



Data-driven inference of snowmelt from streamflow in 110 catchments across three continents

Pau Wiersma¹, Ralf Loritz², Sanika Baste², Katharina Hafner^{1,3}, and Grégoire Mariéthoz^{1,3}

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

²Institute for Water and Environment, Karlsruhe Institute of Technology, Karlsruhe, Germany

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

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

Abstract

Direct observations of snow-related processes, and especially snowmelt rates, are difficult to acquire. In contrast, streamflow is continuously measured in many snow-influenced catchments around the world, where it contains valuable information on the timing and rate of catchment-wide snowmelt. Importantly, streamflow observations are often available in regions and periods that lack the observations commonly used for snow reanalyses, such as in-situ snow measurements and space-based observations. Here, we investigate how much information streamflow contains on snowmelt, and whether this information can be used to reconstruct snowmelt in catchments with streamflow observations but without more direct snow observations. To this end, we train bidirectional long short-term memory (BiLSTM) networks on catchment-scale snowmelt estimates derived from state-of-the-art snow water equivalent (SWE) reanalysis products across 163 snow-influenced catchments in the Western United States, Switzerland, and Chile, and evaluate them on a subset of 110 snow-dominated catchments. We compare three input configurations: a streamflow-only model, a meteorology-only model driven by ERA5-Land temperature and precipitation, and a combined model using both streamflow and meteorological forcing. We compare the results against both a climatology baseline and a benchmark derived directly from ERA5-Land SWE estimates. We find that the streamflow-only model reconstructs snowmelt substantially better than both the climatological baseline and the ERA5-Land benchmark, demonstrating that streamflow contains considerable and extractable information on catchment-scale snowmelt. This result holds especially in catchments used during training, but also in out-of-sample catchments not seen during training. However, the meteorology-only model performs considerably better, showing the strong capacity of the BiLSTM to correct the biases in the ERA5-Land forcing. Adding streamflow alongside the meteorological inputs results in only modest performance improvements, despite the substantial snowmelt information contained in streamflow. Our results demonstrate both the potential of BiLSTMs to transfer snowmelt information from existing reanalyses to new years and catchments, and the significant but limited added value of streamflow beyond meteorological forcing in this task. We recommend future work to explore the potential of streamflow to constrain catchment-wide melt rates within existing SWE reanalysis frameworks, in order to improve both the reanalyses themselves and their value as training data for future reanalysis extensions.



1 Introduction

25 Seasonal snow in mountainous regions constitutes a critical freshwater resource (Qin et al., 2020; Mankin et al., 2015), while also playing a fundamental role in cold-region ecosystems (Sullender et al., 2023; Phillips et al., 2026) and global climate regulation (Thackeray and Fletcher, 2016; Schwaab et al., 2015). Ongoing climate change is reducing snow amounts across the globe (Notarnicola, 2022; Hock et al., 2019), affecting streamflow seasonality and potentially reducing the total streamflow output of catchments (Gordon et al., 2022; Han et al., 2024; Wang et al., 2025b). Accurately characterizing its magnitude and
30 interannual variability is therefore essential, requiring reliable reanalyses of seasonal snow mass, commonly expressed as snow water equivalent (SWE). Given the sparsity and limited spatial representativeness of direct SWE observations (Mortimer and Vionnet, 2025; Meromy et al., 2013), SWE reanalysis approaches typically rely on snow models driven by meteorological forcing and constrained by indirect SWE observations, such as remotely sensed snow cover disappearance dates (Margulis et al., 2015; Fang et al., 2022b), in-situ snow depth observations (Mott, 2023), or both (Avanzi et al., 2023).

35 However, in many instances those observations are not available (Pritchard, 2021). In such settings, data-driven approaches provide an alternative by transferring information from well-observed to poorly observed conditions using predictor variables that are available in both. For example, Bair et al. (2018), Pflug et al. (2022) and Besso et al. (2026) reproduced SWE reanalysis estimates in near real time using snow cover observations as input. Zakeri et al. (2025) extended a SWE reanalysis temporally using climate analogues, while Michel et al. (2023) learned the bias correction between a simple snow model and a SWE
40 reanalysis during overlapping years and applied this correction to the same model in pre-reanalysis decades. Pflug et al. (2025) demonstrated spatial transfer by extending a Western US SWE reanalysis to the European Alps using snow cover observations, multidecadal precipitation patterns, and daily air temperature. The success of such temporal and spatial transfer approaches is aided by the strong interannual persistence of snow cover and SWE patterns, which tend to recur within a given region regardless of year-to-year variations in meteorological conditions (Wayand et al., 2018; Mendoza et al., 2020; Pflug et al.,
45 2021).

Streamflow observations also have the potential to extend SWE reanalysis in space and time. By integrating the catchment-wide response to snowmelt, streamflow encodes information on the timing of melt onset (Lundquist and Flint, 2006), as well as the timing and magnitude of individual melt events and the total accumulation of mass within a basin (Schaeffli, 2016; Horner et al., 2020), including the often poorly monitored higher-elevation areas. Streamflow is particularly helpful in reconstructing
50 catchment-wide snowmelt (Wiersma et al., 2026), a variable of direct and high relevance for downstream water resources and hydrological forecasting (Guo et al., 2025). Streamflow records are also widely available across snow-dominated regions, both in catchments and periods covered by SWE reanalyses, and in data-sparse regions or historical periods where such products do not exist (See Sect. A for an illustrative quantification). We therefore hypothesize that snowmelt-streamflow relationships learned during periods of SWE reanalysis availability can be transferred to reconstruct snowmelt in catchments and years
55 beyond the spatial and temporal domain of existing SWE reanalyses.

Inferring catchment states and fluxes from the streamflow output is known as "doing hydrology backwards" (Kirchner, 2009). Probabilistic Bayesian frameworks are commonly used for this task, where a prior distribution of e.g. rainfall or SWE



realizations is fed through a forward process-based hydrological model, and the mismatch between the resulting streamflow output and streamflow observations is then used to produce a posterior distribution (Koskela et al., 2012; Henn et al., 2015; Wiersma et al., 2026). However, such approaches often rely on formal likelihood definitions, are computationally demanding, and are sensitive to uncertainties in the modeling chain. Manoj et al. (2025) proposed an alternative "doing hydrology backward" approach that instead uses a time-agnostic Long Short-term Memory (LSTM) network able to process catchment inputs and outputs simultaneously as model inputs to reconstruct extreme rainfall events. LSTMs lack explicit process-representation and do not provide a probabilistic characterization of uncertainty by default, but they substantially reduce methodological complexity and computational demand, and avoid the need to explicitly formulate and propagate uncertainties through each component of the hydrological modeling chain. LSTMs have been used extensively and successfully in hydrological modeling in recent years (Kratzert et al., 2018; Klotz et al., 2022; Baste et al., 2025), and have been shown to be able to account for snow processes internally (Lees et al., 2022).

Beyond their predictive capabilities, machine-learning models can also be used to estimate the amount of information that observations contain about a target variable (Nearing et al., 2020). A sufficiently flexible model can provide a practical upper bound on the predictive information that can be extracted from the available data. Quantifying how much information streamflow contains about snow processes, and how this information varies across climatic and physiographic conditions, remains an open question (Horner et al., 2020; Wiersma et al., 2026). We expect the snowmelt information content of streamflow to depend on factors such as geology, storage dynamics, catchment size and soil hydraulic properties (Brooks et al., 2025; Wang et al., 2025b). Training an LSTM to reconstruct snowmelt from streamflow observations therefore provides not only a reconstruction framework, but also a means of assessing how much snowmelt information is contained in streamflow under real-world conditions.

In this study, we investigate whether streamflow observations can be used to transfer snowmelt information from existing SWE reanalysis products to catchments and periods not covered by those reanalyses. To do so, we employ bidirectional LSTM (BiLSTM) networks, which we hypothesize to be well suited to reconstruction problems since they are able to simultaneously exploit information from both before and after the prediction date. We consider three input configurations. First, we force the BiLSTM only with streamflow observations, thereby isolating the information content of snowmelt in streamflow. Second, we use only coarse-resolution temperature and precipitation forcing. Finally, we combine streamflow and meteorological forcing within a single BiLSTM model. We train the models on a dataset combining SWE reanalyses and large-sample hydrological data from 163 snow-influenced catchments across three regions and evaluate them on a subset of 110 snow-dominated catchments, assessing their ability to generalize both temporally to years preceding the reanalyses and spatially to catchments not covered by them. We compare the results against a climatology baseline model and melt estimates derived from the global ERA5-Land SWE reanalysis. We address three key questions. First, how much information does streamflow contain about snowmelt across different catchments and regions? Second, to what extent can BiLSTMs transfer snowmelt information from SWE reanalyses to years and catchments not covered by those reanalyses? Third, how relevant is streamflow in this task compared to meteorological forcing, and does this relevance vary among catchments? Answering these questions will clarify when and where streamflow provides meaningful information on snowmelt dynamics and will inform the integration of stream-



flow observations into future snow reconstruction efforts, particularly in data-scarce regions and periods preceding satellite observations.

95 2 Data

2.1 SWE reanalysis training data

The Western United States Snow Reanalysis (WUS-SR, Fang et al. (2022b)) dataset is a Landsat satellite-era SWE reanalysis product that provides daily SWE at 16 arcseconds (~500m) spatial resolution across the entire western United States for water years 1985–2021. It is generated using a Bayesian framework in which an ensemble of land surface model simulations
100 forced by bias-corrected MERRA-2 meteorology is constrained by Landsat fractional snow-covered area (fSCA) observations via batch data assimilation. The Andes snow reanalysis (Andes-SR, Cortés and Margulis (2017)) is produced using the same underlying data assimilation methodology but at finer spatial resolution (180m) and over a smaller regional domain, covering the Andean cordillera between 27 and 37 degrees south. Also, contrary to WUS-SR, glaciers and elevations below 1500m are masked. Compared to in-situ peak SWE measurements, WUS-SR and Andes-SR showed correlation coefficients of 0.77 and
105 0.73 respectively (Fang et al., 2023). Among 10 SWE reanalysis products, Lundquist et al. (2026) have shown WUS-SR to have the lowest total SWE volume bias (1.6%) compared to 26 aerial lidar acquisitions in California, Colorado and Washington.

The SPAtial Snow climatology for Switzerland (SPASS, Marty et al. (2025a)) differs from satellite-constrained SWE reanalyses in that it does not assimilate remotely sensed snow cover. Instead, it relies on a temperature-index snow model driven by gridded meteorological forcing and constrained using in-situ snow observations within an ensemble Kalman filter frame-
110 work for the period 1998–2023 (Mott, 2023). The dataset is then extended back until 1962 by applying quantile mapping to align the non-assimilated simulation (1962-1998) with the higher-quality reference period (1998-2023). This approach yields near-zero bias and good agreement with observations at climatological scales. Performance is generally strongest at mid to high elevations and for monthly to seasonal timescales, while larger relative errors occur at low elevations where snow cover is intermittent. Overall, the two periods are broadly consistent, supporting the use of SPASS as a unified dataset for evaluation
115 of streamflow-informed SWE reconstruction.

An important caveat is that none of the three SWE reanalysis products provide explicit estimates of snowmelt, and therefore do not distinguish between SWE losses due to snowmelt and sublimation. Sublimation from mountain snowpacks is known to be highly uncertain (Wang et al., 2025a) and strongly variable across events and regions (Hogan et al., 2026; Mott et al., 2018). In particular, sublimation can be a dominant loss term in semiarid Andean catchments, where substantial snowmelt may
120 occur only in localized hotspots (Ayala et al., 2023). Nonetheless, given the large-scale scope of this study, the complexity and uncertainty associated with sublimation estimates, and the limited availability of key drivers such as radiation and wind speed, we assume that snowmelt represents the majority of SWE losses and approximate all SWE decreases as snowmelt. All results are interpreted with this assumption in mind.



2.2 Caravan catchment input data

125 Streamflow observations, meteorological forcing, and catchment static attributes are drawn from the Caravan large-sample hydrological dataset (Kratzert et al., 2023). Caravan builds on multiple CAMELS (Catchment Attributes and Meteorology for Large-sample Studies) datasets from individual countries, such as CAMELS-US (Addor et al., 2017), CAMELS-CH (Höge et al., 2023) and CAMELS-CL (Alvarez-Garreton et al., 2018) (see Knoben et al. (2025) for a recent overview), and provides a standardized, quality-controlled global collection of catchments. Such large-sample datasets are particularly well suited for machine learning applications, which require diverse and extensive training data. Here, we use data from the original Caravan dataset for the continental United States (US) and Chile (CL), as well as from the Swiss Caravan extension (CH; Höge et al. (2025)).

We retain only small- to medium-sized, snow-influenced catchments with near-continuous streamflow availability. We require that at least 5% of precipitation falls as snow and a minimum average annual snow cover of 5%, in order to exclude snow-free catchments while still retaining regions with intermittent snow cover. We further require that the annual maximum streamflow is above 1 mm d^{-1} for at least 10 years within the study period (1985-2015), to exclude extremely arid catchments where streamflow is a priori unlikely to contain snowmelt information. Catchments with more than 10% glacier cover are excluded, as reconstructing glacier melt is outside the scope of this study. Catchments with more than 5% lake cover are removed to avoid strongly dampened snowmelt-streamflow relationships. The analysis is restricted to catchments smaller than 2000 km^2 and within the spatial domains of the SWE reanalysis products. With the exception of two southern catchments, Chilean catchments with elevations below the 1500m lower limit of the Andes-SR SWE reanalysis were retained, as areas below this threshold did not contribute substantially to snowmelt generation. Eleven Swiss catchments were omitted due to water extraction or hydropower disturbance. Finally, only catchments with less than 10% missing streamflow data over the study period are retained; remaining gaps are treated as described in Sect. 3.3. This results in a total of 163 catchments (84 US, 59 CH, 20 CL) out of 1472 initial candidates (671 US, 296 CH, 505 CL) (See Fig. 1 for maps containing the retained catchments and the Supplementary material for a detailed list).

One catchment from each study region is highlighted throughout the manuscript for illustrative purposes (Fig. 1 and Table 1), representing contrasting snowmelt-streamflow relationships. From this point onward, we refer to them simply as the Merced (Western US), Dischma (Switzerland), and Grande (Chile) catchments. The Merced catchment in California is a fast-responding snow-dominated basin under a Mediterranean climate, where dry summers and shallow soils overlying granite lead to a strong and direct connection between snowmelt and streamflow (Tague and Grant, 2009; Henn et al., 2015). The Dischma catchment in the Swiss Alps meanwhile is much smaller and receives precipitation year-round, leading to high snowmelt-rainfall overlap (Magnusson et al., 2024). The Grande catchment is located in the semi-arid Andes characterized by low and highly seasonal precipitation, with high sublimation losses and much of the effective snowmelt production concentrated in localized melt hotspots (Ayala et al., 2023).

The meteorological forcing data provided in Caravan are derived from the ERA5-Land climate reanalysis (Muñoz-Sabater et al., 2021), available globally at 9 km spatial resolution from 1950 onwards. Catchment-scale forcing is obtained by spatial



aggregation and temporal resampling from hourly to daily resolution. We use temperature and precipitation as meteorological inputs, representing the minimal forcing required and most widely available for snow process modeling. In addition, we include a set of 25 static catchment attributes, broadly following Kratzert et al. (2019), describing climate, vegetation, anthropogenic influence, geology, and soil properties. These attributes are derived from HydroATLAS (Linke et al., 2019) and ERA5-Land climatologies and span a wide range of hydro-climatic conditions across the selected catchments (Fig. C1). The three regions are also representative for most snow classes found globally (Johnston et al., 2026) (Fig. B1).

Table 1. Characteristics of the three highlighted catchments shown in Fig. 1. P_{mean} is mean annual precipitation (mm), f_{snow} indicates the fraction of P_{mean} falling as snow.

Catchment name	Caravan ID	Area (km ²)	Elev. (m)	f_{snow}	P_{mean} (mm)	Gauge lat./lon.
Merced River at Pohono Bridge	camels_11266500	836	1476	0.71	1075	37.72 / -119.67
Dischmabach at Davos-Kriegsmatte	camelsch_2327	43	1512	0.40	1421	46.78 / 9.88
Rio Grande en Las Ramadas	camelscl_4511002	569	1962	0.70	621	-31.01 / -70.58

3 Methodology

3.1 Model architecture

A long short-term memory (LSTM) network is a type of recurrent neural network designed to capture both short- and long-term dependencies in sequential data (Hochreiter and Schmidhuber, 1997). It achieves this by maintaining an internal cell state that represents slowly evolving processes, alongside a hidden state that captures more rapidly varying signals (Fig. 2a). Information flow is regulated through gating mechanisms that control which information is retained, updated, and passed forward through the network. This model architecture is particularly well suited to hydrological applications, as it processes input data sequentially while retaining memory of past states, enabling the representation of processes operating across a wide range of timescales, such as snow accumulation and melt or soil moisture dynamics (Lees et al., 2022). This is critical for our application, where snowmelt estimates potentially depend on both antecedent meteorological conditions and delayed streamflow responses. For a more detailed description of LSTMs in the context of hydrological modeling, we refer to Kratzert et al. (2018, 2019).

In this work we use an adapted LSTM architecture called a bidirectional LSTM (BiLSTM) (Graves and Schmidhuber, 2005). While a classical LSTM processes a sequence only in the forward direction and bases its predictions on past information, a BiLSTM consists of two separate LSTM networks operating on the same sequence. One processes the sequence in chronological order, while the other processes a time-reversed version of the sequence (Fig. 2b). At each timestep, the hidden states of both networks are combined, allowing the model to use information from both before and after the prediction time. This makes the architecture particularly suitable for our application, where snowmelt is reconstructed using both antecedent meteorolog-

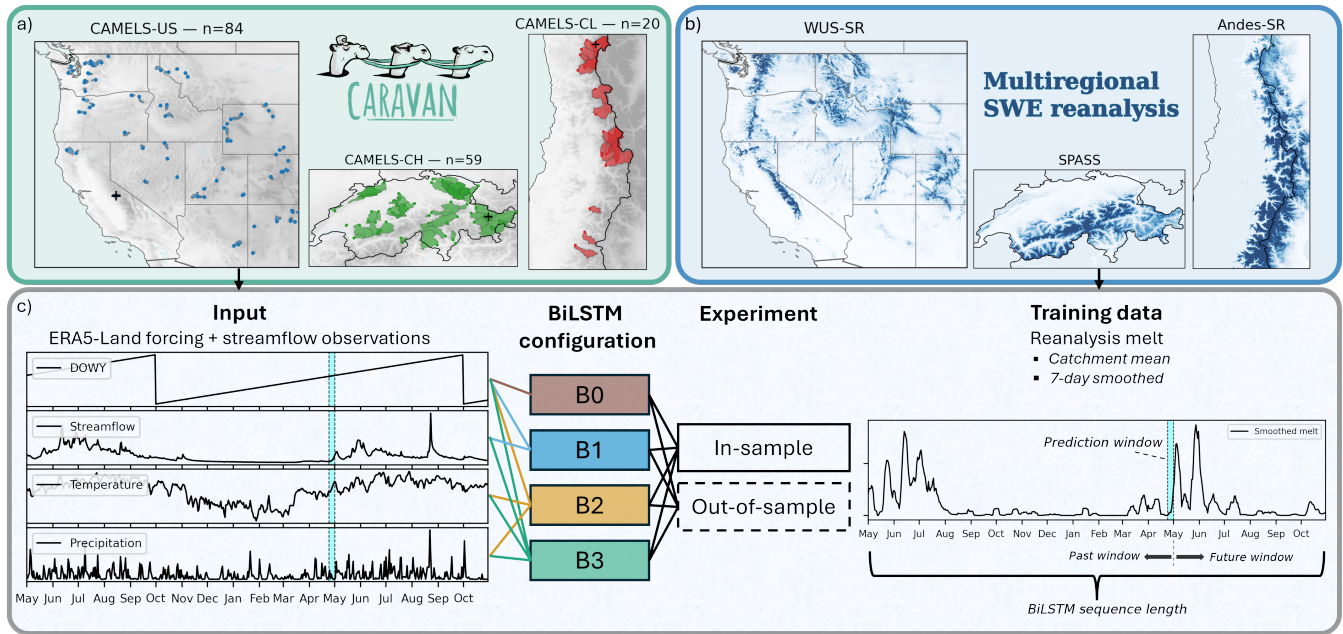


Figure 1. Experiment setup for testing streamflow-assisted snowmelt reconstruction. Panel a shows the delineation of the selected Caravan catchments and their gauge locations in the US, Switzerland (CH) and Chile (CL), with the titles showing the number of catchments per region and the plus signs marking the catchments highlighted throughout the manuscript (Sect. 2.2). Panel b shows the mean annual maximum SWE within the spatial bounds of the region-specific SWE reanalysis products (Sect. 2.1), with dark blue representing values exceeding 700 mm. Panel c shows the experiment workflow. All BiLSTM configurations differ only in their input data combination. DOWY stands for day-of-water-year. Each of the BiLSTM configurations is tested in both in-sample and out-of-sample catchments. The melt training data are derived from the SWE reanalysis products and consequently aggregated per catchment and smoothed with a 7-day window. The region and BiLSTM configuration color-coding, as well as the experiment line styles are consistent throughout the manuscript.

ical conditions and subsequent streamflow observations. We implement the BiLSTM using the PyTorch and Hy2DL Python libraries (Imambi et al., 2021; Espinoza et al., 2024).

3.2 Experimental design

185 We consider three experimental settings (B1-B3) and one baseline (B0) that differ only in their input data configuration (Fig. 1). Together, they allow us to disentangle the respective contributions of meteorological forcing and streamflow observations to snowmelt reconstruction. All configurations are trained under identical conditions and evaluated on the same target variable, enabling direct comparison of their predictive performance. The target variable is catchment-aggregated mean snowmelt, smoothed with a 7-day moving window (hereafter referred to as smoothed melt). This temporal smoothing reduces the impact of high-frequency variability in the snowmelt signal, which otherwise led to underfitting when using the raw melt time series as a prediction target. This effectively transforms the prediction at a single time step into a prediction over a temporal window

190

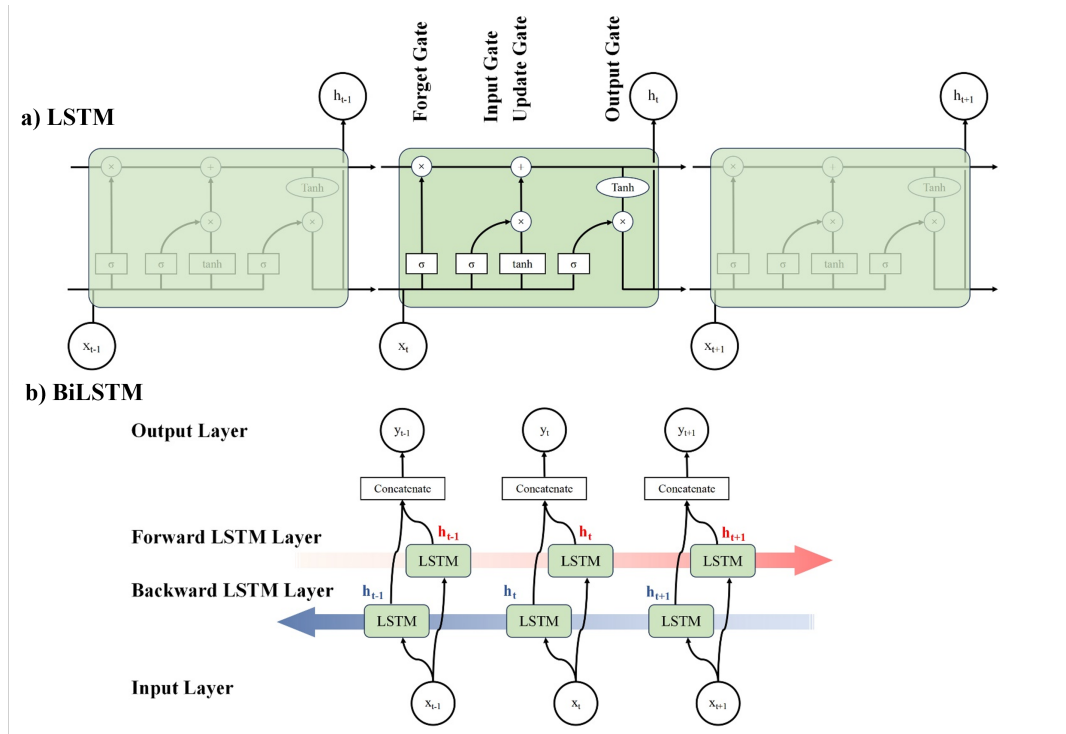


Figure 2. Schematic diagram of a) LSTM and b) BiLSTM architecture, adapted from Liu et al. (2024). x and y represent the model input and output, h represents the hidden cell state of the two separate LSTM layers. For details on the internal LSTM cell operations and BiLSTM architecture, we refer to Kratzert et al. (2018) and Graves and Schmidhuber (2005).

spanning the 7 days preceding the prediction date. All BiLSTM models use a one-year past window and a six-month future window around the prediction date, resulting in a total sequence length of 548 days. The one-year past window provides a full hydrological year of context and follows common practice in hydrological deep learning (e.g. Kratzert et al., 2019). The six-month future window is intended to capture both short- and long-term streamflow responses to snowmelt. This setup differs from that of Manoj et al. (2025), who incorporated future streamflow information through a short 7-day lag within a classical LSTM architecture.

The first configuration, B1, uses streamflow and DOWY as inputs. It is intended to provide an estimate of how much information on catchment-scale snowmelt dynamics is contained in streamflow alone and is used to investigate how this information content varies across catchments and hydro-climatic conditions (Sect. 3.4.3). While future streamflow observations are expected to contain most of the information relevant for the inversion, the past window may provide additional information on antecedent wetness and storage conditions, which are known to influence snowmelt runoff efficiency (Taherian and Ameli, 2026; Wang et al., 2025b).

The second configuration, B2, uses ERA5-Land precipitation, air temperature, and DOWY as inputs. This configuration represents the classical snow modeling setting and serves as a benchmark for the information content of meteorological forcing



alone. The choice of variables reflects the minimum data requirements needed to run a snow model in data-scarce and pre-satellite settings (Tang et al., 2021). We do not expect the future meteorological inputs to substantially improve reconstruction performance, but include them to maintain consistency with the sequence structure used by the other experiments.

The third configuration, B3, combines streamflow, meteorological forcing, and DOWY in a single model. This configuration is used to assess whether streamflow provides additional information beyond meteorological forcing alone. In this setting, future meteorological observations may not directly inform snowmelt reconstruction, but may help the model interpret future streamflow observations by distinguishing rainfall-driven from snowmelt-driven runoff responses. Comparison of B3 against B1 and B2 therefore allows us to quantify both the information content of streamflow and its added value relative to meteorological forcing.

Finally, the B0 configuration serves as a baseline model and uses only day-of-water-year (DOWY) as input. DOWY represents the seasonal timing within a hydrological year, spanning from October 1 to September 30 in the Northern Hemisphere and from April 1 to March 31 in the Southern Hemisphere. This configuration therefore predicts the snowmelt climatology and provides a lower benchmark against which the other models can be compared.

In addition to the baseline model, we use catchment-wide melt derived from the global ERA5-Land SWE reanalysis as a benchmark (Muñoz-Sabater et al., 2021). Mudryk et al. (2024) showed the ERA5-Land to be most accurate among 23 global SWE products with varying modeling and data assimilation strategies. While they did not explicitly evaluate melt rates, the ERA5-Land SWE melt-season variability was also most accurate in both mountainous and non-mountainous regions. It therefore represents a suitable benchmark for this study, as it is currently among the most reliable globally available SWE products and is available worldwide from 1950 onwards, making it directly applicable to the data-scarce settings that motivate this work.

3.3 Training strategy

The models are trained over water years (WY) 1986-2014. This corresponds to October 1985 to September 2014 in the Northern Hemisphere and April 1986 to March 2015 in the Southern Hemisphere. We keep a margin of close to one year within the full data availability period (1985-2015) to accommodate the past and future input windows. Any remaining gaps in streamflow input are filled following the input replacing strategy from Gauch et al. (2025), where gaps are set to a default value and flagged as such in an auxiliary binary input to the model. All input variables (dynamic forcings and static catchment attributes) and training targets are z-score standardized using mean and standard deviation estimated from the training period only. The same training statistics are used to normalize inputs at validation and test time; predicted targets are rescaled to physical units for evaluation.

We evaluate model performance using both temporal and spatial splitting strategies. We use the temporal splitting to test in-sample (IS) inversion performance. This will tell us how well we can reconstruct snowmelt in unseen years (e.g. pre-Landsat years) in catchments where present-day state-of-the-art SWE reanalysis products are available and the snowmelt-streamflow relationship is therefore known to a large degree. We divide the full period into five contiguous temporal blocks (four 6-year blocks and one 5-year block covering WY2010–WY2014). We then construct five folds by rotating these blocks, each time



240 assigning three blocks to training, one to validation, and one to testing. This rotating scheme reduces the risk of unrepresentative splits and improves robustness to non-stationary climatic conditions, such as recent decadal droughts in both the western United States (Hogan and Lundquist, 2024) and Chile (Garreaud et al., 2025). During IS training, all 163 catchments are used in each rotation.

245 Spatial splitting is carried out to test out-of-sample (OOS) inversion performance, which will be indicative of how well we can reconstruct snowmelt in catchments with streamflow observations but without an observationally-constrained, present-day SWE reanalysis. Similar to IS training, we generate ten blocks with equal proportions of Swiss, western US, and Chilean catchments, and use six blocks for training, one for validation and three for testing in each of the 5 folds. During OOS training, all 29 years are used in each rotation.

250 To account for stochasticity in network initialization and optimization, each rotation in both IS and OOS training and in each of the 4 configurations is repeated using five different random seeds to generate a small stochastic ensemble (Espinoza et al., 2024; Baste et al., 2025). Model training uses the Adam optimizer (Kingma and Ba, 2014) with an initial learning rate of 0.001 for the first 10 epochs, followed by 0.0005 for the remaining epochs. Training is performed for 20 epochs with a dropout rate of 0.4 and a batch size of 256. All results are based on the model parameters from the final (20th) training epoch. Following Kratzert et al. (2019), we use the catchment-averaged Nash-Sutcliffe Efficiency (NSE*) as the loss function, which is defined 255 as

$$\text{NSE}^* = \frac{1}{C} \sum_{c=1}^C \sum_{n=1}^N \frac{(\hat{M}_{c,n} - M_{c,n}^{\text{ref}})^2}{(s(c) + \epsilon)^2}, \quad (1)$$

where C is the number of catchments in the training set, N is the number of training samples per catchment c , $\hat{M}_{c,n}$ and $M_{c,n}^{\text{ref}}$ are simulated and reference snowmelt for sample n of catchment c , $s(c)$ is the standard deviation of reference melt for catchment c computed over the training period, and $\epsilon = 0.1$ is a small constant for numerical stability. Minimizing NSE* is 260 equivalent to minimizing a catchment-balanced, variance-normalized squared error, or maximizing NSE itself (Sect. 3.4.1).

3.4 Analysis

3.4.1 Model performance assessment

We evaluate the reconstructed smoothed melt by assessing how well it reproduces the reference reanalysis melt on a daily basis. Unlike the training procedure, we use the unadapted NSE defined as

$$265 \quad \text{NSE} = 1 - \frac{\sum_{t=1}^T (\hat{M}_t - M_t^{\text{ref}})^2}{\sum_{t=1}^T (M_t^{\text{ref}} - \overline{M^{\text{ref}}})^2} \quad (2)$$

where $\hat{M}_t$ and M_t^{ref} denote simulated and reference snowmelt at time step t , and $\overline{M^{\text{ref}}}$ is the mean reference melt over the evaluation period (Nash and Sutcliffe, 1970). The NSE ranges from $-\infty$ to 1, where 1 indicates perfect agreement, values greater than 0 indicate performance better than a mean benchmark, and negative values indicate poorer performance. We



compute the NSE for each individual catchment–year rather than averaging over all training windows within a catchment as
270 during training. To focus the evaluation on meaningful melt signals, metrics are calculated only for days where both reference
and simulated melt are non-zero. This avoids artificially inflating performance scores through correct prediction of no-melt
conditions, which dominate large parts of the water year.

Because our focus is on snow-dominated regions, the evaluation is restricted to catchments with a strong snow signal.
Specifically, we tighten the catchment selection criteria from a minimum of 5% snowfall fraction and 5% mean annual snow
275 cover during training to 25% for both indicators during testing. This reduces the number of catchments from 163 to 110
(65 WUS, 28 CH, 17 CL), ensuring that performance metrics only reflect climates where snowmelt represents a dominant
hydrological process.

To assess whether two experimental settings produce significantly different reconstruction performance, we compare their
catchment-year NSE distributions using paired Wilcoxon signed-rank tests (Woolson, 2008). The test evaluates whether posi-
280 tive and negative differences between paired catchment-years are balanced, with the null hypothesis being an equal number of
positive and negative differences. Differences are considered significant at the 5% significance level ($p < 0.05$).

3.4.2 LSTM input feature attribution

We expect streamflow to carry information relevant to reconstructing smoothed melt over preceding timesteps, complementing
meteorological inputs that primarily constrain melt over subsequent timesteps. Although the LSTM does not explicitly encode
285 governing physical processes, it may still learn statistical associations between streamflow, meteorology, and melt from data.

To directly address the third research question, which concerns the relative reliance of the model on streamflow versus
meteorological inputs and how this dependence varies across timesteps, we conduct an input feature attribution analysis using
the Integrated Gradients algorithm (IG; Sundararajan et al. (2017)), implemented with the Captum python package (Kokhlikyan
et al., 2020) and previously used for LSTM input feature attribution in hydrology (Kratzert et al., 2021).

290 IG attributes the model output to its inputs by integrating gradients along a continuous path between a user-defined baseline
input x' and the observed input x . For an input feature i , the attribution is defined as:

$$A_i^{\text{IG}}(x) = (x_i - x'_i) \int_{\alpha=0}^1 \frac{\partial f(x' + \alpha(x - x'))}{\partial x_i} d\alpha, \quad (3)$$

where $A_i^{\text{IG}}(x)$ denotes the attribution assigned to input feature i by the IG method, f denotes the model, and α parametrizes
the linear interpolation between baseline and input. In practice, this integral is approximated using a finite number of Riemann
295 steps, during which gradients of the model output with respect to the inputs are evaluated.

The resulting attribution scores quantify how strongly each input variable and timestep contributes to the model output
along this path. Because IG operates directly on the full input sequence, it naturally provides attribution values at the level of
individual timesteps, enabling analysis of how information content varies across past and future input windows. In addition,



IG can be computed for each individual prediction and catchment, allowing for both local (catchment- and event-specific) and aggregated (regional or global) analyses of model behavior.

We aggregate attributions over 50 prediction windows for each test catchment and consider only absolute values, thereby focusing on the magnitude and timing of influence rather than the direction of effect. Because inputs and outputs are normalized prior to training, the resulting attribution scores are comparable across catchments. The analysis is conducted for all seeds and catchment folds of the B3 IS and OOS models, as this configuration includes the full set of dynamic inputs.

While IG is formally an attribution method, the resulting scores can be interpreted as an indicator of model sensitivity in a limited sense, reflecting how responsive the model output is to changes in inputs along the chosen baseline-to-input path. However, this notion of sensitivity differs from classical perturbation-based sensitivity analysis: IG does not measure the effect of explicit input perturbations under physically realistic scenarios. Instead, it quantifies model behavior, i.e. which inputs the learned mapping relies on. As such, attribution patterns may reflect statistical dependencies in the data and do not necessarily imply causal or physically meaningful relationships. We choose IG over alternative attribution methods such as SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017) because it scales efficiently to long sequential inputs and provides temporally resolved attributions without requiring sampling or independence assumptions between inputs.

The choice of baseline is critical, as it determines which aspects of model behavior are emphasized in the attribution analysis. We use a catchment climatology baseline, in which dynamic inputs are replaced by their catchment-specific median values over the test period, while static attributes are kept at their observed values. This baseline preserves catchment identity and quantifies how strongly the model's melt reconstruction depends on deviations of the dynamic inputs from typical conditions at that site. It therefore enables assessment of the relative importance among dynamic inputs, including the contribution of streamflow compared to meteorological variables.

3.4.3 Controls on the snowmelt information content of streamflow

To understand if there are catchment attributes that explain the catchment variability in usefulness of streamflow for snowmelt reconstruction, we assess to what degree physiographic and climatic descriptors explain different measures of the snowmelt information content of streamflow.

The first measure is the melt NSE obtained from the B1 model (NSE_{B1}), which isolates the predictive capacity of streamflow and DOWY in unobserved catchments. The second measure is the difference in NSE (ΔNSE) between models B2 and B3, which quantifies the added (or reduced) value of streamflow relative to meteorological inputs at the catchment scale. The third measure is the relative streamflow attribution obtained from the B3 IG analysis (Sect. 3.4.2) using the catchment climatology baseline, expressed as a fraction of total input attribution ($f_{IG,Q}$, ranging from 0 to 100%). This measure reflects the reliance of the model on streamflow relative to other inputs. For each catchment, each measure is aggregated as the median across all eligible years, seeds, and folds.

We assess the influence of a selection of catchment descriptors that we expect to control the strength of the snowmelt-streamflow relationship. All descriptors are drawn from the Caravan static catchment attributes dataset (Table 2). Correlations between predictors and reconstruction performance are evaluated at the catchment level. The Caravan attributes provide a stan-



335 dardized description of long-term climatic, topographic, and hydrological catchment characteristics. Because these attributes are static, they do not capture year-to-year variability in meteorological and hydrological conditions, which may also influence reconstruction performance. However, deriving such dynamic descriptors consistently across all study catchments was considered beyond the scope of this work.

Table 2. Catchment attributes evaluated as predictors of streamflow information content, derived from the Caravan dataset (Kratzert et al., 2023).

Variable	Description
frac_snow	Snowfall fraction (-)
sgr_dk_sav	Stream gradient (-)
slp_df_sav	Terrain slope (-)
pet_mm_yr	Annual potential evapotranspiration (mm yr ⁻¹)
ttmp_dc_syr	Annual air temperature (°C)
aet_mm_syr	Annual actual evapotranspiration (mm yr ⁻¹)
ari_ix_sav	Aridity index (-)
snw_pc_syr	Annual snow cover fraction (%)
cmi_ix_syr	Climate moisture index (-)
swc_pc_syr	Annual soil water content (%)
area	Catchment area (km ²)
gla_pc_sse	Glacierized area fraction (%)

4 Results

4.1 Melt reconstruction performance across experiments

340 The streamflow-informed B3 models slightly outperform the streamflow-agnostic B2 models in both IS and OOS (Figs. 3-4). Across all temporal (IS) and catchment (OOS) folds, B3 performs better in a majority of catchment-years. This majority is Wilcoxon-significant in 5/5 (IS) and 3/5 (OOS) folds. Median B3 NSE values are 0.01 higher than B2 in both IS and OOS.

345 In IS, the difference between B3 and B2 (Δ NSE) is generally small, with both positive and negative values occurring within most catchments (Fig. 4), suggesting that the added value of streamflow is primarily year-dependent rather than catchment-dependent. In OOS, by contrast, Δ NSE values are larger and exhibit a stronger catchment dependence. Regional differences also become more pronounced in OOS: catchments in the WUS tend to mostly benefit from the inclusion of streamflow, whereas a majority of Swiss catchments perform better when relying on meteorological forcing alone. Overall, streamflow improves melt inversion performance in a majority of catchment-years, but its added value shows strong year-to-year variability and high dependence on the specific data fold, as well as slight regional differences in the melt inversion for out-of-sample catchments.



350 Both the B2 and B3 models outperform the B1 model, all IS models outperform all OOS models, and all models outperform the ERA5-Land benchmark, including the B0 climatology baseline model (Fig. 3). In all models the spread across the 5 cross-validation folds and 5 random seeds is substantial. See Fig. S1 for the performance across all configurations and experiments.

4.2 Example time series of three highlighted catchments

Figure 5 shows example years for the three highlighted catchments. They illustrate the climatic and physiographic differences
355 between the catchments, with the Merced catchment showing a strong melt-discharge correlation (Fig. 5a&b), the Dischma catchment showing high rainfall disturbance and a dampened streamflow response (Fig. 5c&d), and finally the Grande catchment showing a weak streamflow response, likely due to high sublimation losses (Fig. 5e&f). The ERA5-Land melt benchmark performs poorly in all cases, with particularly large deviations in the Grande catchment. Its melt estimates are too late and too high in most cases, suggesting a cold-bias in the ERA5-Land snow model, consequently compensated by excessively high melt
360 rates. The IS and OOS B1 models behave similarly and generally reproduce melt dynamics well in the Merced and Dischma catchments, except during certain periods where streamflow and reanalysis melt are inconsistent with each other. For example, in March 1994, WUS-SR simulates a pronounced melt peak in the Merced catchment without a corresponding streamflow response, leading the B1 models to underestimate melt. A similar situation occurs in April 1992 in the Dischma catchment, except here all models underestimate melt. Conversely, in early June 1992, the SPASS reanalysis indicates melt rates of around
365 7 mm d⁻¹, while streamflow peaks at 17 mm d⁻¹ due to combined rainfall and snowmelt. This results in melt overestimation by the B1 models, which lack explicit rainfall information. In the Grande catchment, the B1 models perform noticeably worse and resemble the B0 climatology baseline more, particularly in 1993. This reflects the weak linkage between SWE loss and discharge. With the exception of Grande 1993, the B2 and B3 models generally reproduce the reanalysis melt well. Merced 1991 exemplifies a case of strongly positive IS Δ NSE, with NSE increasing from 0.81 in the B2 model to 0.92 in the
370 B3, as streamflow information helps prevent late-season melt overestimation. In contrast, for Dischma 1994, the inclusion of streamflow leads to overestimation of mid-May and mid-June melt peaks, resulting in a negative Δ NSE of -0.1.

4.3 LSTM input attribution

Of the four dynamic inputs (temperature, precipitation, streamflow, and DOWY), the OOS B3 model relies most strongly on temperature for snowmelt reconstruction ($f_{IG,T} = 38.4\%$) and least on streamflow ($f_{IG,Q} = 16.8\%$). However, this dependence
375 is strongly catchment-dependent: some catchments exhibit near-zero streamflow attribution (e.g. Grande), while others exceed $f_{IG,Q} = 30\%$ (e.g. Merced). Chilean catchments differ from those in the western United States and Switzerland in exhibiting comparatively weaker temperature and streamflow attribution, and stronger reliance on DOWY and precipitation. The Martin Creek catchment in Nevada (camels_10329500) shows the highest temperature attribution at 64%, while the South Twin River catchment also in Nevada (camels_10249300) shows the highest DOWY attribution at 50%.

380 The per-timestep IG attribution analysis shows that the OOS B3 is most sensitive to dynamic inputs during the prediction window ($t \in [-7, 0]$), with attributions decreasing to near-zero in both past and future directions beyond 1–2 months (Fig. 7a). Temperature and precipitation dominate the attribution in the period leading up to the prediction window, followed by a rapid

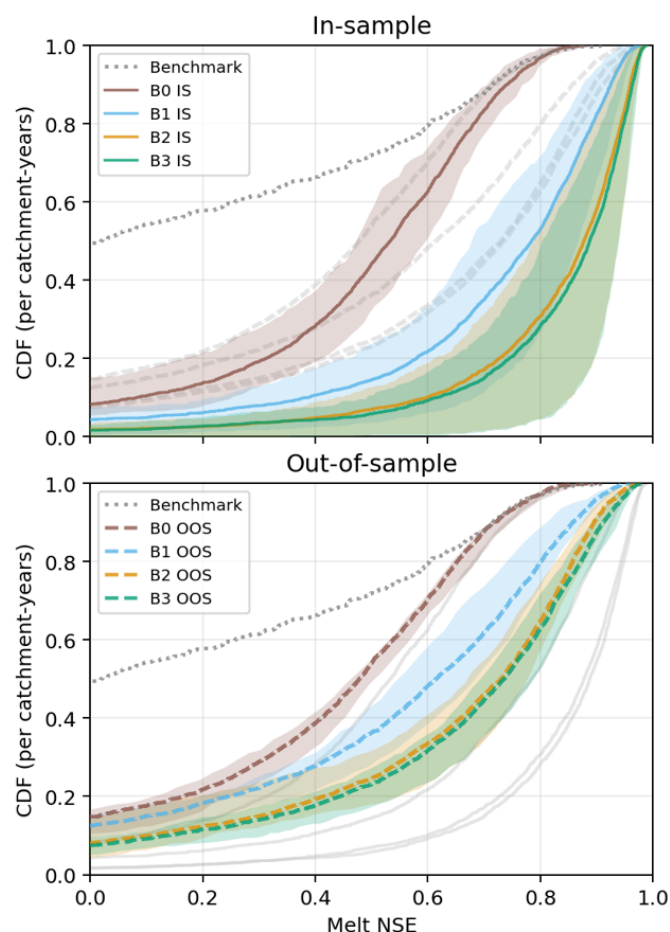


Figure 3. Cumulative distribution functions (CDFs, y-axis) for the reconstructed smoothed melt NSE values (x-axis) per individual year across all catchments (catchment-year), for the in-sample experiment evaluating generalization to new years (IS, panel a) and the out-of-sample experiment evaluating generalization to new catchments (OOS, panel b). Higher NSE values are better. The lines and shaded bands represent the median and minimum/maximum NSE across 5 cross-validation folds with 5 random seeds each. IS and OOS median results are mutually shown in semi-transparent grey lines. See Fig. 1 for the different input configurations among the evaluated models.

decline during the prediction window itself (Fig. 7b). Streamflow attribution peaks at the start of the prediction window and becomes dominant during its final three days, after which it drops sharply and gradually decays. This pattern suggests that the model exploits a strong same-day or near-synchronous melt–discharge relationship, with limited reliance on streamflow observations after the prediction window. The DOWY input shows two distinct attribution peaks at the start and end of the prediction window and dominates in the weeks following it, proving the importance of seasonal context. Precipitation attribution in the months preceding the prediction window is low when aggregated over all years, catchments and ensemble members, but shows distinct peaks corresponding to snowfall events when looking at individual runs (results not shown). Both the wet-summer,

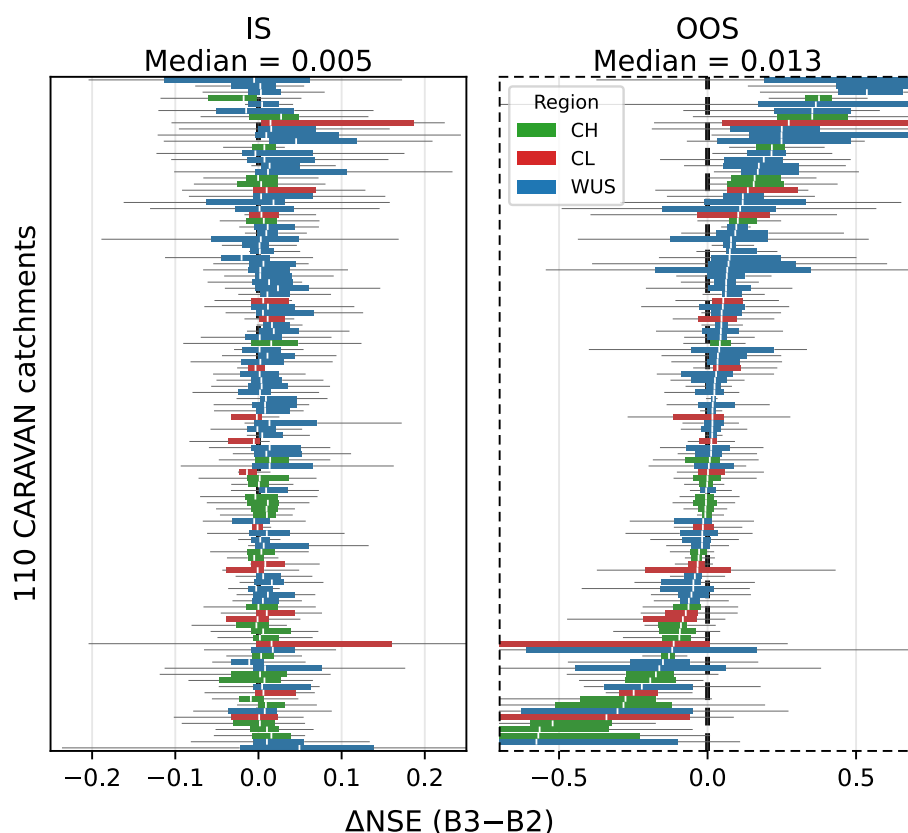


Figure 4. Difference in NSE (Δ NSE, x-axis) between the B3 (including streamflow) and the B2 (excluding streamflow) models, across all 110 test catchments (y-axis). Note the different limits on the x-axis. Boxes show the interquartile (IQR) range, white lines show the median, and whiskers extend to the most extreme values within $1.5 \times \text{IQR}$ of the quartiles.

slow-responding Dischma catchment and the dry-summer, fast-responding Merced catchment show above-average streamflow attribution across all timesteps (Fig. 7c). This suggests that the same-day melt-streamflow link can be dominant even when melt is translated to streamflow relatively inefficiently. The semiarid Grande catchment, by contrast, shows consistently low streamflow attribution, with only a modest peak during the prediction window.

4.4 Controls on the melt information content of streamflow

Catchment physiographic and climatic controls on the three measures of snowmelt information content in streamflow are generally weak and often not statistically significant (Fig. 8). Among the tested predictors, annual temperature, annual snow cover fraction, and snowfall fraction show the strongest relationships with IS NSE_{B1} , explaining 39, 34, and 29% of its variance, respectively. This suggests that streamflow contains more information on snowmelt in catchments with a stronger snow influence. In the OOS experiment, the same predictors remain statistically significant but explain substantially less variance. The

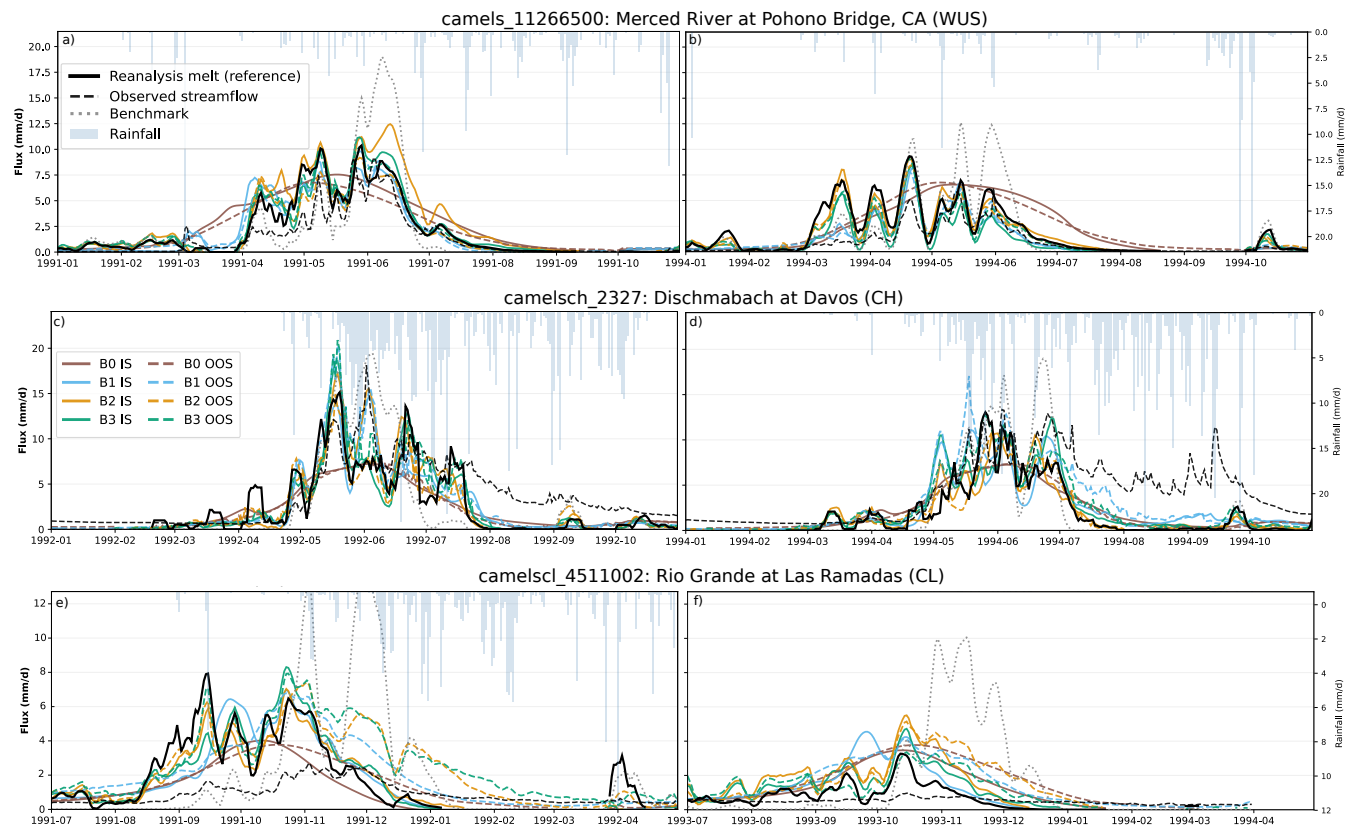


Figure 5. Example time series of observed streamflow (dashed), reference smoothed melt (solid) and reconstructed smoothed melt (color-coded) for several years in the three highlighted catchments. The ERA5-Land-derived melt serving as the benchmark is shown in grey dotted lines. Daily ERA5-Land rainfall hyetographs are shown on the inverted right y-axis. For each catchment, both y-axes share the same range. The time series shown correspond to a single seed and fold (out of five each) and are therefore intended for illustrative purposes only.

relative streamflow IG attribution ($f_{IG,Q}$) is weakly related to annual actual evapotranspiration, indicating that the snowmelt-streamflow link becomes less pronounced under high evaporative demand. Stream gradient explains 10% of the variance in ΔNSE in the OOS experiment, but is not a statistically significant predictor in any other analysis, suggesting that this relationship may be spurious. More generally, no catchment attribute explains a meaningful fraction of the variance in ΔNSE , indicating that the added value of streamflow beyond meteorological forcing is largely independent of the catchment characteristics considered here. The three measures of snowmelt information content are only weakly related to one another and are not explained by the same predictors. This suggests that they capture different aspects of the snow–streamflow relationship and should not be considered interchangeable. Finally, catchment area is not a meaningful predictor of the variance in any of the three measures, indicating that snowmelt information content does not systematically depend on catchment size, at least within the investigated range of catchment areas (10–2000 km²).

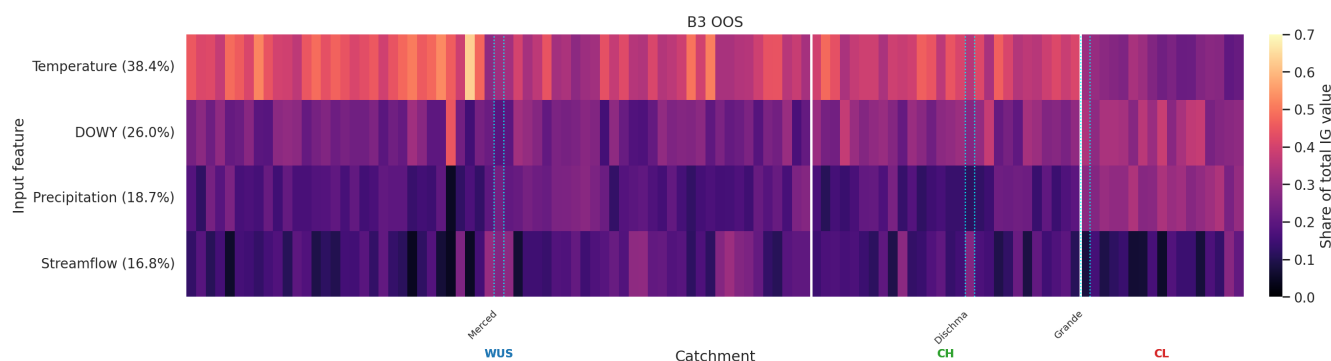


Figure 6. Integrated Gradients (IG) attribution of the B3 model for the four dynamic input variables (y-axis) for each of the 110 catchments (x-axis) in the B3 out-of-sample (OOS) experiment. Values are expressed as fractions of the total per-catchment attribution, with population-mean relative attribution indicated in parentheses on the y-axis. Catchments are grouped by region (Western US, Switzerland and Chile), with cyan dotted columns marking the three highlighted catchments. Results are based on the catchment climatology baseline, in which each dynamic input is replaced by its catchment-specific median over the test period.

410 5 Discussion

5.1 Streamflow contains substantial information on snowmelt

The B1 model, forced only with streamflow and DOWY, shows that streamflow contains substantial information on snowmelt, and that a BiLSTM is a suitable tool for extracting it. Performance is particularly high in the IS experiment, where the model is evaluated on unseen years from catchments included in the training set. This indicates that the BiLSTM is able to learn the relationship between snowmelt and streamflow from the reanalysis period and transfer it to years outside the training period.

415 Even in catchments that were not used during training (OOS experiment), meaningful melt estimates are obtained in most catchment-years. The large number and diversity of catchments used for training likely helped the model learn the snowmelt-streamflow relationship across a range of hydro-climatic conditions, enabling its transfer to new catchments. However, the larger performance variability observed in OOS compared to IS suggests that such transfer remains imperfect and dependent

420 on the characteristics of individual catchments.

The B1 model performs somewhat better in snow-dominated, rain-scarce catchments. This is consistent with Wiersma et al. (2026), who identified melt-season rainfall fraction as a secondary but important control on the effectiveness of snowmelt inference from streamflow. In contrast, we were unable to confirm their finding that catchment buffering capacity is the dominant control. We hypothesize that this discrepancy arises because the static Caravan attributes neither capture interannual variability

425 in storage behavior nor necessarily provide good proxies for catchment buffering capacity. As such, we cannot exclude that stronger relationships would emerge when using more detailed estimates of catchment storage behavior.

The BiLSTM reliance on streamflow differs strongly between catchments (Fig. 6-8), with a slight preference for catchments with low ET losses. These are mostly found in Chile and the interior western US. The BiLSTM is most sensitive to stream-

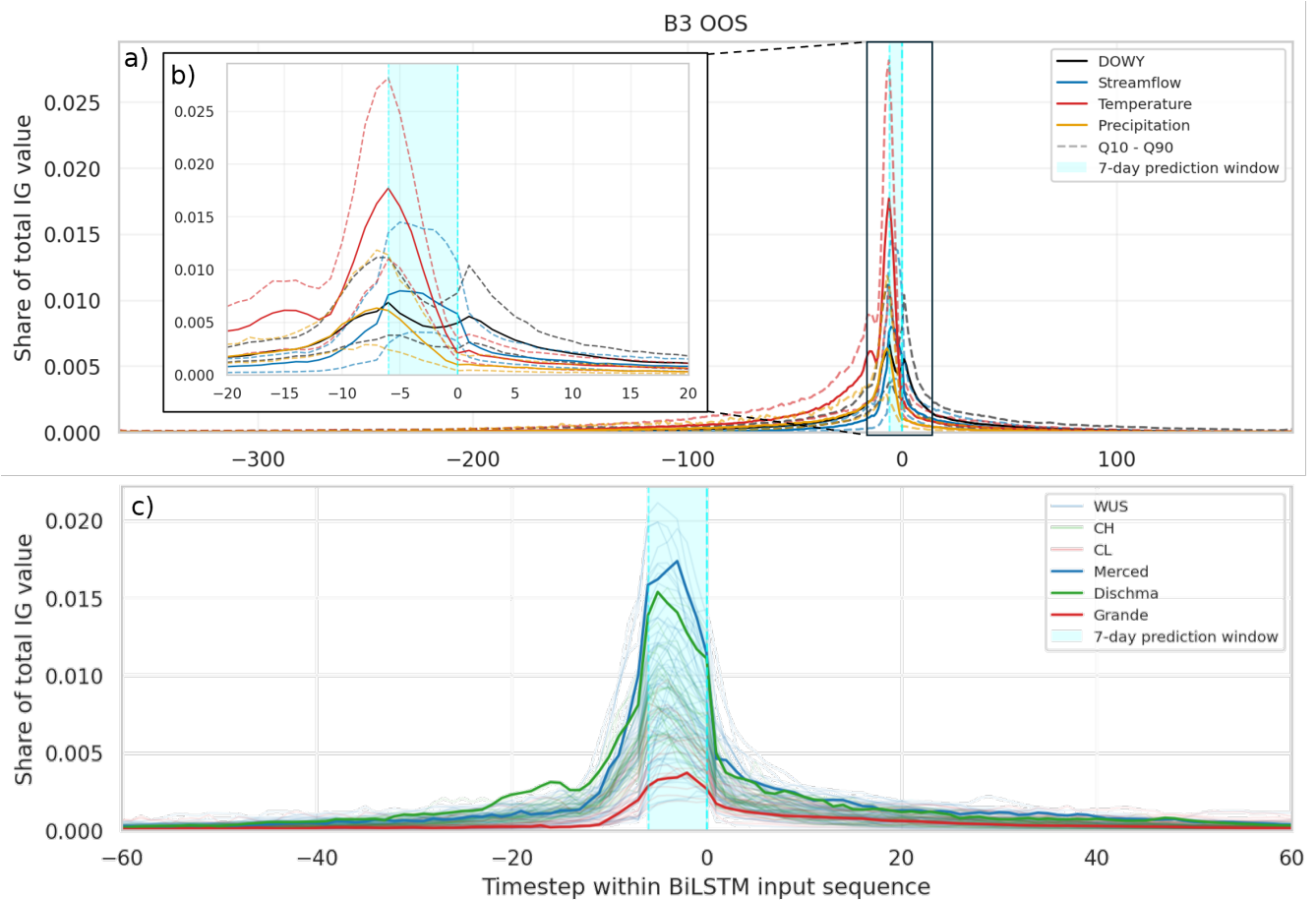


Figure 7. Temporal IG input feature attribution of the four dynamic inputs in the OOS B3 model, computed from the same catchment climatology baseline as in Fig. 6. Y-axes represent the contribution to the total attribution for each feature at each timestep. X-axes show the timestep within the full B3 model input sequence spanning 1.5 years between $t = -365$ and $t = 183$ days, with the prediction window between $t = -7$ and $t = 0$. Panels a & b show the 50th (solid), 10th and 90th (dashed) quantiles among all 110 catchments and all 25 ensemble members. Panel c shows only the relative temporal attribution of the streamflow input, with the three highlighted catchments in low transparency and the remaining 107 catchments in high transparency lines.

flow observations within the 7-day snowmelt prediction window, suggesting that streamflow contains information primarily on snowmelt occurring on the same day, and surprisingly little on snowmelt over recent days and weeks. The considerable BiLSTM sensitivity to antecedent streamflow could be an indication that this information is used to infer the snowmelt runoff efficiency (Taherian and Ameli, 2026).

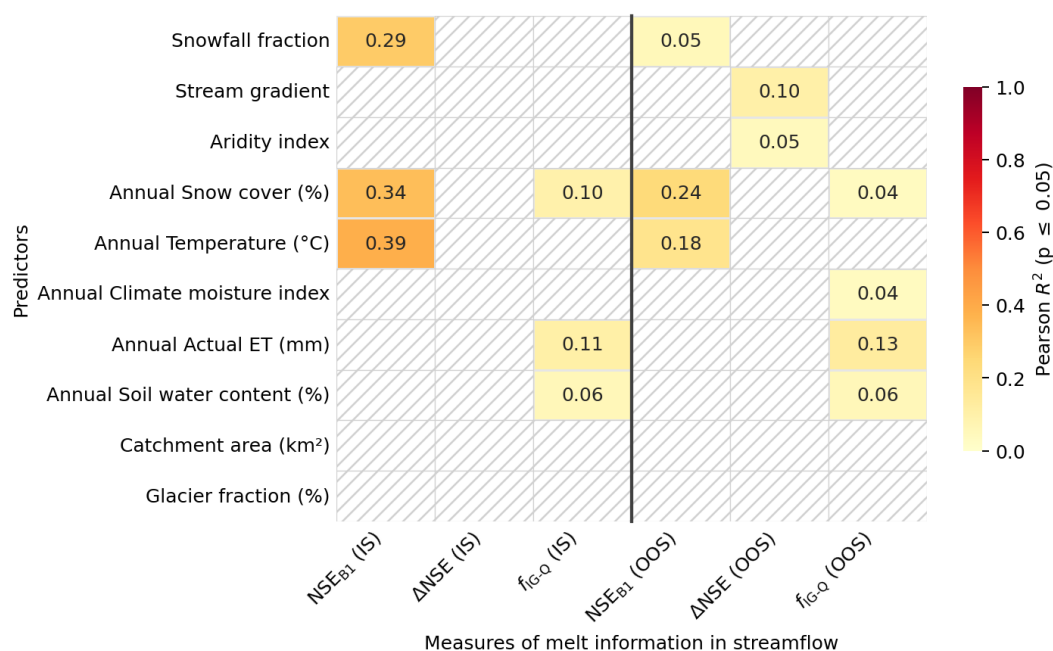


Figure 8. Heatmap showing the coefficient of determination (R^2) between meteorological and physiographic catchment descriptors (y-axis) and measures of the melt information content in streamflow (x-axis), for both IS and OOS experiments. Correlations are computed against catchment-level medians of each measure. Only statistically significant R^2 values are shown ($p \leq 0.05$).

5.2 Streamflow adds little beyond meteorological forcing

The meteorology-informed B2 and fully-informed B3 models both achieve high performance in the IS experiment and somewhat lower, but still generally good performance in the OOS experiment. In both cases, the B3 model only marginally outperforms B2, despite having access to streamflow observations that were shown by the B1 model to contain substantial information on snowmelt. We argue that these findings are not contradictory. Rather, they suggest that while streamflow contains meaningful information on snowmelt, much of the variability represented in the reanalysis target can already be recovered from meteorological inputs and seasonal context alone.

The high performance of the B2 models is somewhat surprising, given that they are forced with the same ERA5-Land meteorological data as the benchmark, which performs much worse (Fig. 3). The B0 baseline model demonstrates that many of the benchmark's cold and fast-melt biases can already be avoided by simply reproducing the average seasonal melt cycle (Fig. 5). This suggests that a substantial fraction of the benchmark errors arises from systematic biases rather than a lack of information in the forcing data. The strong performance gain of B2 can therefore be interpreted as a bias-correction effect, whereby the LSTM learns the systematic snowmelt biases of ERA5-Land during training and effectively corrects them during testing. Other studies have similarly demonstrated the ability of machine-learning models to learn relationships between coarse-



resolution meteorological inputs and snow observations (Cui et al., 2023; Thapa et al., 2020; Umirbekov et al., 2023; Kanda and Fletcher, 2025).

An additional explanation for the strong performance of B2 is that the SWE reanalysis products themselves are largely driven by meteorological forcing. Although they are subsequently constrained by snow observations, meteorological inputs remain a dominant driver of the final melt estimates. Reconstructing meteorologically driven SWE reanalysis melt rates may therefore be substantially easier than reconstructing true catchment-wide melt rates. Evidence for this interpretation comes from the several instances of apparent inconsistencies between melt and streamflow observed in the time series analysis (Fig. 5). Nonetheless, throughout this study we treat the SWE reanalyses as the reference and assume that the assimilated observations sufficiently constrain the reanalysis products for the reconstruction problem to remain meaningful. Investigating the influence of potential melt biases in the SWE reanalyses is beyond the scope of the present study.

In IS, the performance difference between the B2 and B3 models (ΔNSE) is generally small and primarily year-dependent. In certain years, streamflow might lead to positive ΔNSE values by providing information on snowmelt onset, meltout timing or strength of a snowmelt event. However, the fact that a considerable number of years within each catchment show negative ΔNSE values suggests that the bias-correcting capacities of streamflow are less reliable, and thus that the biases are less consistent between training and testing periods. The negative ΔNSE might be years where the reanalysis melt itself is melt-biased and inconsistent with streamflow. Since the SWE reanalysis products are primarily constrained by snow states rather than melt rates, such inconsistencies are plausible and are visible in some of the example years shown in Fig. 5.

The OOS experiment introduces an additional challenge, as both the meteorological bias-correction and the snowmelt-streamflow relationship must be transferred to entirely new catchments. Consequently, OOS ΔNSE values are larger and show a stronger catchment dependence than in the IS experiment. Conceptually, OOS performance depends on the transferability of two separate relationships: (i) the mapping between ERA5-Land forcing and reanalysis melt, and (ii) the mapping between snowmelt and streamflow. When both relationships transfer with similar success, ΔNSE values remain small and centered around 0 as in the IS experiment. Larger positive or negative ΔNSE values arise when one relationship transfers more successfully than the other. However, the generally weak correlations between ΔNSE and the catchment descriptors considered in Sect. 4.4, as well as between ΔNSE and NSE_{B1} , indicate that the controls on streamflow's added value remain difficult to identify and are likely influenced by factors not captured in our analysis.

5.3 Practical implications and future work

For the Western US, Swiss and Chilean catchments used in this study, the IS experiment demonstrates that it is feasible to use a BiLSTM to extend snowmelt reanalyses in time. This effectively represents a transfer of catchment-specific information from recent, observation-constrained years to historical periods for which only coarse-scale meteorological forcing is available, enabled by the existence of high-quality snow reanalysis products. Extension of snowmelt reanalyses to new catchments is less reliable, but still likely to yield better melt estimates than melt derived directly from ERA5-Land SWE, despite it being considered the most reliable global SWE product (Mudryk et al., 2024). Both in-sample and out-of-sample extensions can be performed with or without streamflow as an additional predictor, as the B2 and B3 models achieved broadly similar perfor-



mance. When using the meteorology-informed B2 model, the snowmelt reanalysis extension presented here does not depend on the availability of streamflow measurements anymore, which would greatly increase the training data potential.

While we focused on snowmelt, the primary variable provided by most snow reanalysis products is SWE rather than snowmelt. The strong performance of the B2 and B3 models suggests that they may also prove effective when trained directly on (gridded) SWE. When SWE is the target variable, streamflow may provide information not only on melt rates, but also on the total seasonal melt volume. This could be better exploited by evaluating model performance over longer temporal windows than the 7-day prediction window used here, thereby allowing streamflow to constrain both the timing of snowmelt and the total amount of water released from the seasonal snowpack.

For streamflow to become more informative in the temporal and spatial extension of SWE or snowmelt reanalyses, the reanalyses themselves may need to become more strongly constrained by melt observations (Besso et al., 2026). Currently, the reanalyses used here are constrained primarily by snow states, either through snow disappearance observations (WUS-SR and Andes-SR) or snow depth observations converted to SWE estimates (SPASS). These constrain melt rates only indirectly, leaving potential catchment-wide melt biases undetected. Streamflow observations themselves could in theory be used to constrain melt rates within SWE reanalysis frameworks, but the absence of catchment-scale melt observations makes evaluation of such an approach challenging. Alternatively, Sentinel-1 C-band SAR observations offer a promising source of melt-related information. Through the detection of the different phases of snow ripening (Marin et al., 2020), these observations can provide global, high-resolution estimates of snowmelt onset dates (Gagliano et al., 2026). Premier et al. (2023) presented a methodology of integrating this information into a SWE reanalysis framework that also integrates snow disappearance and in-situ snow depth measurements. Such melt-informed reanalysis frameworks could improve not only contemporary reanalysis products themselves, but also the quality of the training targets used for historical reconstructions such as those developed in this study. In addition, SWE reanalyses with a more realistic representation of snowmelt-streamflow interactions may increase the added value of streamflow for SWE and snowmelt reconstruction beyond contemporary reanalyses.

We further encourage future work to expand the combined SWE reanalysis and large-sample hydrological dataset introduced here. Norway and Austria both contain long-term (since 1957 and 1961 resp.) high-resolution SWE products and are included in the Caravan GRDC extension, making them straightforward candidates for expanding the dataset (Olefs et al., 2020; Saloranta, 2016). Additional opportunities are emerging through ongoing developments in France, Italy, Iceland and other regions where either long-term SWE products, large-sample hydrological datasets, or both are becoming available (Avanzi et al., 2023; Delaigue et al., 2025; Vernay et al., 2022; Helgason and Nijssen, 2024; Gunnarsson and Gardarsson, 2023). Expanding the geographical coverage of the dataset would increase both training diversity and reconstruction potential, while enabling evaluation across a wider range of snow climates. Finally, the dataset opens several additional research directions. For example, it may be possible to investigate whether streamflow contains information on melt originating from different elevation bands, whether regional meteorological products or additional forcing variables such as radiation, humidity and wind speed improve reconstruction skill, and whether streamflow observations can help distinguish between melt and sublimation losses rather than merely inferring total SWE loss.



515 6 Conclusion

Using a novel dataset combining multiregional SWE reanalyses, large-sample hydrological observations, and catchment attributes, we trained BiLSTM models with different input configurations to reconstruct catchment-wide snowmelt derived from SWE reanalyses. The models were evaluated both on unseen years from catchments included during training (in-sample; IS) and on entirely unseen catchments (out-of-sample; OOS). Training used 163 snow-influenced catchments, and evaluation was
520 restricted to a subset of 110 snow-dominated catchments in the western United States, Switzerland, and Chile.

Answering the three research questions posed in the introduction, our results show that streamflow contains substantial and extractable information on catchment-scale snowmelt across a wide range of snow-dominated catchments and hydro-climatic settings. At the same time, meteorological forcing alone proved highly effective in reconstructing melt from existing SWE reanalyses, despite its known biases and coarse spatial resolution. As a result, streamflow provided only limited additional
525 predictive value when transferring snowmelt information from SWE reanalyses to years and catchments not covered by those reanalyses. This limited added value was observed consistently across catchments and could not be explained well by the considered catchment characteristics. From a practical perspective, these findings suggest that BiLSTMs offer a promising approach for extending snowmelt information beyond the spatial and temporal coverage of existing SWE reanalyses. We conclude by encouraging future research to integrate streamflow observations into existing SWE reanalysis frameworks alongside
530 traditional snow observation sources. By providing direct information on catchment-scale melt dynamics, streamflow could help produce SWE reanalyses that are more strongly constrained on melt. In turn, these improved reanalyses would serve as more informative training datasets for future efforts to extend SWE and snowmelt information to regions and periods lacking snow observations.

Code and data availability. All code, data, and instructions necessary to reproduce this work are available at
535 <https://github.com/pauwiersma/BiLSTM4snowmelt> and <https://doi.org/10.5281/zenodo.22283784> (Wiersma, 2026) respectively. The Caravan large-sample hydrological dataset is available at <https://zenodo.org/records/7540792> (Kratzert et al., 2025). The Swiss SPASS SWE reanalysis is available at <https://doi.org/10.16904/envdat.580> (Marty et al., 2025b). The WUS-SR and Andes-SR SWE reanalyses are available at <https://doi.org/10.5067/PP7T2GBI52I2> (Fang et al., 2022a) and <https://ucla.app.box.com/v/ANDES-SWE-REANALYSIS> (Cortés and Margulis, 2017) respectively.

540 Appendix A: Availability of streamflow measurements in snow-dominated catchments

Among snow-influenced catchments included in the Caravan large-sample hydrological dataset and its GRDC extension (Kratzert et al., 2025; Färber et al., 2025), a total of 101,732 years of streamflow data are available across 2,495 snow-influenced catchments. Of these, 69,648 years precede the onset of MODIS observations (2000), and 41,335 predate the launch of Landsat (1985) (see Fig. A1 and Supplementary Material S1 for details on the calculation). Among them, 40,480 pre-MODIS and
545 23,833 pre-Landsat years are located in 1,371 mountainous snow-influenced catchments. Note that the Caravan records start at



1950 and might exclude many years in the preceding decades. Including streamflow observations from additional regions not yet included in Caravan (e.g. Eastern European, Asian, Middle Eastern, French and Italian mountain regions) would further increase these numbers. Leveraging streamflow records to extract information on snow dynamics could therefore substantially improve SWE reanalyses across many years and catchments that are currently under-sampled by in situ or remote sensing observations, and which otherwise rely on high-uncertainty meteorological forcing data alone (Kanda and Fletcher, 2025).

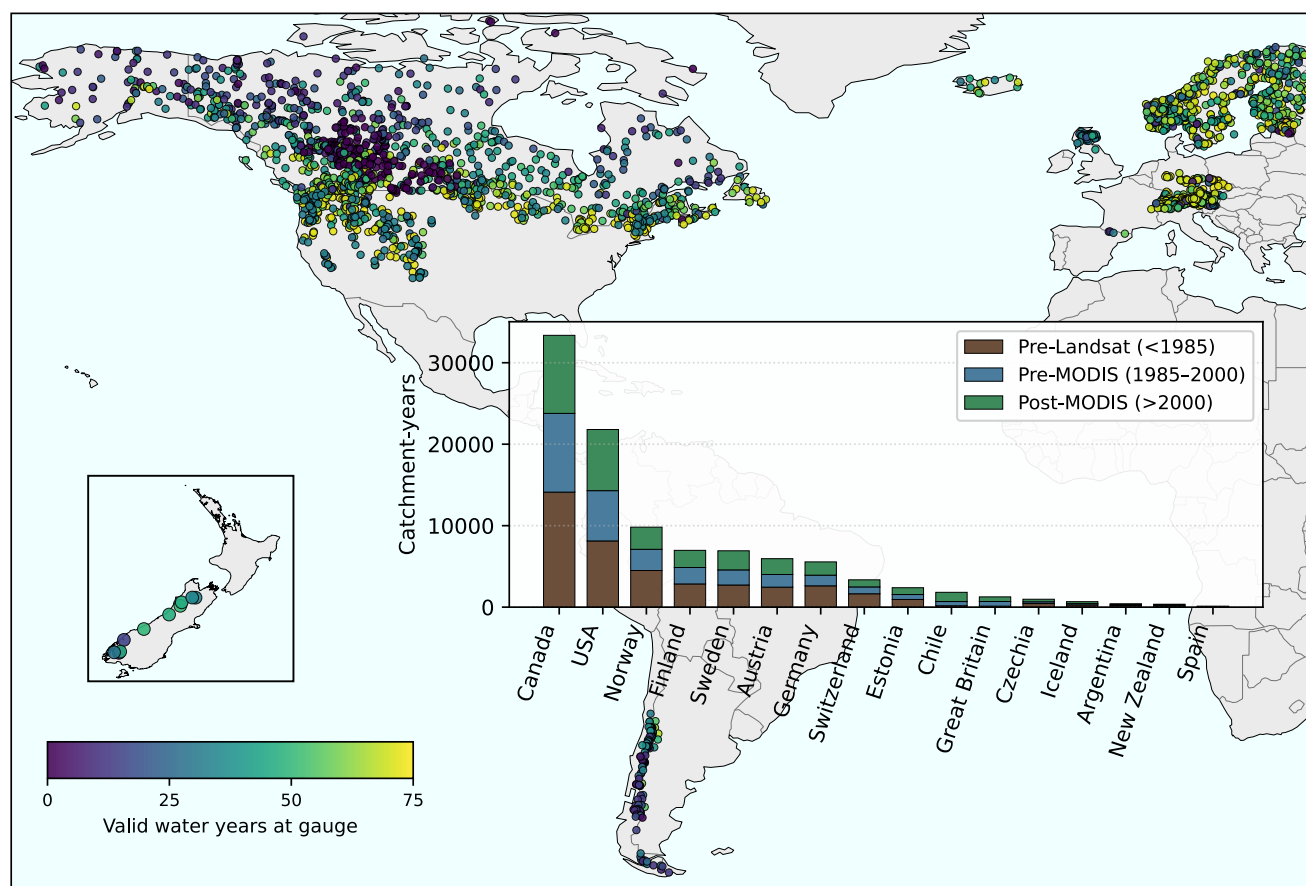


Figure A1. Number of years per snow-dominated catchment with streamflow measurements in the Caravan large-sample hydrological dataset (Kratzert et al., 2025). Only catchments with more than 25% mean annual snow cover and less than 10% glacier cover are shown.

Appendix B: Snow classes

The ephemeral snow class is largely represented in all regions, covering most of the lower elevations within each catchment (Fig. B1). Otherwise, WUS contains nearly all classes, with transitional montane forest dominating. Swiss catchments contain



mostly transitional maritime and seasonal tundra, while Chilean catchments contain mostly transitional prairie and transitional tundra.

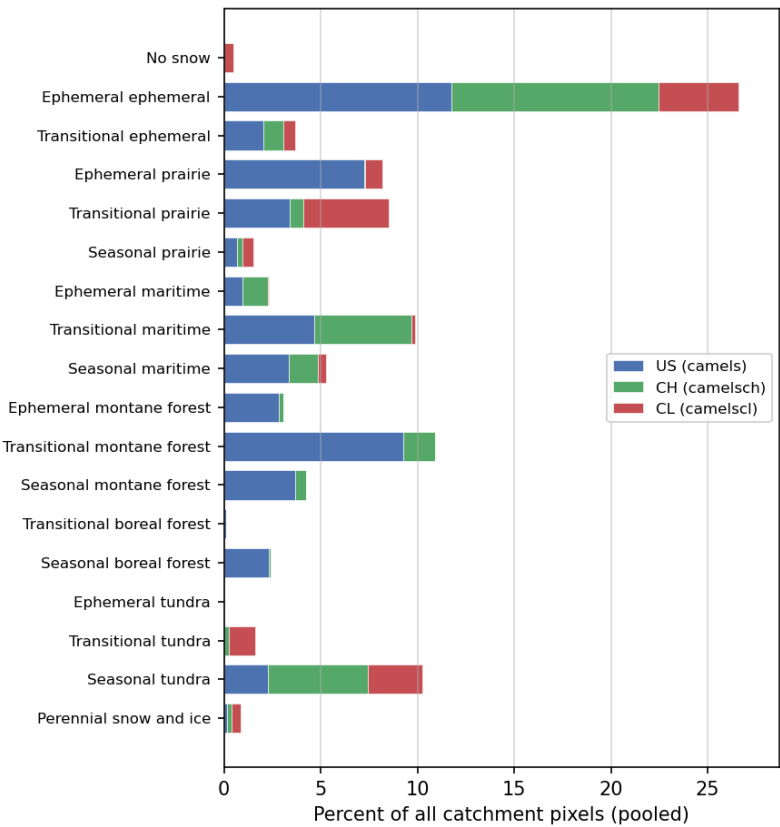


Figure B1. Cumulative percentages of the 163 snow-influenced catchments covered by each of the snow classes described in Johnston et al. (2026). Calculations are based on their gridded 1km SnowMAP dataset. Bar length is the share of all valid catchment pixels (water/no-data excluded) across all catchments. Segments show the contribution of each region (US/CAMELS, Switzerland/CAMELS-CH, Chile/CAMELS-CL), so segment lengths are area-weighted by pixel count rather than equal weight per catchment.

555

Appendix C: Catchment variability

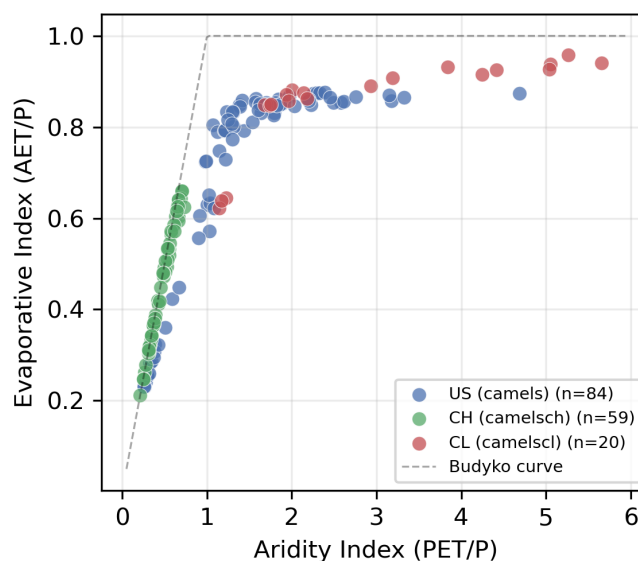


Figure C1. Budyko space for the filtered Caravan catchments used in training (US, Switzerland, Chile). Each point is a basin, plotted with HydroATLAS long-term mean potential and actual evapotranspiration and precipitation as aridity (PET/P) and evaporative index (AET/P). The dashed line is the theoretical energy- and water-limited envelope, $\min(PET/P, 1)$.

Author contributions. Conceptualization: PW, RL, GM. Methodology: PW, RL, KH, GM. Formal analysis: PW. Supervision: RL, GM. Visualization: PW. Software: PW, SB. Writing – original draft preparation: PW. Writing – review and editing: PW, RL, SB, KH, GM.

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

560 *Acknowledgements.* We thank Mathieu Gravey for providing the GPU on which the BiLSTMs were trained. Large language models were used to support the programming and to improve the clarity and phrasing of the manuscript.



References

- Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, *Hydrology and Earth System Sciences*, 21, 5293–5313, <https://doi.org/10.5194/hess-21-5293-2017>, 2017.
- 565 Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset, *Hydrology and Earth System Sciences*, 22, 5817–5846, <https://doi.org/10.5194/hess-22-5817-2018>, 2018.
- Avanzi, F., Gabellani, S., Delogu, F., Silvestro, F., Pignone, F., Bruno, G., Pulvirenti, L., Squicciarino, G., Fiori, E., Rossi, L., Puca, S., Toniazzi, A., Giordano, P., Falzacappa, M., Ratto, S., Stevenin, H., Cardillo, A., Fioletti, M., Cazzuli, O., Cremonese, E., Cella, U. M. d., 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.
- 570 Ayala, , Schauwecker, S., and MacDonell, S.: Spatial distribution and controls of snowmelt runoff in a sublimation-dominated environment in the semiarid Andes of Chile, *Hydrology and Earth System Sciences*, 27, 3463–3484, <https://doi.org/10.5194/hess-27-3463-2023>, 2023.
- Bair, E. H., Calfa, A. A., Rittger, K., and Dozier, J.: Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan, *The Cryosphere*, 12, 1579–1594, <https://doi.org/10.5194/tc-12-1579-2018>, 2018.
- 575 Baste, S., Klotz, D., Espinoza, E. A., Bardossy, A., and Loritz, R.: Unveiling the limits of deep learning models in hydrological extrapolation tasks, *Hydrology and Earth System Sciences*, 29, 5871–5891, <https://doi.org/10.5194/hess-29-5871-2025>, 2025.
- Besso, H., Mower, R., Pflug, J. M., and Lundquist, J. D.: Mapping 1 April SWE in the Western US Using Standardized Anomalies and Quantiles From SWE Reanalysis and In Situ Stations, *Water Resources Research*, 62, <https://doi.org/10.1029/2025wr040902>, 2026.
- 580 Brooks, P. D., Solomon, D. K., Kampf, S., Warix, S., Bern, C., Barnard, D., Barnard, H. R., Carling, G. T., Carroll, R. W. H., Chorover, J., Harpold, A., Lohse, K., Meza, F., McIntosh, J., Neilson, B., Sears, M., and Wolf, M.: Groundwater dominates snowmelt runoff and controls streamflow efficiency in the western United States, *Communications Earth & Environment*, 6, 341, <https://doi.org/10.1038/s43247-025-02303-3>, 2025.
- Cortés, G. and Margulis, S.: Impacts of El Niño and La Niña on interannual snow accumulation in the Andes: Results from a high-resolution 31 year reanalysis, *Geophysical Research Letters*, 44, 6859–6867, <https://doi.org/10.1002/2017gl073826>, 2017.
- 585 Cui, G., Anderson, M., and Bales, R.: Mapping of snow water equivalent by a deep-learning model assimilating snow observations, *Journal of Hydrology*, 616, 128 835, <https://doi.org/10.1016/j.jhydrol.2022.128835>, 2023.
- Delaigue, O., Guimarães, G. M., Brigode, P., Génot, B., Perrin, C., Soubeyroux, J.-M., Janet, B., Addor, N., and Andréassian, V.: CAMELS-FR dataset: a large-sample hydroclimatic dataset for France to explore hydrological diversity and support model benchmarking, *Earth System Science Data*, 17, 1461–1479, <https://doi.org/10.5194/essd-17-1461-2025>, 2025.
- 590 Espinoza, E. A., Loritz, R., Chaves, M. , Bäuerle, N., and Ehret, U.: To bucket or not to bucket? Analyzing the performance and interpretability of hybrid hydrological models with dynamic parameterization, *Hydrology and Earth System Sciences*, 28, 2705–2719, <https://doi.org/10.5194/hess-28-2705-2024>, 2024.
- Fang, Y., Liu, Y., and Margulis, S.: Western United States UCLA Daily Snow Reanalysis, Version 1, <https://doi.org/10.5067/PP7T2GBI52I2>, 2022a.
- 595 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>, 2022b.



- Fang, Y., Liu, Y., Li, D., Sun, H., and Margulis, S. A.: Spatiotemporal snow water storage uncertainty in the midlatitude American Cordillera, *The Cryosphere*, 17, 5175–5195, <https://doi.org/10.5194/tc-17-5175-2023>, 2023.
- 600 Färber, C., Plessow, H., Mischel, S., Kratzert, F., Addor, N., Shalev, G., and Looser, U.: GRDC-Caravan: extending the original dataset with data from the Global Runoff Data Centre, <https://doi.org/10.5281/zenodo.15349031>, 2025.
- Gagliano, E., Shean, D., and Henderson, S.: A global high-resolution dataset of snowmelt runoff onset timing from Sentinel-1 SAR, 2015–2024, *Earth System Science Data Discussions*, 2026, 1–41, <https://doi.org/10.5194/essd-2026-216>, 2026.
- Garreaud, R., Boisier, J. P., Alvarez-Garretón, C., Christie, D. A., Carrasco-Escaff, T., Vergara, I., Chávez, R. O., Aldunce, P., Camus, P., Suazo-Álvarez, M., Masiokas, M., Castro, G., Muñoz, A., Zambrano-Bigiarini, M., Fuster, R., and Godoy, L.: Hyperdroughts in central Chile: drivers, impacts, and projections, *Hydrology and Earth System Sciences*, 29, 5347–5369, [https://doi.org/10.5194/hess-29-5347-](https://doi.org/10.5194/hess-29-5347-2025)
605 2025, 2025.
- Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Cohen, D., and Gilon, O.: How to deal w___ missing input data, *Hydrology and Earth System Sciences*, 29, 6221–6235, <https://doi.org/10.5194/hess-29-6221-2025>, 2025.
- 610 Gordon, B. L., Brooks, P. D., Krogh, S. A., Boismore, 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.
- Graves, A. and Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM and other neural network architectures, *Neural Networks*, 18, 602–610, <https://doi.org/10.1016/j.neunet.2005.06.042>, 2005.
- 615 Gunnarsson, A. and Gardarsson, S. M.: Spatial Estimation of Snow Water Equivalent for Glaciers and Seasonal Snow in Iceland Using Remote Sensing Snow Cover and Albedo, *Hydrology*, 11, 3, <https://doi.org/10.3390/hydrology11010003>, 2023.
- Guo, Y., Yang, Y., Yang, D., Zhang, L., Zheng, H., Xiong, J., Ruan, F., Han, J., and Liu, Z.: Warming leads to both earlier and later snowmelt floods over the past 70 years, *Nature Communications*, 16, 3663, <https://doi.org/10.1038/s41467-025-58832-0>, 2025.
- 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-
620 dwindling world, *Nature*, 629, 1075–1081, <https://doi.org/10.1038/s41586-024-07299-y>, 2024.
- Helgason, H. B. and Nijssen, B.: LamaH-Ice: LARge-SaMPle DATA for Hydrology and Environmental Sciences for Iceland, *Earth System Science Data*, 16, 2741–2771, <https://doi.org/10.5194/essd-16-2741-2024>, 2024.
- 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.
- 625 Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, 9, 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>, 1997.
- Hock, R., Rasul, G., Adler, C., Cáceres, B., Gruber, S., Hirabayashi, Y., Jackson, M., Kääb, A., Kang, S., Kutuzov, S., Milner, A., Molau, U., Morin, S., Orlove, B., and Steltze, H.: The Ocean and Cryosphere in a Changing Climate - Chapter 2: High Mountain Areas, pp. 131–202, <https://doi.org/10.1017/9781009157964.004>, 2019.
- 630 Hogan, D. and Lundquist, J. D.: Recent Upper Colorado River Streamflow Declines Driven by Loss of Spring Precipitation, *Geophysical Research Letters*, 51, <https://doi.org/10.1029/2024gl109826>, 2024.
- Hogan, D., Schwat, E., Gutmann, E., Vano, J., and Lundquist, J. D.: Synoptic-Scale Systems Control Total Winter Sublimation at an Alpine Site, *Journal of Hydrometeorology*, 27, 529–547, <https://doi.org/10.1175/jhm-d-25-0143.1>, 2026.
- 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.
- 635



- 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.: Catchment attributes and hydro-meteorological time series for large-sample studies across hydrologic Switzerland (CAMELS-CH), <https://doi.org/10.5281/zenodo.10354485>, 2023.
- 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.: Catchment attributes and hydro-meteorological time series for large-sample studies across hydrologic Switzerland (CAMELS-CH), <https://doi.org/10.5281/zenodo.15025258>, 2025.
- Imambi, S., Prakash, K. B., and Kanagachidambaresan, G. R.: Programming with TensorFlow, Solution for Edge Computing Applications, EAI/Springer Innovations in Communication and Computing, pp. 87–104, https://doi.org/10.1007/978-3-030-57077-4_10, 2021.
- Johnston, J., Jacobs, J. M., Vardaman, M., and Cho, E.: The snow meteorology and phenology classification (SnowMAP): global snow cover observations enhance snow's representation, *Scientific Reports*, 16, 14 075, <https://doi.org/10.1038/s41598-026-44321-x>, 2026.
- Kanda, N. and Fletcher, C. G.: Evaluating a hierarchy of bias correction methods for ERA5-Land SWE across Canada, *Environmental Research Communications*, 7, 085 015, <https://doi.org/10.1088/2515-7620/aded5a>, 2025.
- Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *arXiv*, <https://doi.org/10.48550/arxiv.1412.6980>, 2014.
- 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.
- Klotz, D., Kratzert, F., Gauch, M., Sampson, A. K., Brandstetter, J., Klambauer, G., Hochreiter, S., and Nearing, G.: Uncertainty estimation with deep learning for rainfall-runoff modeling, *Hydrology and Earth System Sciences*, 26, 1673–1693, <https://doi.org/10.5194/hess-26-1673-2022>, 2022.
- Knoben, W. J. M., Thébault, C., Keshavarz, K., Torres-Rojas, L., Chaney, N. W., Pietroniro, A., and Clark, M. P.: Catchment Attributes and Meteorology for Large-Sample SPATIally distributed analysis (CAMELS-SPAT): streamflow observations, forcing data and geospatial data for hydrologic studies across North America, *Hydrology and Earth System Sciences*, 29, 5791–5833, <https://doi.org/10.5194/hess-29-5791-2025>, 2025.
- Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., Melnikov, A., Kliushkina, N., Araya, C., Yan, S., and Reblitz-Richardson, O.: Captum: A unified and generic model interpretability library for PyTorch, <https://doi.org/10.48550/arXiv.2009.07896>, 2020.
- Koskela, J. J., Croke, B. W. F., Koivusalo, H., Jakeman, A. J., and Kokkonen, T.: Bayesian inference of uncertainties in precipitation-streamflow modeling in a snow affected catchment, *Water Resources Research*, 48, <https://doi.org/10.1029/2011wr011773>, 2012.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks, *Hydrology and Earth System Sciences*, 22, 6005–6022, <https://doi.org/10.5194/hess-22-6005-2018>, 2018.
- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, *Hydrology and Earth System Sciences*, 23, 5089–5110, <https://doi.org/10.5194/hess-23-5089-2019>, 2019.
- Kratzert, F., Klotz, D., Hochreiter, S., and Nearing, G. S.: A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall-runoff modeling, *Hydrology and Earth System Sciences*, 25, 2685–2703, <https://doi.org/10.5194/hess-25-2685-2021>, 2021.
- Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and Matias, Y.: Caravan - A global community dataset for large-sample hydrology, *Scientific Data*, 10, 61, <https://doi.org/10.1038/s41597-023-01975-w>, [tex.pmcid: PMC9887008](https://pubmed.ncbi.nlm.nih.gov/40887008/), 2023.



- Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and
675 Matias, Y.: Caravan - A global community dataset for large- sample hydrology (csv version), <https://doi.org/10.5281/zenodo.15530022>,
2025.
- Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., Bruijn, J. D., Sahu, R. K., Greve, P., Slater, L., and Dadson, S. J.: Hydro-
logical concept formation inside long short-term memory (LSTM) networks, *Hydrology and Earth System Sciences*, 26, 3079–3101,
<https://doi.org/10.5194/hess-26-3079-2022>, 2022.
- 680 Linke, S., Lehner, B., Dallaire, C. O., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F.,
and Thieme, M.: Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution, *Scientific Data*, 6, 283,
<https://doi.org/10.1038/s41597-019-0300-6>, 2019.
- Liu, Z., Sun, Y., Li, Y., Liu, Y., Chen, Y., and Zhang, Y.: Lithium-ion battery health prognosis via electrochemical impedance spectroscopy
using CNN-BiLSTM model, *Journal of Materials Informatics*, 4, <https://doi.org/10.20517/jmi.2024.09>, 2024.
- 685 Lundberg, S. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, <https://arxiv.org/abs/1705.07874>, 2017.
- Lundquist, J. D. and Flint, A. L.: Onset of snowmelt and streamflow in 2004 in the western united states: How shading may affect spring
streamflow timing in a warmer world, *Journal of Hydrometeorology*, 7, 1199–1217, <https://doi.org/10.1175/jhm539.1>, 2006.
- Lundquist, J. D., Abel, M., Currier, W. R., and Jackson, D. L.: Snow what? Comparing snow products to reveal what works, where, and why,
<https://doi.org/10.22541/essoar.177100538.80471917/v1>, 2026.
- 690 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*, 17, 703–717, <https://doi.org/10.5194/essd-17-703-2025>,
2024.
- Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., and Diffenbaugh, N. S.: The potential for snow to supply human water demand in the
present and future, *Environmental Research Letters*, 10, 114 016, <https://doi.org/10.1088/1748-9326/10/11/114016>, 2015.
- 695 Manoj, A. J., Loritz, R., Gupta, H., and Zehe, E.: Can discharge be used to inversely correct precipitation?, *Hydrology and Earth System
Sciences*, 29, 6115–6135, <https://doi.org/10.5194/hess-29-6115-2025>, 2025.
- Margulis, S. A., Giroto, M., Cortés, G., and Durand, M.: A particle batch smoother approach to snow water equivalent estimation, *Journal
of Hydrometeorology*, 16, 1752–1772, <https://doi.org/10.1175/jhm-d-14-0177.1>, 2015.
- Marin, C., Bertoldi, G., Premier, V., Callegari, M., Brida, C., Hürkamp, K., Tschiersch, J., Zebisch, M., and Notarnicola, C.: Use of Sentinel-
700 1 radar observations to evaluate snowmelt dynamics in alpine regions, *The Cryosphere*, 14, 935–956, <https://doi.org/10.5194/tc-14-935-2020>, 2020.
- Marty, C., Michel, A., Jonas, T., Steijn, C., Muelchi, R., and Kotlarski, S.: SPASS – new gridded climatological snow datasets for Switzerland:
potential and limitations, *The Cryosphere*, 19, 4391–4407, <https://doi.org/10.5194/tc-19-4391-2025>, 2025a.
- Marty, C., Michel, A., Jonas, T., Steijn, C., Muelchi, R., and Kotlarski, S.: SPASS - new gridded climatological snow datasets for Switzerland,
705 <https://doi.org/http://dx.doi.org/10.16904/envidat.580>, 2025b.
- Mendoza, P. A., Musselman, K. N., Revuelto, J., Deems, J. S., López-Moreno, J. I., and McPhee, J.: Interannual and Seasonal Variability of
Snow Depth Scaling Behavior in a Subalpine Catchment, *Water Resources Research*, 56, <https://doi.org/10.1029/2020wr027343>, 2020.
- Meromy, L., Molotch, N. P., Link, T. E., Fassnacht, S. R., and Rice, R.: Subgrid variability of snow water equivalent at operational snow
stations in the western USA, *Hydrological Processes*, 27, 2383–2400, <https://doi.org/10.1002/hyp.9355>, 2013.



- 710 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.
- Mortimer, C. and Vionnet, V.: Northern Hemisphere in situ snow water equivalent dataset (NorSWE, 1979–2021), *Earth System Science Data*, 17, 3619–3640, <https://doi.org/10.5194/essd-17-3619-2025>, 2025.
- 715 Mott, R.: Climatological snow data since 1998, OSHD, <https://doi.org/http://dx.doi.org/10.16904/envidat.401>, 2023.
- Mott, R., Vionnet, V., and Grünewald, T.: The Seasonal Snow Cover Dynamics: Review on Wind-Driven Coupling Processes, *Frontiers in Earth Science*, 6, 197, <https://doi.org/10.3389/feart.2018.00197>, 2018.
- 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.
- 720 Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, *Earth System Science Data*, 13, 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021>, 2021.
- Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, *Journal of Hydrology*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- 725 Nearing, G. S., Ruddell, B. L., Bennett, A. R., Prieto, C., and Gupta, H. V.: Does information theory provide a new paradigm for earth science? Hypothesis testing, *Water Resources Research*, 56, <https://doi.org/10.1029/2019wr024918>, 2020.
- Notarnicola, C.: Overall negative trends for snow cover extent and duration in global mountain regions over 1982–2020, *Scientific Reports*, 12, 13 731, <https://doi.org/10.1038/s41598-022-16743-w>, 2022.
- 730 Olefs, M., Koch, R., Schöner, W., and Marke, T.: Changes in snow depth, snow cover duration, and potential snowmaking conditions in austria, 1961–2020—a model based approach, *Atmosphere*, 11, 1330, <https://doi.org/10.3390/atmos11121330>, 2020.
- 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.
- 735 Pflug, J. M., Margulis, S. A., and Lundquist, J. D.: Inferring watershed-scale mean snowfall magnitude and distribution using multidecadal snow reanalysis patterns and snow pillow observations, *Hydrological Processes*, 36, <https://doi.org/10.1002/hyp.14581>, 2022.
- Pflug, J. M., Kumar, S. V., Hall, D. K., Riggs, G. A., Konapala, G., Whitney, K. M., Wrzesien, M. L., Nie, W., Sun, Z., and Arsenaault, K. R.: Efficient and Regionally Transferable Snow Water Equivalent Estimation Using a Long Short-Term Memory Network, *Journal of Geophysical Research: Machine Learning and Computation*, 2, <https://doi.org/10.1029/2025jh000593>, 2025.
- 740 Phillips, K. B. B., Lineman, B., Murray, G. L. D., Nelson, S. J., Sperduto, D. D., and Tourville, J. C.: Refugia or at risk? Alpine snowbank communities in the face of climate change, *Conservation Science and Practice*, 8, <https://doi.org/10.1111/csp2.70079>, 2026.
- 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.
- 745 Pritchard, H. D.: Global data gaps in our knowledge of the terrestrial cryosphere, *Frontiers in Climate*, 3, 689 823, <https://doi.org/10.3389/fclim.2021.689823>, 2021.



- Qin, Y., Abatzoglou, J. T., Siebert, S., Huning, L. S., AghaKouchak, A., Mankin, J. S., Hong, C., Tong, D., Davis, S. J., and Mueller, N. D.: Agricultural risks from changing snowmelt, *Nature Climate Change*, 10, 459–465, <https://doi.org/10.1038/s41558-020-0746-8>, 2020.
- Saloranta, T. M.: Operational snow mapping with simplified data assimilation using the seNorge snow model, *Journal of Hydrology*, 538, 314–325, <https://doi.org/10.1016/j.jhydrol.2016.03.061>, 2016.
- 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>, *tex.rating*: 5, 2016.
- Schwaab, J., Bavay, M., Davin, E., Hagedorn, F., Hüsler, F., Lehning, M., Schneebeli, M., Thürig, E., and Bebi, P.: Carbon storage versus albedo change: radiative forcing of forest expansion in temperate mountainous regions of Switzerland, *Biogeosciences (Online)*, 12, 467–487, <https://doi.org/10.5194/bg-12-467-2015>, 2015.
- Sullender, B. K., Cunningham, C. X., Lundquist, J. D., and Prugh, L. R.: Defining the danger zone: critical snow properties for predator–prey interactions, *Oikos (Copenhagen, Denmark)*, 2023, <https://doi.org/10.1111/oik.09925>, 2023.
- Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic Attribution for Deep Networks, *arXiv*, <https://doi.org/10.48550/arxiv.1703.01365>, 2017.
- Tague, C. and Grant, G. E.: Groundwater dynamics mediate low-flow response to global warming in snow-dominated alpine regions, *Water Resources Research*, 45, <https://doi.org/10.1029/2008wr007179>, 2009.
- Taherian, M. and Ameli, A. A.: Time Variance in Snowmelt Partitioning: A Mechanistic Modeling Approach to Explore the Role of Catchment Structure and Pre-Snow Rainfall, *Water Resources Research*, 62, <https://doi.org/10.1029/2025wr040679>, 2026.
- Tang, G., Clark, M. P., and Papalexiou, S. M.: SC-earth: a station-based serially complete earth dataset from 1950 to 2019, *Journal of Climate*, 34, 6493–6511, <https://doi.org/10.1175/jcli-d-21-0067.1>, 2021.
- Thackeray, C. W. and Fletcher, C. G.: Snow albedo feedback, *Progress in Physical Geography*, 40, 392–408, <https://doi.org/10.1177/0309133315620999>, 2016.
- Thapa, S., Zhao, Z., Li, B., Lu, L., Fu, D., Shi, X., Tang, B., and Qi, H.: Snowmelt-Driven Streamflow Prediction Using Machine Learning Techniques (LSTM, NARX, GPR, and SVR), *Water*, 12, 1734, <https://doi.org/10.3390/w12061734>, 2020.
- Umrbekov, A., Essery, R., and Müller, D.: GEMS v1.0: Generalizable Empirical Model of Snow Accumulation and Melt, based on daily snow mass changes in response to climate and topographic drivers, *Geoscientific Model Development*, 17, 911–929, <https://doi.org/10.5194/gmd-17-911-2024>, 2023.
- Vernay, M., Lafaysse, M., Monteiro, D., Hagenmuller, P., Nheili, R., Samacoïts, R., Verfaillie, D., and Morin, S.: The S2M meteorological and snow cover reanalysis over the French mountainous areas: description and evaluation (1958–2021), *Earth System Science Data*, 14, 1707–1733, <https://doi.org/10.5194/essd-14-1707-2022>, 2022.
- Wang, Z., Baskar, J. V. J., Sai, M. S. S. N., Svoma, B., and Vivoni, E. R.: Spatiotemporal patterns of intermittent snow cover from PlanetScope imagery using deep learning, *Geophysical Research Letters*, 52, <https://doi.org/10.1029/2025gl116582>, 2025a.
- Wang, Z., Berghuijs, W. R., Howden, N. J. K., and Woods, R.: Diverging Impacts of Snow Fraction and Soil Drainage on Seasonal and Annual Water Balances Across Snow-Influenced Catchments, *Water Resources Research*, 61, <https://doi.org/10.1029/2025wr040032>, 2025b.
- Wayand, N. E., Marsh, C. B., Shea, J. M., and Pomeroy, J. W.: Globally scalable alpine snow metrics, *Remote Sensing of Environment*, 213, 61–72, <https://doi.org/10.1016/j.rse.2018.05.012>, 2018.
- Wiersma, P.: Data and Code accompanying "Data-driven inference of snowmelt from streamflow in 110 catchment across three continents", <https://doi.org/10.5281/zenodo.22283784>, 2026.



Wiersma, P., Magnusson, J., Peleg, N., Schaefli, B., and Mariethoz, G.: Can streamflow observations constrain snow mass reconstructions?

Lessons from two synthetic numerical experiments, *Hydrology and Earth System Sciences*, 30, 3331–3350, [https://doi.org/10.5194/hess-](https://doi.org/10.5194/hess-30-3331-2026)

785 30-3331-2026, 2026.

Woolson, R. F.: *Wiley Encyclopedia of Clinical Trials*, pp. 1–3, <https://doi.org/10.1002/9780471462422.eoct979>, 2008.

Zakeri, F., Mariethoz, G., and Girotto, M.: High-resolution snow water equivalent estimation: a data-driven method for localized downscaling of climate data, *Hydrology and Earth System Sciences*, 29, 6935–6958, <https://doi.org/10.5194/hess-29-6935-2025>, 2025.