

Can streamflow observations constrain snow mass reconstructions? Lessons from two synthetic numerical experiments

Pau Wiersma¹, Jan Magnusson², Nadav Peleg^{1,3}, Bettina Schaeffli^{4,5}, and Gregoire Mariethoz^{1,3}

¹Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland

²WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

³Expertise Center for Climate Extremes, University of Lausanne, Lausanne, Switzerland

⁴Institute of Geography, University of Bern, Bern, Switzerland

⁵Oeschger Center on Climate Change Research, University of Bern, Switzerland, Bern, Switzerland

Correspondence: Pau Wiersma (pau.wiersma@unil.ch)

Abstract

Historical estimates of seasonal snow mass are key to understanding snowmelt-driven streamflow and climate change impacts on mountain water resources. However, direct observations of snow mass are sparse in space and time, forcing most reconstructions to rely on snow models driven by uncertain meteorological inputs. While ground-based and satellite snow observations are commonly used to constrain these models, their potential is limited in data-scarce regions and before the onset of satellite monitoring. Here, we investigate the potential of streamflow observations as an additional source of information to improve historical snow mass reconstructions. We introduce an inverse hydrological modeling framework that selects realistic snow mass realizations based on the accuracy of their streamflow response. Before real-world application, we test the framework in two synthetic experiments. Our results demonstrate that streamflow has the potential to constrain snow mass reconstructions, but that non-uniqueness in the snow-streamflow relationship and uncertainties in the inverse modeling chain can easily stand in the way. We also show that streamflow is most helpful in constraining catchment-aggregated properties of snow mass reconstructions, in particular catchment-aggregated melt rates. Future work should assess the potential of streamflow to constrain snow mass reconstruction under real-world conditions and investigate the added value of streamflow when combined with other snow data sources.

1 Introduction

Seasonal snow is essential to hydrology, ecology, tourism, and hydropower in mountainous regions (Beniston et al., 2018). A key variable in understanding snow dynamics is snow water equivalent (SWE), which represents the amount of water stored in the snowpack. Historical SWE estimates are important to understand how snow accumulation and melt have responded to climate change over the past decades (Gottlieb and Mankin, 2024), and to assess the role of changing snowpack dynamics in altering streamflow timing, volume, and drought risk (Berghuijs et al., 2014; Gordon et al., 2022; Brunner et al., 2023; Han

et al., 2024; Hou et al., 2025). However, direct observations of SWE from ground stations are often limited due to sparse station networks and the high logistical and physical cost of manual snow surveys (Haberkorn et al., 2019). In addition, snowfall and snowmelt patterns vary spatially (Grünewald et al., 2010; Mooney and Webb, 2025), making it difficult to generalize available observations. Passive microwave measurements from space provide large-scale SWE estimates, but at a resolution insufficient for mountainous areas (Luo et al., 2021). Measurements of other snow properties are more widespread, such as snow covered area (SCA) (Gascoin et al., 2019) and wet snow maps (Cluzet et al., 2024) from satellites, and snow depth (SD) from both satellites (Lievens et al., 2021; Besso et al., 2024) and ground measurements (Fontrodona-Bach et al., 2023), but their relationship to SWE is indirect; SCA and wet snow measurements only provide information on the presence or the wetness of snow, while SD must be converted to SWE using snow density estimates, which are highly variable in space and time as well (López-Moreno et al., 2013; Raleigh and Small, 2017).

To understand SWE dynamics, numerous studies have performed gridded SWE reconstructions through snow modeling constrained by different sources of indirect SWE observations. Mudryk et al. (2024) benchmarked 23 coarse-resolution, continental-scale SWE products with different inputs and data assimilation approaches. While most analyzed products performed well in capturing SWE climatology and interannual variability over low-relief regions, their performance degraded substantially in mountainous areas. Using a higher resolution to target mountain areas specifically, Margulis et al. (2016) and Fang et al. (2022) reconstructed gridded SWE in the Western US using a land-surface model combined with remotely sensed fractional SCA maps through batch data assimilation. Fiddes et al. (2019) applied a similar approach to Switzerland, while additionally including a grid cell clustering scheme in the land-surface model. Also in Switzerland, Mott et al. (2023) produced gridded SWE reconstructions using two different snow models with forward data assimilation of in-situ SD observations. Similarly, Broxton et al. (2016, 2019) combined in-situ SWE and SD observations with meteorological data to reconstruct SWE since 1981 in the continental United States. Avanzi et al. (2023) reconstructed SWE in Italy, using a snow model with data assimilation of both interpolated SD and SCA maps. Finally, Premier et al. (2023) identified periods of snow accumulation and melt by integrating in-situ SD observations, SCA maps, and snow classification maps from spaceborne synthetic aperture radar. They then reconstructed SWE accumulation by summing degree-day melt estimates during the identified melt phases using an empirical melt factor.

In addition to indirect SWE observations, empirical knowledge on recurring SD patterns can help to reconstruct SWE. Numerous studies have shown that spatial SD distributions can be statistically linked to terrain characteristics such as elevation, slope, and sky view factor (Lehning et al., 2011; Grünewald et al., 2013; Revuelto et al., 2014) and vegetation features such as canopy structure and density (Trujillo et al., 2007; Mazzotti et al., 2019; Helbig et al., 2020). Helbig and van Herwijnen (2017) derived gridded SD estimates from point-scale SD measurements using terrain properties of each grid cell. Pflug and Lundquist (2020) were able to extrapolate SD for an entire catchment from observing only 4% of its surface by leveraging recurring SD patterns. Similarly, Geissler et al. (2025) and Ylönen et al. (2025) used repeated UAV LiDAR surveys to define clusters of locations showing similar snow dynamics. They then used these clusters to spatially extrapolate point snow-depth measurements, producing region-wide maps of SD and SWE. Zakeri et al. (2025) downscaled low resolution SWE estimates by reusing high resolution reanalysis SWE from dates with similar low resolution SWE and climate data patterns. Finally,

Michel et al. (2023) demonstrated that SWE reconstructions for poorly observed years can be constrained by applying bias corrections derived from well-observed years.

However, in scarcely monitored regions and before the onset of remote sensing, the above sources of information are often lacking. In such contexts, streamflow observations offer a complementary source of information for SWE reconstruction. Streamflow gauging stations are relatively abundant due to their cost-effectiveness and importance for flood forecasting (Harrigan et al., 2022), and their observations often predate snow information sources (Do et al., 2018). Streamflow represents the integrated hydrological response of a catchment, in terms of both timing and volume (Kirchner, 2009). As such, it ought to contain information on the snow melt dynamics and the water balance of the entire catchment, including the higher elevations which are typically underrepresented in snow and meteorological observations (Thornton et al., 2021; Dettinger, 2014). However, the SWE information in streamflow is indirect and subject to transformation: the melt signal is delayed and smoothed by processes of water partitioning, storage and transport through the catchment, confounded by rainfall contributions, and affected by sublimation and evaporation losses. Moreover, streamflow is a one-dimensional, catchment-integrated observation, while SWE is a spatially distributed state variable. These complications raise a fundamental question: to what extent can streamflow observations constrain SWE reconstructions?

Three main approaches have been proposed to retrieve SWE information from streamflow. The first is the mass-curve technique, which estimates maximum catchment-wide SWE directly from the maximum seasonal deficit between accumulated precipitation and streamflow. Schaeffli (2016) showed good agreement with the SWE output of a snow model, while Horner et al. (2020) found that although interannual variability was well captured, absolute SWE was overestimated due to unaccounted losses and storage assumptions. A second approach estimates SWE from the difference between total streamflow and baseflow, as applied by Casson et al. (2018) and Whittaker and Leconte (2022) in large boreal catchments. This method is sensitive to baseflow separation uncertainty and assumes that all direct runoff in spring originates from snowmelt, an assumption less valid in smaller and steeper basins. A third strategy involves inverse hydrological modeling, or "doing hydrology backwards" (Kirchner, 2009): Henn et al. (2015, 2018) used Bayesian inversion to infer annual catchment precipitation from streamflow in snow-dominated Californian basins, while Rudisill et al. (2023) applied the same approach in the Upper Colorado river basin. However, they did not evaluate SWE explicitly and did not separate rain from snow, limiting the applicability of the approach in mixed-phase climates. Also using inverse hydrological modeling, Le Moine et al. (2015) and Ruelland (2020) derived multi-year temperature and precipitation gradients in mountainous catchments. While Le Moine et al. (2015) evaluated the resulting SWE estimates against station observations, Ruelland (2020) used binary snow cover maps alongside streamflow in a multi-objective inference. Despite these advances, the potential of inverse hydrological modeling for gridded SWE reconstruction remains largely unexplored, along with the amount and nature of SWE information embedded in streamflow and the conditions under which it can effectively constrain SWE reconstructions.

Here, we present a framework for streamflow-constrained SWE reconstruction that formulates snow inference as an inverse hydrological problem. Similar in concept to the inversion approach of Henn et al. (2015, 2018), our method generates a large ensemble of spatially distributed SWE realizations, propagates them through a distributed hydrological model, and selects a posterior ensemble based on the match between simulated and observed streamflow. To benchmark the core capabilities of

the inversion, we conduct two synthetic numerical experiments. The first is a fully synthetic experiment, where we eliminate all sources of uncertainty to test the theoretical constraining potential of streamflow on SWE. The second is a semi-synthetic experiment, where we test how much the constraining potential is reduced under meteorological forcing and snow model uncertainty. In both experiments, we evaluate which SWE metrics are best constrained by the streamflow and how their identifiability changes across spatial scales. While in real-world settings streamflow observations will often exist alongside spaceborne or in-situ snow observations, the goal of this study is to isolate the constraining potential of streamflow on SWE to reveal how streamflow observations are most effectively exploited in inverse hydrological SWE reconstruction.

2 Methodology

2.1 Streamflow-constrained SWE as an inverse problem

The constraining of SWE reconstructions through streamflow can be framed as an inverse problem, where the known output of a system (streamflow) is used to infer an unknown internal state (SWE). Prior knowledge on snow physics, topographic controls, and meteorological inputs reduce the solution space. Still, the inversion remains ill-posed: we aim to retrieve the space-time evolution of gridded SWE (3-dimensional aspect) from a catchment-integrated streamflow signal (single dimension).

We denote the time series of observed streamflow with Q_{obs} , and the spatio-temporal SWE field as H_{SWE} . In a Bayesian framework, we seek the posterior distribution:

$$P(H_{\text{SWE}} | Q_{\text{obs}}) \propto P(Q_{\text{obs}} | H_{\text{SWE}}) \cdot P(H_{\text{SWE}}). \quad (1)$$

The prior distribution $P(H_{\text{SWE}})$ reflects our initial uncertainty about SWE, and the likelihood $P(Q_{\text{obs}} | H_{\text{SWE}})$ quantifies how well a given SWE realization explains the observed discharge. Since H_{SWE} is not a free variable but the result of snow model simulations, we rather define $P(H_{\text{SWE}})$ as the result of the finite sampling of the informative prior distributions of parameters θ as follows:

$$H_{\text{SWE}}^{(i)} = f_{\text{snow}}(M; \theta_{\text{meteo}}^{(i)}, \theta_{\text{snow}}^{(i)}), \quad \text{with } \theta_{\text{meteo}}^{(i)} \sim P(\theta_{\text{meteo}}), \theta_{\text{snow}}^{(i)} \sim P(\theta_{\text{snow}}), \quad (2)$$

where M is the meteorological forcing (precipitation and temperature), θ_{meteo} are meteorological parameters (e.g., precipitation scaling, lapse rates, phase partitioning), and θ_{snow} are snow model parameters controlling melt rates and snowpack dynamics. Repeating this for $i = 1, \dots, N_{\text{prior}}$ yields an ensemble that approximates the prior distribution $P(H_{\text{SWE}})$.

To be able to compute the likelihood, the resulting SWE and the meteorological forcing are passed to a runoff generation model f_{runoff} :

$$Q_{\text{sim}}^{(i)} = f_{\text{runoff}}(H_{\text{SWE}}^{(i)}, M; \theta_{\text{meteo}}^{(i)}, \theta_{\text{runoff}}^{(i)}), \quad \text{with } \theta_{\text{runoff}}^{(i)} \sim P(\theta_{\text{runoff}}) \quad (3)$$

where Q_{sim} is the simulated streamflow and θ_{runoff} governs surface and subsurface runoff generation, soil storage, and evaporation. The model thus maps each parameter set $\Theta = \{\theta_{\text{meteo}}, \theta_{\text{snow}}, \theta_{\text{runoff}}\}$ to a streamflow simulation Q_{sim} , and the

120 inverse problem becomes one of estimating the posterior distribution:

$$P(\Theta | Q_{\text{obs}}) \propto P(Q_{\text{obs}} | \Theta) \cdot P(\Theta). \quad (4)$$

While it is difficult to compute this posterior distribution analytically, it can be approximated with numerical methods that generate samples of the posterior distribution, such as Importance Sampling (Nott et al., 2012) and Markov Chain Monte Carlo methods (Vrugt, 2016). These methods repeatedly sample parameter sets from their prior distributions, use them to run a simulation model, and evaluate their likelihood against observations. Parameters sets with a high likelihood have more chances of being considered as samples from the posterior (e.g., Vrugt (2016)).

Both formal and informal methods exist in hydrological parameter inference literature: formal methods use a well-defined likelihood function based on an assumed error distribution and combine this with the prior to obtain a well-defined posterior distribution (Kavetski et al., 2006; Renard et al., 2010). Informal methods do not necessitate a formal likelihood function and instead obtain a heuristic approximation of the posterior distribution using performance metrics as proxies for likelihood (Beven and Binley, 1992; Nott et al., 2012). We opt for an informal approach where we select a fixed percentage of the best-performing members among the prior ensemble as the heuristic posterior ensemble. This informal approach has the main advantage that the size of the posterior ensemble remains constant across experiments, which is helpful in assessing whether the posterior ensemble indeed contains the most realistic SWE realizations. Section 2.2.5 presents the sampling strategy, while Sect. 2.4 introduces the posterior ensemble selection and the performance metric used for streamflow evaluation.

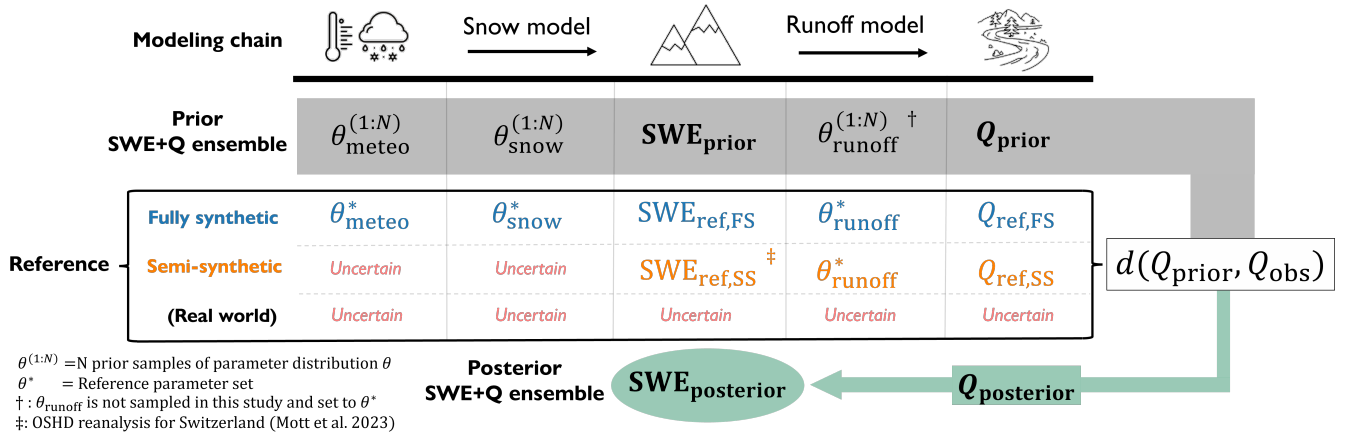


Figure 1. Schematic overview of the streamflow-constrained SWE reconstruction framework and the two synthetic numerical experiments. Q represents streamflow, θ represents the parameters to be sampled, θ^* represents the reference parameter set, and d represents the streamflow performance metric. Color-coding is consistent with the remainder of the study, with grey denoting the prior ensemble, green the posterior ensemble, blue the fully synthetic experiment (FS), and orange the semi-synthetic experiment (SS). See Sect. 2.2 for a detailed explanation of the workflow.

135

2.2 Implementation

Figure 1 illustrates the streamflow-constrained SWE reconstruction framework implemented in this study. Meteorological forcing M (Sect. 2.2.4) is used to drive a snow model f_{snow} (Sect. 2.2.1), producing gridded SWE and snowmelt estimates. These are combined with rainfall inputs and routed through a runoff model f_{runoff} (Sect. 2.2.2) to generate simulated streamflow Q .

140 For each year between 2001 and 2022, 5000 model realizations are generated by randomly sampling parameter sets from uniform prior distributions $\Theta = \{\theta_{\text{meteo}}, \theta_{\text{snow}}, \theta_{\text{soil}}\}$ using Latin Hypercube Sampling (Sect. 2.2.5). The resulting prior ensemble of simulated streamflow Q_{sim} is compared to observed streamflow Q_{obs} using a performance metric $d(Q_{\text{sim}}, Q_{\text{obs}})$, and the 1% best-performing members are selected as the heuristic posterior ensemble (Sect. 2.4).

To test the methodology in a controlled environment, we evaluate it in two synthetic experiments: a fully synthetic case (FS; 145 Sect. 2.3), which eliminates all modeling chain uncertainty, and a semi-synthetic case (SS), which adds meteorological and snow model structural uncertainty. The lower panel of Fig. 1 outlines the anticipated challenges for real-world applications, where additional uncertainty sources, particularly in the runoff model and streamflow observations, further complicate the inversion (Sect. 4.2).

2.2.1 Snow model

150 We use an enhanced temperature-index snow model that includes both air temperature and potential clear-sky radiation as melt drivers (Hock, 1999; Argentin et al., 2025). The model is implemented within the hydrological model wflow_sbm (van Verseveld et al., 2024, Sect. 2.2.2). Precipitation is partitioned into rainfall and snowfall using a melt temperature threshold T_{thresh} and a transition range as follows:

$$P_{\text{snow}} = \begin{cases} P & \text{if } T_{\text{a}} \leq T_{\text{thresh}} - 1 \\ P \cdot \left(\frac{(T_{\text{thresh}} + 1) - T_{\text{a}}}{2} \right) & \text{if } T_{\text{thresh}} - 1 < T_{\text{a}} < T_{\text{thresh}} + 1 \\ 0 & \text{if } T_{\text{a}} \geq T_{\text{thresh}} + 1 \end{cases} \quad (5)$$

155 where P is precipitation and T_{a} represents air temperature in $^{\circ}\text{C}$. We account for elevation-dependent biases in T_{thresh} using a linear lapse-rate correction:

$$T_{\text{thresh}}(x, y) = T_{\text{thresh}} + \frac{\gamma_{T_{\text{thresh}}}}{1000} \cdot (z(x, y) - \bar{z}), \quad (6)$$

where (x, y) denote grid-cell coordinates, $\gamma_{T_{\text{thresh}}}$ is the temperature threshold lapse rate ($^{\circ}\text{C km}^{-1}$), $z(x, y)$ is grid-cell elevation, and \bar{z} is the catchment-mean elevation. Snowfall biases are corrected using a spatially uniform multiplicative correction

160 factor c_{snow} combined with an elevation-dependent modulation γ_{snow} :

$$c_{\text{snow}}(x, y) = c_{\text{snow}} \cdot (1 + \tilde{z}(x, y) (\gamma_{\text{snow}} - 1)), \quad (7)$$

where $\tilde{z}(x, y)$ is a dimensionless elevation coordinate defined as

$$\tilde{z}(x, y) = \frac{z(x, y) - \bar{z}}{z_{\text{max}} - \bar{z}} \quad (8)$$

This formulation ensures non-negativity and equal but opposite adjustments of $c_{\text{snow}}(x, y)$ above and below the catchment mean elevation. Note that a linear precipitation lapse rate does not account for potential capping of high elevation precipitation due to moisture depletion (Napoli et al., 2019). Liquid precipitation is calculated as $P - P_{\text{snow}}$ and corrected separately using a spatially uniform rainfall correction factor c_{rain} (Sect. 2.2.4). The use of separate rainfall and snowfall correction factors is shown to improve meteorological bias correction in snow-dominated catchments (Pulka et al., 2024), and additionally allows us to assess whether the rainfall–snowmelt partitioning can be inferred from streamflow observations. The parameters c_{snow} , γ_{snow} , T_{thresh} , and $\gamma_{T_{\text{thresh}}}$ belong to the meteorological parameter set θ_{meteo} used to generate the prior SWE ensemble (Eq. 2 and Table 1).

Following the melt model introduced by Hock (1999), melt occurs when air temperature exceeds T_{thresh} , following:

$$F_M(t) = \begin{cases} (m + \alpha_{\text{rad}} * I_{\text{pot}})(T_a(t) - T_{\text{thresh}}) & \text{if } T_a(t) > T_{\text{thresh}} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where F_M is the melt rate (mm d^{-1}), m is the melt factor ($\text{mm d}^{-1} \text{ } ^\circ\text{C}^{-1}$), α_{rad} is the radiation factor for snow or ice ($\text{mm d}^{-1} \text{ } ^\circ\text{C}^{-1} \text{ m}^2 \text{ W}^{-1}$) and I_{pot} is the potential clear-sky direct solar radiation (W m^{-2}). We calculate I_{pot} for each grid cell based on the formula by Hock (1999) using the HydroBricks Python package (Horton and Argentin, 2024).

Meltwater is retained within the snowpack until it exceeds a calibratable water holding capacity (C_{ret}) fraction of the total snow mass (default: 0.1), after which drainage occurs. Liquid water may refreeze within the snowpack when $T_a < T_{\text{thresh}}$. Snow density evolution and rain-on-snow thermodynamics are not represented.

To represent sub-grid variability in snow depletion, we apply a fractional snow-covered area (f_{SCA}) parameterization based on Essery and Pomeroy (2004) and Magnusson et al. (2014):

$$f_{\text{SCA}}(t) = \tanh \left(1.26 \cdot \frac{H_{\text{SWE}_{\text{sim}}}(t)}{\beta_{\text{cv}} \cdot H_{\text{SWE}_{\text{max}}}} \right), \quad (10)$$

where $H_{\text{SWE}_{\text{sim}}}$ is the simulated average SWE in the grid cell at time t , β_{cv} is the coefficient of variation, and $H_{\text{SWE}_{\text{max}}}$ is the pre-melt seasonal maximum SWE.

To account for snow redistribution by gravity, we implement a mass wasting scheme adapted from Frey and Holzmann (2015). Snow is redistributed to downhill cells as follows:

$$F_{\text{MW}} = \begin{cases} c_{\text{MW}} \min(0.5, \frac{\gamma}{\alpha}) \min\left(1.0, \frac{H_{\text{SWE}}}{\beta}\right) & \text{if } H_{\text{SWE}} > 500 \wedge \rho_{\text{wet}} < 0.001 \wedge \gamma > 0.3, \\ 0 & \text{otherwise} \end{cases}, \quad (11)$$

where F_{MW} (mm d^{-1}) is the mass wasting flux per grid cell, c_{MW} (d^{-1}) is a correction factor set to 0.5, γ (-) is the slope, α (-) and β (-) are precalibrated factors set to $\alpha = 5.67$ and $\beta = 10000$. ρ_{wet} (-) is the wet snow to dry snow ratio, which is a variable that our model outputs at every time step. m , α_{rad} , C_{ret} , and β_{cv} are retained as snow model parameters θ_{snow} used to generate the prior SWE ensemble (Eq. 2 and Table 1).

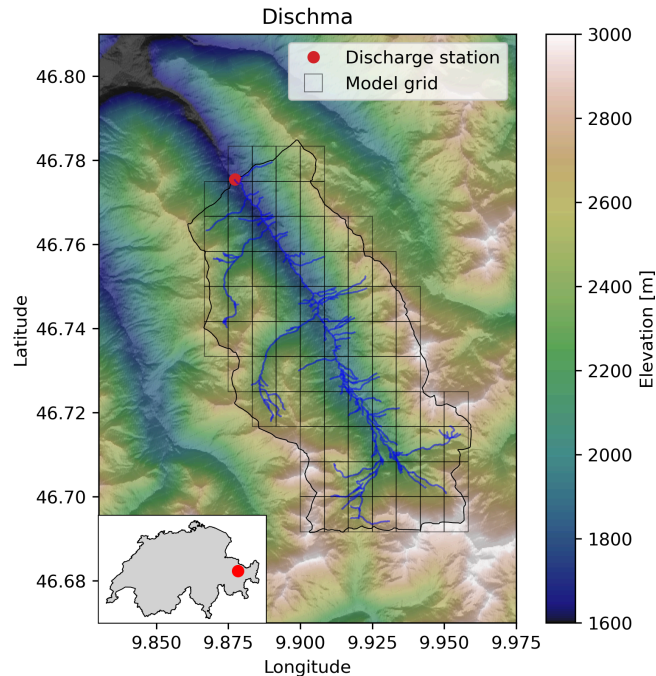


Figure 2. Digital elevation model and delineation of the Dischma catchment, along with its location within Switzerland. The regular model grid has a resolution of 30x30 arcseconds.

2.2.2 Runoff model

wflow_sbm (v0.7.1; van Verseveld et al., 2024) is an open-source, medium-complexity distributed hydrological model. While we adapted the wflow_sbm snow model (Sect. 2.2.1), we kept the runoff model intact. Each grid cell contains a vertically
 195 stratified soil column with up to four unsaturated layers and one saturated layer, allowing for dynamic water table movement. Soil hydraulic properties are inferred from global soil texture maps using pedotransfer functions (Imhoff et al., 2020).

For channel, overland, and lateral subsurface flow, the model uses the kinematic wave approach (van Verseveld et al., 2024). wflow_sbm uses globally available soil, vegetation, and terrain datasets, which are preprocessed using HydroMT (Eilander et al., 2023) (Table A1), and operates on a regular grid set to 30 arcsecond resolution (approximately 900 m \times 700 m at 40°
 200 latitude). We run wflow_sbm through the eWaterCycle hydrological modeling platform (Hut et al., 2022).

2.2.3 Test case: the Dischma catchment

The Dischma catchment in Switzerland spans 42.9 km² with elevations ranging from 1595 to 3180 m a.s.l. (mean: 2372 m) (Fig. 2). The catchment is predominantly alpine, with minimal forest cover (\sim 3%) and limited glacier extent ($<$ 1%). Beside cattle grazing, anthropogenic disturbances are negligible. Precipitation is fairly evenly distributed throughout the year, with roughly
 205 half falling as snow. Average annual discharge is 1229 mm/yr. The catchment has featured in numerous snow hydrological

studies (Berghuijs et al., 2025; Brauchli et al., 2017; Comola et al., 2015; Schaeffli, 2016), is actively monitored by the Swiss Federal Institute for Forest, Snow and Landscape Research (SLF; Magnusson et al., 2024), and is part of the CAMELS-CH dataset (Höge et al., 2023).

2.2.4 Meteorological forcing

210 Meteorological forcing data are obtained from MeteoSwiss and consist of gridded daily temperature (TabsD) and precipitation (RhiresD) estimates at $2 \text{ km} \times 2 \text{ km}$ spatial resolution (MeteoSwiss, Federal Office of Meteorology and Climatology (2024), version 2.0). Both are based on station observations and use interpolation methods that account for topographic effects. The RhiresD dataset is known to suffer from gauge undercatch inherited from the station data (Magnusson et al., 2014). For the Dischma catchment, mean RhiresD precipitation across all grid cells is 1029 mm/yr (1998–2022), which is in contradiction
215 with observed streamflow of 1229 mm/yr over the same period.

220 TabsD and RhiresD were downscaled to the 30 arcsecond model grid using area-weighted regridding with ESMValTool (Eyring et al., 2020). TabsD was first adjusted to sea level using a fixed lapse rate of $6.5 \text{ }^\circ\text{C km}^{-1}$ before regridding and then reprojected back to the original terrain elevation, while precipitation was regridded directly. For both coarse and fine resolution DEM we used the MERIT digital elevation model (Yamazaki et al., 2017). Potential and actual evapotranspiration were estimated using the semi-empirical method of de Bruin et al. (2016), which relies on shortwave radiation and near-surface air temperature.

Table 1. Overview of meteorological and snow model parameters used in the synthetic experiments. For details on the synthetic true parameter values, see Sect. 2.3.

Parameter	Class	Description	Unit	Prior range	Synthetic true value θ^*
c_{snow}	θ_{meteo}	Snowfall correction factor	–	0.9–1.5	1.0–1.4
γ_{snow}	θ_{meteo}	Elevation scaling of snowfall	–	0.7–1.3	1.0
c_{rain}	θ_{meteo}	Rainfall correction factor	–	0.7–1.3	1.0
T_{thresh}	θ_{meteo}	Melt temperature threshold	$^\circ\text{C}$	–1 to 1	0.0
$\gamma_{T_{\text{thresh}}}$	θ_{meteo}	Elevation scaling of T_{thresh}	$^\circ\text{C km}^{-1}$	–2 to 2	0.0
m	θ_{snow}	Degree-day melt factor	$\text{mm } ^\circ\text{C}^{-1} \text{ d}^{-1}$	1–4	2.5
α_{rad}	θ_{snow}	Radiation multiplier	$\text{mm W}^{-1} \text{ m}^2 \text{ } ^\circ\text{C}^{-1} \text{ d}^{-1}$	0.005–0.04	0.025
C_{ret}	θ_{snow}	Snowpack water holding capacity	–	0.1–0.4	0.25
β_{cv}	θ_{snow}	Snow cover depletion curve shape	–	0.1–0.5	0.3

2.2.5 Sampling strategy

As defined in Sect. 2.1, each parameter set Θ consists of meteorological parameters (θ_{meteo}), snow model parameters (θ_{snow}), and runoff model parameters (θ_{runoff}) (Eq. 2 and 3). We restrict our analysis to synthetic experiments with complete knowledge

225 of the runoff model structure and parameters. Consequently, θ_{runoff} is not subject to calibration and is fixed at default values as defined in the wflow_sbm documentation (Imhoff et al., 2020; van Verseveld et al., 2024) (Table B1). To generate the prior SWE and streamflow ensemble, we thus only sample from meteorological and snow model parameters θ_{meteo} and θ_{snow} (Table 1). Note that this approach is unsuitable when including θ_{runoff} , whose values likely vary little between years. A two-step sampling is then more suited, separating constant and annually varying parameters (Henn et al., 2015).

230 For each year between 2001 and 2022, 5000 parameter combinations are sampled from the joint prior parameter distributions of the 9 retained parameters using Latin Hypercube Sampling (LHS) (McKay et al., 2000), implemented through the SPOTPY Python package (Houska et al., 2015). While 5000 samples do not densely populate the prior parameter space, it is considered adequate for this study, as increasing the number of samples did not alter the results. We do not use an actual optimization algorithm or Markov Chain Monte Carlo sampling since the objective of our study is to explore the information content of streamflow for SWE inference by efficiently exploring the full parameter space rather than identifying the posterior distribution.

2.3 Synthetic numerical experiment design

To evaluate the constraining potential of streamflow for SWE reconstruction, we perform two synthetic experiments, both use the same prior ensemble of 5000 parameters described above.

2.3.1 Experiment 1: Fully synthetic (FS)

240 The first is an “inverse crime” experiment (Wirgin, 2004): we generate synthetic SWE and streamflow using the same snow and runoff model structures as those used for inversion, ensuring consistency between forward and inverse models. In doing so, we aim to quantify the theoretical potential of streamflow-constrained SWE inversion by eliminating any model structural error or observation uncertainty. The synthetic true parameters θ^* used to generate synthetic SWE ($\text{SWE}_{\text{ref,FS}}$) are given in Table 1. The snowfall correction factor oscillates over all years between 1 and 1.4, with annual changes of 0.1. This mimics the full potential extent of seasonal meteorological forcing bias. For the remaining parameters, θ^* is set to the midpoint between the lower and upper prior bounds. Because LHS ensures uniform coverage of each parameter’s range, the median of the sampled parameter set Θ will approximate θ^* . Consequently, the ensemble mean of the resulting prior SWE simulations $\text{SWE}_{\text{prior}}$ is expected to roughly approximate the reference simulation $\text{SWE}_{\text{ref,FS}}$.

2.3.2 Experiment 2: Semi-synthetic (SS)

250 The second experiment is a semi-synthetic experiment, where we use the Swiss temperature-index SWE reanalysis product OSHD (Mott et al., 2023; Mott, 2023) as the synthetic SWE reference $\text{SWE}_{\text{ref,SS}}$. This product combines a temperature-index snow model with data assimilation of in-situ SD observations for both snowfall and SWE state correction. It is available for all of Switzerland since 1998 at 1 km resolution. Although the underlying meteorological forcing is comparable to that used in this study, the combination of an alternative model structure and assimilation-induced SWE corrections introduces both snow model and meteorological deviations relative to the base snow model and forcing. This introduces artificial snow-related uncertainty

in the inversion, thereby making it closer to real-world conditions (Fig. 1). The semi-synthetic experiment thus allows us to examine the degradation in inversion performance when realistic discrepancies exist between the “true” and assumed snow processes. To establish the coupling between OSHD and wflow_sbm, the OSHD output is first resampled to the wflow_sbm grid and then inserted in the wflow_sbm model by modifying the meteorological forcing: all snowfall events (i.e., P when $T_{\text{air}} < 0^{\circ}\text{C}$) are removed, OSHD-derived snowmelt is added as precipitation, and air temperature is capped at a minimum of 0°C to ensure this precipitation falls as rain. Testing of this coupling approach showed that the secondary effects of capping the temperature at 0°C are negligible. The runoff model and streamflow observations remain without uncertainty, isolating the impact of snow-related uncertainties. Rainfall correction factor c_{rain} is not dictated by OSHD and is still inferred, with the true c_{rain} (c_{rain}^*) set to 1. The rainfall correction is applied only to the RhiresD forcing, not to the OSHD-derived snowmelt implemented as rainfall.

2.4 Posterior ensemble selection

We use the Nash–Sutcliffe Efficiency (NSE, Nash and Sutcliffe, 1970) as the streamflow performance metric to quantify agreement between model output and observations (denoted as $E_{Q-\text{NSE}}$). An NSE of 1 indicates perfect agreement, while an NSE of 0 implies no improvement over using the observed mean as a predictor. We calculate NSE over the snowmelt season (March to July) to focus on snowmelt-driven discharge. Although NSE can give inflated values in catchments with strong seasonality, such as the Dischma (Schaeffli and Gupta, 2007), our focus is on relative differences in NSE, reflecting variations in squared error magnitudes.

We adopt a rank-based heuristic posterior selection. All prior ensemble members are evaluated against observed streamflow using NSE, and the top 1% are selected as the posterior ensemble, yielding a posterior size of $N_{\text{posterior}} = 50$. The quality of this posterior ensemble is then evaluated on different SWE metrics (Sect. 2.5).

2.5 Posterior SWE evaluation

2.5.1 SWE metrics and scales

We evaluate SWE reconstructions using a set of performance metrics that target different physical properties of the seasonal snowpack. We follow the concept of the “snow triangle” metrics from Trujillo and Molotch (2014) and Rhoades et al. (2018), with modifications. Unlike Rhoades et al. (2018), who reduce snowfall and melt to seasonal means, we use the daily time series of snowfall and melt to better evaluate temporal dynamics and individual events. For snow accumulation, we use the sum of seasonal snow accumulation rather than peak SWE volume, to reflect the total snow contribution to the catchment water balance. We include the dates of SWE onset and melt-out but omit other timing metrics, such as date of peak SWE and melt season length, as their information is assumed to be embedded in the remaining metrics. Each performance metric E is computed annually at two spatial scales:

- Catchment-aggregated (AGG): E_{AGG} metrics are calculated from the spatially averaged SWE time series across the catchment.

- Distributed (GRID): E_{GRID} metrics are computed per grid cell and averaged over space.

This allows assessment of whether streamflow informs the spatial structure or only the integrated behavior of the snowpack.
 290 Such multi-scale evaluation is enabled by full spatio-temporal availability of the reference SWE.

Each performance metric matches the nature of the evaluated variable (Table 2). For the evaluation of time series (melt and snowfall), we use the NSE (Sect. 2.4) in AGG mode, and the grid-mean NSE in GRID mode. For total accumulation, we use Absolute Percentage Error (APE) in AGG mode and Mean Absolute Percentage Error (MAPE) in GRID mode. For timing metrics evaluating SWE onset and melt-out dates, we express the dates in day-of-year and use Absolute Error (AE) in AGG
 295 mode and Mean Absolute Error (MAE) in GRID mode.

Table 2. Overview of SWE performance metrics used to evaluate the streamflow-derived posterior SWE ensemble. Error types are given for catchment-aggregated (AGG) and distributed (GRID) modes.

Metric	Description	Performance metric (AGG/GRID)
E^{melt}	NSE of daily snowmelt time series ($-d\text{SWE}/dt$)	NSE / grid-mean NSE
E^{snowfall}	NSE of daily snowfall time series ($+d\text{SWE}/dt$)	NSE / grid-mean NSE
E^{onset}	First day SWE exceeds 10% of seasonal max	AE / MAE
$E^{\text{melt-out}}$	First day SWE drops below 10% of seasonal max	AE / MAE
$E^{\text{accumulation}}$	Total water year snowfall	APE / MAPE

2.5.2 Posterior rank evaluation

To assess how well streamflow constrains SWE, we apply a rank-based diagnostic. All 5000 prior members are ranked on each performance metric. We then identify the ranks of the 50 posterior ensemble members in this list and compute their median rank, denoted $R_{\text{post,median}}$.

300 If streamflow perfectly selects the best SWE scenarios, we expect $R_{\text{post,median}} = 25$, corresponding to the median of 50 samples (rounded down from 25.5). Conversely, if streamflow offers no useful constraint, posterior members will be randomly distributed throughout the prior, and $R_{\text{post,median}} = 2500$, corresponding to the median rank among 5000 samples (rounded down from 2500.5). A median rank significantly higher than 2500 would suggest streamflow-based selection degrades performance for that metric. Note that this rank-based summary neglects the distribution shape of posterior ranks, focusing solely on
 305 the median.

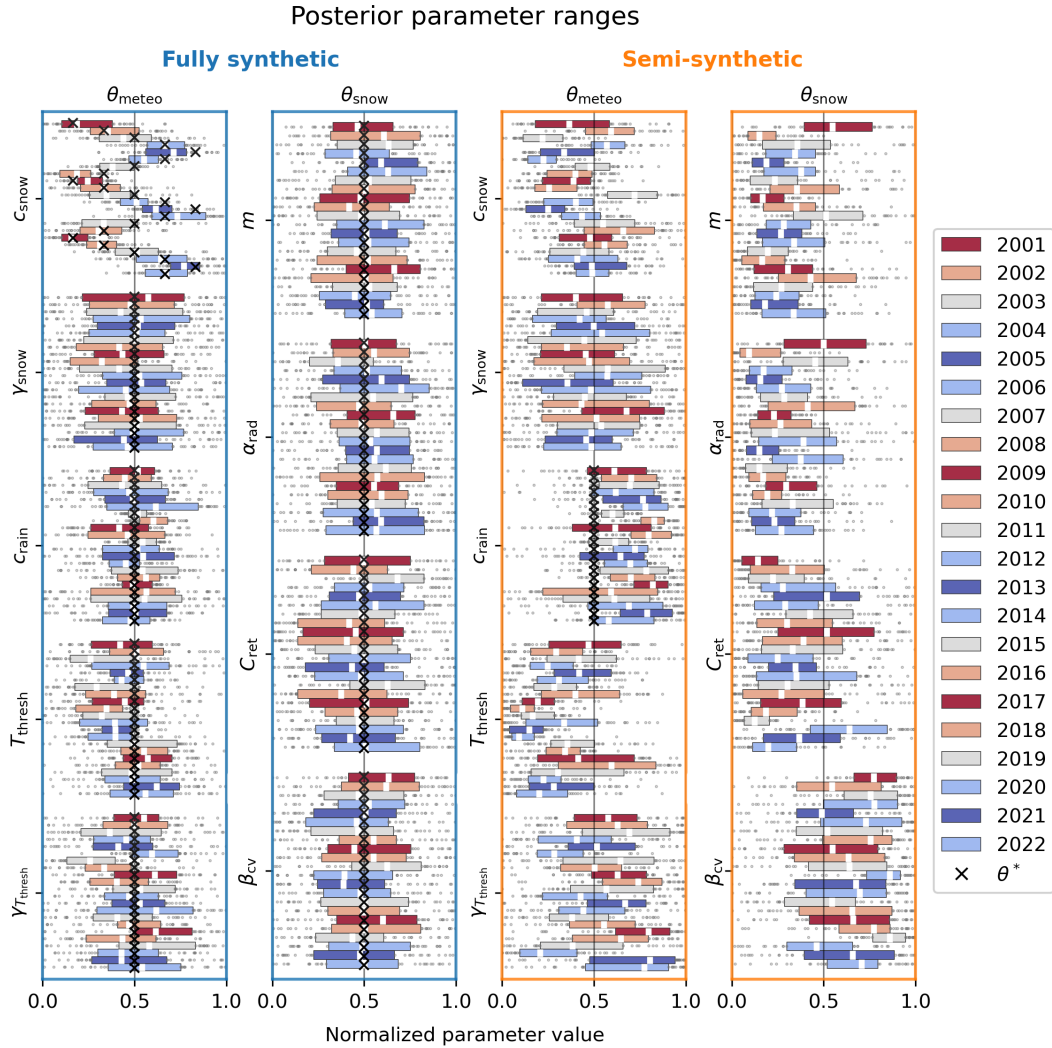


Figure 3. Annual posterior parameter ensembles for the FS and SS experiments, expressed relative to the normalized prior range. θ_{meteo} and θ_{snow} represent the meteorological and snow model parameters. Medians (white squares), interquartile ranges (boxes), and lower- and upper-quartile values (grey dots) are shown for the 50 posterior parameter values. The true parameter values used to generate the synthetic observations (θ^*) are represented by black crosses. The color-coding is based on the annually fluctuating values of c_{snow}^* in FS.

3 Results

3.1 Posterior parameter ensembles

The inferred posterior parameter ensembles do not consistently align with the true parameter values in the FS experiment (Fig. 3a & b). Among all parameters, c_{snow} is the most sensitive. Its annual posterior values generally reflect the imposed

310 artificial bias fluctuations c_{snow}^* , but still span a wide range. For example, in 2001 where $c_{\text{snow}}^* = 1$, posterior values range from approximately 0.9 to 1.3, implying that both a 10% underestimation and a 30% overestimation of total snowfall can result in high streamflow skill. The remaining parameters are considerably less sensitive and generally span most of their prior ranges, equally indicating that a wide range of SWE realizations can produce similarly high-performing streamflow responses and suggesting compensating behavior among parameters.

315 In the SS experiment, of all true parameter values Θ^* only c_{rain}^* is imposed. The true value of the remaining parameters is unknown, as the reference SWE is an external product. Figure 3c shows that c_{rain} is consistently overestimated, which is compensated by an annual underestimation of catchment-wide SWE accumulation of $6.6 \pm 4.4\%$. The posterior c_{snow} values vary widely across the prior range, suggesting annually varying biases in the snowfall forcing and confirming the need for annual rather than multiannual inversion. The values of T_{thresh} , m , r and C_{ret} are generally on the lower edge of the prior
320 range, while $\gamma_{T_{\text{thresh}}}$ and β_{cv} are generally on the higher edge. This suggests slower melt taking place preferentially at higher elevations, lower water holding capacity, and faster snow cover depletion of $\text{SWE}_{\text{ref,SS}}$ (i.e. OSHD-TI product) compared to our prior assumptions.

3.2 Streamflow and SWE performance

In the previous section, we showed that we cannot reliably recover the true parameter values from streamflow alone, with NSE
325 as the streamflow performance metric. To better understand this result, we analyze the model performances associated with the best ranked parameter sets. Figure 4 shows the $E_{\text{Q-NSE}}$ results and posterior ensemble selection (Fig. 4a and 4c) and the subsequent evaluation of this selection on $E_{\text{GRID}}^{\text{accumulation}}$ (Fig. 4b & d), for both FS and SS. We show $E_{\text{GRID}}^{\text{accumulation}}$ results as it is arguably the most relevant performance metric for gridded SWE reconstructions. The results for other target SWE metrics are presented in Figs. S1-10.

330 The $E_{\text{Q-NSE}}$ results confirm strong agreement between simulated and synthetic streamflow in both experiments, with an overall mean posterior NSE of 0.99 ± 0.01 for FS, and 0.94 ± 0.03 for SS (Fig. 4a & c), compared to an overall mean prior NSE of 0.67 ± 0.11 for FS and 0.56 ± 0.16 for SS. $E_{\text{GRID}}^{\text{accumulation}}$ results reach maximum scores of near 0% in FS in some years, while not exceeding 10% in most other FS and SS years. This is likely due to a combination of high sensitivity of the highest and lowest elevation grid cells to the γ_{snow} and $\gamma_{T_{\text{thresh}}}$ parameters, and the fact that in the SS experiment a perfect
335 approximation of the external SWE product is generally not possible.

Across both experiments, posterior members generally occupy the lower-error portion of the prior distribution, indicating that streamflow provides a meaningful constraint on gridded SWE accumulation. However, many prior members outperform the posterior ensemble, demonstrating that high streamflow skill does not uniquely translate into high spatial SWE accumulation skill. The strength of this constraint varies between years, with some years showing a narrow spread and others showing a wide
340 spread among the posterior ensemble.

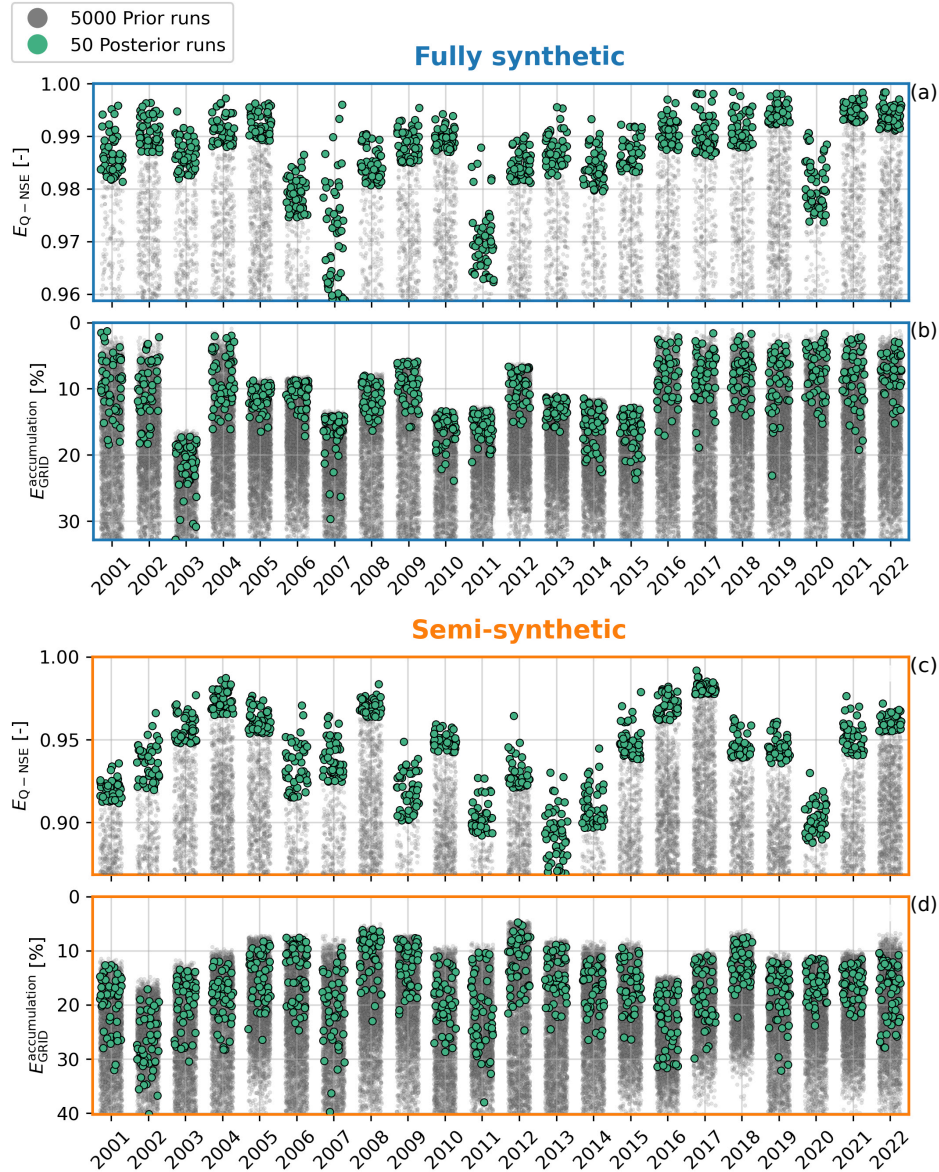


Figure 4. NSE-based posterior selection (a, c) and grid-mean total snowfall mismatch ($E_{GRID}^{accumulation}$) of all model runs (b, d), for both FS (a, b) and SS (c, d) experiments. Grey points represent the 5000 annual prior members (above the y-axis cutoff), while green points represent the posterior ensemble, i.e. the 50 members with the best streamflow performance.

3.3 Posterior rank evaluation across SWE metrics

Across both experiments, E_{AGG}^{melt} is the most strongly constrained metric among all SWE metrics, indicating that streamflow most effectively constrains catchment-scale melt dynamics. In FS, $R_{post,median}$ values lie close to the perfect-constraint limit in

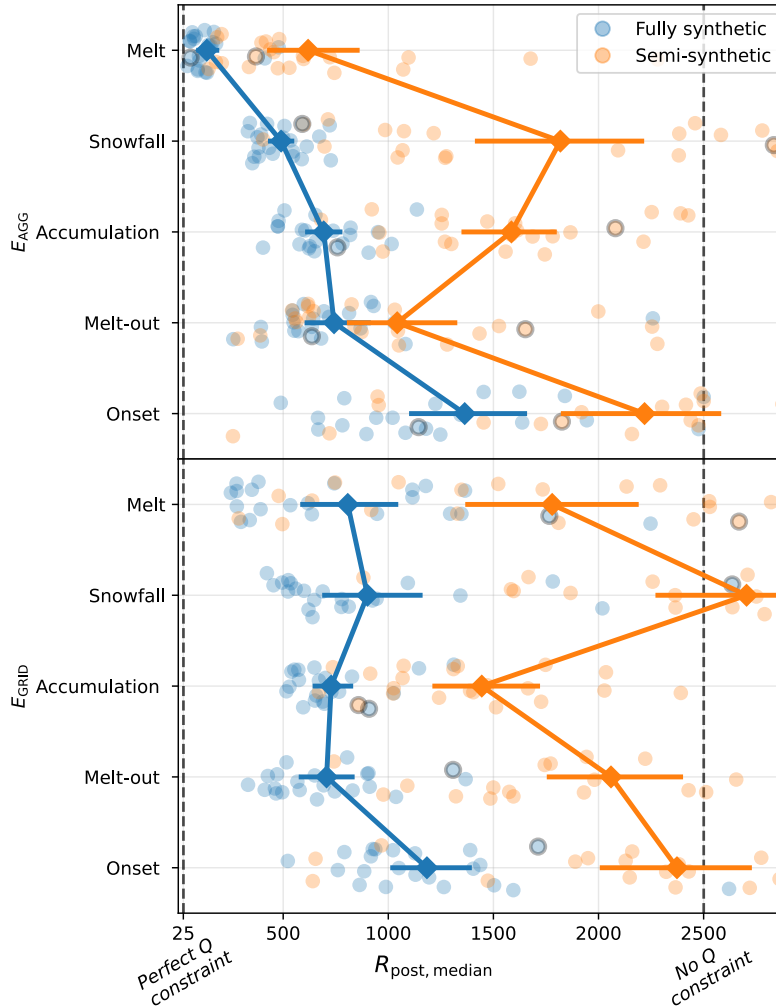


Figure 5. Annual median SWE metric ranks of the streamflow-derived posterior ensembles, relative to all 5000 prior members. The top figure shows the results for catchment-aggregated SWE metrics (E_{AGG}), while the bottom shows grid-averaged metrics (E_{GRID}), sorted based on the FS E_{AGG} ranks. Each point represents the annual median posterior rank between 2001–2022, with the year 2003 in thick outline as an example. The diamonds represent the mean of all median posterior ranks, and the error bars represent the 95% confidence interval. The fully and semi-synthetic experiments are represented in blue and orange, respectively. The definitions of the error metrics are given in Sect. 2.5.1.

most years, whereas in SS they are substantially higher and more variable, reflecting reduced identifiability under added forcing
 345 and model uncertainty. In contrast, E_{GRID}^{melt} is only weakly constrained in both experiments, indicating that while streamflow
 constrains integrated melt production, it provides limited information on its spatial origin.

Snowfall-related metrics are more weakly and inconsistently constrained. In FS, both $E_{AGG}^{snowfall}$ and $E_{GRID}^{snowfall}$ show moderate
 constraint, likely benefiting from the same parameter sets that favor melt performance. In SS however, both metrics are weakly

constrained, confirming the limited ability of streamflow to inform snowfall dynamics when accumulation and melt biases differ.

Accumulation metrics show intermediate constraint. $E_{\text{GRID}}^{\text{accumulation}}$ is more strongly constrained than $E_{\text{AGG}}^{\text{accumulation}}$ in both experiments, suggesting that the spatial distribution of snow accumulation is equally or better constrained by streamflow than the total catchment-wide accumulation. Note, however, that a different streamflow performance metric than NSE (e.g. seasonal streamflow bias) might favor the constraint of catchment-aggregated accumulation more (Sect. 4.3).

Among timing metrics, melt-out dates are relatively well constrained, particularly in AGG mode. This is consistent with their physical link to the cessation of snowmelt-driven streamflow. In contrast, SWE onset dates are weakly constrained across experiments and scales.

Overall, constraints are systematically weaker and more heterogeneous in SS than in FS. $R_{\text{post,median}}$ increases by 989 on average across all metrics, corresponding to 20% of N_{prior} , while its standard deviation increases on average by 414. This confirms that the added structural and input uncertainty in SS reduce the ability of streamflow to constrain SWE. Additionally, the increased spread in $R_{\text{post,median}}$ across different SWE performance metrics in SS indicates that members performing well on one metric no longer consistently perform well on others. This suggests a decoupling of performance among metrics and growing trade-offs between competing aspects of SWE reconstruction under added uncertainty. Nonetheless, except for $E_{\text{GRID}}^{\text{snowfall}}$ in SS, most median ranks remain above the no-constraint threshold.

3.4 Correlation among metrics

To assess the complementarity of the retained SWE metrics, Fig 6 shows the Spearman rank correlation between all SWE performance metrics and streamflow NSE across the full prior ensemble for each year. Consistent with previous results, $E_{\text{AGG}}^{\text{melt}}$ is the SWE metric most strongly associated with $E_{\text{Q-NSE}}$ in both experiments ($\rho_{\text{FS}} = 0.89$, $\rho_{\text{SS}} = 0.63$). Overall, correlations are systematically stronger in FS than in SS, and catchment-aggregated metrics generally correlate well with their gridded counterparts. While $E_{\text{AGG}}^{\text{melt}}$ correlates the strongest with $E_{\text{Q-NSE}}$, correlation of other SWE metrics with catchment-aggregated melt does not translate to strong correlation with $E_{\text{Q-NSE}}$ as well. It is the case for $E^{\text{melt-out}}$, but not for $E_{\text{GRID}}^{\text{melt}}$. These results suggest that, in general, streamflow contains information about different SWE metrics independently. In other words, the posterior members performing well on one SWE metric do not necessarily perform well on others. This supports the use of multiple, complementary SWE metrics and cautions against drawing conclusions from any single metric.

4 Discussion

4.1 Streamflow constraining potential under idealized conditions

The fully synthetic experiment confirms that streamflow-constrained SWE inversion works in theory, but near-perfect constraint is only achieved for catchment-aggregated melt. Other SWE properties generally remain well-constrained, but far from perfectly constrained. We hereby demonstrate that even under highly idealized conditions, streamflow does not consistently

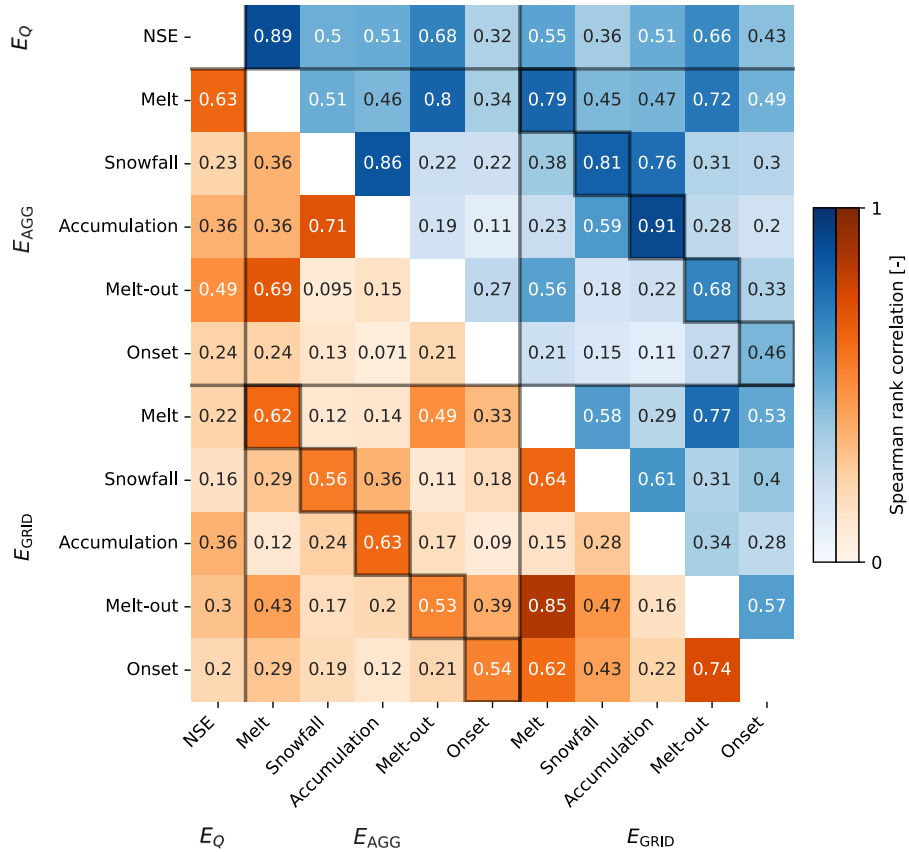


Figure 6. Correlations among all retained streamflow and SWE performance metrics, expressed as the median of annual Spearman rank correlations over all 5000 yearly prior members between 2001 and 2022. The upper values in blue represent the fully synthetic experiment, while the lower values in orange represent the semi-synthetic experiment. Black lines delineate streamflow (Q), catchment-aggregated SWE (E_{AGG}), and spatially distributed SWE (E_{GRID}) performance metrics. Black squares emphasize E_{AGG} - E_{GRID} diagonals.

380 identify the best-performing SWE scenarios across all performance metrics. This finding is mainly explained by physical non-uniqueness in the SWE-streamflow relationship (Beaton et al., 2024): different SWE and rainfall scenarios can lead to equivalent streamflow responses. Catchment-aggregated melt being better constrained than distributed melt is a first indication of this, by showing that biased spatial melt distributions can lead to an accurate aggregated melt output. A second indication is given by the imperfect constraint on catchment-aggregated accumulation: the best-performing streamflow performance can be

385 achieved with biased catchment-wide SWE accumulation estimates. Finally, we show that multiple distinct SWE and rainfall combinations can yield similar streamflow responses (Fig. 3 and Sect. 3.1).

A second source of uncertainty in the FS experiment is structural non-uniqueness or equifinality (Beven and Freer, 2001; Günther et al., 2020), whereby multiple parameter sets yield similar SWE outcomes. However, since this study focuses on SWE performance rather than parameter convergence, such equifinality is not of major concern. A third potential source is

390 parameter estimation uncertainty, i.e., the failure to identify optimal parameter combinations by the sampling algorithm. Yet this is also of minor importance, as the posterior simulations already achieve high streamflow skill, and further optimization or a different $N_{\text{posterior}}/N_{\text{prior}}$ ratio would not affect the SWE-streamflow relationships central to our analysis.

The semi-synthetic experiment shows that adding meteorological and snow model uncertainty significantly reduces the ability of streamflow to constrain SWE across all performance metrics. This reduction suggests a mismatch between our
395 meteorological forcing and snow model versus the OSHD reference that the current inversion framework is unable to correct. Alongside physical non-uniqueness, uncertainties throughout the modeling chain thus present an additional barrier to accurately identifying realistic SWE scenarios from streamflow. These added uncertainties also introduce stronger trade-offs between SWE metrics: even when the best-performing ensemble members for one SWE metric are correctly identified, they are less likely to perform well on other SWE performance metrics.

400 **4.2 Additional challenges under real-world conditions**

Under real-world conditions, constraining SWE reconstructions using streamflow presents additional challenges beyond the idealized setup explored in this study. These challenges include both substantially increased uncertainty and reduced opportunities for performance evaluation (Fig. 1). One major source of uncertainty not addressed here is runoff model uncertainty. This encompasses both uncertainty in static catchment properties and uncertainty in the representation of water transport processes
405 through the catchment (Beven, 2006). Such uncertainty can introduce persistent timing biases in the translation of snowmelt into streamflow, thereby complicating efforts to infer SWE dynamics from streamflow observations (Henn et al., 2018). Additional uncertainty arises from streamflow observations, including errors in stage measurements, discharge gauging, rating curve estimation (Di Baldassarre and Montanari, 2009), and ice-related effects (Burrell et al., 2023), which can propagate into both event-scale timing errors and biases in seasonal water balance estimates. Finally, the meteorological and snow model un-
410 certainties imposed in the semi-synthetic experiment are likely to be conservative relative to real-world conditions. In practice, meteorological forcing errors are expected to be larger, not least due to the addition of evaporation estimation uncertainty, and snowpack dynamics are more heterogeneous and complex than represented by the OSHD model. Taken together, these additional sources of uncertainty are expected to further diminish the constraining potential of streamflow on SWE reconstruction beyond the reduction observed here between the fully synthetic and semi-synthetic experiments.

415 An additional challenge in real-world applications is the lack of long-term, temporally continuous and catchment-scale SWE observations against which to evaluate inversion results (Revuelto et al., 2025). Unlike in our synthetic experiments, real-world evaluations forcibly rely on spatially or temporally incomplete observations, making it inherently difficult to assess whether the SWE inversion was successful. At present, the best evaluation dataset is arguably the biweekly gridded SWE product of the Airborne Snow Observatory (Painter et al., 2016), available for a limited number of catchments and years in the Western
420 US. The lack of evaluation data equally implies a lack of training data for data-driven methods, thereby limiting the potential of machine learning methods as an alternative link between SWE and streamflow in inverse hydrological SWE reconstruction. A practical way forward is to continue refining idealized experiments by further adding controlled sources of uncertainty, such

as runoff model and streamflow observation errors, thereby approximating real-world complexity while retaining the ability to assess inversion effectiveness quantitatively.

425 **4.3 Potential inversion framework adaptations**

Several elements of the inversion framework proposed here may require adaptation under real-world conditions, where uncertainty is higher and evaluation opportunities are more limited. One key limitation is the use of NSE as the streamflow performance metric. NSE is highly sensitive to timing errors, potentially penalizing simulations that reproduce melt events with small temporal shifts more strongly than simulations that miss them entirely. This sensitivity is particularly relevant given
430 our finding that streamflow most strongly constrains catchment-aggregated melt, which is inherently timing-dependent. In addition, as a residual-based metric, NSE may preferentially favor parameter sets that perform well on residual-based SWE metrics, while disadvantaging those that perform better on bias-based SWE metrics. Alternative streamflow metrics targeting hydrological signatures, such as variability or seasonal volume (Schaeffli, 2016), may therefore provide more robust constraints under real-world uncertainty. Consequently, the results presented here should be considered strictly in light of the use of NSE
435 as the streamflow performance metric. A full analysis of the streamflow and SWE performance metric interactions is outside the scope of this work, but is recommended for future research.

Given the increased spatial and temporal variability of meteorological forcing in real-world applications, the simple bias-correction factors used here (θ_{meteo}) are likely insufficient. In reality, snowfall, rainfall and melt biases vary at sub-seasonal and event time scales, and elevation dependencies are often non-linear in complex terrain. Allowing for greater temporal flexibility
440 and spatial heterogeneity in the correction of meteorological biases may therefore improve the identifiability of relevant SWE processes from streamflow observations, albeit at increased computational cost.

Finally, the choice of snow and runoff models is likely to influence inversion performance. While relatively simple snow models can perform well at the catchment scale (Magnusson et al., 2015), more physically based formulations may be better suited where meteorological data are abundant (Mott et al., 2023). In contrast, the inversion could benefit from decreased
445 complexity in the runoff model. Since the primary function of the runoff model in this framework is to translate spatial melt into streamflow, semi-distributed or lumped models could reduce computational costs and allow for larger ensembles compared to the fully distributed runoff model used here. More broadly, employing multiple snow and runoff models within the inversion framework could enhance robustness by better accounting for structural model uncertainty and increasing the likelihood of capturing realistic SWE evolution and snowmelt runoff.

450 **4.4 Outlook on the added value of streamflow in SWE reconstructions**

Under real-world conditions, streamflow alone may fail to reliably distinguish biased from unbiased SWE simulations. In the absence of reliable SWE evaluation data, the streamflow-based selection of biased SWE simulations might even go undetected. This implies that streamflow may, in some cases, not provide added value compared to simply running a snow model with uncorrected meteorological forcing. Several factors determine whether streamflow is likely to provide added value for SWE
455 reconstruction. First, the quality of meteorological observations is crucial. Low meteorological biases result in low biases

in SWE reconstructions, reducing the need for streamflow to constrain them. Secondly, the size, shape, and climate of the target catchment play a role. Smaller, elongated catchments (e.g. the Dischma catchment of this study) exhibit lower non-uniqueness than large, round catchments (Rinaldo et al., 1995), while snow-dominated catchments offer better identifiability of snowmelt than snow-scarce catchments (Griessinger et al., 2016). Dry spring and summer climates particularly benefit streamflow-assisted SWE inversion as they limit the confounding between rainfall and snowmelt signals (Henn et al., 2015). The same logic likely also applies to inter-annual variability within each catchment. In years with higher snowfall fractions and less spring rainfall, streamflow likely has greater constraining potential on SWE reconstructions. The above factors favor the application of streamflow-assisted SWE inversion as far back as streamflow observations allow, as meteorological forcing products have become less biased (Kouki et al., 2023), and snowfall dominance has decreased with time (Han et al., 2024). They also favor its application to meteorologically under-observed mountain regions such as the Himalayas and the Andes, where forcing products equally tend to be more biased and SWE evaluation is scarcer (Beck et al., 2019; Thornton et al., 2021). While this study focuses on mountainous catchments, it would be valuable to assess the constraining potential of streamflow in lower-relief, less topographically complex environments such as boreal catchments, where different controls on snow accumulation and melt may lead to different identifiability of SWE dynamics in streamflow.

In this study, we isolate the constraining potential of streamflow alone. However, streamflow is likely most effective when used in combination with other sources of snow information. These can be direct observations of different snow properties (Revuelto et al., 2025), but we particularly encourage future work to explore the use of recurring spatial patterns of snow dynamics (Vögeli et al., 2016; Pflug and Lundquist, 2020; Geissler et al., 2025). Such patterns represent catchment-specific prior knowledge that, once characterized, can be reused across time, including periods predating satellite observations but overlapping streamflow observations. Prior SWE ensembles could be defined as scaled versions of present-day observed spatial SWE patterns, thereby reducing the number of parameters to be inferred and mitigating part of the physical non-uniqueness identified in this study. Alternatively, the recurring patterns could be used to filter posterior SWE members based on physical plausibility. So far, these recurring patterns have been characterized in a limited number of catchments, but recent advances in automated UAV and LiDAR technologies are likely to increase this number in the near future (Revuelto et al., 2021). Combined with long streamflow records, this opens the possibility of extending SWE reconstructions back in time by decades while maintaining realistic spatial SWE patterns.

Regardless of the accompanying information source, streamflow remains a unique source of snow information in its ability to capture catchment-integrated SWE dynamics, most notably the timing and total volume of snowmelt runoff. Our finding that streamflow most effectively constrains catchment-aggregated melt supports its potential role in this context. In light of results by Rhoades et al. (2018), who showed that many SWE products systematically misrepresent average melt rates in mountainous terrain, streamflow is the only observational source capable of directly constraining such errors at the catchment scale. We therefore propose that future studies investigate the integration of streamflow with other snow data sources to constrain SWE reconstructions as much as practically possible.

5 Conclusion

490 We presented a framework for streamflow-constrained SWE reconstruction at the catchment scale using inverse hydrological modeling. We tested the methodology in two synthetic numerical experiments and across five target SWE metrics calculated on both catchment-aggregated and spatially distributed scales. The fully synthetic experiment showed that, even in the absence of all modeling chain uncertainty, a range of different SWE realizations and snowmelt/rainfall combinations can lead to equivalent and very high-performing streamflow estimates. The semi-synthetic experiment showed that the addition of artificial meteorological and snow model uncertainty leads to a considerable reduction in the constraining potential of streamflow across all SWE properties. In both experiments, streamflow has the most constraining potential on catchment-aggregated melt, although this finding is conditional to the use of NSE as the streamflow performance metric. Overall, this study showed that even in synthetic experiments devoid of observation and runoff model uncertainty, the relationship between streamflow and SWE properties is complex and non-linear, and streamflow alone can only constrain SWE reconstructions to a limited degree. We therefore expect streamflow-constrained SWE reconstructions using the presented framework to be challenging in many real-world cases, when the issues of non-uniqueness and uncertainties across the modeling chain are further amplified. We suggest future studies to explore the combined use of streamflow with other sources of snow information, across diverse catchments and using a wider range of streamflow performance metrics.

Code and data availability. All code and supporting files used in this study are available at 10.5281/zenodo.16146617. The latest wflow_sbm code can be found at <https://zenodo.org/records/15722493>. The Python wrapper for the wflow_sbm Julia code as part of eWaterCycle can be found at <https://github.com/eWaterCycle/ewatercycle-wflowjl>. The version of wflow_sbm used in this study including our snow model adjustments can be found at <https://github.com/pauwiersma/Wflow.jl>.

Appendix A: HydroMT global datasets

Table A1. Global datasets used to setup wflow through the HydroMT package

Dataset	Function	Reference
Chelsa	Precipitation climatology	Karger et al. (2017)
Köppen-Geiger	Climate classification	Kottek et al. (2006)
MERIT DEM (90m)	Topography	Yamazaki et al. (2017)
MERIT Hydro	Hydrography	Yamazaki et al. (2019)
MODIS LAI	Leaf Area Index	Myeni et al. (2015)
Soilgrids (v2020)	Gridded soil classification	de Sousa et al. (2020)
Vito	Land use classification	Buchhorn and smets (2020)

Appendix B: wflow_sbm default parameters

Table B1. Key wflow_sbm parameters used in this study. All parameters are unitless. For the remaining parameter values, we refer to van Verseveld et al. (2024).

Parameter	Function	Default value
khfrac	Multiplication factor applied to vertical hydraulic conductivity to obtain horizontal hydraulic conductivity	100
f	Scaling parameter controlling the decline of vertical hydraulic conductivity with depth	1
kvfrac	Multiplication factor applied to the vertical hydraulic conductivity	5
C	Brooks-Corey power coefficient controlling soil water pressure for each soil layer	1

510 *Author contributions.* Conceptualization: PW, GM. Methodology: PW, GM. Formal analysis: PW. Methodology: PW, GM. Supervision: GM. Visualization: PW. Writing – original draft preparation: PW. Writing – review and editing: PW, GM, JM, NP, BS.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Hydrology and Earth System Sciences.

Acknowledgements. We thank Joschka Geissler and Simon Gascoin for their constructive and helpful reviews. In particular, we acknowledge the insightful comments of Joschka Geissler on the use of present-day spatial SWE patterns in historical streamflow-informed SWE reconstruction (Geissler, 2026). We thank MeteoSwiss for providing open access to the meteorological datasets. We are also grateful to Willem van Verseveld and Bart Schilperoort for their technical support and advice on the wflow_sbm model and its implementation in eWaterCycle. AI-assisted tools were used to improve the clarity and phrasing of the manuscript.

515

References

- Argentin, A.-L., Horton, P., Schaeffli, B., Shokory, J., Pitscheider, F., Repnik, L., Gianini, M., Bizzi, S., Lane, S. N., and Comiti, F.: Scale
520 Dependency in Modeling Nivo-Glacial Hydrological Systems: The Case of the Arolla Basin, Switzerland, *Hydrology and Earth System Sciences*, 29, 1725–1748, <https://doi.org/10.5194/hess-29-1725-2025>, 2025.
- Avanzi, F., Gabellani, S., Delogu, F., Silvestro, F., Pignone, F., Bruno, G., Pulvirenti, L., Squicciarino, G., Fiori, E., Rossi, L., Puca, S.,
Toniazzo, A., Giordano, P., Falzacappa, M., Ratto, S., Stevenin, H., Cardillo, A., Fioletti, M., Cazzuli, O., Cremonese, E., di Cella, U. M.,
and Ferraris, L.: IT-SNOW: A Snow Reanalysis for Italy Blending Modeling, in Situ Data, and Satellite Observations (2010–2021), *Earth*
525 *System Science Data*, 15, 639–660, <https://doi.org/10.5194/essd-15-639-2023>, 2023.
- Beaton, A. D., Han, M., Tolson, B. A., Buttle, J. M., and Metcalfe, R. A.: Assessing the Impact of Distributed Snow Water Equivalent Cali-
bration and Assimilation of Copernicus Snow Water Equivalent on Modelled Snow and Streamflow Performance, *Hydrological Processes*,
38, <https://doi.org/10.1002/hyp.15075>, 2024.
- Beck, H. E., Wood, E. F., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O. M., Sheffield, J., and Karger,
530 D. N.: Bias Correction of Global High-Resolution Precipitation Climatologies Using Streamflow Observations from 9372 Catchments
Bias Correction of Global High-Resolution Precipitation Climatologies Using Streamflow Observations from 9372 Catchments, *Journal*
of Climate, 33, 1299–1315, <https://doi.org/10.1175/jcli-d-19-0332.1>, 2019.
- Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N., Fantini, A., Giacona, F., Hauck, C., Huss, M., Huwald,
H., Lehning, M., López-Moreno, J.-I., Magnusson, J., Marty, C., Morán-Tejeda, E., Morin, S., Naaim, M., Provenzale, A., Rabatel, A.,
535 Six, D., Stötter, J., Strasser, U., Terzago, S., and Vincent, C.: The European Mountain Cryosphere: A Review of Its Current State, Trends,
and Future Challenges, *The Cryosphere*, 12, 759–794, <https://doi.org/10.5194/tc-12-759-2018>, 2018.
- Berghuijs, W. R., Woods, R. A., and Hrachowitz, M.: A Precipitation Shift from Snow towards Rain Leads to a Decrease in Streamflow,
Nature Climate Change, 4, 583–586, <https://doi.org/10.1038/nclimate2246>, 2014.
- Berghuijs, W. R., Hale, K., and Beria, H.: Technical Note: Streamflow Seasonality Using Directional Statistics, *EGUsphere*, 2025, 1–16,
540 <https://doi.org/10.5194/egusphere-2024-4117>, 2025.
- Besso, H., Shean, D., and Lundquist, J. D.: Mountain Snow Depth Retrievals from Customized Processing of ICESat-2 Satellite Laser
Altimetry, *Remote Sensing of Environment*, 300, 113 843, <https://doi.org/10.1016/j.rse.2023.113843>, 2024.
- Beven, K.: A Manifesto for the Equifinality Thesis, *Journal of Hydrology*, 320, 18–36, <https://doi.org/10.1016/j.jhydrol.2005.07.007>, 2006.
- Beven, K. and Binley, A.: The Future of Distributed Models: Model Calibration and Uncertainty Prediction, *Hydrological Processes*, 6,
545 279–298, <https://doi.org/10.1002/hyp.3360060305>, 1992.
- Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental
systems using the GLUE methodology, *Journal of Hydrology*, 249, 11–29, [https://doi.org/https://doi.org/10.1016/S0022-1694\(01\)00421-](https://doi.org/https://doi.org/10.1016/S0022-1694(01)00421-8)
8, 2001.
- Brauchli, T., Trujillo, E., Huwald, H., and Lehning, M.: Influence of Slope-scale Snowmelt on Catchment Response Simulated with the
550 *Alpine3D Model*, *Water Resources Research*, 53, 10 723–10 739, <https://doi.org/10.1002/2017wr021278>, 2017.
- Broxton, P. D., Dawson, N., and Zeng, X.: Linking Snowfall and Snow Accumulation to Generate Spatial Maps of SWE and Snow Depth,
Earth and Space Science, 3, 246–256, <https://doi.org/10.1002/2016ea000174>, 2016.
- Broxton, P. D., Leeuwen, W. J. D., and Biederman, J. A.: Improving Snow Water Equivalent Maps with Machine Learning of Snow Survey
and Lidar Measurements, *Water Resources Research*, 55, 3739–3757, <https://doi.org/10.1029/2018wr024146>, 2019.

- 555 Brunner, M. I., Götze, J., Schlemper, C., and Loon, A. F. V.: Hydrological Drought Generation Processes and Severity Are Changing in the Alps, *Geophysical Research Letters*, 50, <https://doi.org/10.1029/2022gl101776>, 2023.
- Buchhorn, M. and smets, B.: Copernicus Global Land Service: Land Cover 100m: Collection 3: Epoch 2019: Globe (V3.0.1), <https://doi.org/10.5281/zenodo.3939050>, 2020.
- Burrell, B., Beltaos, S., and Turcotte, B.: Effects of climate change on river-ice processes and ice jams, *International Journal of River Basin Management*, 21, 421–441, 2023.
- 560 Casson, D. R., Werner, M., Weerts, A., and Solomatine, D.: Global Re-Analysis Datasets to Improve Hydrological Assessment and Snow Water Equivalent Estimation in a Sub-Arctic Watershed, *Hydrology and Earth System Sciences*, 22, 4685–4697, <https://doi.org/10.5194/hess-22-4685-2018>, 2018.
- Cluzet, B., Magnusson, J., Quéno, L., Mazzotti, G., Mott, R., and Jonas, T.: Exploring How Sentinel-1 Wet-Snow Maps Can Inform Fully Distributed Physically Based Snowpack Models, *The Cryosphere*, 18, 5753–5767, <https://doi.org/10.5194/tc-18-5753-2024>, 2024.
- 565 Comola, F., Schaeffli, B., Rinaldo, A., and Lehning, M.: Thermodynamics in the Hydrologic Response: Travel Time Formulation and Application to Alpine Catchments, *Water Resources Research*, 51, 1671–1687, <https://doi.org/10.1002/2014wr016228>, 2015.
- de Bruin, H. A. R., Trigo, I. F., Bosveld, F. C., and Meirink, J. F.: A Thermodynamically Based Model for Actual Evapotranspiration of an Extensive Grass Field Close to FAO Reference, Suitable for Remote Sensing Application, *Journal of Hydrometeorology*, 17, 1373–1382, <https://doi.org/10.1175/jhm-d-15-0006.1>, 2016.
- 570 de Sousa, L. M., Poggio, L., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Riberio, E., and Rossiter, D.: SoilGrids 2.0: Producing Quality-Assessed Soil Information for the Globe, *SOIL Discussions*, 2020, 1–37, <https://doi.org/10.5194/soil-2020-65>, 2020.
- Dettinger, M.: Impacts in the third dimension, *Nature Geoscience*, 7, 166–167, <https://doi.org/10.1038/ngeo2096>, 2014.
- Di Baldassarre, G. and Montanari, A.: Uncertainty in river discharge observations: a quantitative analysis, *Hydrology and Earth System Sciences*, 13, 913–921, 2009.
- 575 Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The Production of a Daily Streamflow Archive and Metadata, *Earth System Science Data*, 10, 765–785, <https://doi.org/10.5194/essd-10-765-2018>, 2018.
- Eilander, D., Boisgontier, H., Bouaziz, L. J. E., Buitink, J., Couasnon, A., Dalmijn, B., Hegnauer, M., de Jong, T., Loos, S., Marth, I., and van Verseveld, W.: HydroMT: Automated and Reproducible Model Building and Analysis, *Journal of Open Source Software*, 8, 4897, <https://doi.org/10.21105/joss.04897>, 2023.
- 580 Essery, R. and Pomeroy, J.: Implications of Spatial Distributions of Snow Mass and Melt Rate for Snow-Cover Depletion: Theoretical Considerations, *Annals of Glaciology*, 38, 261–265, <https://doi.org/10.3189/172756404781815275>, 2004.
- Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., Arnone, E., Bellprat, O., Brötz, B., Caron, L.-P., Carvalhais, N., Cionni, I., Cortesi, N., Crezee, B., Davin, E. L., Davini, P., Debeire, K., Mora, L. d., Deser, C., Docquier, D., Earnshaw, P., Ehbrecht, C., Gier, B. K., Gonzalez-Reviriego, N., Goodman, P., Hagemann, S., Hardiman, S., Hassler, B., Hunter, A., Kadow, C., Kindermann, S., Koirala, S., Koldunov, N., Lejeune, Q., Lembo, V., Lovato, T., Lucarini, V., Massonnet, F., Müller, B., Pandde, A., Pérez-Zanón, N., Phillips, A., Predoi, V., Russell, J., Sellar, A., Serva, F., Stacke, T., Swaminathan, R., Torralba, V., Vegas-Regidor, J., Hardenberg, J. v., Weigel, K., and Zimmermann, K.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP, *Geoscientific Model Development*, 13, 3383–3438, <https://doi.org/10.5194/gmd-13-3383-2020>, 2020.
- 590

- Fang, Y., Liu, Y., and Margulis, S. A.: A Western United States Snow Reanalysis Dataset over the Landsat Era from Water Years 1985 to 2021, *Scientific Data*, 9, 677, <https://doi.org/10.1038/s41597-022-01768-7>, 2022.
- Fiddes, J., Aalstad, K., and Westermann, S.: Hyper-Resolution Ensemble-Based Snow Reanalysis in Mountain Regions Using Clustering, *Hydrology and Earth System Sciences*, 23, 4717–4736, <https://doi.org/10.5194/hess-23-4717-2019>, 2019.
- Fontrodona-Bach, A., Schaefli, B., Woods, R., Teuling, A. J., and Larsen, J. R.: NH-SWE: Northern Hemisphere Snow Water Equivalent Dataset Based on in Situ Snow Depth Time Series, *Earth System Science Data*, 15, 2577–2599, <https://doi.org/10.5194/essd-15-2577-2023>, 2023.
- Frey, S. and Holzmann, H.: A Conceptual, Distributed Snow Redistribution Model, *Hydrology and Earth System Sciences*, 19, 4517–4530, <https://doi.org/10.5194/hess-19-4517-2015>, 2015.
- Gascoïn, S., Grizonnet, M., Bouchet, M., Salgues, G., and Hagolle, O.: Theia Snow Collection: High-Resolution Operational Snow Cover Maps from Sentinel-2 and Landsat-8 Data, *Earth System Science Data*, 11, 493–514, <https://doi.org/10.5194/essd-11-493-2019>, 2019.
- Geissler, J.: Referee 1 Comment on "Can streamflow observations constrain snow mass reconstructions? Lessons from two synthetic numerical experiments", *EGUsphere interactive comment*, <https://doi.org/10.5194/egusphere-2025-3610-RC1>, comment on *EGUsphere preprint egusphere-2025-3610*, 2026.
- Geissler, J., Mazzotti, G., Rathmann, L., Webster, C., and Weiler, M.: Forest Snow Patterns Derived Using ClustSnow Are Temporally Persistent Under Variable Environmental Conditions, *Water Resources Research*, 61, <https://doi.org/10.1029/2024wr038442>, 2025.
- Gordon, B. L., Brooks, P. D., Krogh, S. A., Boisrime, G. F. S., Carroll, R. W. H., McNamara, J. P., and Harpold, A. A.: Why Does Snowmelt-Driven Streamflow Response to Warming Vary? A Data-Driven Review and Predictive Framework, *Environmental Research Letters*, 17, 053004, <https://doi.org/10.1088/1748-9326/ac64b4>, 2022.
- Gottlieb, A. R. and Mankin, J. S.: Evidence of Human Influence on Northern Hemisphere Snow Loss, *Nature*, 625, 293–300, <https://doi.org/10.1038/s41586-023-06794-y>, 2024.
- Griessinger, N., Seibert, J., Magnusson, J., and Jonas, T.: Assessing the Benefit of Snow Data Assimilation for Runoff Modeling in Alpine Catchments, *Hydrology and Earth System Sciences*, 20, 3895–3905, <https://doi.org/10.5194/hess-20-3895-2016>, 2016.
- Grünewald, T., Schirmer, M., Mott, R., and Lehning, M.: Spatial and temporal variability of snow depth and ablation rates in a small mountain catchment, *The Cryosphere*, 4, 215–225, <https://doi.org/10.5194/tc-4-215-2010>, 2010.
- Grünewald, T., Stötter, J., Pomeroy, J. W., Dadić, R., Baños, I. M., Marturià, J., Spross, M., Hopkinson, C., Burlando, P., and Lehning, M.: Statistical modelling of the snow depth distribution in open alpine terrain, *Hydrology and Earth System Sciences*, 17, 3005–3021, <https://doi.org/10.5194/hess-17-3005-2013>, *tex.rating: 5*, 2013.
- Günther, D., Hanzer, F., Warscher, M., Essery, R., and Strasser, U.: Including parameter uncertainty in an intercomparison of physically-based snow models, *Front. Earth Sci.*, 8, 2020.
- Haberkorn, A., López-Moreno, J. I., Helmert, J., Pirazzini, R., and Leppänen, L.: European Snow Booklet, <https://doi.org/10.16904/envidat.59>, 2019.
- Han, J., Liu, Z., Woods, R., McVicar, T. R., Yang, D., Wang, T., Hou, Y., Guo, Y., Li, C., and Yang, Y.: Streamflow Seasonality in a Snow-Dwindling World, *Nature*, 629, 1075–1081, <https://doi.org/10.1038/s41586-024-07299-y>, 2024.
- Harrigan, S., Zsoter, E., Cloke, H., Salamon, P., and Prudhomme, C.: Daily Ensemble River Discharge Reforecasts and Real-Time Forecasts from the Operational Global Flood Awareness System, *Hydrology and Earth System Sciences*, 27, 1–19, <https://doi.org/10.5194/hess-27-1-2023>, 2022.

- Helbig, N. and van Herwijnen, A.: Subgrid Parameterization for Snow Depth over Mountainous Terrain from Flat Field Snow Depth, *Water Resources Research*, 53, 1444–1456, <https://doi.org/10.1002/2016wr019872>, 2017.
- 630 Helbig, N., Bühler, Y., Eberhard, L., Deschamps-Berger, C., Gascoin, S., Dumont, M., Revuelto, J., Deems, J. S., and Jonas, T.: Fractional Snow-Covered Area: Scale-Independent Peak of Winter Parameterization, *The Cryosphere*, 15, 615–632, <https://doi.org/10.5194/tc-15-615-2021>, 2020.
- Henn, B., Clark, M. P., Kavetski, D., and Lundquist, J. D.: Estimating Mountain Basin-mean Precipitation from Streamflow Using Bayesian Inference, *Water Resources Research*, 51, 8012–8033, <https://doi.org/10.1002/2014wr016736>, 2015.
- 635 Henn, B., Clark, M. P., Kavetski, D., Newman, A. J., Hughes, M., McGurk, B., and Lundquist, J. D.: Spatiotemporal Patterns of Precipitation Inferred from Streamflow Observations across the Sierra Nevada Mountain Range, *Journal of Hydrology*, 556, 993–1012, <https://doi.org/10.1016/j.jhydrol.2016.08.009>, 2018.
- Hock, R.: A Distributed Temperature-Index Ice- and Snowmelt Model Including Potential Direct Solar Radiation, *Journal of Glaciology*, 45, 101–111, <https://doi.org/10.3189/s0022143000003087>, 1999.
- 640 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 Discussions*, 2023, 1–46, <https://doi.org/10.5194/essd-2023-127>, 2023.
- 645 Horner, I., Branger, F., McMillan, H., Vannier, O., and Braud, I.: Information Content of Snow Hydrological Signatures Based on Streamflow, Precipitation and Air Temperature, *Hydrological Processes*, 34, 2763–2779, <https://doi.org/10.1002/hyp.13762>, 2020.
- Horton, P. and Argentin, A.-L.: hydrobricks: v0.7.2, <https://doi.org/10.5281/zenodo.11082505>, 2024.
- Hou, Y., Han, J., Woods, R., Guo, Y., and Yang, Y.: Understanding Long-term Streamflow Response to Snowfall Change: Insights from a Multivariate Analysis, *Water Resources Research*, 61, <https://doi.org/10.1029/2024wr038215>, 2025.
- 650 Houska, T., Kraft, P., Chamorro-Chavez, A., and Breuer, L.: SPOTting Model Parameters Using a Ready-Made Python Package, *PLoS ONE*, 10, e0145180, <https://doi.org/10.1371/journal.pone.0145180>, 2015.
- Hut, R., Drost, N., van de Giesen, N., van Werkhoven, B., Abdollahi, B., Aerts, J., Albers, T., Alidoost, F., Andela, B., Camphuijsen, J., Dzigan, Y., van Haren, R., Hutton, E., Kalverla, P., van Meersbergen, M., van den Oord, G., Pelulessy, I., Smeets, S., Verhoeven, S., de Vos, M., and Weel, B.: The eWaterCycle Platform for Open and FAIR Hydrological Collaboration, *Geoscientific Model Development*, 15, 5371–5390, <https://doi.org/10.5194/gmd-15-5371-2022>, 2022.
- 655 Imhoff, R. O., van Verseveld, W. J., van Osnabrugge, B., and Weerts, A. H.: Scaling Point-scale (Pedo)Transfer Functions to Seamless Large-domain Parameter Estimates for High-resolution Distributed Hydrologic Modeling: An Example for the Rhine River, *Water Resources Research*, 56, <https://doi.org/10.1029/2019wr026807>, 2020.
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M.: Climatologies at High Resolution for the Earth's Land Surface Areas, *Scientific Data*, 4, 170122, <https://doi.org/10.1038/sdata.2017.122>, 2017.
- 660 Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian Analysis of Input Uncertainty in Hydrological Modeling: 1. Theory, *Water Resources Research*, 42, <https://doi.org/10.1029/2005wr004368>, 2006.
- Kirchner, J. W.: Catchments as Simple Dynamical Systems: Catchment Characterization, Rainfall-runoff Modeling, and Doing Hydrology Backward, *Water Resources Research*, 45, <https://doi.org/10.1029/2008wr006912>, 2009.
- 665

- Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger Climate Classification Updated, *Meteorologische Zeitschrift*, 15, 259–263, <https://doi.org/10.1127/0941-2948/2006/0130>, 2006.
- Kouki, K., Luojus, K., and Riihelä, A.: Evaluation of Snow Cover Properties in ERA5 and ERA5-Land with Several Satellite-Based Datasets in the Northern Hemisphere in Spring 1982–2018, *The Cryosphere*, 17, 5007–5026, <https://doi.org/10.5194/tc-17-5007-2023>, 2023.
- 670 Le Moine, N., Hendrickx, F., Gailhard, J., Garçon, R., and Gottardi, F.: Hydrologically Aided Interpolation of Daily Precipitation and Temperature Fields in a Mesoscale Alpine Catchment, *Journal of Hydrometeorology*, 16, 2595–2618, <https://doi.org/10.1175/jhm-d-14-0162.1>, 2015.
- Lehning, M., Grünewald, T., and Schirmer, M.: Mountain Snow Distribution Governed by an Altitudinal Gradient and Terrain Roughness, *Geophysical Research Letters*, 38, n/a–n/a, <https://doi.org/10.1029/2011gl048927>, 2011.
- 675 Lievens, H., Brangers, I., Marshall, H.-P., Jonas, T., Olefs, M., and Lannoy, G. D.: Sentinel-1 Snow Depth Retrieval at Sub-Kilometer Resolution over the European Alps, *The Cryosphere*, 16, 159–177, <https://doi.org/10.5194/tc-16-159-2022>, 2021.
- Luojus, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Mortimer, C., Derksen, C., Mudryk, L., Moisander, M., Hiltunen, M., Smolander, T., Ikonen, J., Cohen, J., Salminen, M., Norberg, J., Veijola, K., and Venäläinen, P.: GlobSnow v3.0 Northern Hemisphere Snow Water Equivalent Dataset, *Scientific Data*, 8, 163, <https://doi.org/10.1038/s41597-021-00939-2>, 2021.
- 680 López-Moreno, J., Fassnacht, S., Heath, J., Musselman, K., Revuelto, J., Latron, J., Morán-Tejeda, E., and Jonas, T.: Small scale spatial variability of snow density and depth over complex alpine terrain: Implications for estimating snow water equivalent, *Advances in Water Resources*, 55, 40–52, <https://doi.org/10.1016/j.advwatres.2012.08.010>, 2013.
- Magnusson, J., Gustafsson, D., Hüsler, F., and Jonas, T.: Assimilation of Point SWE Data into a Distributed Snow Cover Model Comparing Two Contrasting Methods, *Water Resources Research*, 50, 7816–7835, <https://doi.org/10.1002/2014wr015302>, 2014.
- 685 Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A., and Jonas, T.: Evaluating Snow Models with Varying Process Representations for Hydrological Applications, *Water Resources Research*, 51, 2707–2723, <https://doi.org/10.1002/2014wr016498>, 2015.
- Magnusson, J., Bühler, Y., Quéno, L., Cluzet, B., Mazzotti, G., Webster, C., Mott, R., and Jonas, T.: High-Resolution Hydrometeorological and Snow Data for the Dischma Catchment in Switzerland, *Earth System Science Data Discussions*, 2024, 1–24, <https://doi.org/10.5194/essd-2024-374>, 2024.
- 690 Margulis, S. A., Cortés, G., Giroto, M., Huning, L. S., Li, D., and Durand, M.: Characterizing the Extreme 2015 Snowpack Deficit in the Sierra Nevada (USA) and the Implications for Drought Recovery, *Geophysical Research Letters*, 43, 6341–6349, <https://doi.org/10.1002/2016gl068520>, 2016.
- Mazzotti, G., Currier, W. R., Deems, J. S., Pflug, J. M., Lundquist, J. D., and Jonas, T.: Revisiting Snow Cover Variability and Canopy Structure within Forest Stands: Insights from Airborne Lidar Data, *Water Resources Research*, 55, 6198–6216, <https://doi.org/10.1029/2019wr024898>, 2019.
- 695 McKay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 42, 55–61, 2000.
- MeteoSwiss, Federal Office of Meteorology and Climatology: MeteoSwiss RhiresD, 2024.
- Michel, A., Aschauer, J., Jonas, T., Gubler, S., Kotlarski, S., and Marty, C.: SnowQM 1.0: A Fast R Package for Bias-Correcting Spatial Fields of Snow Water Equivalent Using Quantile Mapping, *Geoscientific Model Development*, 17, 8969–8988, <https://doi.org/10.5194/gmd-17-8969-2024>, 2023.
- 700 Mooney, K. L. and Webb, R. W.: Aspect controls on the spatial redistribution of snow water equivalence through the lateral flow of liquid water in a subalpine catchment, *The Cryosphere*, 19, 2507–2526, <https://doi.org/10.5194/tc-19-2507-2025>, 2025.

- Mott, R.: Climatological snow data since 1998, OSHD, <https://doi.org/http://dx.doi.org/10.16904/envidat.401>, 2023.
- 705 Mott, R., Winstral, A., Cluzet, B., Helbig, N., Magnusson, J., Mazzotti, G., Quéno, L., Schirmer, M., Webster, C., and Jonas, T.: Operational Snow-Hydrological Modeling for Switzerland, *Frontiers in Earth Science*, 11, 1228–1258, <https://doi.org/10.3389/feart.2023.1228158>, 2023.
- Mudryk, L., Mortimer, C., Derksen, C., Chereque, A. E., and Kushner, P.: Benchmarking of Snow Water Equivalent (SWE) Products Based on Outcomes of the SnowPEX+ Intercomparison Project, *The Cryosphere*, 19, 201–218, <https://doi.org/10.5194/tc-19-201-2025>, 2024.
- 710 Myeni, R., Knyazikhin, Y., and Park, T.: MCD15A3H MODIS/Terra+qua Leaf Area Index/FPAR 4-Day L4 Global 500m SIN Grid V006, 2015.
- Napoli, A., Crespi, A., Ragone, F., Maugeri, M., and Pasquero, C.: Variability of orographic enhancement of precipitation in the Alpine region, *Scientific Reports*, 9, 13 352, <https://doi.org/10.1038/s41598-019-49974-5>, 2019.
- Nash, J. and Sutcliffe, J.: River Flow Forecasting through Conceptual Models Part I — A Discussion of Principles, *Journal of Hydrology*, 715 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- Nott, D. J., Marshall, L., and Brown, J.: Generalized Likelihood Uncertainty Estimation (GLUE) and Approximate Bayesian Computation: What’s the Connection?, *Water Resources Research*, 48, <https://doi.org/10.1029/2011wr011128>, 2012.
- Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M., Seidel, F. C., and Winstral, A.: The Airborne Snow Observatory: 720 Fusion of Scanning Lidar, Imaging Spectrometer, and Physically-Based Modeling for Mapping Snow Water Equivalent and Snow Albedo, *Remote Sensing of Environment*, 184, 139–152, <https://doi.org/10.1016/j.rse.2016.06.018>, 2016.
- Pflug, J. M. and Lundquist, J. D.: Inferring Distributed Snow Depth by Leveraging Snow Pattern Repeatability: Investigation Using 47 Lidar Observations in the Tuolumne Watershed, Sierra Nevada, California, *Water Resources Research*, 56, <https://doi.org/10.1029/2020wr027243>, 2020.
- 725 Premier, V., Marin, C., Bertoldi, G., Barella, R., Notarnicola, C., and Bruzzone, L.: Exploring the Use of Multi-Source High-Resolution Satellite Data for Snow Water Equivalent Reconstruction over Mountainous Catchments, *The Cryosphere*, 17, 2387–2407, <https://doi.org/10.5194/tc-17-2387-2023>, 2023.
- Pulka, T., Herrnegger, M., Ehrendorfer, C., Lüicking, S., Avanzi, F., Formayer, H., Schulz, K., and Koch, F.: Evaluating Precipitation Corrections to Enhance High-Alpine Hydrological Modeling for Hydropower, <https://doi.org/10.2139/ssrn.4823086>, 2024.
- 730 Raleigh, M. S. and Small, E. E.: Snowpack Density Modeling Is the Primary Source of Uncertainty When Mapping Basin-wide SWE with Lidar, *Geophysical Research Letters*, 44, 3700–3709, <https://doi.org/10.1002/2016gl071999>, 2017.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding Predictive Uncertainty in Hydrologic Modeling: The Challenge of Identifying Input and Structural Errors, *Water Resources Research*, 46, <https://doi.org/10.1029/2009wr008328>, 2010.
- Revuelto, J., López-Moreno, J. I., Azorin-Molina, C., and Vicente-Serrano, S. M.: Topographic Control of Snowpack Distribution in a Small Catchment in the Central Spanish Pyrenees: Intra- and Inter-Annual Persistence, *The Cryosphere*, 8, 1989–2006, 735 <https://doi.org/10.5194/tc-8-1989-2014>, 2014.
- Revuelto, J., López-Moreno, J. I., and Alonso-González, E.: Light and Shadow in Mapping Alpine Snowpack With Unmanned Aerial Vehicles in the Absence of Ground Control Points, *Water Resources Research*, 57, <https://doi.org/10.1029/2020wr028980>, 2021.
- Revuelto, J., Alonso-González, E., Deschamps-Berger, C., Gutmann, E. D., and López-Moreno, J. I.: Recent Advances in Snow Monitoring 740 from Local to Global Scales, *Current Climate Change Reports*, 11, 10, <https://doi.org/10.1007/s40641-025-00207-0>, 2025.

- Rhoades, A. M., Jones, A. D., and Ullrich, P. A.: Assessing Mountains as Natural Reservoirs with a Multimetric Framework, *Earth's Future*, 6, 1221–1241, <https://doi.org/10.1002/2017ef000789>, 2018.
- Rinaldo, A., Vogel, G. K., Rigon, R., and Rodriguez-Iturbe, I.: Can One Gauge the Shape of a Basin?, *Water Resources Research*, 31, 1119–1127, <https://doi.org/10.1029/94wr03290>, 1995.
- 745 Rudisill, W., Flores, A., and Carroll, R.: Evaluating 3 decades of precipitation in the Upper Colorado River basin from a high-resolution regional climate model, *Geoscientific Model Development*, 16, 6531–6552, <https://doi.org/10.5194/gmd-16-6531-2023>, 2023.
- Ruelland, D.: Should Altitudinal Gradients of Temperature and Precipitation Inputs Be Inferred from Key Parameters in Snow-Hydrological Models?, *Hydrology and Earth System Sciences*, 24, 2609–2632, <https://doi.org/10.5194/hess-24-2609-2020>, 2020.
- Schaeffli, B.: Snow Hydrology Signatures for Model Identification within a Limits-of-acceptability Approach, *Hydrological Processes*, 30, 750 4019–4035, <https://doi.org/10.1002/hyp.10972>, 2016.
- Schaeffli, B. and Gupta, H. V.: Do Nash Values Have Value?, *Hydrological Processes*, 21, 2075–2080, <https://doi.org/10.1002/hyp.6825>, 2007.
- Thornton, J., Brauchli, T., Mariethoz, G., and Brunner, P.: Efficient Multi-Objective Calibration and Uncertainty Analysis of Distributed Snow Simulations in Rugged Alpine Terrain, *Journal of Hydrology*, 598, 126 241, <https://doi.org/10.1016/j.jhydrol.2021.126241>, 2021.
- Trujillo, E. and Molotch, N. P.: Snowpack Regimes of the Western United States, *Water Resources Research*, 50, 5611–5623, 755 <https://doi.org/10.1002/2013wr014753>, 2014.
- Trujillo, E., Ramírez, J. A., and Elder, K. J.: Topographic, Meteorologic, and Canopy Controls on the Scaling Characteristics of the Spatial Distribution of Snow Depth Fields, *Water Resources Research*, 43, <https://doi.org/10.1029/2006wr005317>, 2007.
- van Verseveld, W. J., Weerts, A. H., Visser, M., Buitink, J., Imhoff, R. O., Boisgontier, H., Bouaziz, L., Eilander, D., Hegnauer, M., ten Velden, C., and Russell, B.: Wflow_sbm v0.7.3, a Spatially Distributed Hydrological Model: From Global Data to Local Applications, 760 *Geoscientific Model Development*, 17, 3199–3234, <https://doi.org/10.5194/gmd-17-3199-2024>, 2024.
- Vögeli, C., Lehning, M., Wever, N., and Bavay, M.: Scaling Precipitation Input to Spatially Distributed Hydrological Models by Measured Snow Distribution, *Frontiers in Earth Science*, 4, 108, <https://doi.org/10.3389/feart.2016.00108>, 2016.
- Vrugt, J. A.: Markov Chain Monte Carlo Simulation Using the DREAM Software Package: Theory, Concepts, and MATLAB Implementation, *Environmental Modelling & Software*, 75, 273–316, <https://doi.org/10.1016/j.envsoft.2015.08.013>, 2016.
- 765 Whittaker, C. and Leconte, R.: A Hydrograph-Based Approach to Improve Satellite-Derived Snow Water Equivalent at the Watershed Scale, *Water*, 14, 3575, <https://doi.org/10.3390/w14213575>, 2022.
- Wirgin, A.: The Inverse Crime, arXiv, <https://doi.org/10.48550/arxiv.math-ph/0401050>, 2004.
- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D.: A 770 High-accuracy Map of Global Terrain Elevations, *Geophysical Research Letters*, 44, 5844–5853, <https://doi.org/10.1002/2017gl072874>, 2017.
- Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M.: MERIT Hydro: A High-resolution Global Hydrography Map Based on Latest Topography Dataset, *Water Resources Research*, 55, 5053–5073, <https://doi.org/10.1029/2019wr024873>, 2019.
- Ylönen, M., Marttila, H., Geissler, J., Kuzmin, A., Korpelainen, P., Kumpula, T., and Ala-Aho, P.: UAV LiDAR surveys and machine learning improve snow depth and water equivalent estimates in boreal landscapes, *The Cryosphere*, 19, 4585–4610, <https://doi.org/10.5194/tc-19-4585-2025>, 2025.
- 775 Zakeri, F., Mariethoz, G., and Giroto, M.: High-resolution snow water equivalent estimation: a data-driven method for localized downscaling of climate data, *Hydrology and Earth System Sciences*, 29, 6935–6958, <https://doi.org/10.5194/hess-29-6935-2025>, 2025.