https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

Multiple-objective calibration of a conceptual hydrological model using satellite data of snow cover, soil moisture and limited streamflow observations

Asma Khalil[1], Rui Tong[2], Zahra Majid[1], Borbála Széles[1], Miriam Bertola[1], Juraj Parajka[1]

[1]Institute of Hydraulic Engineering and Water Resources Management, Faculty of Civil and Environmental Engineering, Vienna University of Technology, Vienna, 1040, Austria
[2]State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing, 100084, China

*Correspondence to: Asma Khalil (asmakhalil201@gmail.com)

Abstract. One way to improve hydrological predictions in data-sparse regions is to assimilate satellite data of water cycle components into the calibration of hydrological models. This study evaluates the value of combining satellite snow cover (MODIS) and soil moisture (ASCAT) data with limited streamflow observations to improve hydrological model calibration at the regional scale. The study compares model performance of eleven calibration variants that differ in (a) whether they use satellite data only, (b) the type and temporal distribution of streamflow observations used, and (c) whether satellite and streamflow data are combined. The streamflow sampling strategies cover two scenarios: regularly spaced observations distributed over multiple years (one value per season or per month), and event-based strategies that mimic a single short-term gauging campaign in the wettest, driest, or average year (a peak-flow event with recession plus six bimonthly background samples). The analysis is performed for 213 catchments in Austria, grouped into 119 alpine and 94 lowland catchments. The results show that calibration to satellite data only provides reliable runoff simulations primarily in lower-elevation, drier, and more agricultural lowland catchments. In alpine catchments, adding any limited streamflow data substantially improves model efficiency. The combination of monthly streamflow observations and satellite data (Vmo+sat) results in the best overall runoff performance in lowland catchments, with a median validation runoff efficiency of 0.67. In alpine catchments, event-based streamflow-only strategies achieve median validation runoff efficiencies of 0.69–0.71, close to the regular monthly variant (Vmo, median 0.75) and substantially better than satellite-only calibration (Vsat, 0.26). In lowland catchments, Vmo+sat (0.67) outperforms all event-based variants. Adding satellite data to any streamflow-based variant reduces median snow-cover errors in alpine catchments by approximately a factor of five and consistently improves the simulated soil moisture, although it can reduce runoff efficiency in alpine catchments compared to streamflow-only calibration. These results support the practical value of short, targeted gauging campaigns combined with satellite remote sensing for hydrological modeling in data-sparse regions.

Keywords: Multiple-objective calibration, MODIS snow cover, ASCAT soil moisture, limited streamflow data, event-based sampling, ungauged catchments

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

1 Introduction

Estimating streamflow in ungauged or poorly gauged catchments remains a central challenge in operational hydrology. Most locations where streamflow estimates are needed are either ungauged or only poorly gauged (Hrachowitz et al., 2013). Rainfall-runoff models are widely used to address this challenge, but their performance depends critically on reliable model parameter estimation. Two main groups of modeling approaches have emerged in the past: (1) transfer of hydrological model parameters (i.e., parameter regionalization) from gauged (donor) to ungauged catchments (Blöschl et al., 2013; Parajka et al., 2013); or (2) calibration of hydrologic models using alternative data sources that do not require continuous discharge records.

Recent advances in satellite remote sensing have expanded the suite of data available for constraining hydrological models. Snow cover (such as from MODIS, Hall and Riggs, 2016), soil moisture from ASCAT (Wagner et al., 1999; Brocca et al., 2017) or Sentinel, evapotranspiration, terrestrial water storage anomalies, and other variables have all been used to improve internal model consistency and reduce parameter uncertainty (Lettenmaier et al., 2015; Wagner et al., 2025). A recent comprehensive review by Wagner et al. (2025) concluded that while alternative satellite datasets improve the internal consistency of model states, they often result in only a modest improvement, or even a slight decrease, in runoff simulation performance compared to traditional calibration to runoff alone. This underlines a persistent limitation of remote sensing products, which cannot fully substitute for direct runoff observations in constraining hydrological models.

An alternative, or complementary, strategy is to collect a small number of direct streamflow measurements in the catchment of interest. The PUB (Predictions in Ungauged Basins) initiative identified the value of limited discharge data as a key research priority (Sivapalan et al., 2003). Several subsequent studies have demonstrated that even a few strategically collected measurements can substantially constrain model parameters and reduce simulation uncertainty. Seibert and Beven (2009) showed that approximately sixteen randomly selected measurements within a hydrological year could already yield an acceptable model calibration. Seibert and McDonnell (2015) and Singh and Bárdossy (2012) demonstrated that measurements taken during hydrologically important events, such as peak flows, recessions, or unusual flow conditions, carry substantially more information than measurements taken at fixed intervals or during low-flow periods. Pool et al. (2017) systematically evaluated thirteen sampling strategies using twelve measurements per year and found that strategies capturing high-flow conditions and recessions provided the best runoff simulation performance.

A key practical question is how such measurements would realistically be collected. In practice, initial gauging campaigns in a previously ungauged catchment are typically short and often limited to a single season or a year due to financial and logistical constraints (Melsen et al., 2014; Pool et al., 2019). The most promising scenario is therefore a targeted short-term campaign designed to capture one or a few high-flow events, as these carry the most information for model calibration (Seibert and Beven, 2009; Singh and Bárdossy, 2012; Seibert and McDonnell, 2015). Scheller et al. (2024) showed that even temporary stream observations from a single campaign can be useful for calibrating lumped hydrological models, particularly when

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

combined with additional information. Viviroli and Seibert (2015) also demonstrated that a limited number of runoff measurements can improve upon regionalization performance when used to constrain model parameters transferred from gauged catchments.

A second practically motivated scenario involves the potential use of satellite-based discharge estimates for calibrating hydrologic models. The recent SWOT satellite mission provides global water surface elevation observations with a revisit time of approximately 21 days, enabling discharge estimation at this frequency (Biancamaria et al., 2016). Similarly, citizen science platforms can provide intermittent water-level observations that may approximate monthly or seasonal frequencies (Etter et al., 2020). These contexts motivate evaluating regularly spaced but infrequent streamflow observations, such as one per month or one per season, as a calibration strategy, since they mimic what might realistically be available from these sources over longer periods. Such regular strategies are also often used as a benchmark for evaluating other irregularly spaced strategies (Pool et al., 2017).

Most previous studies have evaluated limited streamflow data in isolation, without combining it with remote sensing data. Tong et al. (2021) showed that using a multi-objective calibration to MODIS snow cover and ASCAT soil moisture provides an alternative to regionalization approaches for predicting runoff in ungauged sites in Austria. Tong et al. (2022) further demonstrated the value of this approach for parameter transfer to ungauged sites. However, the question of how best to complement satellite data with limited direct streamflow observations, and what form these observations should take, remains open.

This study addresses this gap by evaluating eleven calibration variants that differ systematically in the type and temporal distribution of streamflow data (regularly spaced monthly or seasonal observations, and event-based observations from a single high-flow campaign in the wettest, driest, or average year), and whether or not these are combined with satellite data. The specific objectives are: (1) to cluster the regional differences in model performance when no streamflow data are used in calibration; (2) to assess and compare model performance across variants using regularly spaced limited observations; and (3) to evaluate whether event-based streamflow strategies inspired by short-term gauging campaigns can match or approach the performance of regularly spaced observations, particularly when combined with satellite data. The methodology is tested across 213 catchments in Austria, covering a wide range of alpine and lowland hydrological regimes. This study directly follows Tong et al. (2021), who used continuous streamflow records; in contrast, this work evaluates what can be achieved with a realistic minimum of direct discharge information.

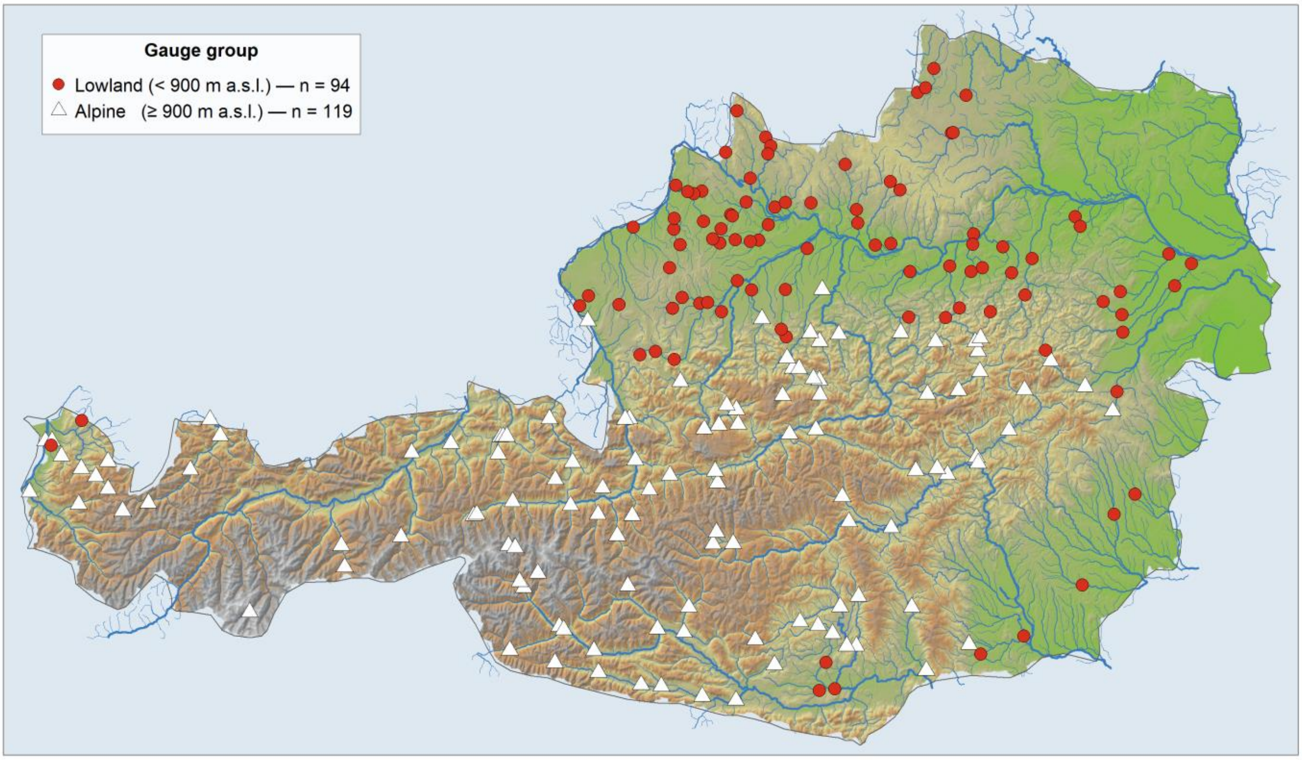
2 Data

2.1 Study Region

The study region is Austria. Austria is characterized by lowlands and hilly regions in the eastern and northern parts, and alpine topography in the western and central parts (Fig. 1). Elevations range from 115 m a.s.l. to almost 3800 m a.s.l. Runoff generation in the alpine region is strongly affected by snowmelt, while the lowlands, particularly in the eastern part, are the

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

driest part of Austria where evapotranspiration plays an important role. The climate is classified as temperate humid, with mean annual precipitation of less than 400 mm yr^{-1} in the lowlands and more than 2500 mm yr^{-1} in the central and western Alps. Land use is predominantly agricultural in the lowlands and forested in mid-elevation areas.



n = 213 gauges | 119 alpine (△) | 94 lowland (●)

Figure 1: Topography of Austria and locations of 213 gauges with streamflow observations. Red and white symbols show two groups of catchments: lowland (red circles, mean catchment elevation below 900 m a.s.l.) and alpine (white triangles, above 900 m a.s.l.), respectively.

2.2 Hydrological Data

The analysis is performed in 213 catchments (Fig. 1). This dataset has been used in numerous previous modeling studies in Austria (please see e.g., Viglione et al., 2013; Sleziak et al., 2020; Tong et al., 2021, 2022). Catchment size ranges from 13.7 to 6214 km², and mean catchment elevation ranges from 353 to 2940 m a.s.l. We split the dataset into two groups: (1) lowland and hilly catchments (94 catchments, mean elevation below 900 m a.s.l.); and (2) alpine catchments (119 catchments with mean elevation above 900m a.s.l.). The basic hydrological characteristics of these two groups are presented in Table 1.

Table 1: Physiographic attributes of the two groups of analysed catchments. The attributes represent the median over 119 alpine and 94 lowland catchments, respectively. Mean annual values cover the period 2000–2014.

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

Attributes	Alpine	Lowland
Mean annual air temperature (°C)	5.8	9.0
Mean annual precipitation (mm yr^{-1})	1477.0	999.5
Mean annual potential evaporation (mm yr^{-1})	563.0	690.1
Mean catchment aridity index (–)	0.4	0.7
Mean catchment elevation (m a.s.l.)	1391	636
Catchment size (km^2)	165	169

Hydrological modelling uses mean daily values of precipitation, air temperature, and potential evaporation as inputs, derived for 200-m elevation zones from the SPARTACUS database (Hiebl and Frei, 2016, 2018). Mean daily potential evaporation is calculated from air temperature and solar radiation index using a modified Blaney–Criddle approach (Parajka et al., 2003). Daily streamflow records are provided by the Hydrographic Service of Austria (https://ehyd.gv.at, last access: 31 December 2025), available for 1 September 1999 to 31 August 2014.

2.3 Remote Sensing Data

2.3.1 MODIS snow cover

Snow cover data are derived from daily 500-m Terra (MOD10A1) and Aqua (MYD10A1) MODIS (version 6) products (Hall and Riggs, 2016). We applied seasonally varying NDSI thresholds that vary with altitude and land cover (Tong et al., 2020). To reduce cloud impacts, Terra and Aqua products are combined following Parajka and Blöschl (2008). More details about the dataset are presented in Tong et al. (2021).

2.3.2 ASCAT soil moisture

Satellite soil moisture estimates are extracted from the ASCAT-DIREX soil moisture product (Tong et al., 2021), which combines ASCAT and Sentinel-1 observations to provide daily gridded maps of the soil water index (SWI) at 500-m resolution. ASCAT-DIREX data are available from 2007 onwards, and details on processing and accuracy are provided in Tong et al. (2021) and Sleziak et al. (2024).

3 Methods

3.1 Clustering of hydrological model performance

Catchments with similar model performance are grouped using the k-means clustering method (MacQueen, 1967). The basic idea of k-means clustering is to group the catchments into a pre-specified number of clusters so that the total intra-cluster

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

variation of model performance is minimized. The model performance is defined by the seasonal difference between simulated and observed streamflow in the validation period. The simulated streamflow is calculated using the calibration variant (termed Vsat here), which represents multiple-objective calibration to satellite snow cover (MODIS) and soil moisture (ASCAT) without using any streamflow data (i.e., runoff weight wQ = 0; see Sect. 3.4 and Table 3 for a full description of this and all other calibration variants). This variant corresponds to var3 in Tong et al. (2021). The input to the k-means clustering is the mean monthly difference between simulated and observed streamflow (in mm month^{-1}) calculated in each catchment, yielding 12 values per catchment.

The optimal number of clusters is selected using the Elbow method. The Elbow method estimates the total within-cluster sum of squares (WSS) of inputs. With an increasing number of clusters, WSS decreases due to improved cluster compactness; however, beyond a certain number of clusters, the rate of improvement decreases. The optimal number of clusters is identified at the point where the rate of decrease in WSS shows a marked change, forming an "elbow" in the plot of WSS versus the number of clusters. The k-means clustering and the Elbow method are calculated within R software using the k-means base function (R Core Team, 2024). The number of clusters evaluated by the Elbow method ranges from 2 to 10.

3.2 Conceptual hydrological model

The hydrological model used is the TUWmodel (Viglione and Parajka, 2020), a conceptual model that follows the HBV structure (Jansen et al., 2021). Inputs are daily precipitation, air temperature, and potential evaporation, applied for 200-m elevation zones. The model has three subroutines (snow, soil moisture, and runoff generation) and 15 parameters (Table 2). All variants tested in this study are calibrated using DEoptim in R (Mullen et al., 2011). The model is calibrated from September 2000 to August 2010 and validated from September 2010 to August 2014, each with a one-year warm-up period. Since ASCAT is available only from January 2007, the soil moisture component of the calibration objective is computed over the period 2007–2010.

Table 2: TUWmodel parameters and calibration ranges.

Sub-routine	Parameter	Description	Range
Snow	SCF	Snow correction factor [–]	0.9–1.5
	DDF	Degree-day melt factor [mm °C^{-1} d^{-1}]	0–5
	Tr	Rain/snow threshold temperature [°C]	1–3
	Ts	Snow threshold temperature [°C]	–3–1
	Tm	Snowmelt onset temperature [°C]	–2–2
Soil moisture	LP	Evapotranspiration limit [–]	0–1
	FC	Field capacity [mm]	0–600
	Beta	Soil moisture shape coefficient [–]	0–20

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

Runoff generation	K0	Surface storage coefficient [days]	0–2
	K1	Sub-surface storage coefficient [days]	2–30
	K2	Baseflow storage coefficient [days]	30–250
	Lsuz	Fast surface runoff threshold [mm]	1–100
	Cperc	Percolation rate [mm d^{-1}]	0–8
	Bmax	Maximum base parameter [day]	0–30
	Croute	Free scaling parameter [day^2 mm^{-1}]	0–50

3.3 Calibration objective functions

The general form of the calibration objective function is:

$$\mathrm{FB} = \mathrm{wQ} \cdot (1 - \mathrm{OQ}) + \mathrm{wSC} \cdot \mathrm{OSC} + \mathrm{wSM} \cdot (1 - \mathrm{OSM}) \quad (1)$$

where OQ, OSC, and OSM are individual objectives related to streamflow, snow cover, and soil moisture, weighted by wQ, wSC, and wSM, respectively. The streamflow efficiency OQ combines the Nash–Sutcliffe efficiency (NSE) and its log-transformed variant (NSElog):

$$\mathrm{OQ} = 0.5 \cdot \mathrm{NSE} + 0.5 \cdot \mathrm{NSElog} \quad (2)$$

NSE emphasizes high-flow performance; NSElog gives more weight to low and medium flows. Combining both prevents calibration from being biased toward either extreme of the flow regime, which is particularly important when only a few observations are available.

The snow cover error OSC counts the fraction of days with poor agreement between simulated snow water equivalent (SWE) and MODIS snow cover area (SCA), excluding days with mean catchment cloud cover exceeding 60%. A poor simulation day is defined when (a) simulated SWE exceeds 1 mm but MODIS SCA is zero, or (b) simulated SWE is zero but MODIS SCA exceeds 5%.

The soil moisture efficiency OSM is the Pearson correlation coefficient between daily ASCAT SWI and model-simulated relative root-zone soil moisture.

3.4 Calibration variants and design rationale

This study evaluates eleven calibration variants (Table 3), motivated by two distinct practical scenarios.

Scenario 1: Regularly spaced infrequent observations (Vmo, Vse, Vmo+sat, Vse+sat). These variants use one streamflow observation per month (Vmo, Vmo+sat) or per season (Vse, Vse+sat) throughout the entire calibration period (2000–2010). This scenario is motivated by two contexts: (i) emerging satellite-based discharge estimation products, such as those derived

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

from the SWOT mission (revisit time ~21 days; Biancamaria et al., 2016) or citizen science streamflow observations (Etter et al., 2020), which could provide discharge estimates at approximately monthly frequency over multi-year periods; and (ii) the need for reference baselines consistent with earlier literature (e.g., Pool et al., 2017; Seibert and Beven, 2009), to facilitate comparison.

Scenario 2: Event-based short-term gauging campaigns (Vwet, Vdry, Vavg, Vwet+sat, Vdry+sat, Vavg+sat). These variants are inspired by the CMax_Rec_Dom strategy from Pool et al. (2017), which was shown to be among the most informative for hydrograph simulation. Each event-based variant uses twelve observations from a single hydrological year: the highest peak flow with its first five subsequent recession days, plus six observations at the 15th day of every other month. Three variants differ in which year is selected for the campaign: the wettest year (Vwet, Vwet+sat), the driest year (Vdry, Vdry+sat), or the year closest to the long-term mean annual precipitation (Vavg, Vavg+sat). These three variants bound the realistic range of outcomes and explore sensitivity of calibration performance to hydrological conditions during the campaign. The +sat variants test whether satellite constraints compensate for the limited duration of streamflow data.

The baseline variant (Vsat) uses only satellite snow cover and soil moisture data (wQ = 0), corresponding to var3 in Tong et al. (2021).

Table 3: Description of calibration variants. Seasons are defined as: winter (Dec–Feb), spring (Mar–May), summer (Jun–Aug), autumn (Sep–Nov). The event-based strategy follows Pool et al. (2017): highest peak in the selected year, plus five recession days and six bimonthly background samples.

Variant	Description	wQ	wSC	wSM
Vsat	Calibration to OSC and OSM only. No streamflow.	0.0	0.5	0.5
Vmo	Calibration to OQ only. Streamflow: first day of each month.	1.0	0.0	0.0
Vse	Calibration to OQ only. Streamflow: Jan 1, Apr 1, Jul 1, Oct 1 each year.	1.0	0.0	0.0
Vwet	Calibration to OQ only. Event strategy from wettest year.	1.0	0.0	0.0
Vdry	Calibration to OQ only. Event strategy from driest year.	1.0	0.0	0.0
Vavg	Calibration to OQ only. Event strategy from average year.	1.0	0.0	0.0
Vmo+sat	Combined calibration. Streamflow: first day of each month.	0.33	0.33	0.33
Vse+sat	Combined calibration. Streamflow: once per season.	0.33	0.33	0.33
Vwet+sat	Combined calibration. Event strategy from wettest year.	0.33	0.33	0.33
Vdry+sat	Combined calibration. Event strategy from driest year.	0.33	0.33	0.33
Vavg+sat	Combined calibration. Event strategy from average year.	0.33	0.33	0.33

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

4 Results

4.1 Clustering of model performance of calibration to satellite data only

The previous study of Tong et al (2021) showed that hydrologic model simulations based on calibrations to satellite data only (Vsat) can provide an alternative to regionalization approaches for predicting runoff hydrographs in ungauged sites. This study therefore builds upon the previous analysis and evaluates factors that control the model performance of simulations based on calibrations to MODIS snow cover and ASCAT soil moisture datasets. Figure 2 shows grouping of runoff gauges based on the difference between simulated and observed mean monthly streamflow into four clusters. The identified four clusters represent the optimal number of clusters according to the Elbow method. The left panel shows the mean monthly difference between simulated (estimated by Vsat variant) and observed runoff for the cluster centres in the validation period 2010-2014. The right panel presents the spatial arrangement of identified clusters within Austria. The results indicate that four identified clusters represent catchments with a distinct difference between model streamflow simulations and observations.

Cluster 1, the largest group (63% of gauges), contains catchments where simulations agree well with observations throughout the year. The mean monthly streamflow difference ranges from approximately –20 mm month^{-1} in May to +80 mm month^{-1} in September. Cluster 2 (20% of gauges) shows systematic overestimation of streamflow between June and September, when the mean monthly difference can exceed 200 mm month^{-1}. Clusters 3 and 4 show the opposite: systematic underestimation between May and August, with the largest differences in Cluster 4 (five stations only).

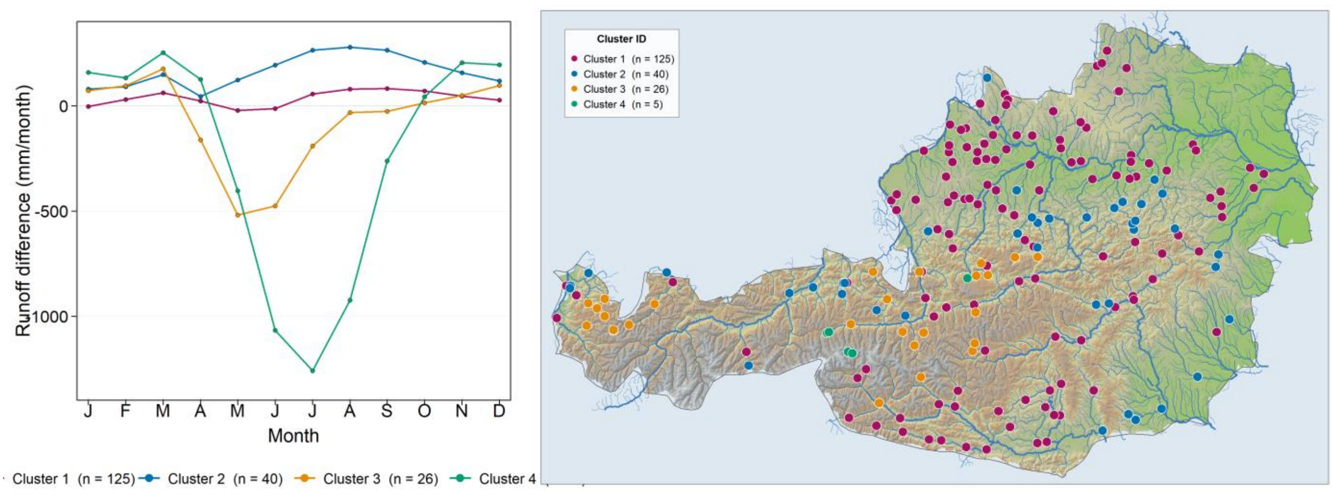


Figure 2: Clusters of model performance for calibration to satellite snow cover and soil moisture (Vsat). The left panel shows the mean monthly difference between simulated and observed runoff for the centres of four clusters in the validation period 2010–2014. The right panel shows the spatial distribution of the 213 gauges grouped into the four clusters.

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

The identified clusters provide not only four distinct groups of gauges in terms of mean monthly streamflow difference, but the catchments in individual clusters have very distinct physiographic characteristics. Figure 3 shows the distribution of selected physiographic characteristics of catchments grouped within four clusters. It is clear that the best model simulations, grouped in cluster 1, are for catchments with larger size, lower mean catchment elevation and mean annual precipitation, higher aridity, and a larger proportion of arable land. Clusters 2 to 4 represent gauges with decreasing model performance. Fig. 3 indicates that the decreasing model performance in clusters 2 to 4 is associated with increasing mean catchment elevation and mean annual precipitation, as well as a decreasing proportion of arable land, decreasing aridity, and catchment size. Interestingly, the best performance in cluster 1 is associated with catchments that have a larger number of stations used for interpolating model inputs (i.e., the precipitation SPARTACUS dataset). In contrast, model inputs for all catchments in cluster 4 are based on interpolations from stations that are not located within these catchments.

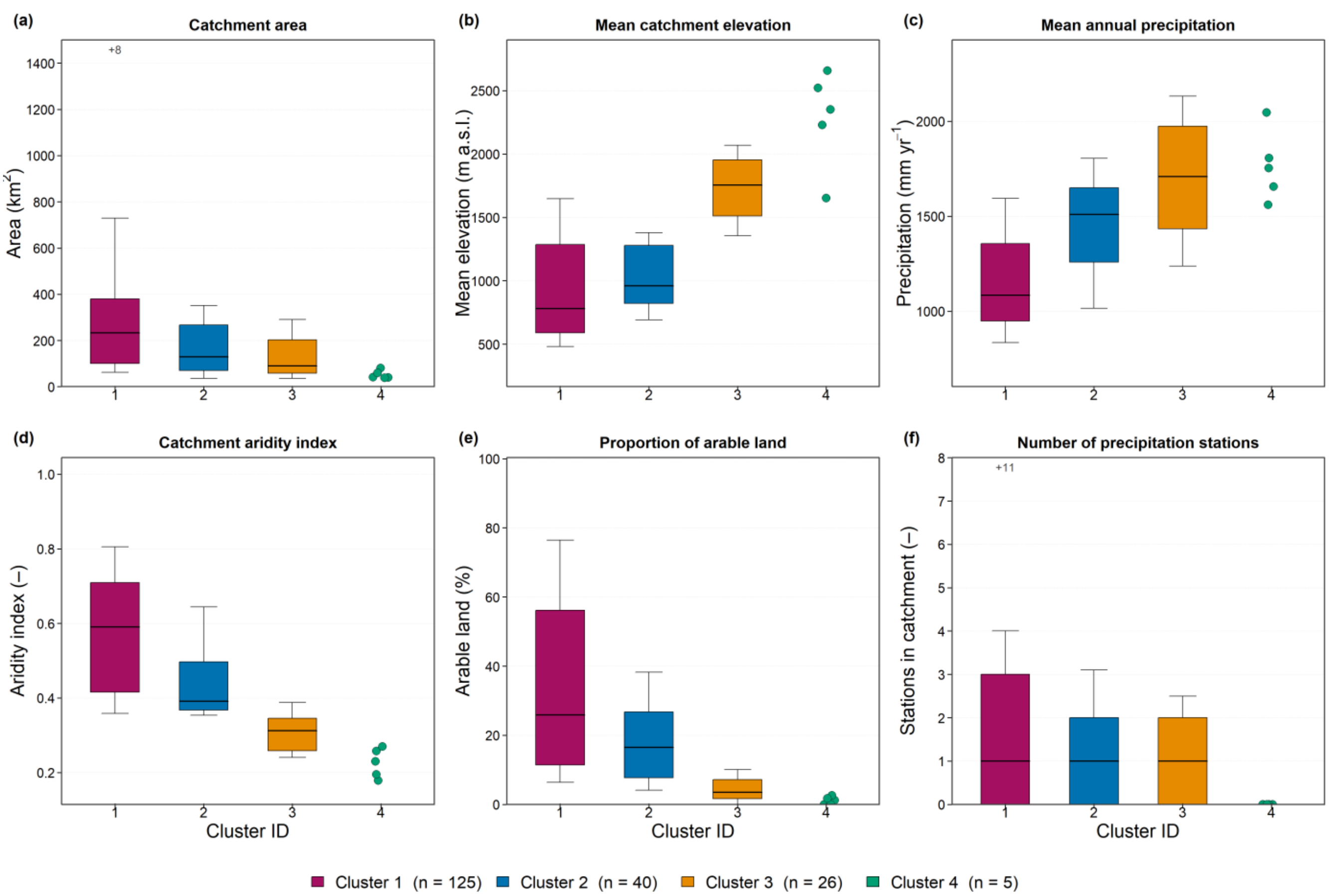


Figure 3: Distribution of selected physiographic attributes of catchments grouped into four clusters (Fig. 2). Panels: (a) catchment area, (b) mean catchment elevation, (c) mean annual precipitation, (d) catchment aridity index, (e) proportion of arable land, and (f) number of precipitation stations located within each catchment. Boxes show the 25th–75th percentile range with the horizontal line indicating the median; whiskers extend to the 10th and 90th percentiles. Cluster 4 (n = 5) is shown as individual data points

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

rather than as a box. The y-axes of panels (a) and (f) are clipped at 1500 km² and 10 stations, respectively; labels "+N" above each cluster's column indicate the number of catchments with values exceeding the cap.

Overall, the clustering of model performance in Austria shows clear spatio-temporal patterns. While in lowland region the simulations are seasonally balanced, in alpine regions, the model simulations based on calibration to satellite data only tend to significantly underestimate river flows in the summer season. The following assessment evaluates whether and to what extent can limited streamflow observations improve multiple-objective calibration that use satellite data of snow cover and soil moisture.

4.2 Hydrological model performance using limited streamflow data in model calibration

4.2.1 Hydrological model performance in the calibration period

The model performance of all eleven calibration variants in the calibration period (September 2000 – August 2010) is summarized in Table 4. The runoff efficiency OQ is shown in Figure 4, the snow cover error OSC in Figure 5, and the soil moisture correlation OSM in Figure 6. Each figure shows the alpine catchments (n = 119) in the top panel and the lowland catchments (n = 94) in the bottom panel.

The runoff efficiency results (Fig. 4) show that adding any streamflow observations to the calibration substantially improves runoff simulations compared to using satellite data alone. In alpine catchments, the median OQ increases from 0.30 for Vsat to between 0.65 and 0.78 for all other variants. In lowland catchments, the improvement is even larger in relative terms, with the median OQ increasing from 0.15 for Vsat to between 0.59 and 0.69. The largest increase is observed for the variants that use one streamflow value per month (Vmo and Vmo+sat). Among variants that use streamflow only, Vmo variant reaches the highest median OQ in both alpine (0.78) and lowland (0.68) catchments. Variants that use streamflow only at one value per season (Vse) reach somewhat lower medians (0.72 alpine, 0.65 lowland). The event-based variants from a single year (Vwet, Vdry, Vavg) reach median OQ values of 0.74–0.75 in alpine catchments. This is only slightly below Vmo and very close to Vse, even though they use only twelve observations from one year. In lowland catchments, the event-based variants perform clearly less well, with medians of 0.59 to 0.65.

Combining the limited streamflow data with satellite snow cover and soil moisture data has different effects in alpine and lowland catchments. In alpine catchments, the “+sat” variants achieve median runoff efficiencies that are lower than their streamflow-only counterparts (e.g. Vmo+sat 0.73 vs. Vmo 0.78; Vwet+sat 0.65 vs. Vwet 0.74). This reflects a trade-off in multi-objective calibration: matching the satellite snow and soil moisture data pulls the parameters away from the runoff optimum. In lowland catchments, in contrast, the “+sat” variants perform either similarly to or better than the streamflow-only variants (e.g. Vmo+sat 0.69 vs. Vmo 0.68; Vavg+sat 0.64 vs. Vavg 0.59).

The snow cover results (Fig. 5) show the opposite pattern. In alpine catchments, the streamflow-only variants produce median snow cover errors of 0.22 to 0.34. These are much larger than the value of 0.06 obtained for Vsat. This means that calibrating only to streamflow allows the model to simulate runoff well, but the simulated snow dynamics differ from the MODIS snow

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

cover. When the same streamflow data are combined with the satellite data (the +sat variants), the median snow cover errors drop back to 0.05–0.08. This is the same range as Vsat and reflects the strong constraining role of MODIS snow cover in the calibration objective. In lowland catchments, snow cover errors are much smaller for all variants (medians 0.03 to 0.07), as expected because lowland catchments have less snow.

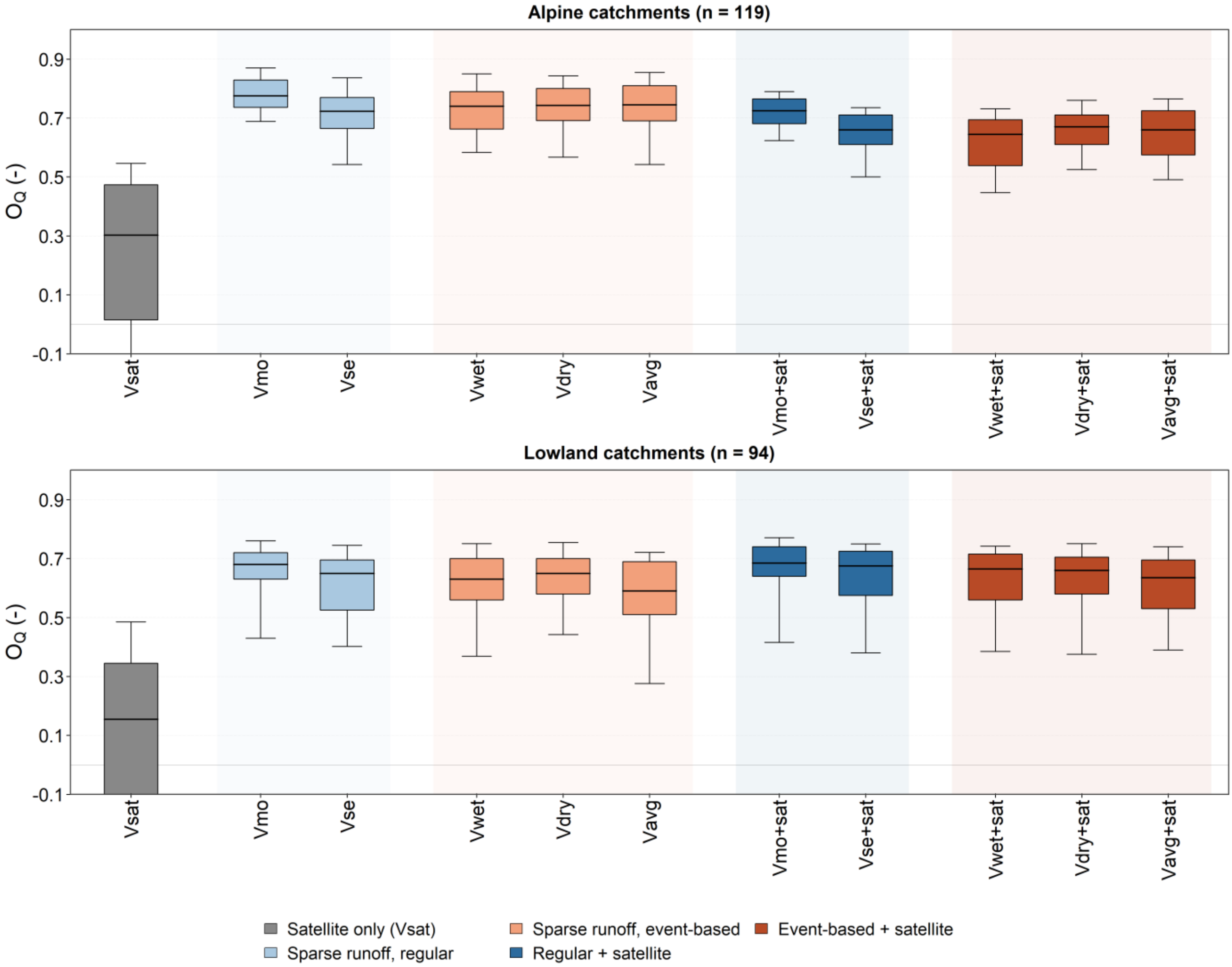
The soil moisture results (Fig. 6) show a similar pattern. In alpine catchments, the median soil moisture correlation OSM decreases from 0.45 for Vsat to 0.23–0.25 for all streamflow-only variants. The "+sat" variants recover most of this loss, with medians of 0.42–0.43. In lowland catchments, the soil moisture correlation is high for all variants. Vsat reaches a median of 0.77, the +sat variants 0.74–0.75, and the streamflow-only variants 0.64–0.67.

Overall, the calibration results show that limited streamflow observations alone are enough to obtain good runoff simulations, but at the cost of poorer snow and soil moisture simulations. Combining the limited streamflow data with satellite data preserves the snow and soil moisture performance while still strongly improving runoff simulations compared to Vsat. The choice of variant therefore depends on whether the user needs only good runoff simulations or also internally consistent snow and soil moisture simulations.

Table 4: Median model performance of eleven calibration variants (defined in Table 3) in the calibration period 2000–2010. Medians are estimated separately for 119 alpine and 94 lowland catchments. The runoff (OQ), snow cover (OSC) and soil moisture (OSM) efficiencies are defined in section 3.3.

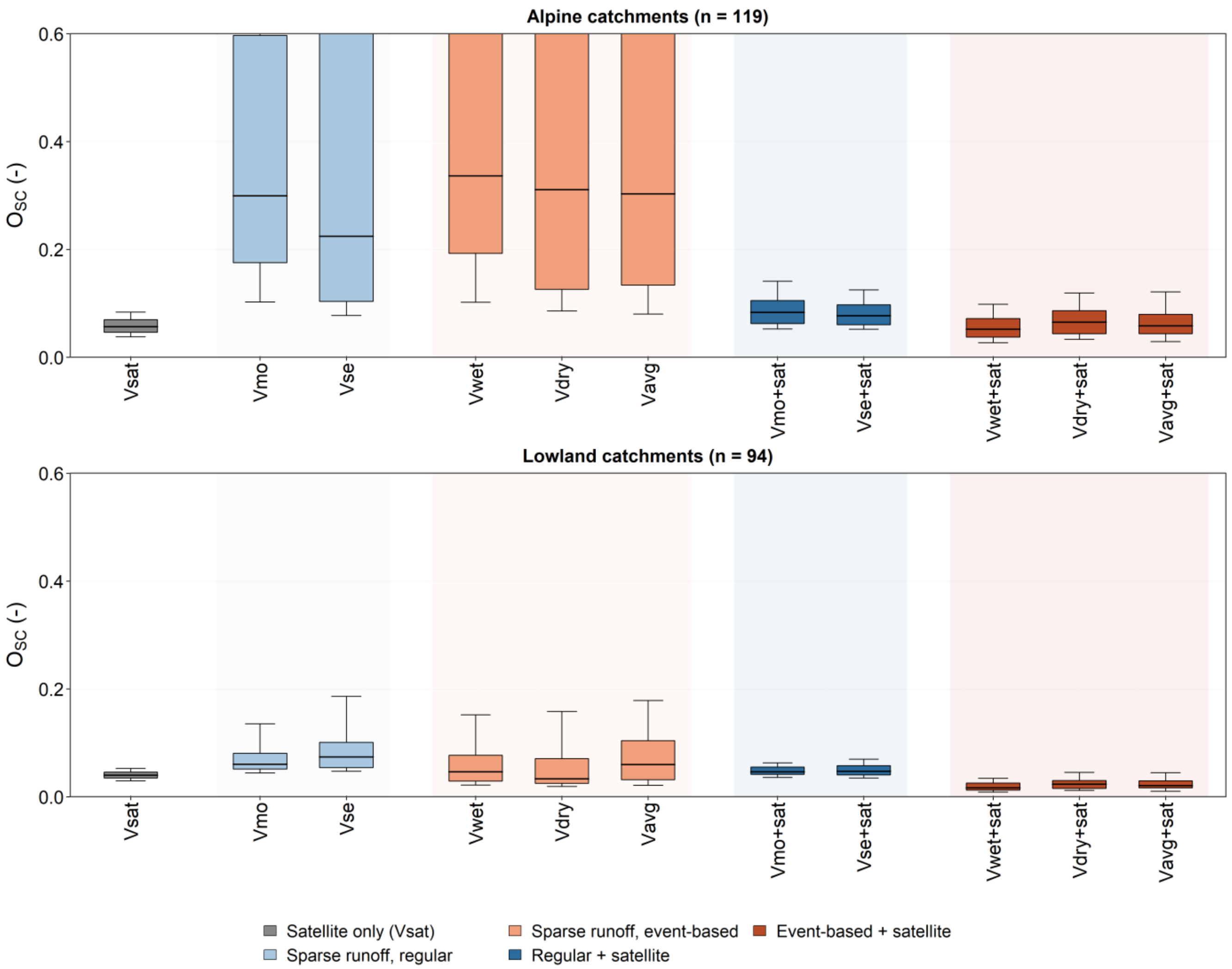
Variant	Alpine OQ	Alpine OSC	Alpine OSM	Lowland OQ	Lowland OSC	Lowland OSM
Vsat	0.30	0.06	0.45	0.15	0.04	0.77
Vmo+sat	0.72	0.08	0.42	0.69	0.05	0.74
Vse+sat	0.66	0.08	0.42	0.68	0.05	0.74
Vwet+sat	0.65	0.05	0.43	0.67	0.02	0.75
Vdry+sat	0.67	0.07	0.43	0.66	0.02	0.74
Vavg+sat	0.66	0.06	0.43	0.64	0.02	0.74
Vmo	0.78	0.30	0.24	0.68	0.06	0.65
Vse	0.72	0.22	0.24	0.65	0.07	0.66
Vwet	0.74	0.34	0.23	0.63	0.05	0.66
Vdry	0.74	0.31	0.25	0.65	0.03	0.64
Vavg	0.74	0.30	0.23	0.59	0.06	0.67

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026



290 **Figure 4. Runoff model efficiency O_Q of the eleven calibration variants (Table 3) in the calibration period 2000–2010. The top and bottom panels show the runoff efficiency in the alpine (n = 119) and lowland (n = 94) catchments, respectively. Boxes show the 25th–75th percentile range with the horizontal line indicating the median; whiskers extend to the 10th and 90th percentiles. Box fill colour indicates the calibration class: satellite only (Vsat); regular sparse runoff (Vmo, Vse); event-based runoff (Vwet, Vdry, Vavg); and the corresponding multi-objective variants combining each runoff strategy with satellite snow cover and soil moisture data (+sat).**

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026



295 **Figure 5. Snow cover error OSC of the eleven calibration variants (Table 3) in the calibration period 2000–2010. The top and bottom panels show the snow cover error in the alpine (n = 119) and lowland (n = 94) catchments, respectively. Boxes show the 25th–75th percentile range with the horizontal line indicating the median; whiskers extend to the 10th and 90th percentiles. Box fill colour indicates the calibration class as defined in Fig. 4. The y-axis is clipped at 0.6 for readability; a small number of catchments have OSC. OSC values above this range for the calibration variants that do not use satellite data.**

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

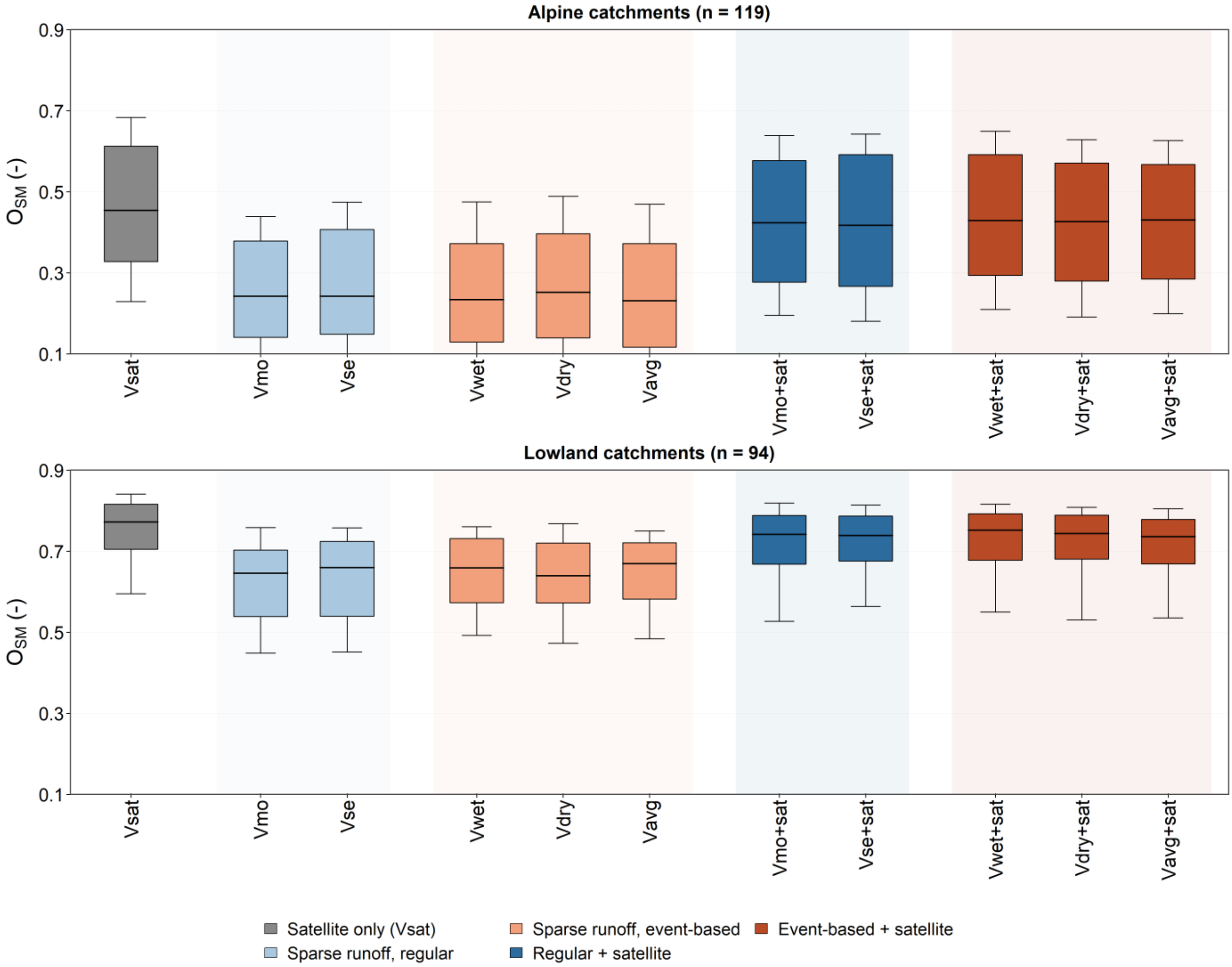


Figure 6. Soil moisture correlation OSM of the eleven calibration variants (Table 3) in the calibration period 2000–2010. The top and bottom panels show the soil moisture correlation in the alpine (n = 119) and lowland (n = 94) catchments, respectively. Boxes show the 25th–75th percentile range with the horizontal line indicating the median; whiskers extend to the 10th and 90th percentiles. Box fill colour indicates the calibration class as defined in Fig. 4.

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

4.2.2 Hydrological model performance in the validation period

The model performance in the independent validation period (September 2010 – August 2014) is summarised in Table 5. Figure 7 shows the runoff efficiency OQ, Figure 8 the snow cover error OSC, and Figure 9 the soil moisture correlation OSM. The general patterns are very similar to the calibration period. All variants that use streamflow data clearly outperform Vsat, and the "+sat" variants preserve the snow and soil moisture performance while strongly improving runoff simulations compared to Vsat.

In alpine catchments, Vmo is again the strongest variant for runoff simulation, with a median validation OQ of 0.75. The event-based streamflow-only variants follow closely, with median OQ of 0.71 for Vdry and Vavg and 0.69 for Vwet. These values are 5-8% lower than median efficiency of Vmo and confirm that, in alpine catchments, a single well-chosen gauging campaign in an average or dry year can produce runoff simulations that are nearly as good as those obtained from one streamflow value every month over the full calibration period. The "+sat" variants are somewhat lower in alpine catchments: Vmo+sat reaches 0.69, while the event-based "+sat" variants achieve medians of 0.55 (Vwet+sat) to 0.62 (Vavg+sat).

In lowland catchments, the situation is reversed and Vmo+sat is the best variant overall, with a median validation OQ of 0.67. This is slightly above Vmo (0.64) and the event-based "+sat" variants (0.60–0.64). The event-based streamflow-only variants perform less well in lowland catchments, with medians of 0.57–0.59. The advantage of Vmo+sat in lowland catchments is large and consistent across most cases, as discussed in Section 4.2.3 below.

The snow cover errors in validation (Fig. 8) show a clear separation between variants with and without satellite data. In alpine catchments, the "+sat" variants all have median validation OSC between 0.05 and 0.07. The streamflow-only variants have much larger snow cover errors, with medians between 0.25 (Vse) and 0.36 (Vwet). The difference is approximately a factor of five between an event-based variant with satellite data (e.g. Vwet+sat = 0.06) and the same variant without (Vwet = 0.36). In lowland catchments, snow cover errors are smaller for all variants, as in calibration, but the same pattern is visible: the "+sat" variants are slightly better than the streamflow-only variants (medians 0.02–0.03 vs. 0.03–0.05).

The soil moisture correlations in validation (Fig. 9) show the same pattern. In alpine catchments, all "+sat" variants reach a median OSM of 0.52–0.53, equal to Vsat. The streamflow-only variants are clearly lower, with medians of 0.34–0.39. In lowland catchments, the "+sat" variants reach medians of 0.73–0.74 (close to Vsat at 0.77), while the streamflow-only variants drop to 0.64–0.67.

A practical implication of the validation results is that the choice between streamflow-only and "+sat" variants in alpine catchments depends on what the model will be used for. If runoff is the only output of interest, the streamflow-only event-based variants are competitive with Vmo and considerably better than Vsat. If the model is also used to simulate snow cover or soil moisture, the "+sat" variants are clearly preferable, even though their runoff efficiency is somewhat lower. In lowland catchments, the choice is simpler: Vmo+sat is the best variant for all three model outputs and should be preferred whenever monthly streamflow observations are available.

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

Table 5: Median model performance of eleven calibration variants (defined in Table 3) in the validation period 2010–2014. Medians are estimated separately for alpine and lowland catchments. The runoff (OQ), snow cover (OSC) and soil moisture (OSM) efficiencies are defined in section 3.3.

Variant	Alpine OQ	Alpine OSC	Alpine OSM	Lowland OQ	Lowland OSC	Lowland OSM
Vsat	0.26	0.04	0.52	0.12	0.03	0.77
Vmo+sat	0.69	0.07	0.52	0.67	0.03	0.74
Vse+sat	0.63	0.06	0.53	0.60	0.03	0.74
Vwet+sat	0.55	0.06	0.53	0.64	0.03	0.74
Vdry+sat	0.61	0.06	0.53	0.61	0.02	0.73
Vavg+sat	0.62	0.06	0.53	0.60	0.03	0.74
Vmo	0.75	0.31	0.38	0.64	0.04	0.64
Vse	0.68	0.25	0.39	0.56	0.05	0.65
Vwet	0.69	0.36	0.35	0.59	0.04	0.67
Vdry	0.71	0.27	0.37	0.58	0.03	0.66
Vavg	0.71	0.28	0.34	0.57	0.04	0.66

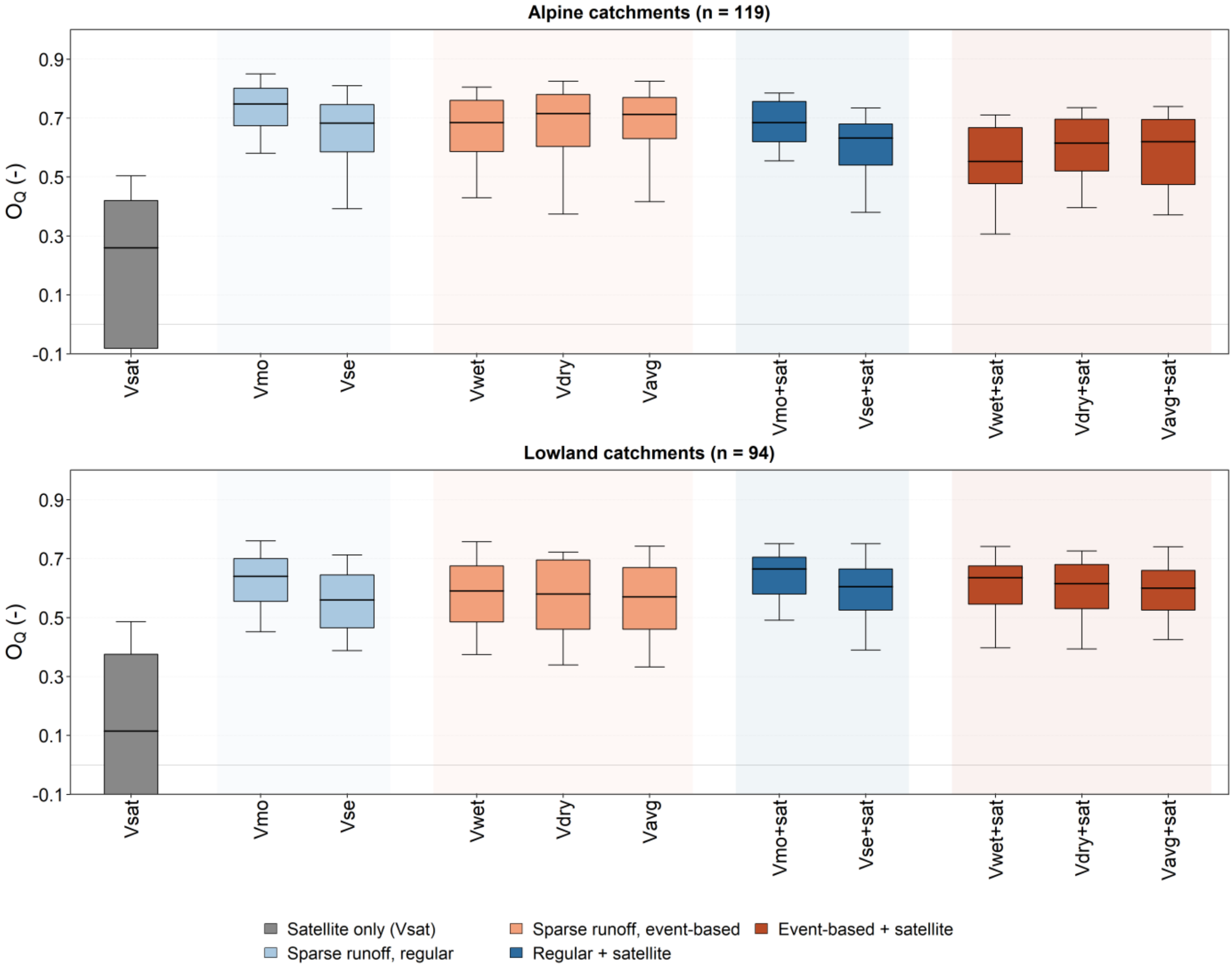
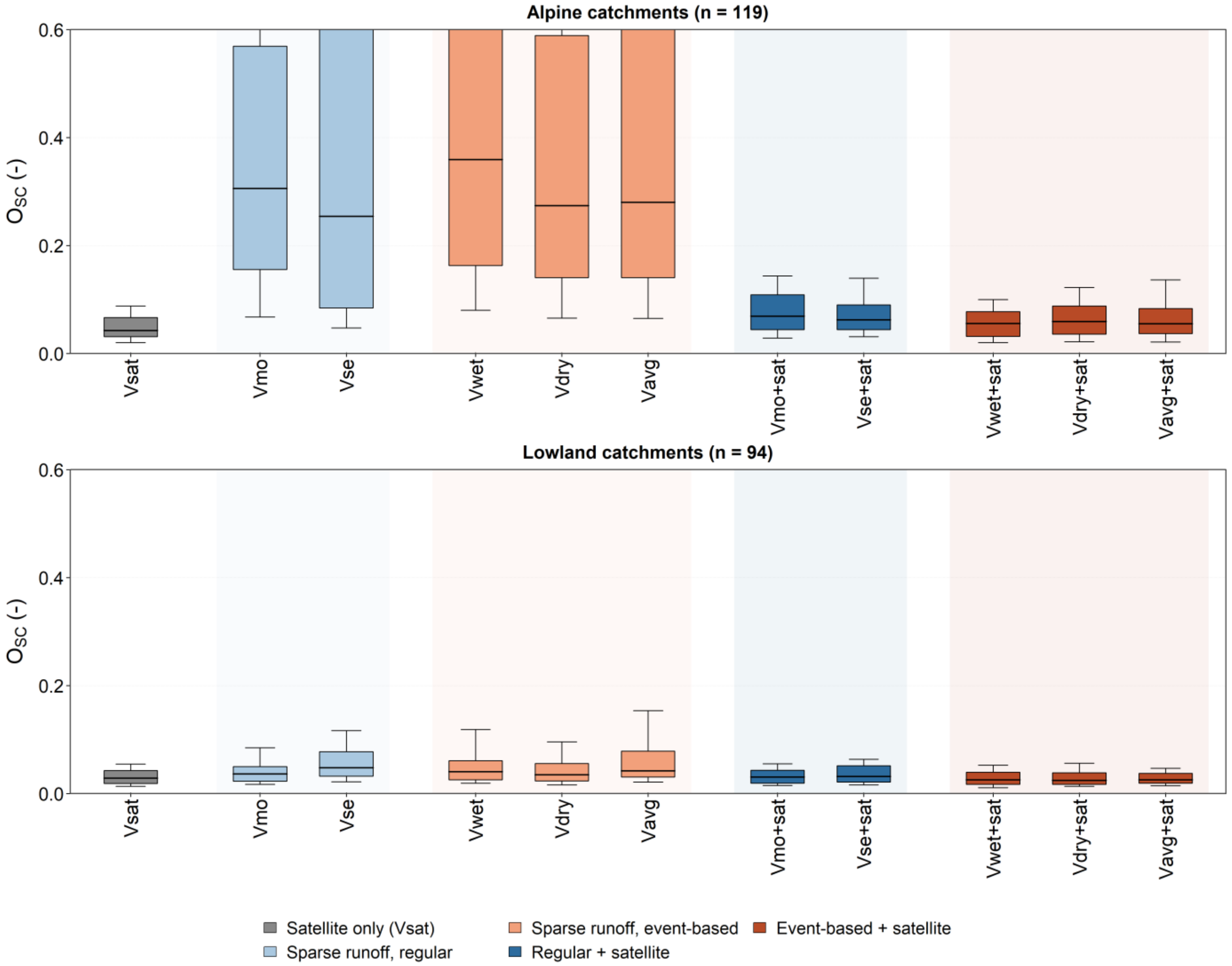


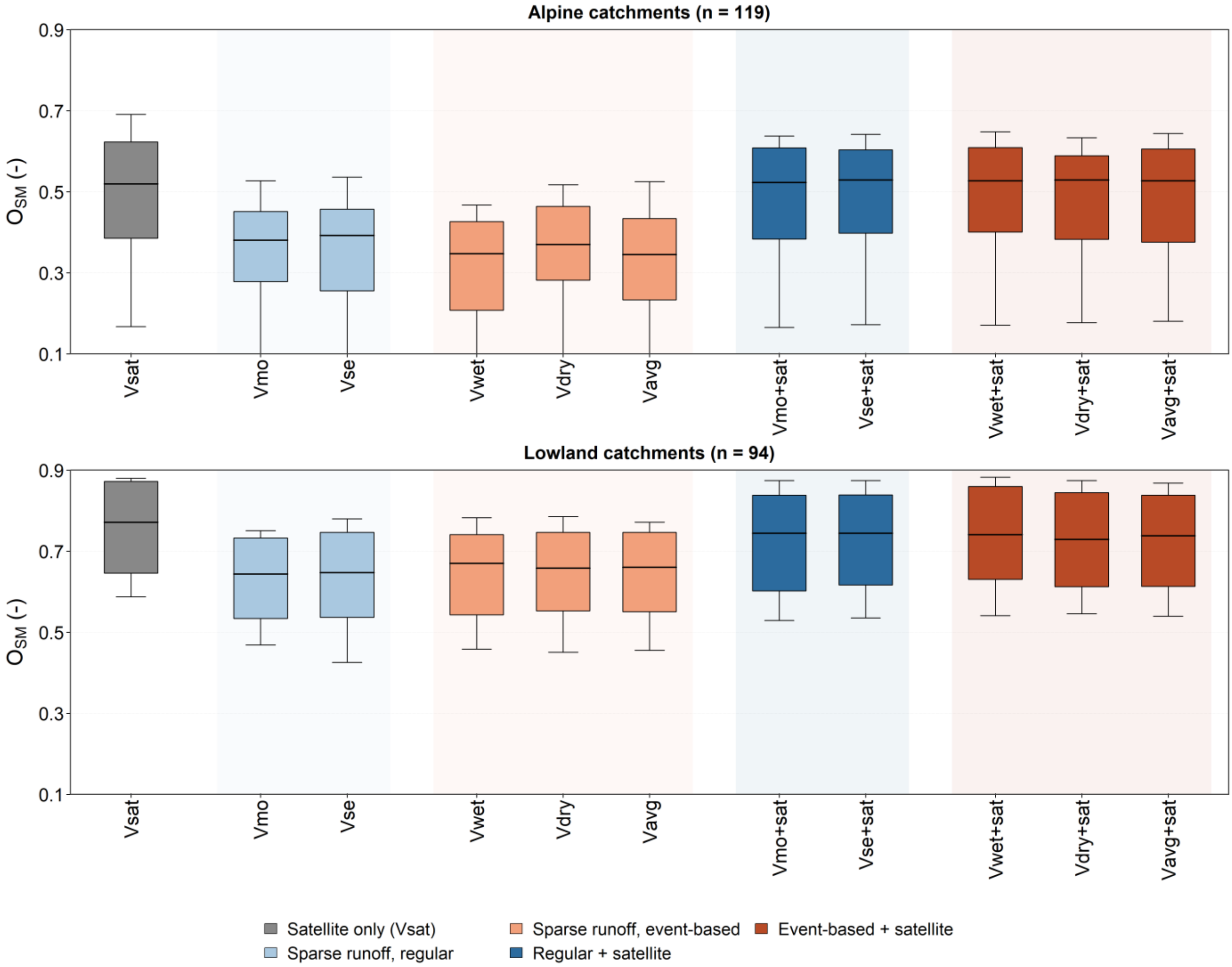
Figure 7. Runoff model efficiency OQ of the eleven calibration variants (Table 3) in the independent validation period 2010–2014. The top and bottom panels show the runoff efficiency in the alpine (n = 119) and lowland (n = 94) catchments, respectively. Boxes show the 25th–75th percentile range with the horizontal line indicating the median; whiskers extend to the 10th and 90th percentiles.
350 **Box fill colour indicates the calibration class as defined in Fig. 4.**

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026



355 **Figure 8. Snow cover error OS of the eleven calibration variants (Table 3) in the independent validation period 2010–2014. The top and bottom panels show the snow cover error in the alpine (n = 119) and lowland (n = 94) catchments, respectively. Boxes show the 25th–75th percentile range with the horizontal line indicating the median; whiskers extend to the 10th and 90th percentiles. Box fill colour indicates the calibration class as defined in Fig. 4. The y-axis is clipped at 0.6 for readability; a small number of catchments have OSC values above this range for the calibration variants that do not use satellite data.**

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026



360 **Figure 9. Soil moisture correlation OSM of the eleven calibration variants (Table 3) in the independent validation period 2010–2014. The top and bottom panels show the soil moisture correlation in the alpine (n = 119) and lowland (n = 94) catchments, respectively. Boxes show the 25th–75th percentile range with the horizontal line indicating the median; whiskers extend to the 10th and 90th percentiles. Box fill colour indicates the calibration class as defined in Fig. 4.**

365

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

4.2.3 Comparison of variant pairs with and without satellite data

The results in Sections 4.2.1 and 4.2.2 show that adding satellite data to a calibration with limited streamflow records reduces snow cover errors and improves soil moisture correlations consistently. The results for the runoff efficiency show a regionally different pattern. To examine this, we compare the runoff efficiency in the validation period for each pair of variants that use the same streamflow data with and without satellite data (e.g. Vmo+sat against Vmo, Vwet+sat against Vwet, and so on). Table 6 shows the relative frequency of catchments in which adding satellite data improves the runoff efficiency.

The results show a strong and consistent regional difference. In lowland catchments, adding satellite data to any streamflow strategy improves runoff simulation in approximately 54–56% of cases. This pattern is remarkably stable across all five variant pairs, regardless of whether the streamflow data are regularly spaced (Vmo, Vse) or come from a single-year event-based campaign (Vwet, Vdry, Vavg). The median runoff efficiency in lowland catchments improves slightly when satellite data are added (by +0.01 to +0.03), confirming that combining the two data sources is the preferred approach in this region.

In alpine catchments, the opposite is observed. Adding satellite data to a streamflow-based calibration improves the runoff efficiency in only 15–29% of catchments. The percentages are highest for the variants that use regularly spaced streamflow records (Vmo+sat improves over Vmo in 24% of catchments and Vse+sat improves over Vse in 29%) and lowest for the event-based variants (Vwet+sat improves over Vwet in only 15% of catchments). The median change in runoff efficiency in alpine catchments is negative for all five pairs, ranging from −0.05 (Vmo+sat) to −0.11 (Vwet+sat).

While Table 6 quantifies how often the addition of satellite data improves the runoff efficiency in each catchment, it does not show the magnitude or seasonal pattern of these effects. To visualise where in the seasonal cycle and in what magnitude the eleven calibration variants differ, and how this varies with the level of bias of the satellite-only model, we examine the simulated monthly runoff differences grouped by the four clusters identified in Section 4.1. Figure 10 shows the mean monthly runoff difference between simulated and observed streamflow for each cluster and for all eleven variants in the validation period. The variants are colour-coded by class as in Figure 4, and for each class with two or more variants, the min-to-max envelope across class members is shown to indicate the typical range of behaviour within each class.

In cluster 1, which contains 125 catchments with seasonally balanced runoff simulations, all calibration variants produce monthly differences close to zero throughout the year. Vsat shows a moderate overestimation in late summer and autumn (up to about +80 mm month^{-1} in September), while all variants that include streamflow data, with or without satellite data, reduce this overestimation to within approximately ±60 mm month^{-1}. The envelopes of the four variant classes are narrow and largely overlap, indicating that for these well-behaved catchments, the choice of calibration strategy has only a minor effect on the simulated monthly water balance.

In cluster 2, which contains 40 catchments characterised by strong summer overestimation, Vsat reaches a peak monthly difference of approximately +280 mm month^{-1} in August. All variants that include streamflow data substantially compress this summer bias. The streamflow-only variants (classes 2 and 3) keep the monthly difference within approximately ±30 mm month^{-1}, while the +sat variants (classes 4 and 5) leave residual summer overestimation of up to +80 mm month^{-1}. This pattern

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

shows that adding satellite data to the streamflow calibration partially limits the bias correction, because the parameters are also constrained to match the satellite snow and soil moisture observations.

In clusters 3 and 4, which contain 26 and 5 catchments with strong and extreme summer underestimation, respectively, the patterns are similar in shape but very different in magnitude. Vsat reaches monthly underestimations of about −520 mm month^{-1} in cluster 3 and −1260 mm month^{-1} in cluster 4. Adding any streamflow data substantially reduces this underestimation. The event-based streamflow-only variants (class 3) produce the smallest residual bias in both clusters, with worst-month differences of about −150 mm month^{-1} in cluster 3 and −360 mm month^{-1} in cluster 4 — a reduction of three- to fourfold compared to Vsat. The regularly spaced streamflow-only variants (class 2) reach somewhat larger residuals (−305 mm month^{-1} in cluster 3 and −350 mm month^{-1} in cluster 4). The +sat variants achieve intermediate performance, with worst-month differences of about −400 to −420 mm month^{-1} in cluster 3 and −1000 to −1090 mm month^{-1} in cluster 4. This means that in the alpine clusters where the satellite-only calibration is biased the most, the simultaneous matching of satellite snow cover and soil moisture limits the extent to which the streamflow data can correct the model. The visual ranking of variant classes in Figure 10 is therefore reversed compared to clusters 1 and 2: in clusters 3 and 4, streamflow-only event-based variants visibly outperform the corresponding +sat variants for monthly bias correction.

These patterns are consistent with the catchment-level comparison in Table 6 and with the alpine-versus-lowland trade-off identified above. Where the satellite-only model is already close to observed runoff (cluster 1) or only moderately biased (cluster 2), adding satellite data is either neutral or beneficial. Where the satellite-only model is strongly biased (clusters 3 and 4, predominantly alpine), event-based streamflow observations alone produce the strongest correction, and the addition of satellite data, while improving snow and soil moisture consistency, comes at the cost of part of the runoff bias correction.

Table 6: Relative frequency (%) of catchments where the multiple-objective variant (combining streamflow with satellite snow cover and soil moisture data) outperforms its single-objective counterpart (calibration to limited streamflow only) in the validation period 2010–2014. Performance is measured by the runoff efficiency OQ.

Variant pair	Alpine catchments	Lowland catchments
Vmo+sat vs. Vmo	24.4 % (29/119)	56.4 % (53/94)
Vse+sat vs. Vse	28.6 % (34/119)	55.3 % (52/94)
Vwet+sat vs. Vwet	15.1 % (18/119)	54.3 % (51/94)
Vdry+sat vs. Vdry	16.0 % (19/119)	54.3 % (51/94)
Vavg+sat vs. Vavg	21.8 % (26/119)	55.3 % (52/94)

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

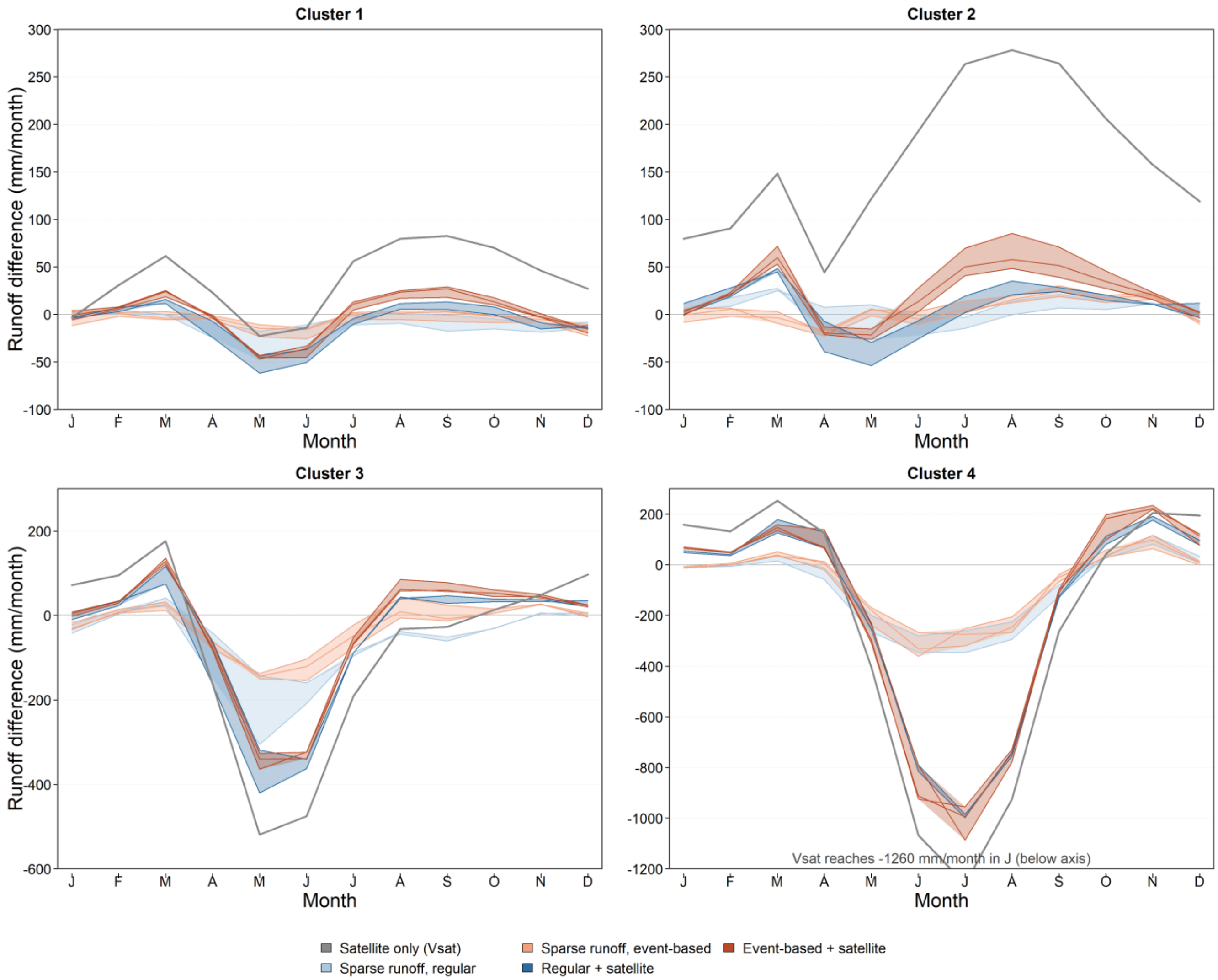


Figure 10: Mean monthly difference between simulated and observed runoff (mm/month) for the four clusters (Fig. 2) in the validation period 2010–2014. Each panel shows one cluster. Within each panel, the eleven calibration variants are plotted as thin lines, with line colour indicating the variant class as in Fig. 4. Semi-transparent polygons fill the minimum-to-maximum envelope across the variants in each class with two or more members (i.e. all classes except Vsat). The Vsat baseline (calibration to satellite data only) is shown as a thicker grey line for reference. Note that the y-axis range differs between panels: cluster 1 and cluster 2 use −100 to +300 mm/month, cluster 3 uses −600 to +300 mm/month, and cluster 4 uses −1200 to +300 mm/month. In cluster 4, the Vsat curve in July reaches −1260 mm/month and falls slightly below the panel's lower axis limit.

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

5 Discussion and conclusions

5.1 Performance of regularly spaced observations combined with satellite data

The first part of the analysis characterizes spatial patterns of model performance when calibration relies solely on satellite data (Vsat). In line with Tong et al. (2021), the best satellite-only performance is achieved in larger lowland catchments with a higher proportion of arable land and good meteorological station coverage. In alpine catchments, satellite-only calibration systematically underestimates summer streamflow, partly due to uncertain orographic precipitation inputs in catchments without in-situ stations.

Among regularly spaced variants, monthly observations (Vmo, Vmo+sat) consistently outperform seasonal observations (Vse, Vse+sat). The Vmo+sat variant achieves the best overall performance in lowland catchments, with a median validation runoff efficiency of 0.67 — only slightly below traditional calibration to continuous streamflow records (0.69 in Tong et al., 2021) and comparable to the best regionalization methods in Austria (Tong et al., 2022). The advantage of adding satellite data to monthly streamflow observations is especially clear in lowland catchments, where Vmo+sat outperforms Vmo in approximately 56% of catchments and delivers substantially better snow and soil moisture simulation in the large majority.

A consistent benefit of combining limited streamflow records with satellite data, beyond the regular-spacing case, is the improvement of internal model consistency across the entire set of variant pairs. When the same streamflow data are combined with MODIS snow cover and ASCAT soil moisture data, the snow cover error in the validation period decreases in approximately 85–92% of alpine catchments and 67–82% of lowland catchments, depending on the variant pair (Sect. 4.2.3). The improvement in soil moisture correlation is even more consistent, with the +sat variants outperforming the streamflow-only counterpart in approximately 87–91% of catchments in both alpine and lowland regions. These improvements are achieved across all five variant pairs (Vmo+sat vs. Vmo, Vse+sat vs. Vse, and the three event-based pairs), confirming that the benefit of satellite data for internal model consistency is robust to the type and temporal distribution of the available streamflow records.

5.2 Performance of event-based strategies

The event-based streamflow sampling strategies (Vwet, Vdry, Vavg and their +sat counterparts) are the novel contribution of this study compared to earlier published results. They are motivated by the practical reality that initial gauging in an ungauged catchment most commonly takes the form of a short, targeted campaign designed to capture one or a few high-flow events (Pool et al., 2017; Seibert and McDonnell, 2015). The strategies follow the design of CMax_Rec_Dom from Pool et al. (2017), which was among the most informative for hydrograph simulation in humid catchments.

In alpine catchments, the event-based streamflow-only variants achieve median validation runoff efficiencies of 0.69–0.71, closely approaching Vmo (0.75) and substantially exceeding Vse+sat (0.63). This suggests that a single well-designed gauging campaign can capture most of the hydrograph information needed for model calibration in alpine catchments. The driest and average-year variants (Vdry, Vavg) consistently outperform the wettest-year variant (Vwet) in alpine catchments, likely

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

because high-flow events in very wet years may not represent the dominant runoff generation mechanisms. In lowland catchments, event-based variants perform somewhat lower (0.57–0.59 vs. 0.64 for Vmo), reflecting the greater importance of average and low-flow conditions for water balance estimation in these drier regions (Pool et al., 2017).

When satellite data are added to event-based strategies, the most important gain is in internal model consistency. The +sat variants reduce median snow cover errors in alpine catchments from 0.27–0.36 to approximately 0.06 in the validation period — about a fivefold improvement (Sect. 4.2.2). The modest reduction in runoff efficiency in alpine catchments (e.g., Vdry+sat = 0.61 vs. Vdry = 0.71) reflects the inherent tension in multi-objective calibration: constraining the snow routine toward satellite observations may shift parameters away from the runoff-optimal configuration. This tension is region-specific: in lowland catchments, the same combination of event-based streamflow with satellite data improves runoff efficiency in more than half of the catchments (Table 6). For applications where both streamflow prediction and snow or soil moisture monitoring matter, the +sat variants are clearly preferable in both regions.

The choice of year for the event-based campaign is an important practical consideration. Our results show that campaigns in the driest or average year tend to outperform those in the wettest year, particularly in the +sat setting. When choosing between possible field seasons, a year close to the long-term precipitation mean may therefore be preferable to an exceptional high-precipitation year. This is consistent with the general recommendation that model calibration performs best when calibration conditions are representative of long-term hydrological variability (Pool et al., 2017; Vrugt et al., 2006).

5.3 Comparison with previous studies

The model performance values achieved here are broadly consistent with previous calibration and regionalization studies in Austria. Tong et al. (2021) reported a median validation runoff efficiency of 0.78 for alpine catchments using continuous streamflow calibration, compared to 0.75 (Vmo) and 0.71 (Vdry, Vavg) obtained here with far fewer observations. In lowland catchments, Vmo+sat (0.67) matches the performance of the best regionalization approaches in Austria (spatial proximity and local similarity, both 0.67–0.69; Tong et al., 2022). The results extend Pool et al. (2017) and Pool et al. (2019) by confirming that event-based strategies perform best for hydrograph simulation and by demonstrating that combining these strategies with satellite data substantially improves internal model consistency.

5.4 Limitations and outlook

Several limitations should be acknowledged. First, the event-based strategies are defined by selecting the year based on full knowledge of the record; in practice, this information is not available a priori. Future studies could explore adaptive strategies using climatological proxies or remote sensing to characterize catchment wetness at campaign design time. Second, we did not evaluate the sensitivity of results to the exact timing of the peak flow event, as explored in Pool et al. (2017). Third, we assumed equal weights for streamflow, snow cover, and soil moisture in multi-objective variants.

Crowd-sourced or citizen-science streamflow observations (Etter et al., 2020) could provide event-based data at low cost and could be combined with satellite data using the approach demonstrated here. New satellite products — such as higher-

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

resolution snow cover fraction retrievals or dedicated mountain soil moisture products — could further improve model performance in alpine catchments.

5.5 Conclusions

This study demonstrates that limited streamflow observations can substantially improve hydrological model calibration when combined with satellite remote sensing data, and that the benefit depends on both the amount of streamflow data and the hydrological setting. The main conclusions are:

1. Satellite-only calibration (Vsat) provides reliable model simulations primarily in lowland catchments with lower elevation, higher aridity, more arable land cover, and good precipitation station coverage. In alpine catchments, simulations based on satellite data alone systematically underestimate summer streamflow.
2. Any addition of streamflow data — whether regularly spaced or event-based — substantially improves runoff model efficiency in both alpine and lowland catchments, with the largest gains in alpine catchments.
3. Regularly spaced monthly observations combined with satellite data (Vmo+sat) provide the best overall performance balance. In lowland catchments, Vmo+sat achieves a median validation efficiency of 0.67, comparable to the best regionalization methods in Austria.
4. Event-based streamflow strategies based on a single gauging campaign achieve median validation runoff efficiencies of 0.57–0.71, closely approaching monthly observations in alpine catchments. Campaigns from the driest or average year outperform those from the wettest year.
5. Combining event-based streamflow data with satellite constraints (+sat) dramatically improves snow cover and soil moisture simulation consistency in alpine catchments, reducing snow cover errors by approximately a factor of five compared to streamflow-only calibration.

These results support the use of short, targeted gauging campaigns combined with satellite remote sensing for hydrological predictions in data-sparse regions, with clear implications for campaign planning in ungauged alpine and lowland catchments.

Code and data availability

R scripts used for the analysis and visualisations are available upon request from the corresponding authors. MODIS snow cover data are available from the National Snow and Ice Data Center (https://nsidc.org). Streamflow data are available from the Hydrographic Service of Austria (https://ehyd.gv.at, last access: 31 December 2025). The SPARTACUS dataset is available from the Geosphere Austria data portal (https://data.hub.geosphere.at, last access: 31 December 2025). Processed data can be provided by the corresponding author upon reasonable request.

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

Author contributions

AK and JP conceptualized and designed the study, developed the methodology, designed the code, performed the analyses, and prepared the manuscript. RT was responsible for dataset preparation and management. ZM, BS, and MB contributed to code refinement and assisted in manuscript revision. All authors discussed the results and commented on the manuscript.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Disclaimer

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

Acknowledgements

The study was supported by the Higher Education Commission of Pakistan (HRD/OSS-III/2022/HEC), administered under the Austrian Agency for Education and Internationalization (OeAD), and the FFG ASAP project Digital Twin for Austria – Alpine Hydrology and Future Hazards (DTA-Hydro), project No. FO999918403. We declare the use of Claude 4.7 AI to improve the readability of manuscripts, including grammar and spelling checks, and to prepare scripts for visualization.

Financial support

AK was funded by the Higher Education Commission of Pakistan (HRD/OSS-III/2022/HEC) administered under OeAD. The FFG ASAP project DTA-Hydro (No. FO999918403) also provided financial support.

References

Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M.: The SWOT Mission and Its Capabilities for Land Hydrology, Surv. Geophys., 37, 307–337, https://doi.org/10.1007/s10712-015-9346-y, 2016.

Blöschl, G. (Ed.): Runoff prediction in ungauged basins: synthesis across processes, places and scales, Cambridge University Press, Cambridge, UK, 465 pp., 2013.

Brocca, L., Crow, W.T., Ciabatta, L., Massari, C., De Rosnay, P., Enenkel, M., Hahn, S., Amarnath, G., Camici, S., Tarpanelli, A., and Wagner, W.: A Review of the Applications of ASCAT Soil Moisture Products, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 10, 2285–2306, https://doi.org/10.1109/JSTARS.2017.2651140, 2017.

Etter, S., Strobl, B., van Meerveld, I., and Seibert, J.: Quality and timing of crowd-based water level class observations, Hydrol. Process., 34, 4365–4378, https://doi.org/10.1002/hyp.13864, 2020.

Hall, D. and Riggs, G.: MODIS/Aqua Snow Cover Daily L3 Global 500m SIN Grid, Version 6, NASA NSIDC DAAC, https://doi.org/10.5067/MODIS/MYD10A1.006, 2016.

Hiebl, J. and Frei, C.: Daily temperature grids for Austria since 1961 — concept, creation and applicability, Theor. Appl. Climatol., 124, 161–178, https://doi.org/10.1007/s00704-015-1411-4, 2016.

Hiebl, J. and Frei, C.: Daily precipitation grids for Austria since 1961 — development and evaluation of a spatial dataset for hydroclimatic monitoring and modelling, Theor. Appl. Climatol., 132, 327–345, https://doi.org/10.1007/s00704-017-2093-x, 2018.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of predictions in ungauged basins (PUB) – a review, Hydrol. Sci. J., 58, 1198–1255, https://doi.org/10.1080/02626667.2013.803183, 2013.

Jansen, K. F., Teuling, A. J., Craig, J. R., Dal Molin, M., Knoben, W. J., Parajka, J., Vis, M., and Melsen, L. A.: Mimicry of a conceptual hydrological model (HBV): what's in a name?, Water Resour. Res., 57, e2020WR029143, https://doi.org/10.1029/2020WR029143, 2021.

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., and Wood, E. F.: Inroads of remote sensing into hydrologic science during the WRR era, Water Resour. Res., 51, 7309–7342, https://doi.org/10.1002/2015WR017616, 2015.

MacQueen, J. B.: Some methods for classification and analysis of multivariate observations, in: Proc. Fifth Berkeley Symp. Math. Stat. Prob., Vol. 1, University of California Press, 281–297, 1967.

Melsen, L. A., Teuling, A. J., van Berkum, S. W., Torfs, P. J. J. F., and Uijlenhoet, R.: Catchments as simple dynamical systems: a case study on methods and data requirements for parameter identification, Water Resour. Res., 50, 5577–5596, https://doi.org/10.1002/2013WR014720, 2014.

Mullen, K. M., Ardia, D., Gil, D. L., Windover, D., and Cline, J.: DEoptim: An R package for global optimization by differential evolution, J. Stat. Softw., 40, 1–26, https://doi.org/10.18637/jss.v040.i06, 2011.

Parajka, J. and Blöschl, G.: The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, J. Hydrol., 358, 240–258, https://doi.org/10.1016/j.jhydrol.2008.06.006, 2008.

Parajka, J., Merz, R., and Blöschl, G.: Estimation of daily potential evapotranspiration for regional water balance modeling in Austria, in: 11th International Poster Day, Slovak Academy of Sciences, 20 November 2003, Bratislava, 299–306, 2003.

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies, Hydrol. Earth Syst. Sci., 17, 1783–1795, https://doi.org/10.5194/hess-17-1783-2013, 2013.

Pool, S., Viviroli, D., and Seibert, J.: Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration?, J. Hydrol., 554, 613–622, https://doi.org/10.1016/j.jhydrol.2017.09.037, 2017.

Pool, S., Viviroli, D., and Seibert, J.: Value of a limited number of discharge observations for improving regionalization: A large-sample study across the United States, Water Resour. Res., 55, 363–377, https://doi.org/10.1029/2018WR023855, 2019.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2024.

Scheller, M., van Meerveld, I., Sauquet, E., Vis, M., and Seibert, J.: Are temporary stream observations useful for calibrating a lumped hydrological model?, J. Hydrol., 632, 130686, https://doi.org/10.1016/j.jhydrol.2024.130686, 2024.

Seibert, J. and Beven, K. J.: Gauging the ungauged basin: how many discharge measurements are needed?, Hydrol. Earth Syst. Sci., 13, 883–892, https://doi.org/10.5194/hess-13-883-2009, 2009.

Seibert, J. and McDonnell, J. J.: Gauging the ungauged basin: relative value of soft and hard data, J. Hydrol. Eng., 20, A4014004, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000861, 2015.

Singh, S. K. and Bárdossy, A.: Calibration of hydrological models on hydrologically unusual events, Adv. Water Resour., 38, 81–91, https://doi.org/10.1016/j.advwatres.2011.12.006, 2012.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences, Hydrol. Sci. J., 48, 857–880, https://doi.org/10.1623/hysj.48.6.857.51421, 2003.

Sleziak, P., Szolgay, J., Hlavčová, K., Danko, M., and Parajka, J.: The effect of the snow weighting on the temporal stability of hydrologic model efficiency and parameters, J. Hydrol., 583, 124639, https://doi.org/10.1016/j.jhydrol.2020.124639, 2020.

Sleziak, P., Danko, M., Jančo, M., Holko, L., Greimeister-Pfeil, I., Vreugdenhil, M., and Parajka, J.: Accuracy of ASCAT-DIREX Soil Moisture Mapping in a Small Alpine Catchment, Water, 17, 49, https://doi.org/10.3390/w17010049, 2024.

Tong, R., Parajka, J., Komma, J., and Blöschl, G.: Mapping snow cover from daily Collection 6 MODIS products over Austria, J. Hydrol., 590, 125548, https://doi.org/10.1016/j.jhydrol.2020.125548, 2020.

Tong, R., Parajka, J., Salentinig, A., Pfeil, I., Komma, J., Széles, B., Kubán, M., Valent, P., Vreugdenhil, M., Wagner, W., and Blöschl, G.: The value of ASCAT soil moisture and MODIS snow cover data for calibrating a conceptual hydrologic model, Hydrol. Earth Syst. Sci., 25, 1389–1410, https://doi.org/10.5194/hess-25-1389-2021, 2021.

https://doi.org/10.5194/egusphere-2026-2880
Preprint. Discussion started: 18 June 2026

Tong, R., Parajka, J., Széles, B., Greimeister-Pfeil, I., Vreugdenhil, M., Komma, J., Valent, P., and Blöschl, G.: The value of satellite soil moisture and snow cover data for the transfer of hydrological model parameters to ungauged sites, Hydrol. Earth Syst. Sci., 26, 1779–1799, https://doi.org/10.5194/hess-26-1779-2022, 2022.

Viglione, A. and Parajka, J.: TUWmodel: Lumped/Semi-Distributed Hydrological Model for Education Purposes, R package version 1.1-1, https://CRAN.R-project.org/package=TUWmodel, 2020.

Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria, Hydrol. Earth Syst. Sci., 17, 2263–2279, https://doi.org/10.5194/hess-17-2263-2013, 2013.

Viviroli, D. and Seibert, J.: Can a regionalized model parameterisation be improved with a limited number of runoff measurements?, J. Hydrol., 529, 49–61, https://doi.org/10.1016/j.jhydrol.2015.07.009, 2015.

Vrugt, J. A., Gupta, H. V., Nualáin, B.Ó., and Bouten, W.: Real-time data assimilation for operational ensemble streamflow forecasting, J. Hydrometeorol., 7, 548–565, https://doi.org/10.1175/JHM504.1, 2006.

Wagner, W., Lemoine, G., and Rott, H.: A method for estimating soil moisture from ERS scatterometer and soil data, Remote Sens. Environ., 70, 191–207, https://doi.org/10.1016/S0034-4257(99)00036-X, 1999.

Wagner, P. D., Duethmann, D., Kiesel, J., Pool, S., Hrachowitz, M., Ceola, S., Herzog, A., Houska, T., Loritz, R., Spieler, D., Staudinger, M., Tarasova, L., Thober, S., Fohrer, N., Tetzlaff, D., Wagener, T., and Guse, B.: The Unexploited Treasures of Hydrological Observations Beyond Streamflow for Catchment Modeling, WIREs Water, 12, e70018, https://doi.org/10.1002/wat2.70018, 2025.