



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 snow mass estimates are key to understanding snowmelt-driven streamflow and climate change impacts on snow water resources. However, snow mass observations are scarce, and SWE reconstructions rely largely on snow models forced with meteorological inputs. Ground-based and satellite observations are often used to constrain the typically high uncertainty of modeled snow mass reconstructions, but their constraining potential is limited in data-scarce regions and prior to the onset of satellite monitoring. Here, we suggest using streamflow information as an additional information source to better reconstruct snow mass. 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 modelling chain can easily stand in the way. We also show that streamflow is most helpful in estimating catchment-aggregated properties of snow mass reconstructions, in particular catchment-aggregated melt rates. Future work should assess the potential of streamflow-constrained 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, spatial patterns of snowfall and snowmelt are highly heterogeneous (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 provide only binary 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).

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, including snow model outputs—with and without assimilation of indirect snow observations—and passive microwave retrievals, some of which incorporate in-situ SD observations. While most analyzed products performed well in capturing SWE climatology and interannual variability over low-relief regions, their performance degraded substantially in mountainous areas. Several SWE reconstruction methods have been developed specifically for mountain areas. 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 using 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 satellite-based 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, SWE reconstructions can be constrained by empirical knowledge on recurring snow patterns. Numerous studies have shown that spatial snow depth 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 like canopy structure and density (Trujillo et al., 2007; Mazzotti et al., 2019; Helbig et al., 2020). Helbig and van Herwijnen (2017) derived gridded snow depth estimates from point-scale snow depth measurements using terrain properties of each grid cell. Helbig et al. (2021) in turn used similar terrain properties to relate gridded snow depth to sub-grid fractional snow cover, which is used in snow models to constrain melt rates. Similarly, Mazzotti et al. (2022) used canopy structure parameters to improve fractional snow cover and SWE estimates in forest sites. Pflug et al. (2021) instead leveraged interannual similarity in snow patterns to infer snow deposition patterns based on corresponding information from



55 better-informed years. 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.

Nonetheless, the above sources of information are sometimes insufficient to constrain SWE reconstructions, notably before the onset of satellite observations or in scarcely monitored regions. An additional, relatively abundant data source that can inform the temporal and spatial SWE dynamics is streamflow. However, until now it has not been given a lot of attention in
60 SWE reconstructions. 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). 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
65 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 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 assumes that all direct runoff in spring originates from snowmelt, an assumption less valid in smaller, more complex basins, and is sensitive
75 to baseflow separation uncertainty. 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. However, they did not evaluate SWE directly and did not separate rain from snow, limiting the applicability of the approach in mixed-phase climates. Also using inverse hydrological modeling, Ruelland (2020) accurately derived temperature and precipitation gradients, but did not evaluate SWE directly either. All three approaches focus on seasonal, catchment-integrated SWE estimates and provide little insight into temporal or spatial snowpack dynamics. Moreover, they are confounded by key uncertainties—assumptions about catchment storage (Horner et al., 2020), baseflow separation (Whittaker and Leconte, 2022), and model structure (Henn et al., 2015, 2018), leaving open the question of the amount and nature of SWE information theoretically embedded in streamflow, and under what conditions it can be used to constrain SWE reconstructions.

85 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.

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:

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

The prior distribution $P(H_{SWE})$ reflects our initial uncertainty about SWE, and the likelihood $P(Q_{obs} | H_{SWE})$ 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 $P(H_{SWE})$ as the result of the finite sampling of the informative prior distributions of parameters θ as follows:

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

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 $i = 1, \dots, N_{prior}$ yields an ensemble that approximates the prior distribution $P(H_{SWE})$.

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

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

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 $\Theta = \{\theta_M, \theta_{snow}, \theta_{runoff}\}$ to a streamflow simulation Q_{sim} , and the inverse problem becomes one of estimating the posterior distribution:

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



While it is difficult to compute this posterior distribution analytically, it can be approximated with numerical methods that generate samples of the posterior distribution, the most popular methods in hydrology being 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.

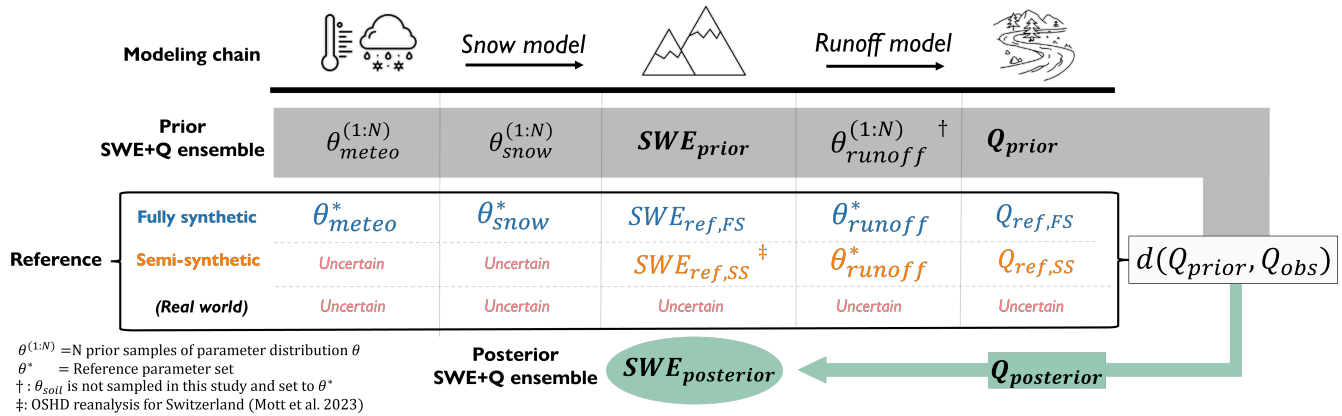


Figure 1. Schematic overview of the streamflow-constrained SWE reconstruction framework and the two synthetic numerical experiments. Q represents streamflow, θ represents the parameters to be sampled, θ^* represents the reference parameter set, and d represents the streamflow performance metric. Color-coding 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.



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 .

For each year, 5000 model realizations are generated by randomly sampling parameter sets from uniform prior distributions $\Theta = \{\theta_{meteo}, \theta_{snow}, \theta_{soil}\}$ 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 $d(Q_{sim}, Q_{obs})$, and the top 1% of 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; 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 process (Sect. 4.2).

2.2.1 Snow model

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). Precipitation is partitioned into rainfall and snowfall using a temperature threshold TT and a transition range as follows:

$$P_{snow} = \begin{cases} P & T_a \leq TT - 1^\circ\text{C} \\ P \cdot \left(\frac{(TT + 1^\circ\text{C}) - T_a}{2^\circ\text{C}} \right) & TT - 1^\circ\text{C} < T_a < TT + 1^\circ\text{C} \\ 0 & T_a \geq TT + 1^\circ\text{C} \end{cases} \quad (5)$$

where P is precipitation and T_a represents air temperature. P_{snow} is then adjusted using a spatially uniform yearly multiplicative correction factor $SFCF$ and a linear elevation lapse rate $SFCF_{ELEV}$. TT is defined for each year and applied uniformly in space and adjusted with an elevation lapse rate TT_{ELEV} . $SFCF$, $SFCF_{ELEV}$, TT , and TT_{ELEV} all belong to meteorological parameters θ_{meteo} used to generate the prior SWE ensemble (Eq. 2 & Table 1). Liquid precipitation is calculated as $P - P_{snow}$, and is corrected separately with rainfall correction factor $RFCF$ (Pulka et al., 2024) (Sect. 2.2.4).

Melt occurs when air temperature exceeds TT , following:

$$M(t) = \begin{cases} (m + r_j * I_{pot})(T_a(t) - TT) & \text{if } T_a(t) > TT \\ 0 & \text{otherwise} \end{cases} \quad (6)$$



where M is the melt rate (mm d^{-1}), m is the melt factor ($\text{mm d}^{-1} \text{ } ^\circ\text{C}^{-1}$), r_j 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 calculated 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 (WHC) fraction of the total snow mass (default: 0.1), after which drainage occurs. Liquid water may refreeze within the snowpack when $T_a < TT$. 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 (fSCA) parameterization based on Essery and Pomeroy (2004) and Magnusson et al. (2014):

$$fSCA(t) = \tanh \left(1.26 \cdot \frac{SWE_{SIM}(t)}{CV \cdot SWE_{MAX}} \right) \quad (7)$$

where SWE_{SIM} is the simulated average SWE in the grid cell at time t , CV is the coefficient of variation, and SWE_{MAX} is the pre-melt seasonal maximum SWE.

To account for snow redistribution by wind and gravity, we implement a mass wasting scheme adapted from Frey and Holzmann (2015). Snow is redistributed to downhill cells if three precalibrated criteria are met: (i) SWE exceeds 500 mm, (ii) wet snow to dry snow ratio does not exceed 0.001, and (iii) slope exceeds 0.3:

$$MW = \min \left(0.5, \frac{slope}{5.67} \right) \cdot \min \left(1.0, \frac{SWE}{10000} \right) \cdot MWF \cdot \mathbf{1}_{\text{criteria met}} \quad (8)$$

where MW is the mass wasting per grid cell, $\mathbf{1}_{\text{criteria met}}$ is a binary mask indicating whether the redistribution conditions are satisfied, and MWF is a mass wasting factor with a precalibrated value of 0.5. m_{hock} , r_{hock} , WHC , and CV are retained as snow model parameters θ_{snow} used to generate the prior SWE ensemble (Eq. 2 and Table 1).

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 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 1 arcsecond resolution (approximately $900 \text{ m} \times 700 \text{ m}$ at 40° 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 (Fig. 2) in Switzerland spans 42.9 km^2 with elevations ranging from 1595 to 3180 m a.s.l. (mean: 2372 m). 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

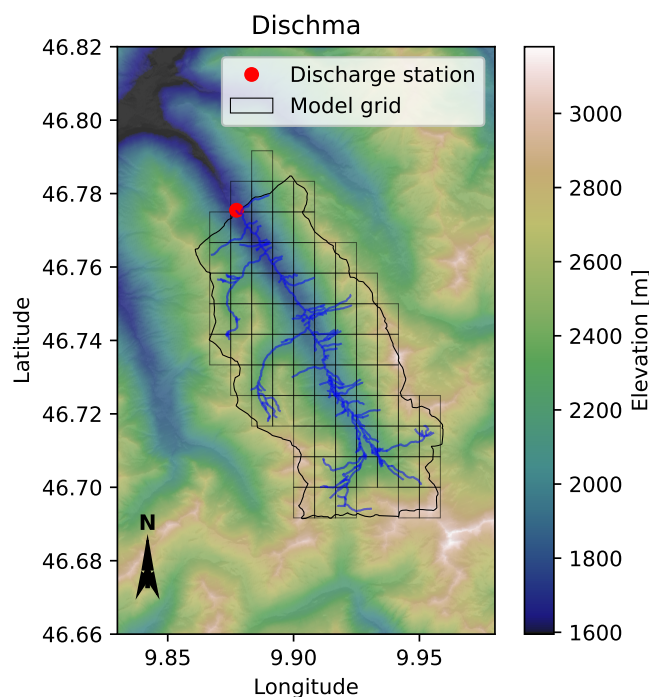


Figure 2. Digital elevation model and delineation of the Dischma catchment, located in the eastern Swiss alps. The regular model grid has a resolution of 1x1 arcsecond.

190 half falling as snow. Average annual discharge is 1240 mm/year. The catchment has featured in numerous snow hydrological studies (Berghuijs et al., 2025; Brauchli et al., 2017; Comola et al., 2015; Schaefli, 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

195 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 estimated precipitation across all grid cells is 1029 mm/year (1998–2022), which is in contradiction
 200 with observed streamflow of 1229 mm/year over the same period.

TabsD and RhiresD were downscaled to the 1 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. 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. Alongside
 205 snowfall correction factors $SFCF$ and $SFCF_{ELEV}$ (Sect. 2.2.1), a flat seasonal rainfall correction factor $RFCF$ is also
 applied to the MeteoSwiss forcing to allow for seasonal rainfall bias correction (Eq. 3 & Table 1).

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^*
$SFCF$	θ_{meteo}	Snowfall correction factor	–	0.9–1.5	1.0–1.4
$SFCF_{ELEV}$	θ_{meteo}	Elevation scaling of snowfall	–	0.7–1.3	1.0
$RFCF$	θ_{meteo}	Rainfall correction factor	–	0.7–1.3	1.0
TT	θ_{meteo}	Rain–snow temperature threshold	°C	–1 to 1	0.0
TT_{ELEV}	θ_{meteo}	Elevation scaling of TT	–	–2 to 2	0.0
M_{HOCK}	θ_{snow}	Degree-day melt factor	mm °C ^{–1} d ^{–1}	1–4	2.5
R_{HOCK}	θ_{snow}	Radiation multiplier	mm W ^{–1} m ² °C ^{–1} d ^{–1}	0.005–0.04	0.025
WHC	θ_{snow}	Snowpack water holding capacity	–	0.1–0.4	0.25
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 knowl-
 210 edge 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 his 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).

215 For each year, 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) using 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 a Markov Chain
 Monte Carlo sampling algorithm since the objective of our study is to explore the information content of streamflow for SWE
 220 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)

225 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 ($SW E_{ref,FS}$) 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
230 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 $SW E_{prior}$ is expected to roughly approximate the reference simulation $SW E_{ref,FS}$.

2.3.2 Experiment 2: Semi-Synthetic (SS)

235 The second experiment is a Semi-Synthetic experiment, where we use the OSHD temperature-index SWE reanalysis product (Mott et al., 2023; Mott, 2023) as the synthetic SWE reference $SW E_{ref,SS}$. This product combines a temperature-index snow model with data assimilation of in-situ snow depth 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
240 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
245 (i.e., when $T_{air} < 0^\circ\text{C}$) are removed, air temperature is capped at a minimum of 0°C , and OSHD-derived snowmelt is added as liquid precipitation. This enables integration without altering the wflow_sbm source code. While this method introduces some physical inconsistencies, such as the omission of refreezing in the soil, these effects are deemed negligible. The runoff model and streamflow observations remain free of uncertainty, isolating the impact of snow-related uncertainties. $RFCF$ is not dictated by OSHD and is still inferred, with the true $RFCF$ ($RFCF^*$) set to 1. The rainfall correction is applied only to
250 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-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 $N_{posterior} = 50$. 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 full daily time series of snowfall and melt rates to better evaluate temporal dynamics and individual events. For snow accumulation, we use the total seasonal snow accumulation, rather than peak SWE volume, to reflect the total snow contribution to the catchment water balance. Several timing metrics, such as date of peak SWE and melt season length, are omitted as their information is assumed to be embedded in other 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. 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 such as 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:

$$E_{AGG}^{Accumulation} = APE = \left| \frac{Accumulation_{sim} - Accumulation_{obs}}{Accumulation_{obs}} \right| \cdot 100 \quad (9)$$



and the Mean Absolute Percentage Error (MAPE) in *GRID* mode:

$$280 \quad E_{GRID}^{Accumulation} = MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Accumulation_{sim,i} - Accumulation_{obs,i}}{Accumulation_{obs,i}} \right| \cdot 100 \quad (10)$$

where i denotes individual grid cells. For timing metrics evaluating SWE onset and melt-out dates, we use Absolute Error (AE) in *AGG* mode:

$$E_{AGG}^{Melt-out/Onset} = AE = |T_{sim} - T_{obs}| \quad (11)$$

and Mean Absolute Error (MAE) in *GRID* mode:

$$285 \quad E_{GRID}^{Melt-out/Onset} = MAE = \frac{1}{n} \sum_{i=1}^n |T_{sim,i} - T_{obs,i}| \quad (12)$$

where T_{sim} and T_{obs} are the simulated and observed event dates (in day-of-year).

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	Error Type (<i>AGG/GRID</i>)
E^{Melt}	NSE of daily snowmelt time series ($-dSWE/dt$)	NSE / grid-mean NSE
$E^{Snowfall}$	NSE of daily snowfall time series ($+dSWE/dt$)	NSE / grid-mean NSE
E^{Onset}	First day SWE exceeds 10% of seasonal max	AE / MAE
$E^{Melt-out}$	First day SWE drops below 10% of seasonal max	AE / MAE
$E^{Accumulation}$	Total snowfall (or melt) over season	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

290 rank, denoted $R_{post,median}$.

If streamflow perfectly selects the best SWE scenarios, we expect $R_{post,median} = 25$, 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 $R_{post,median} = 2500$, 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

295 the median.

3 Results

3.1 Posterior parameter ensembles

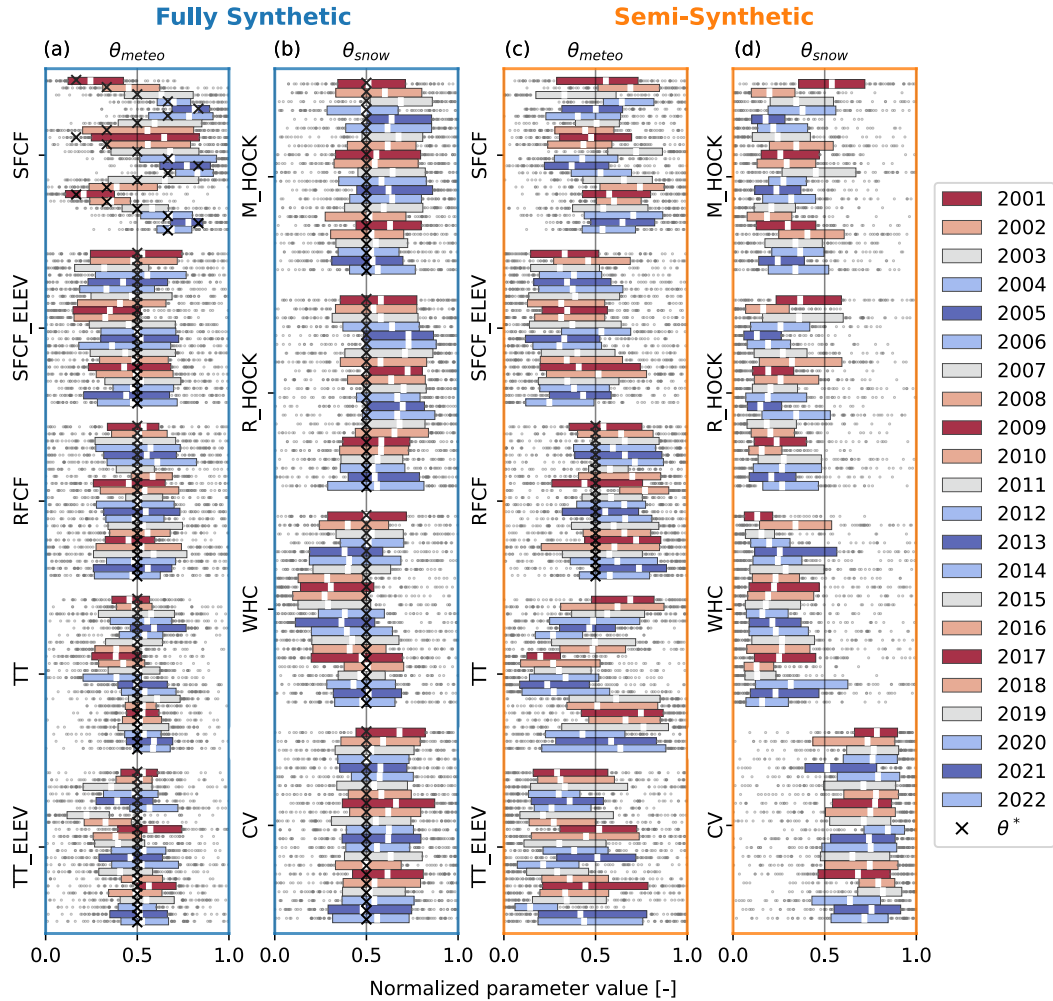


Figure 3. Annual posterior parameter ensembles for the FS and SS experiments, expressed relative to the normalized prior range. θ_M and θ_{snow} represent the meteorological and snow model parameters. Medians (white squares), interquartile ranges (boxes), and outliers (grey dots) are shown for the 50 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 $SFCF^*$ in FS.

Inferred posterior parameter ensembles Θ_{post} do not consistently align with the true parameter values Θ^* in the FS experiment (see Fig. 3a & b). Among all parameters, $SFCF$ shows the highest sensitivity. Its annual posterior ensembles generally reflect the imposed artificial bias fluctuations $SFCF^*$, although they tend to exceed $SFCF^*$ on average, except during years



with the highest $SFCF^*$ (2005, 2013, and 2021). $SFCF_{ELEV}$ and WHC are often underestimated, while M_{HOCK} , R_{HOCK} and CV are overestimated in most years. This suggests the occurrence of numerous posterior SWE scenarios with higher over-
all SWE ($SFCF$), preferential accumulation at lower elevations ($SFCF_{ELEV}$), faster overall melt (M_{HOCK} and R_{HOCK}),
yet slower snow cover depletion (CV), compared to the synthetic SWE observations. The inferred posterior parameter ensem-
bles thus show that biased SWE scenarios can lead to the best-performing streamflow simulations.

In the SS experiment, of all Θ^* only $RFCF^*$ is known, as $SW E_{ref,SS}$ consists of an external SWE product generated
with different forcing and a different snow model (Sect. 2.3.2). Fig. 3c shows that $RFCF$ is consistently overestimated, which
is compensated by an annual underestimation of catchment-wide SWE accumulation of $6.6 \pm 4.4\%$ to close the melt season
water balance (not shown). The annual posterior ensembles of $SFCF$ vary considerably across the prior range, suggesting
annually varying biases in the snowfall forcing and confirming the need for annual over multiannual inversion. The values of
 $SFCF_{ELEV}$, TT_{ELEV} , M_{HOCK} , R_{HOCK} , and WHC are consistently on the lower edge of the prior range, while the CV
values are consistently shifted to the higher edge. This suggests preferential accumulation at lower elevations, slower melt,
and slower snow cover depletion of $SW E_{ref,SS}$ (i.e. OSHD-TI product) compared to our prior assumptions expressed as
parameter ranges.

3.2 Streamflow and SWE performance

The posterior parameter ensembles for the FS experiment show that we cannot recover the true parameter values from stream-
flow alone, with NSE as the streamflow performance metric (Fig. 3). To better understand this result, we analyze the model
performances associated with the best ranked parameter sets. Figure 4 shows the E_{Q-NSE} results and posterior ensemble
member selection (Fig. 4a and 4c) and the subsequent evaluation of this selection on $E_{GRID}^{Accumulation}$ as an example SWE
performance metric (Fig. 4b and 4d), for both FS and SS. The results for other target SWE metrics are presented in Figs.
S1-10.

The E_{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 and 4c), compared to an overall mean prior
NSE of 0.67 ± 0.11 for FS and 0.56 ± 0.16 for SS. The mean annual NSE range of the posterior ensembles equals 0.01 ± 0.01
for FS and 0.03 ± 0.01 for SS, compared to prior ranges of 2.4 ± 1.2 for FS and 2.4 ± 1 for SS.

Figures 4b and 4d show the $E_{GRID}^{Accumulation}$ error results for each of the 5000 prior ensemble members, with the retained
posterior members (i.e. the best 1% under E_{Q-NSE}) subset overlaid. The best-performing prior members show near-zero error
in FS for some years (2001, 2002, 2004, 2016–2022), suggesting a good approximation of the existing solution $SW E_{ref,FS}$. In
other years, however, the overall minimum prior error is higher (e.g. 2003). This is likely caused by both insufficient sampling
and a high sensitivity of the highest and lowest elevation grid cells to $SFCF_{ELEV}$ and TT_{ELEV} fluctuations, leading to high
accumulation biases. In the SS case, a perfect approximation of $SW E_{ref,SS}$ is likely non-existent due to added meteorological
and snow model uncertainty, leading to an increase in the lowest achievable accumulation error compared to FS and resulting
in a higher overall prior error.

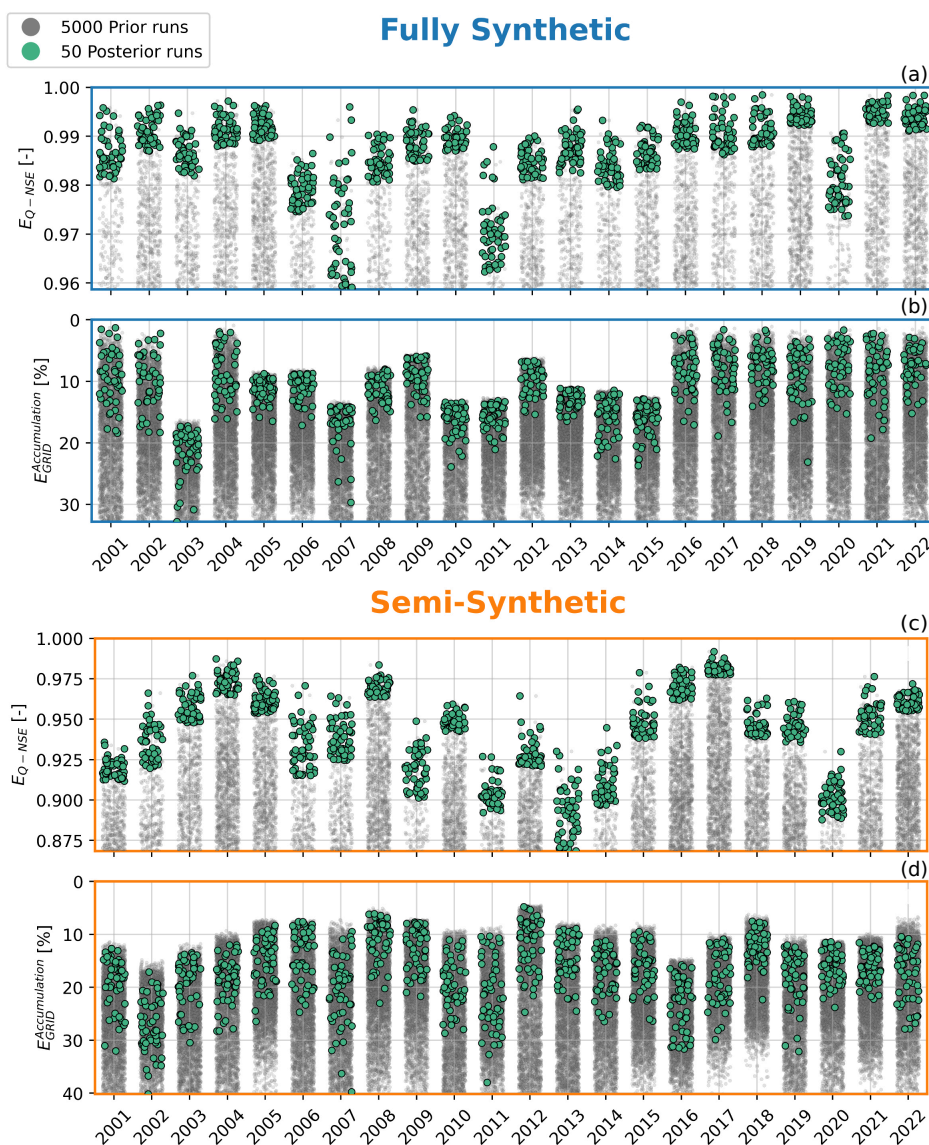


Figure 4. Streamflow-based posterior selection (a, c) and evaluation of the sampled parameter sets on the SWE performance metric $E_{GRID}^{Accumulation}$ (b, d), for both FS (a, b) and SS (c, d) experiments. Grey points represent the 5000 annual prior members (above the y-axis cutoff), while green points represent the posterior ensemble: the 50 members with the best streamflow performance.

335 The posterior subset is located among the better performing prior members in most years for both FS and SS, suggesting that streamflow provides meaningful constraint on SWE properties (in this case, $E_{GRID}^{Accumulation}$). At the same time, numerous prior members outperform the posterior ensemble, which shows that high streamflow skill does not guarantee high gridded

SWE accumulation skill ($E_{GRID}^{Accumulation}$). Additionally, the degree of constraint varies substantially across years, with some years showing a narrow spread (e.g. 2013 in FS) and others showing a wide spread (e.g. 2019) among the posterior ensemble.

3.3 Posterior rank evaluation across SWE metrics

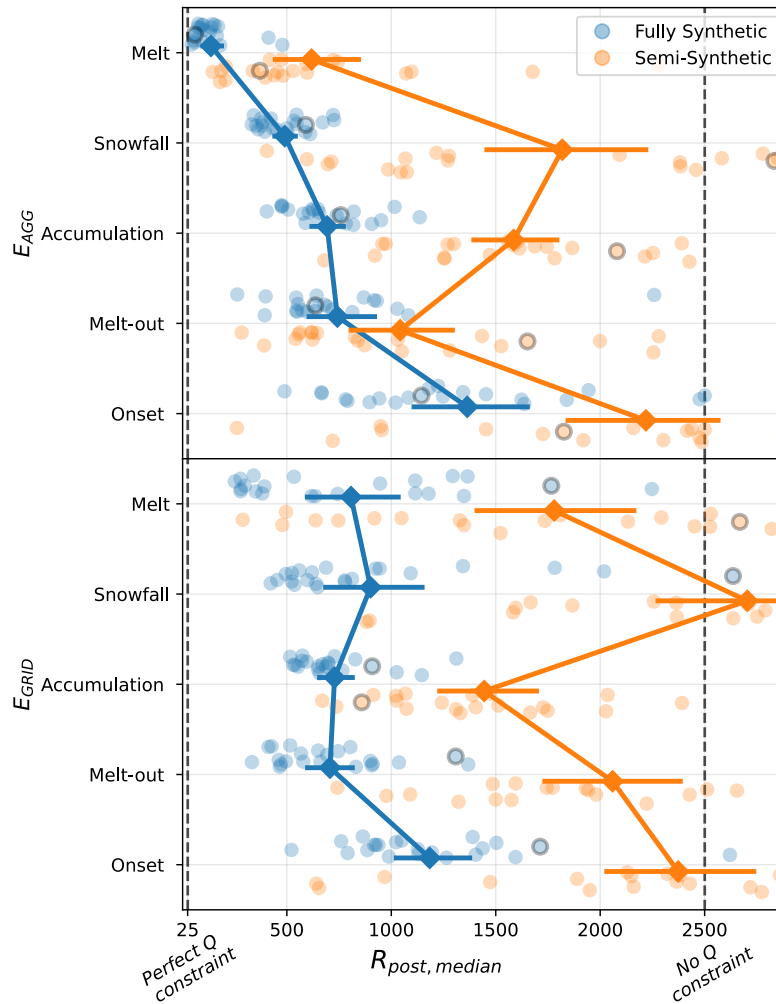


Figure 5. Annual median SWE metric ranks of the streamflow-derived posterior ensembles, relative to all 5000 prior members. The top figure shows the results for catchment-aggregated SWE metrics (E_{AGG}), while the bottom shows grid-averaged metrics (E_{GRID}), sorted based on the FS E_{AGG} ranks. Each 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



To further understand the constraining power of streamflow, Figure 5 shows the annual median posterior ranks $R_{post,median}$ across all SWE performance metrics computed for the streamflow-derived posterior ensembles, with the ranks being relative to the prior ensemble. Results are shown separately for catchment-aggregated (E_{AGG}) and grid-averaged (E_{GRID}) metrics. Each point represents a year and shows the median rank of the 50 posterior members. High ranks (i.e. low values) indicate a strong constraint of streamflow on SWE with a perfect constraint corresponding to $R_{post,median} = 25$ and no constraint corresponding to $R_{post,median} = 2500$ (Sect. 2.5.1).

E_{AGG}^{Melt} emerges as the overall highest ranking SWE performance metric in both experiments. In the FS case, $R_{post,median}$ approaches the perfect constraint limit of 25 in most years ($R_{post,median} = 137 \pm 108$), indicating that the same simulations that best reproduce streamflow NSE also tend to best reproduce catchment-scale melt dynamics. In SS, E_{AGG}^{Melt} remains the overall highest-ranking AGG metric, though with considerably reduced ranks and more year-to-year variability ($R_{post,median} = 619 \pm 518$). In contrast, E_{GRID}^{Melt} ranks considerably lower and shows high interannual spread in both FS ($R_{post,median} = 807 \pm 557$) and SS ($R_{post,median} = 1779 \pm 951$) (see year 2003 in Fig. 5). This suggests that streamflow can constrain the catchment-aggregated meltwater production, but to a lesser degree its spatial origin.

$E^{Snowfall}$ presents somewhat different behavior. In FS, both $E_{AGG}^{Snowfall}$ and $E_{GRID}^{Snowfall}$ rank relatively high, likely because the same parameter sets that benefit E^{Melt} also benefit $E^{Snowfall}$ under unbiased forcing. In SS, the different forcing biases during accumulation and melt periods reduce this effect, and both $E_{AGG}^{Snowfall}$ and $E_{GRID}^{Snowfall}$ rank low. This confirms the expected difficulties of inferring snowfall dynamics from discharge data, due to the high variability of snowfall in space and time (Mott et al., 2014).

$E_{GRID}^{Accumulation}$ ranks slightly higher than $E_{AGG}^{Accumulation}$, both in FS and SS. In SS, $E_{GRID}^{Accumulation}$ is the highest-ranking $GRID$ performance metric ($R_{post,median} = 1444$). This suggests 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.

Among the timing metrics, the melt-out dates ($E^{Melt-out}$) are relatively well constrained in both FS and SS, particularly in AGG mode. This is consistent with their physical link to melt cessation and the end of the snow-driven streamflow season. Compared to $E_{AGG}^{Melt-out}$, $E_{GRID}^{Melt-out}$ performs similarly in FS but deteriorates more substantially in SS. Conversely, SWE onset dates (E^{Onset}) are poorly constrained in all cases, mainly resulting from the abovementioned large variability in snowfall dynamics.

Across all metrics, $R_{post,median}$ values are systematically higher in SS compared to FS (orange vs. blue points in Fig. 5), with an average increase of 989 across all metrics, corresponding to 20% of N_{prior} . The standard deviation of $R_{post,median}$ values for each performance metric also increases on average by 414. This confirms that structural and input uncertainty reduce the ability of streamflow to constrain SWE. Additionally, the increased spread in $R_{post,median}$ 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 highlights growing trade-offs between competing aspects of



375 SWE reconstruction under added uncertainty. Nonetheless, except for $E_{GRID}^{Snowfall}$ in SS, most median ranks remain above the no-constraint threshold, indicating some retained information.

3.4 Correlation among metrics

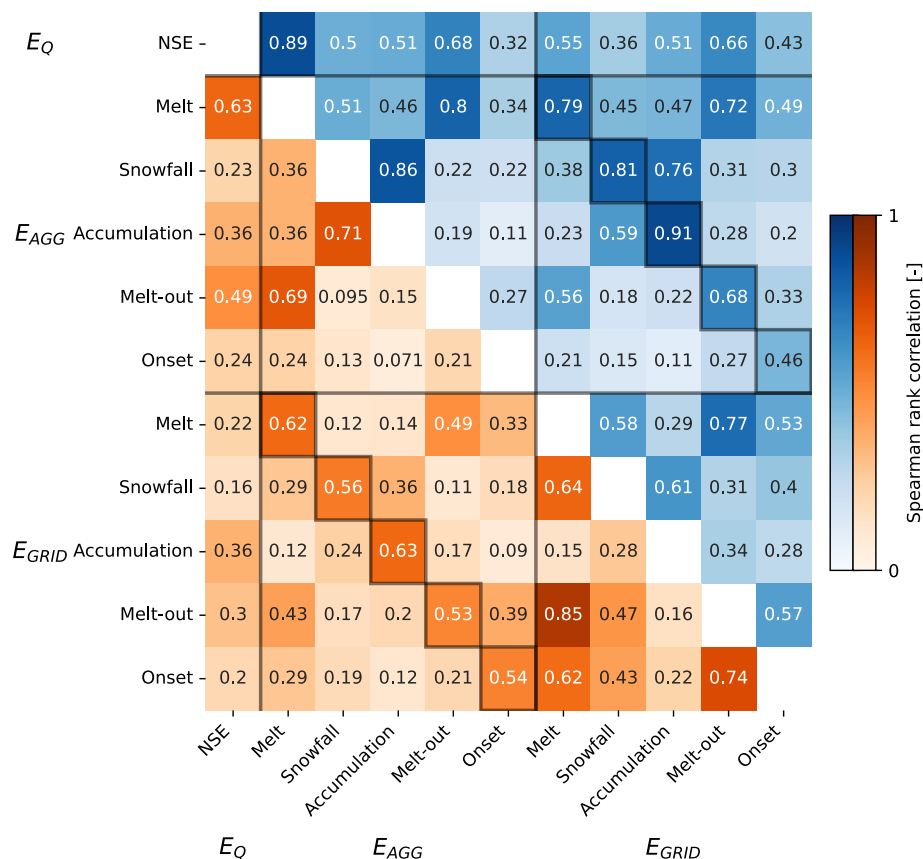


Figure 6. Correlations among all retained streamflow and SWE performance metrics, expressed as the median of annual Spearman rank correlations over all 5000 yearly prior members between 2001 and 2022. The upper values in blue represent the fully synthetic experiment, while the lower values in orange 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.

To complete the picture, we analyze the complementarity of the retained metrics (Fig. 6). The top row and left-most column show the Spearman Rank correlation (ρ) of all SWE metrics with streamflow NSE, which is similar to Fig. 5 but instead considering the correlation among all prior members for each year. Again, E_{AGG}^{Melt} appears as the SWE performance metric most strongly correlated with E_{Q-NSE} ($\rho_{FS} = 0.89$ & $\rho_{SS} = 0.63$). Among the different SWE performance metrics, FS correlations are generally stronger than SS, and AGG metrics correlate well with their $GRID$ counterparts. Strong correlation of



other SWE performance metrics with E_{AGG}^{Melt} is not necessarily a guarantee of strong correlation with E_{Q-NSE} . It is the case for Melt-out, which mirrors the correlation of Melt with E_{Q-NSE} across both AGG and $GRID$ metrics. However, this does
385 not hold for E_{GRID}^{Melt} , which, despite a moderate correlation with E_{AGG}^{Melt} ($\rho_{SS} = 0.62$), shows little correlation with E_{Q-NSE} ($\rho_{SS} = 0.22$). Similarly, $E_{GRID}^{Accumulation}$ correlates only weakly with E_{AGG}^{Melt} ($\rho_{SS} = 0.12$), yet shows a slightly stronger correlation with E_{Q-NSE} ($\rho_{SS} = 0.36$).

These results suggest that, in general, streamflow contains information about different SWE metrics independently. In other words, the members with high E_{AGG}^{Melt} rank might not be the same as the members with high $E_{GRID}^{Accumulation}$ rank. This supports
390 the use of multiple SWE metrics and highlights the risks of drawing conclusions from a single metric.

4 Discussion

4.1 Streamflow constraining potential under idealized conditions

The FS experiment confirms that streamflow-constrained SWE inversion works in theory, but perfect constraint is only achieved for catchment-aggregated melt (E_{AGG}^{Melt}). Other SWE properties remain well-constrained, but far from perfectly constrained. For
395 example, $E_{AGG}^{Accumulation}$ has a mean $R_{post,median}$ of 693; far above the no-constraint benchmark (2500), but still implying an average of 668 ($693 - 0.5 * N_{posterior}$) prior members that match snowfall better but streamflow worse. This demonstrates that even under highly idealized conditions, streamflow does not consistently identify the best-performing SWE scenarios across all performance metrics.

This finding can primarily be explained by physical non-uniqueness in the SWE-streamflow relationship (Beaton et al.,
400 2024): different SWE and rainfall scenarios can lead to equivalent streamflow responses. Catchment-aggregated melt (E_{AGG}^{Melt}) being better constrained than distributed melt (E_{GRID}^{Melt}) is a first indication of this, by showing that biased spatial melt distributions can lead to the accurate aggregated melt output. A second indication is given by the imperfect constraint on catchment-aggregated accumulation ($E_{AGG}^{Accumulation}$): the best-performing streamflow performance can be achieved with biased (i.e., over- or underestimated) catchment-wide SWE accumulation estimates (i.e., $R_{post,median}$ far from the perfect constraint limit). Fi-
405 nally, posterior parameter values of $SFCF$ and $RFCF$ diverge from the known truth (Fig. 3 and Sect. 3.1), confirming that multiple distinct SWE and rainfall combinations can yield similar streamflow responses.

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
410 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 $N_{posterior}/N_{prior}$ 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
415 meteorological forcing and snow model versus the OSHD reference, which the current inversion framework is unable to correct.



Increasing model complexity, enhancing the temporal and spatial resolution of θ_{meteo} , or drawing from a larger prior ensemble could potentially improve inversion outcomes. However, these strategies would quickly render the inversion computationally infeasible. Combined with the aforementioned non-uniqueness in the SWE–streamflow relationship, uncertainties throughout the modeling chain present a major barrier to accurately identifying realistic SWE scenarios from streamflow. These added
 420 uncertainties also introduce stronger trade-offs between SWE metrics: even when the best-performing ensemble members for E_{AGG}^{Melt} are correctly identified, they are increasingly unlikely to also perform well on other SWE performance metrics.

4.2 Additional challenges under real-world conditions

Under real-world conditions, constraining SWE reconstructions using streamflow presents additional challenges and requires methodological adaptations beyond the idealized setup explored in this study. These challenges include both substantially
 425 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—such as lithology, pedology, and hydraulic conductivity—and uncertainty in the representation of water transport processes through the catchment (Beven, 2006). Such uncertainty can introduce persistent timing biases in the translation of snowmelt into streamflow, complicating efforts to infer SWE dynamics from streamflow observations. Henn et al. (2018) have already demonstrated that
 430 model structural choices critically influence the success of precipitation inversion, and similar effects are expected for SWE. Another source of uncertainty excluded from this study is streamflow observation error. This includes uncertainty in river stage measurements, discharge gauging procedures, rating curve estimation (Di Baldassarre and Montanari, 2009), and potentially ice damming (Burrell et al., 2023). These errors can propagate into both event-scale misattribution of meltwater timing and biases in the estimation of seasonal water inputs in mass-conserving models. A further category of uncertainty concerns the me-
 435 teorological and snow model uncertainties imposed in the semi-synthetic experiment, which we deem conservative compared to real-world conditions. Real-world meteorological errors likely exceed the mismatch between our forcing and the OSHD forcing, not least due to the addition of evaporation estimation uncertainty. Likewise, snowpack dynamics are more heterogeneous and complex in reality than what is represented in the OSHD model, leading to additional structural snow model uncertainty. Taken together, these additional sources of uncertainty are expected to further diminish the constraining potential of streamflow
 440 on SWE reconstruction beyond the reduction observed here between the fully synthetic and semi-synthetic experiments.

An additional challenge in real-world applications is the absence of true SWE observations against which to evaluate inversion results. Unlike synthetic experiments, where the true SWE is known, real-world evaluations rely on incomplete and uncertain observations. Ground SWE measurements offer accurate data but are sparse and represent only single points, making comparisons with spatially averaged model outputs problematic (Horner et al., 2020; Magnusson et al., 2019). Ground snow
 445 depth observations require conversion to SWE using snow density estimates, which are often uncertain and spatially variable (López-Moreno et al., 2013). Satellite-derived snow cover products provide complete spatial snow duration information, but lack information on SWE magnitude. State-of-the-art SWE reconstructions, such as the OSHD product, offer gridded reconstructions that can be used for benchmarking: they allow for evaluating whether streamflow can provide a similar degree of constraint compared to ground or remote sensing observations. However, they remain model-based reconstructions themselves,



450 equally suffering from multiple sources of uncertainty. Airborne Snow Observatory (ASO) data in the U.S. are likely the most accurate catchment-scale SWE estimates currently available (Painter et al., 2016), but they remain limited to Western-US catchments and do not provide temporal continuity. On a related note, the lack of SWE evaluation data equally implies a lack of training data for data-driven methods, thereby limiting the potential of machine learning methods as an alternative in streamflow-constrained SWE reconstruction.

455 Taken together, these limitations imply that verifying whether SWE inversion has succeeded is inherently difficult under real-world conditions. 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.

4.3 Potential inversion framework adaptations

460 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 sensitive to timing errors, potentially penalizing simulations that replicate melt events with slight temporal shifts more than those that miss them entirely. This is problematic in light of our finding that streamflow has the strongest constraint on catchment-aggregated melt, which is inherently timing-dependent. Directional optimization sampling algorithms (as opposed to the LHS random sampling applied here) risk further exacerbating the sensitivity of NSE to timing errors. Alternative metrics targeting hydrological signatures, such as the coefficient of variation or the sum of streamflow over the melt season (Schaeffli, 2016), might be more robust when errors are more pronounced.

Real-world uncertainty may additionally alter which SWE properties are best constrained and should therefore be the target of the inversion. While catchment-aggregated melt was the best-constrained SWE property in our synthetic experiments, its sensitivity to timing errors could reduce its identifiability under real-world conditions. SWE accumulation may be less sensitive to timing errors, though it might be more sensitive to mass balance errors. Other SWE properties, such as snowfall rates and SWE onset dates, showed limited constraint even under ideal conditions and are unlikely to improve with added uncertainty.

The interactions between streamflow and SWE performance metrics are also likely to affect the outcome of the inversion, in addition to the individual choice of each metric. NSE is a residual-based streamflow performance metric, that might favor the selection of members performing well on residual-based SWE performance metrics (such as E_{AGG}^{Melt}). A bias-based streamflow performance metric, such as the error on the accumulated discharge over the melt season, might instead favor the selection of members performing well on bias-based SWE performance metrics, such as $E_{AGG}^{Accumulation}$ (which ranked relatively low under E_{Q-NSE}). Consequently, the results presented here should be considered in light of the use of NSE 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.

480 Given the increased spatial and temporal variability and uncertainty in meteorological forcing under real-world conditions, the correction factors currently applied in this study are likely too simplistic to adequately capture real-world meteorological biases. The current application of θ_{meteo} assumes constant biases within each season, whereas in reality, snowfall and melt-



related biases may vary at sub-seasonal and event timescales. Furthermore, the assumed linear dependence of snowfall and melt
 485 threshold corrections on elevation ($SFCF_{ELEV}$ and TT_{ELEV}) likely oversimplifies the actual non-linear lapse-rate behavior
 observed in complex terrain. While our findings suggest that temporal snowfall biases have a limited effect on streamflow
 simulations, the representation of melt dynamics and SWE accumulation is more sensitive. Increasing the temporal resolution
 of temperature-based parameters (TT , TT_{ELEV}) and introducing greater spatial flexibility for snowfall and melt correc-
 tions ($SFCF$, $SFCF_{ELEV}$, TT , and TT_{ELEV}) could improve the identifiability of relevant processes from streamflow
 490 observations. Despite the associated increase in computational cost, such refinements may enhance the potential of streamflow
 to constrain SWE properties.

Finally, the choice of snow and runoff models can be critical. Radiation-enhanced temperature-index models, such as the
 snow model used here, have been shown to perform well on the catchment scale and require minimal meteorological input
 (Magnusson et al., 2015). However, when more abundant meteorological data is available, the more complex energy-balance-
 495 based models might be better suited to capture the true SWE evolution (Mott et al., 2023). In contrast, the inversion could
 benefit from decreased 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 formulations could reduce computational costs and allow
 for larger ensembles compared to the fully distributed runoff model used here. More broadly, the use of multiple snow and
 runoff models within the inversion framework could enhance robustness by increasing the likelihood of capturing the true
 500 SWE evolution and SWE–Q transformation, respectively.

4.4 Outlook on the added value of streamflow in SWE reconstructions

We demonstrate that streamflow can constrain certain aspects of SWE reconstructions under idealized conditions, but also that
 both non-uniqueness and added uncertainty significantly reduce that ability. 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
 505 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 influence whether streamflow can provide added value in SWE reconstructions, independent of the method-
 ological considerations in Sect. 4.3. 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 or bias-correct them. Secondly, the
 510 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, while snow-dominated catchments offer better identi-
 fiability of snowmelt than snow-scarce catchments (Griessinger et al., 2016). Dry spring and summer climates are particularly
 beneficial for streamflow-assisted SWE inversion as they limit the confounding between rainfall and snowmelt signals.

The same logic likely also applies to inter-annual variability within each catchment. In years with higher snowfall fractions
 515 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, 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).

520 While streamflow alone may not fully constrain SWE reconstructions, it can serve as a valuable complement to other sources of snow information, such as remotely sensed snow cover (e.g., Margulis et al., 2016), wet snow maps (Cluzet et al., 2024), ground observations (e.g., Mott et al., 2023), and spatially predictable accumulation and melt patterns governed by topography (e.g., Pflug et al., 2021). Among these, streamflow is unique in that it captures catchment-integrated snow dynamics, most notably the timing and total volume of snowmelt runoff. Our finding that streamflow most effectively constrains catchment-
525 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.

4.5 Broader implications in snow hydrology

530 Our results have important implications beyond the scope of SWE reconstructions. Streamflow observations have been used as an evaluation measure of different snow routines (e.g., Griessinger et al., 2019; Valéry et al., 2014; Follum et al., 2019; Clemenzi et al., 2023) or of snow data assimilation schemes (Roy et al., 2010; Griessinger et al., 2016; Metref et al., 2023). While these studies mention observation uncertainties and uncertainties in the modeling chain, the inherent non-uniqueness of the SWE–streamflow relationship is a critical limitation often unacknowledged. The results presented here demonstrate that
535 multiple combinations of gridded SWE distributions and snowmelt-to-rainfall ratios can produce similar streamflow outputs. This implies a trade-off between snow model and meteorological forcing: different snow model structures or data assimilation approaches may yield comparable streamflow results under differing meteorological forcings. Consequently, when only the snow model structure or assimilation method is varied but the meteorological input remains biased, it becomes difficult to attribute improvements in streamflow performance to genuine advances in snow process representation. Instead, such im-
540 provements may simply reflect more effective compensation for input biases. This confounding effect can lead to incorrect conclusions about model quality or skill. Furthermore, the choice of both streamflow and SWE evaluation metrics is often overlooked in the literature, despite their critical influence on interpreting the relationship between streamflow and snow dynamics, as demonstrated in this study. Notably, catchment-aggregated melt, which is the SWE component most directly linked to streamflow, is seldom the explicit focus of SWE model inter-comparisons using streamflow as an evaluation target. We there-
545 fore recommend future studies using streamflow for snow model inter-comparison to take note of the effects of non-uniqueness, potential modeling chain uncertainties, and the choice of metrics in the interpretation of their results.

5 Conclusion

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 well-performing streamflow results. 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. 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 further test streamflow-constrained SWE reconstruction under real-world conditions or heightened artificial uncertainty (i.e. streamflow observation errors and runoff model imperfections), across diverse catchments, using multiple streamflow performance metrics and in combination with other indirect SWE observations. More broadly, we advise future studies relying on streamflow for snow model inter-comparison to carefully consider the complications in the SWE-streamflow relationship identified here.

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 the 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)



570 **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

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.

575 *Acknowledgements.* 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.



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
 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
 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-Garretón, C., Baez-Villanueva, O. M., Sheffield, J., and Karger,
 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.,
 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,
<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,
 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\(01\)00421-8](https://doi.org/https://doi.org/10.1016/S0022-1694(01)00421-8), 2001.
- Brauchli, T., Trujillo, E., Huwald, H., and Lehning, M.: Influence of Slope-scale Snowmelt on Catchment Response Simulated with the
 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.



- 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.
- 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.
- Clemenzi, I., Gustafsson, D., Marchand, W.-D., Norell, B., Zhang, J., Pettersson, R., and Pohjola, V. A.: Impact of Snow Distribution Modelling for Runoff Predictions, *Hydrology Research*, 54, 633–647, <https://doi.org/10.2166/nh.2023.043>, 2023.
- 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.
- 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.
- 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.
- Di Baldassarre, G. and Montanari, A.: Uncertainty in river discharge observations: a quantitative analysis, *Hydrology and Earth System Sciences*, 13, 913–921, 2009.
- 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.
- 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.
- 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.
- Follum, M. L., Niemann, J. D., and Fassnacht, S. R.: A Comparison of Snowmelt-derived Streamflow from Temperature-index and Modified-temperature-index Snow Models, *Hydrological Processes*, 33, 3030–3045, <https://doi.org/10.1002/hyp.13545>, 2019.
- 655 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.
- 660 Gascoin, 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.
- 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, 053 004, <https://doi.org/10.1088/1748-9326/ac64b4>, 2022.
- 665 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.
- Griessinger, N., Schirmer, M., Helbig, N., Winstral, A., Michel, A., and Jonas, T.: Implications of Observation-Enhanced Energy-
670 Balance Snowmelt Simulations for Runoff Modeling of Alpine Catchments, *Advances in Water Resources*, 133, 103 410, <https://doi.org/10.1016/j.advwatres.2019.103410>, 2019.
- 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., Dadic, 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,
680 <https://doi.org/10.16904/envdat.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.
- 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.
- 685 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.



- Helbig, N., Schirmer, M., Magnusson, J., Mäder, F., van Herwijnen, A., Quéno, L., Bühler, Y., Deems, J. S., and Gascoin, S.: A Seasonal
Algorithm of the Snow-Covered Area Fraction for Mountainous Terrain, *The Cryosphere*, 15, 4607–4624, [https://doi.org/10.5194/tc-15-](https://doi.org/10.5194/tc-15-4607-2021)
690 4607-2021, 2021.
- 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.
- Henn, B., Clark, M. P., Kavetski, D., Newman, A. J., Hughes, M., McGurk, B., and Lundquist, J. D.: Spatiotemporal Patterns of Pre-
cipitation Inferred from Streamflow Observations across the Sierra Nevada Mountain Range, *Journal of Hydrology*, 556, 993–1012,
695 <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.
- 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
700 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.
- 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.
- 705 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.
- Houska, T., Kraft, P., Chamorro-Chavez, A., and Breuer, L.: SPOTting Model Parameters Using a Ready-Made Python Package, *PLoS ONE*,
10, e0145 180, <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.,
710 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.
- 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*
715 *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, 170 122, <https://doi.org/10.1038/sdata.2017.122>,
2017.
- Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian Analysis of Input Uncertainty in Hydrological Modeling: 1. Theory, *Water Resources*
720 *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.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger Climate Classification Updated, *Meteorolo-
gische Zeitschrift*, 15, 259–263, <https://doi.org/10.1127/0941-2948/2006/0130>, 2006.



- 725 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.
- 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.
- Luoju, 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.
- 730 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.
- 735 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.
- 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., Eisner, S., Huang, S., Lussana, C., Mazzotti, G., Essery, R., Saloranta, T., and Beldring, S.: Influence of Spatial Resolution on Snow Cover Dynamics for a Coastal and Mountainous Region at High Latitudes (Norway), *Water Resources Research*, 55, 5612–5630, <https://doi.org/10.1029/2019wr024925>, 2019.
- 740 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.
- 745 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.
- 750 Mazzotti, G., Webster, C., Quéno, L., Cluzet, B., and Jonas, T.: Canopy Structure, Topography and Weather Are Equally Important Drivers of Small-Scale Snow Cover Dynamics in Sub-Alpine Forests, *Hydrology and Earth System Sciences Discussions*, 2022, 1–32, <https://doi.org/10.5194/hess-2022-273>, 2022.
- 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.
- 755 MeteoSwiss, Federal Office of Meteorology and Climatology: MeteoSwiss RhiresD, 2024.
- Metref, S., Cosme, E., Lay, M. L., and Gailhard, J.: Snow Data Assimilation for Seasonal Streamflow Supply Prediction in Mountainous Basins, *Hydrology and Earth System Sciences*, 27, 2283–2299, <https://doi.org/10.5194/hess-27-2283-2023>, 2023.
- 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.
- 760



- 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/envdat.401>, 2023.
- 765 Mott, R., Scipión, D., Schneebeli, M., Dawes, N., Berne, A., and Lehning, M.: Orographic Effects on Snow Deposition Patterns in Mountainous Terrain, *Journal of Geophysical Research: Atmospheres*, 119, 1419–1439, <https://doi.org/10.1002/2013jd019880>, 2014.
- 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 158, <https://doi.org/10.3389/feart.2023.1228158>, 2023.
- 770 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.
- Myeni, R., Knyazikhin, Y., and Park, T.: MCD15A3H MODIS/Terra+aquia Leaf Area Index/FPAR 4-Day L4 Global 500m SIN Grid V006, 2015.
- Nash, J. and Sutcliffe, J.: River Flow Forecasting through Conceptual Models Part I — A Discussion of Principles, *Journal of Hydrology*, 775 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: 780 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., Hughes, M., and Lundquist, J. D.: Downscaling Snow Deposition Using Historic Snow Depth Patterns: Diagnosing Limitations from Snowfall Biases, Winter Snow Losses, and Interannual Snow Pattern Repeatability, *Water Resources Research*, 57, <https://doi.org/10.1029/2021wr029999>, 2021.
- 785 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., Lücking, 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.
- 790 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, <https://doi.org/10.5194/tc-8-1989-2014>, 2014.
- 795 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.
- Roy, A., Royer, A., and Turcotte, R.: Improvement of Springtime Streamflow Simulations in a Boreal Environment by Incorporating Snow-Covered Area Derived from Remote Sensing Data, *Journal of Hydrology*, 390, 35–44, <https://doi.org/10.1016/j.jhydrol.2010.06.027>, 2010.



- 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.
- Schaepli, B.: Snow Hydrology Signatures for Model Identification within a Limits-of-acceptability Approach, *Hydrological Processes*, 30, 4019–4035, <https://doi.org/10.1002/hyp.10972>, 2016.
- Schaepli, 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, <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.
- Valéry, A., Andréassian, V., and Perrin, C.: ‘As Simple as Possible but Not Simpler’: What Is Useful in a Temperature-Based Snow-Accounting Routine? Part 1 – Comparison of Six Snow Accounting Routines on 380 Catchments, *Journal of Hydrology*, 517, 1166–1175, <https://doi.org/10.1016/j.jhydrol.2014.04.059>, 2014.
- 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, *Geoscientific Model Development*, 17, 3199–3234, <https://doi.org/10.5194/gmd-17-3199-2024>, 2024.
- 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.
- 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 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.