T.and Preziosi, E., Wilder, A., Whitaker, D., and Girotto, M.: Intensive irrigation buffers groundwater declines in key European breadbasket, 3, 683–692, https://doi.org/10.1038/s44221-025-00445-4, 2025.

Charlier, J., Desprats, J., and Ladouche, B.: Appui au SCHAPI 2014 – Module 1 – Rôle et contribution des eaux souterraines d'origine karstique dans les crues de la Loue à Chenecey-Buillon, Tech. Rep. BRGM/RP-63844-FR, BRGM, https://infoterre.brgm.fr, 2014.

Chauvin, G., Flerchinger, G., Link, T., Marks, D., Winstral, A., and Seyfried, M.: Long-term water balance and conceptual model of a semi-arid mountainous catchment, Journal of Hydrology, 400, 133–143, https://doi.org/https://doi.org/10.1016/j.jhydrol.2011.01.031, 2011.

Chávez García Silva, R., Reinecke, R., Copty, N. K., Barry, D. A., Heggy, E., Labat, D., Roggero, P. P., Borchardt, D., Rode, M., Gómez-Hernández, J. J., and Jomaa, S.: Multi-decadal Groundwater Observations Reveal Surprisingly Stable Levels in Southwestern Europe, Communications Earth & Environment, 5, 387, https://doi.org/10.1038/s43247-024-01554-w, 2024.

CHMI, C., Krejčí, J., and Nearing, G.: [CAMELS-CZ] Catchment Attributes and Meteorology for Large-Sample Studies – Czech Republic, https://doi.org/10.5281/zenodo.17338511, 2025.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

Clerc-Schwarzenbach, F., Astagneau, P., Castro, E. M., van Meerveld, I., Seibert, J., and Andréassian, V.: Cheeky Cheating or a Sensible Strategy? 'Sweep Parameters' in Bucket-Type Hydrological Models, Hydrological Processes, 40, e70 375, https://doi.org/10.1002/hyp.70375, hYP-25-1190, 2026.

Copernicus Climate Change Service, C.: Multi-model seasonal forecasts of river discharge for Europe from January 2021 to present - Hydrological model specification: A multimodel ensemble of hydrological models at the European scale (5 km grid and catchment) (E-HYPE and VIC-WUR) - Dataset: Hydrological Seasonal forecasts for Europe from January 2021, Climate Data Store, https://doi.org/10.24381/cds.52f45864, accessed on 16/03/2025, 2021.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. D.: An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets, Journal of Geophysical Research: Atmospheres, 123, 9391–9409, https://doi.org/https://doi.org/10.1029/2017JD028200, 2018.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resources Research, 48, https://doi.org/https://doi.org/10.1029/2011WR011721, 2012.

Coron, L., Andréassian, V., Perrin, C., and Le Moine, N.: 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, 2015.

Coutagne, A. and de Martonne, E.: De l'eau qui tombe à l'eau qui coule : évaporation et déficit d'écoulement, in: Proc. Congrès de Lisbonne 1933, International Union of Geodesy and Geophysics (IUGG), International Association of Scientific Hydrology, Lisbonne, Portugal, 1934.

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., and Smith, P. J.: A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations, Water Resources Research, 51, 5531–5546, https://doi.org/https://doi.org/10.1002/2014WR016532, 2015.

de Lavenne, A. and Andréassian, V.: Impact of climate seasonality on catchment yield: A parameterization for commonly-used water balance formulas, Journal of Hydrology, 558, 266–274, https://doi.org/https://doi.org/10.1016/j.jhydrol.2018.01.009, 2018.

de Lavenne, A., Andréassian, V., Crochemore, L., Lindström, G., and Arheimer, B.: Quantifying multi-year hydrological memory with Catchment Forgetting Curves, Hydrology and Earth System Sciences, 26, 2715–2732, https://doi.org/10.5194/hess-26-2715-2022, 2022.

Degenne, A., Bourgin, F., Perrin, C., and Andréassian, V.: Testing the Temporal and Spatial Transferability of a Water Balance Model using a Parameter Learning Approach, EGUsphere, 2026, 1–39, https://doi.org/10.5194/egusphere-2026-1197, 2026.

Delaigue, O., Guimarães, G. M., Brigode, P., Génot, B., Perrin, C., Soubeyroux, J.-M., Janet, B., Addor, N., and Andréassian, V.: CAMELS-FR dataset: a large-sample hydroclimatic dataset for France to explore hydrological diversity and support model benchmarking, Earth System Science Data, 17, 1461–1479, https://doi.org/10.5194/essd-17-1461-2025, 2025.

do Nascimento, T., Rudlang, J., Höge, M., van der Ent, R., Chappon, M., Seibert, J., Hrachowitz, M., and Fenicia, F.: EStreams: An integrated dataset and catalogue of streamflow, hydro-climatic and landscape variables for Europe, 11, 879, https://doi.org/10.1038/s41597-024-03706-1.

Dura, V., Evin, G., Favre, A.-C., and Penot, D.: Spatial variability in the seasonal precipitation lapse rates in complex topographical regions – application in France, Hydrology and Earth System Sciences, 28, 2579–2601, https://doi.org/10.5194/hess-28-2579-2024, 2024.

Dussaillant, I., Hugonnet, R., Huss, M., Berthier, E., Bannwart, J., Paul, F., and Zemp, M.: Annual mass change of the world's glaciers from 1976 to 2024 by temporal downscaling of satellite data with in situ observations, Earth System Science Data, 17, 1977–2006, https://doi.org/10.5194/essd-17-1977-2025, 2025.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

ECMWF, E. C. f. M.-R. W. F.: EFAS v5.0 – Hydrological Model Performances (Documentation), https://confluence.ecmwf.int/display/CEMS/EFAS+v5.0+-+Hydrological+Model+Performances, accessed: 2026-05-22, 2024.

Evangelista, G., Mazzoglio, P., Ganora, D., Pianigiani, F., and Claps, P.: Features of Italian large dams and their upstream catchments, Earth System Science Data, 17, 1407–1426, https://doi.org/10.5194/essd-17-1407-2025, 2025.

Fan, Y.: Are catchments leaky?, WIREs Water, 6, e1386, https://doi.org/https://doi.org/10.1002/wat2.1386, 2019.

Fleig, A. K., Andreassen, L. M., Barfod, E., Haga, J., Haugen, L. E., Hisdal, H., Melvold, K., and Saloranta, T.: Norwegian Hydrological Reference Dataset for Climate Change Studies, 2013.

Flerchinger, G. N. and Cooley, K. R.: A ten-year water balance of a mountainous semi-arid watershed, Journal of hydrology, 237, 86–99, 2000.

GLIMS and NSIDC: Global Land Ice Measurements from Space Glacier Database, https://doi.org/10.7265/N5V98602, updated 2018. Compiled and made available by the international GLIMS community and the National Snow and Ice Data Center (NSIDC), Boulder, Colorado, USA, 2005.

Gnjato, S., Popov, T., Ivanišević, M., and Trbić, G.: Long-term streamflow trends in Bosnia and Herzegovina (BH), 82, 356, https://doi.org/10.1007/s12665-023-11040-9, 2023.

GRDC: GRDC Climate Sensitive Stations Dataset, https://www.bafg.de/GRDC/EN/02_srvcs/22_gslrs/221_GRDC_CSS/grdc_css_node.html, accessed: 2026-06-03, 2026.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martínez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Harrigan, S., Hannaford, J., Muchan, K., and Marsh, T.: Designation and trend analysis of the updated UK Benchmark Network of river flow stations: the UKBN2 dataset, 49, 552–567, https://doi.org/10.2166/nh.2017.058, _eprint: https://iwaponline.com/hr/article-pdf/49/2/552/196278/nh0490552.pdf, 2017.

Helgason, H. B. and Nijssen, B.: LamaH-Ice: LArge-SaMple DAta for Hydrology and Environmental Sciences for Iceland, Earth System Science Data, 16, 2741–2771, https://doi.org/10.5194/essd-16-2741-2024, 2024.

Herrnegger, M., Nachtnebel, H. P., and Schulz, K.: From runoff to rainfall: inverse rainfall–runoff modelling in a high temporal resolution, Hydrology and Earth System Sciences, 19, 4619–4639, https://doi.org/10.5194/hess-19-4619-2015, 2015.

Hiebl, J. and Frei, C.: Daily precipitation grids for Austria since 1961-development and evaluation of a spatial dataset for hydroclimatic monitoring and modelling, 132, 327+, https://link.gale.com/apps/doc/A533217945/AONE?u=anon~24b35909&sid=googleScholar&xid=c6a8ddf3, 327.

Hofstra, N., New, M., and McSweeney, C.: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data, 35, 841–858, https://doi.org/10.1007/s00382-009-0698-1, 2010.

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: CAMELS-CH: hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland, Earth System Science Data, 15, 5755–5784, https://doi.org/10.5194/essd-15-5755-2023, 2023.

Hughes, D. and Gray, R.: Correcting bias in rainfall inputs to a semi-distributed hydrological model using downstream flow simulation errors, Hydrological Sciences Journal, 62, 2427–2439, https://doi.org/10.1080/02626667.2017.1391388, 2017.

Hutchinson, M. F.: Interpolating mean rainfall using thin plate smoothing splines, International journal of geographical information systems, 9, 385–403, https://doi.org/10.1080/02693799508902045, 1995.

Hutchinson, M. F. and Gessler, P. E.: Splines — more than just a smooth interpolator, Geoderma, 62, 45–67, https://doi.org/https://doi.org/10.1016/0016-7061(94)90027-2, 1994.

Kane, D. L. and Yang, D.: Northern research basins water balance, 290, International Association of Hydrological Science Press, 2004.

Kavetski, D., Kuczera, G., and Franks, S.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resources Research, 42, https://doi.org/https://doi.org/10.1029/2005WR004368, 2006a.

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, Water Resources Research, 42, https://doi.org/https://doi.org/10.1029/2005WR004376, 2006b.

Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, 45, https://doi.org/https://doi.org/10.1029/2008WR006912, 2009.

Klingler, C., Schulz, K., and Herrnegger, M.: LamaH-CE: LArge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe, Earth System Science Data, 13, 4529–4565, https://doi.org/10.5194/essd-13-4529-2021, 2021.

Konikow, L. F.: Long-Term Groundwater Depletion in the United States, Groundwater, 53, 2–9, https://doi.org/https://doi.org/10.1111/gwat.12306, 2015.

Koutsoyiannis, D.: Climate change, the Hurst phenomenon, and hydrological statistics, Hydrological Sciences Journal, 48, 3–24, https://doi.org/10.1623/hysj.48.1.3.43481, 2003.

Kuczera, G., Renard, B., Thyer, M., and Kavetski, D.: There are no hydrological monsters, just models and observations with large uncertainties!, Hydrological Sciences Journal, 55, 980–991, https://doi.org/10.1080/02626667.2010.504677, 2010.

Landerer, F. W., Dickey, J. O., and Güntner, A.: Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2003–2009, Journal of Geophysical Research: Atmospheres, 115, 2010.

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 Resources Research, 43, https://doi.org/https://doi.org/10.1029/2006WR005608, 2007.

Lebecherel, L., Andréassian, V., and Perrin, C.: On regionalizing the Turc-Mezentsev water balance formula, Water Resources Research, 49, 7508–7517, https://doi.org/https://doi.org/10.1002/2013WR013575, 2013.

Lehmann, F., Vishwakarma, B. D., and Bamber, J.: How well are we able to close the water budget at the global scale?, Hydrology and Earth System Sciences, 26, 35–54, https://doi.org/10.5194/hess-26-35-2022, 2022.

Li, D., Pan, M., Cong, Z., Zhang, L., and Wood, E.: Vegetation control on water and energy balance within the Budyko framework, Water Resources Research, 49, 969–976, https://doi.org/https://doi.org/10.1002/wrcr.20107, 2013.

Liu, J., Koch, J., Stisen, S., Troldborg, L., Højberg, A. L., Thodsen, H., Hansen, M. F. T., and Schneider, R. J. M.: CAMELS-DK: hydrometeorological time series and landscape attributes for 3330 Danish catchments with streamflow observations from 304 gauged stations, Earth System Science Data, 17, 1551–1572, https://doi.org/10.5194/essd-17-1551-2025, 2025.

Loritz, R., Dolich, A., Acuña Espinoza, E., Ebeling, P., Guse, B., Götte, J., Hassler, S. K., Hauffe, C., Heidbüchel, I., Kiesel, J., Mälicke, M., Müller-Thomy, H., Stölzle, M., and Tarasova, L.: CAMELS-DE: hydro-meteorological time series and attributes for 1582 catchments in Germany, Earth System Science Data, 16, 5625–5642, https://doi.org/10.5194/essd-16-5625-2024, 2024.

Lugeon, J.: Le cycle des précipitations atmosphériques: essai d'une formule donnant l'écoulement en fonction des précipitations, Ph.D. thesis, ETH Zurich, 1928.

Luo, Z., Fu, H., Shao, Q., Dong, W., Chen, X., Ding, X., Wang, L., Gu, X., Sarukkalige, R., Huang, H., and Li, H.: A novel method for correcting water budget components and reducing their uncertainties by optimally distributing the imbalance residual without full closure, Hydrology and Earth System Sciences, 29, 4607–4635, https://doi.org/10.5194/hess-29-4607-2025, 2025.

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., and Andersen, J.: seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day, Earth System Science Data, 10, 235–249, https://doi.org/10.5194/essd-10-235-2018, 2018.

Maier, P., Ehrendorfer, C., Lücking, S., Pulka, T., Lehner, F., Herrnegger, M., Formayer, H., and Koch, F.: Undercatch corrected gridded precipitation data to improve hydrological modeling in high-alpine orography, EGUsphere, 2026, 1–35, https://doi.org/10.5194/egusphere-2025-6382, 2026.

Manoj, J. A., Loritz, R., Gupta, H., and Zehe, E.: Can discharge be used to inversely correct precipitation?, Hydrology and Earth System Sciences, 29, 6115–6135, https://doi.org/10.5194/hess-29-6115-2025, 2025.

Marx, A., Kumar, R., Thober, S., Rakovec, O., Wanders, N., Zink, M., Wood, E. F., Pan, M., Sheffield, J., and Samaniego, L.: Climate change alters low flows in Europe under global warming of 1.5, 2, and 3C, Hydrology and Earth System Sciences, 22, 1017–1032, https://doi.org/10.5194/hess-22-1017-2018, 2018.

MeteoSwiss: Documentation of MeteoSwiss grid-data products-daily precipitation (final analysis): RhiresD, https://ethz.ch/content/dam/ethz/special-interest/baug/ifu/hydrology-dam/documents/research-data/ifu-hydrologie-data-proddocrhiresd.pdf, accessed April 2026, 2013.

Milly, P. C. D.: Climate, soil water storage, and the average annual water balance, Water Resources Research, 30, 2143–2156, https://doi.org/https://doi.org/10.1029/94WR00586, 1994.

Moccia, B., Buonora, L., Bertini, C., Ridolfi, E., Russo, F., and Napolitano, F.: What is our pick? Assessment of satellite and reanalysis precipitation datasets over Italy, Journal of Hydrology: Regional Studies, 60, 102 487, https://doi.org/https://doi.org/10.1016/j.ejrh.2025.102487, 2025.

Nijzink, J., Loritz, R., Gourdol, L., Zoccatelli, D., Iffly, J. F., and Pfister, L.: CAMELS-LUX: Highly Resolved Hydro-Meteorological and Atmospheric Data for Physiographically Characterized Catchments around Luxembourg, Earth System Science Data Discussions, 2025, 1–34, https://doi.org/10.5194/essd-2024-482, 2025.

O'Connor, P., Murphy, C., and Cahill, N.: The Irish Hydrometric Reference Network Version 2.0, EPA Research Report 2022-NE-1124, Environmental Protection Agency, Wexford, Ireland, prepared by the Irish Climate Analysis and Research Units (ICARUS), Department of Geography, Maynooth University, 2022.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., and Michel, C.: Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, Journal of Hydrology, 320, 62–83, https://doi.org/10.1016/j.jhydrol.2005.07.016, 2006.

Padrón, R., Gudmundsson, L., Greve, P., and Seneviratne, S.: Large-Scale Controls of the Surface Water Balance Over Land: Insights From a Systematic Review and Meta-Analysis, Water Resources Research, 53, 9659–9678, https://doi.org/https://doi.org/10.1002/2017WR021215, 2017.

Papa, K.-M. and Koutroulis, A. G.: Evaluation of precipitation datasets over Greece. Insights from comparing multiple gridded products with observations, Atmospheric Research, 315, 107 888, https://doi.org/https://doi.org/10.1016/j.atmosres.2024.107888, 2025.

Penman, H.: Evaporation over the British Isles, Quarterly Journal of the Royal Meteorological Society, 76, 372–383, https://doi.org/https://doi.org/10.1002/qj.49707633003, 1950.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

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

Pintore, A., Speckman, P., and Holmes, C. C.: Spatially adaptive smoothing splines, Biometrika, 93, 113–125, https://doi.org/10.1093/biomet/93.1.113, 2006.

Prein, A. F. and Gobiet, A.: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis, International Journal of Climatology, 37, 305–327, https://doi.org/https://doi.org/10.1002/joc.4706, 2017.

Radevski, I., Gorin, S., Taleska, M., and Dimitrovska, O.: Natural regime of streamflow trends in Macedonia, Geografie, 123, 1–20, https://doi.org/10.37040/geografie2018123010001, 2018.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, Journal of Hydrometeorology, 17, 287 – 307, https://doi.org/10.1175/JHM-D-15-0054.1, 2016.

Raup, B., Racoviteanu, A., Khalsa, S. J. S., C., H., R., A., and Arnaud, Y.: The GLIMS geospatial glacier database: A new tool for studying glacier change, Global and Planetary Change, 56, 101–110, https://doi.org/https://doi.org/10.1016/j.gloplacha.2006.07.018, climate Change Impacts on Mountain Glaciers and Permafrost, 2007.

Reaver, N. G. F., Kaplan, D. A., Klammler, H., and Jawitz, J. W.: Theoretical and empirical evidence against the Budyko catchment trajectory conjecture, Hydrology and Earth System Sciences, 26, 1507–1525, https://doi.org/10.5194/hess-26-1507-2022, 2022.

Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., and Franks, S. W.: Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, Water Resources Research, 47, https://doi.org/https://doi.org/10.1029/2011WR010643, 2011.

Ridal, M., Bazile, E., Le Moigne, P., Randriamampianina, R., Schimanke, S., Andrae, U., Berggren, L., Brousseau, P., Dahlgren, P., Edvinsson, L., El-Said, A., Glinton, M., Hagelin, S., Hopsch, S., Isaksson, L., Medeiros, P., Olsson, E., Unden, P., and Wang, Z.: CERRA, the Copernicus European Regional Reanalysis system, Quarterly Journal of the Royal Meteorological Society, 150, 3385–3411, https://doi.org/https://doi.org/10.1002/qj.4764, 2024.

Rouhier, L.: Régionalisation d'un modèle hydrologique distribué pour la modélisation de bassins non jaugés. Application aux vallées de la Loire et de la Durance, Ph.D. thesis, Sorbonne Université, https://theses.hal.science/tel-02409965, 2018.

Safeeq, M., Bart, R. R., Pelak, N. F., Singh, C. K., Dralle, D. N., Hartsough, P., and Wagenbrenner, J. W.: How realistic are water-balance closure assumptions? A demonstration from the southern sierra critical zone observatory and kings river experimental watersheds, Hydrological Processes, 35, e14 199, https://doi.org/https://doi.org/10.1002/hyp.14199, 2021.

Sahoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., and Wood, E. F.: Reconciling the global terrestrial water budget using satellite remote sensing, Remote Sensing of Environment, 115, 1850–1865, https://doi.org/https://doi.org/10.1016/j.rse.2011.03.009, 2011.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, https://doi.org/https://doi.org/10.1029/2008WR007327, 2010.

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, Hydrology and Earth System Sciences, 21, 4323–4346, https://doi.org/10.5194/hess-21-4323-2017, 2017.

Samaniego, L., Thober, S., Wanders, N., Pan, M., Rakovec, O., Sheffield, J., Wood, E., Prudhomme, C., Rees, G., Houghton-Carr, H., Fry, M., Smith, K., Watts, G., Hisdal, H., Estrela, T., Buontempo, C., Marx, A., and Kumar, R.: Hydrological Forecasts and Projections for Improved Decision-Making in the Water Sector in Europe, Bulletin of the American Meteorological Society, 100, 2451–2472, https://doi.org/10.1175/BAMS-D-17-0274.1, 2019.

Sarrazin, F., Hartmann, A., Pianosi, F., Rosolem, R., and Wagener, T.: V2Karst V1.1: a parsimonious large-scale integrated vegetation–recharge model to simulate the impact of climate and land cover change in karst regions, Geoscientific Model Development, 11, 4933–4964, https://doi.org/10.5194/gmd-11-4933-2018, 2018.

Sauquet, E., Evin, G., Siauve, S., Bornancin-Plantier, A., Jacquin, N., Arnaud, P., Berel, M., Bernus, S., Bonneau, J., Branger, F., Caballero, Y., Colleoni, F., Collet, L., Corre, L., Drouin, A., Ducharne, A., Fournier, M., Gailhard, J., Habets, F., Hendrickx, F., Héraut, L., Hingray, B., Huang, P., Jaouen, T., Jeantet, A., Lanini, S., Le Lay, M., Loudin, S., Magand, C., Marson, P., Mimeau, L., Monteil, C., Munier, S., Perrin, C., Robin, Y., Rousset, F., Soubeyroux, J.-M., Strohmenger, L., Thirel, G., Tocquer, F., Tramblay, Y., Vergnes, J.-P., Vidal, J.-P., and Vrac, M.: Messages et enseignements du projet Explore2, Tech. rep., INRAE ; OiEau ; OFB ; MTECT ; BRGM ; CNRS ; IRD ; EDF ; Météo-France, https://doi.org/10.57745/J3XIPW, 2024.

Schaller, M. F. and Fan, Y.: River basins as groundwater exporters and importers: Implications for water cycle and climate modeling, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/https://doi.org/10.1029/2008JD010636, 2009.

SCV: HydroPortail : Données hydrométriques nationales, https://hydro.eaufrance.fr, consulté le 12 août 2026, 2026.

Senent-Aparicio, J., Castellanos-Osorio, G., Segura-Méndez, F., López-Ballesteros, A., Jimeno-Sáez, P., and Pérez-Sánchez, J.: BULL Database – Spanish Basin attributes for Unravelling Learning in Large-sample hydrology, 11, 737, https://doi.org/10.1038/s41597-024-03594-5, 2024.

Seppä, I., Gonzales Inca, C., Uusikivi, J., and Alho, P.: CAMELS-FI: hydrometeorological time series and landscape properties for 320 catchments in Finland, Earth System Science Data Discussions, 2025, 1–33, https://doi.org/10.5194/essd-2025-578, 2025.

Strohmenger, L., Thirel, G., and Sauquet, É.: Chroniques de débits journaliers du réseau de référence hydrologique Explore2, https://doi.org/10.57745/MXGCX5, 2024.

Tan, X., Liu, B., Tan, X., and Chen, X.: Long-term water imbalances of watersheds resulting from biases in hydroclimatic data sets for water budget analyses, Water Resources Research, 58, e2021WR031 209, 2022.

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, https://doi.org/10.1080/02626667.2020.1839080, 2021.

Troy, T. J., Wood, E. F., and Sheffield, J.: An efficient calibration method for continental-scale land surface modeling, Water Resources Research, 44, https://doi.org/https://doi.org/10.1029/2007WR006513, 2008.

Troy, T. J., Sheffield, J., and Wood, E. F.: Estimation of the Terrestrial Water Budget over Northern Eurasia through the Use of Multiple Data Sources, Journal of Climate, 24, 3272 – 3293, https://doi.org/10.1175/2011JCLI3936.1, 2011.

Turner, S., Hannaford, J., Barker, L. J., Suman, G., Killeen, A., Armitage, R., Chan, W., Davies, H., Griffin, A., Kumar, A., Dixon, H., Albuquerque, M. T. D., Almeida Ribeiro, N., Alvarez-Garreton, C., Amoussou, E., Arheimer, B., Asano, Y., Berezowski, T., Bodian, A., Boutaghane, H., Capell, R., Dakhlaoui, H., Daňhelka, J., Do, H. X., Ekkawatpanit, C., El Khalki, E. M., Fleig, A. K., Fonseca, R., Giraldo-Osorio, J. D., Goula, A. B. T., Hanel, M., Horton, S., Kan, C., Kingston, D. G., Laaha, G., Laugesen, R., Lopes, W., Mager, S., Rachdane, M., Markonis, Y., Medeiro, L., Midgley, G., Murphy, C., O'Connor, P., Pedersen, A. I., Pham, H. T., Piniewski, M., Renard, B., Saidi, M. E., Schmocker-Fackel, P., Stahl, K., Thyer, M., Toucher, M., Tramblay, Y., Uusikivi, J., Venegas-Cordero, N., Visessri, S., Watson, A., Westra, S., and Whitfield, P. H.: ROBIN: Reference observatory of basins for international hydrological climate change detection, 12, 654, https://doi.org/10.1038/s41597-025-04907-y, 2025.

Venegas-Cordero, N., Kundzewicz, Z., Jamro, S., and Piniewski, M.: Detection of trends in observed river floods in Poland, Journal of Hydrology: Regional Studies, 41, 101 098, https://doi.org/https://doi.org/10.1016/j.ejrh.2022.101098, 2022.

Verrelle, A., Glinton, M., Bazile, E., Le Moigne, P., Randriamampianina, R., Ridal, M., Berggren, L., Undén, P., Schimanke, S., Mladek, R., and Soci, C.: CERRA-Land sub-daily regional reanalysis data for Europe from 1984 to present, https://doi.org/10.24381/cds.a7f3cd0b, accessed: 04-08-2026, 2022.

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.

Vogt, J., Soille, P., De Jager, A., Rimaviciute, E., Mehl, W., Foisneau, S., Bodis, K., Dusart, J., Paracchini, M.-L., Haastrup, P., et al.: A pan-European river and catchment database, Report EUR, 22920, 120, 2007.

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, Water Resources Research, 44, https://doi.org/https://doi.org/10.1029/2007WR006720, 2008.

Wagener, T., Reinecke, R., and Pianosi, F.: On the evaluation of climate change impact models, WIREs Climate Change, 13, e772, https://doi.org/https://doi.org/10.1002/wcc.772, 2022.

Wallén, A.: L'Eau Tombée Dans La Haute Montagne De La Suède, Geografiska Annaler, 5, 72–104, https://doi.org/10.1080/20014422.1923.11881070, 1923.

Wallén, A.: Le débit des fleuves suédois et le rapport de ce débit avec l'eau tombée, Geografiska Annaler, 6, 42–55, http://www.jstor.org/stable/519606, 1924.

Wang, S., McKenney, D. W., Shang, J., and Li, J.: A national-scale assessment of long-term water budget closures for Canada's watersheds, Journal of Geophysical Research: Atmospheres, 119, 8712–8725, 2014.

Wang, X., Du, P., and Shen, J.: Smoothing splines with varying smoothing parameter, Biometrika, 100, 955–970, https://doi.org/10.1093/biomet/ast031, 2013.

Wood, R. R., Janzing, J., van Hamel, A., Götte, J., Schumacher, D. L., and Brunner, M. I.: Comparison of high-resolution climate reanalysis datasets for hydro-climatic impact studies, EGUsphere, 2024, 1–41, https://doi.org/10.5194/egusphere-2024-2905, 2024.

Wood, S. N.: Thin Plate Regression Splines, Journal of the Royal Statistical Society Series B: Statistical Methodology, 65, 95–114, https://doi.org/10.1111/1467-9868.00374, 2003.

WSL: Beurteilung der Beeinflussung des Abflusses an NAWA-Messstellen: Detailbericht des BAFU-Projekts HydCHeck, Technical report, Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft (WSL), Birmensdorf, Switzerland, https://www.bafu.admin.ch/dam/it/sd-web/0Jky1IPKkB9C/beurteilung-der-beeinflussung-des-abflusses-an-nawa-messstellen.pdf, im Auftrag des Bundesamtes für Umwelt (BAFU), 2024.

Xiong, L. and Guo, S.: Appraisal of Budyko formula in calculating long-term water balance in humid watersheds of southern China, Hydrological Processes, 26, 1370–1378, https://doi.org/https://doi.org/10.1002/hyp.8273, 2012.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J., Sampson, C., Kanae, S., and Bates, P.: A high-accuracy map of global terrain elevations, Geophysical Research Letters, 44, 5844–5853, https://doi.org/https://doi.org/10.1002/2017GL072874, 2017.

https://doi.org/10.5194/egusphere-2026-5491
Preprint. Discussion started: 17 September 2026

Yang, X., Peralta, C., Amstrup, B., Hintz, K. S., Thorsen, S. B., Denby, L., Kamuk Christiansen, S., Schulz, H., Pelt, S., and Schreiner, M.: DANRA: the kilometer-scale Danish regional atmospheric reanalysis, Earth System Science Data, 18, 2251–2264, https://doi.org/10.5194/essd-18-2251-2026, 2026.

Zhang, L., Dawes, W., and Walker, G.: Response of mean annual evapotranspiration to vegetation changes at catchment scale, Water Resources Research, 37, 701–708, https://doi.org/https://doi.org/10.1029/2000WR900325, 2001.

Zhang, S., Yang, H., Yang, D., and Jayawardena, A. W.: Quantifying the effect of vegetation change on the regional water balance within the Budyko framework, Geophysical Research Letters, 43, 1140–1148, https://doi.org/https://doi.org/10.1002/2015GL066952, 2016.

Zhang, Y., Pan, M., Sheffield, J., Siemann, A. L., Fisher, C. K., Liang, M., Beck, H. E., Wanders, N., MacCracken, R. F., Houser, P. R., Zhou, T., Lettenmaier, D. P., Pinker, R. T., Bytheway, J., Kummerow, C. D., and Wood, E. F.: A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010, Hydrology and Earth System Sciences, 22, 241–263, https://doi.org/10.5194/hess-22-241-2018, 2018.

Zheng, X., Liu, D., Huang, S., Wang, H., and Meng, X.: Achieving water budget closure through physical hydrological process modelling: insights from a large-sample study, Hydrology and Earth System Sciences, 29, 627–653, https://doi.org/10.5194/hess-29-627-2025, 2025.

Zhuravin, S. A.: Features of Water Balance for Small Mountainous Watersheds in East Siberia: Kolyma Water Balance Station Case Study, in: IAHS Publications : Northern Research Basins Water Balance, vol. 290 of *Proceedings of a workshop held at Victoria, Canada, March 2004*, pp. 28–40, International Association of Hydrological Science Press, 2004.