SWECAST: ENHANCING SNOW WATER EQUIVALENT FORECASTING THROUGH PRIOR MODULATION AND PERIODICITY MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Snow water equivalent (SWE) directly measures the amount of water stored in snowpack. Accurate SWE forecasting is essential for water resource management and flood risk assessment in regions relying on seasonal snowpack. However, SWE evolution varies across locations and years and is sensitive to abrupt weather changes during the forecast horizon, which makes accurate SWE forecasting particularly challenging. We therefore propose SWECast, a novel model that improves SWE forecasting by combining periodicity modeling with prior modulation. We select informative meteorological inputs and construct an input configuration comprising complete historical snow seasons, current snow season history, and future meteorological inputs. We design a Periodicity Modeling Module (PMM) that analyzes SWE and meteorological inputs in the frequency domain, inferring the current SWE state and characterizing its response to weather. Based on this information, a gating mechanism dynamically weights each feature at each time step to incorporate meteorological information over the forecast horizon. In parallel, a Prior Conditioner (PC) derives priors from complete historical snow seasons in the time domain, capturing how SWE evolves at each location and thereby addressing the variability in SWE evolution across locations. Finally, the priors from the PC modulate the representations generated by the PMM. The evaluation results at stations excluded from training demonstrate that SWECast consistently outperforms all baselines. Over a forecast horizon of two weeks, SWECast reduces MAE by 13.1% and RMSE by 14.5% compared with the strongest baseline. By providing accurate SWE forecasting, SWECast can inform reservoir operations and flood preparedness in snow-dependent regions.

1 INTRODUCTION

Seasonal snowpack acts as a natural reservoir, storing cold season precipitation and releasing water as snowmelt during warmer months (Immerzeel et al., 2020). Snowmelt is an important contributor to river flow worldwide and supports downstream water supply, hydropower generation, and agriculture (Beniston et al., 2018; Barnett et al., 2005; Qin et al., 2020). Snow water equivalent (SWE) is a measure of the water stored in snowpack and indicates subsequent snowmelt runoff (Tsang et al., 2022; Jonas et al., 2009). Climate warming is altering snow accumulation and melt timing and intensifying extreme weather events, increasing the risks of snowmelt floods and seasonal water shortages (Musselman et al., 2017; Wieder et al., 2022). These risks motivate us to focus on SWE forecasting over one to two weeks, providing actionable lead time to prepare for floods and optimize reservoir operations (Cowherd et al., 2024; Delaney et al., 2020; Painter et al., 2016).

Accurate SWE forecasting remains challenging because SWE evolution varies substantially across locations and years and can respond abruptly to weather. Across locations, this variability arises from differences in climate and topography; across years, the timing and magnitude of SWE changes can differ considerably even at the same location. Abrupt weather events pose an additional challenge by triggering rapid accumulation or melt.

Process based snow models simulate snow evolution through physical processes driven by meteorological forcing and have long been used for operational SWE forecasting. Because future forcing

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

is typically obtained from weather forecasts, their practical forecast horizons range from a few days to approximately two weeks, while their predictions remain sensitive to forcing and parameter uncertainties (Lafaysse et al., 2017; Price et al., 2025; Günther et al., 2019; Raleigh et al., 2015). Growing records of SWE observations and meteorological products have made data driven models increasingly viable, but their reliability remains uncertain when future weather information is limited (Thapa et al., 2026; Liu et al., 2025).

Existing data driven models remain limited in representing variability in SWE evolution across locations and using meteorological information from different periods. Static geographic variables provide only general location context (Umirbekov et al., 2024; Thapa et al., 2026), while spatial correlations require sufficiently dense station networks (Thapa et al., 2024; 2026). Existing models also lack a systematic organization of meteorological information from different periods, limiting how they represent the effects of meteorological conditions at different times (Malin et al., 2025; Cui et al., 2023).

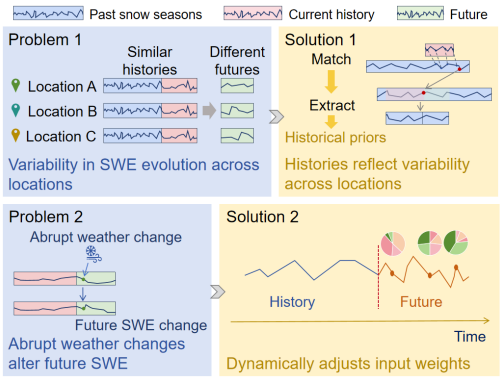


Figure 1: The key challenges in SWE forecasting and how SWECast addresses them. Historical priors derived from matched segments in complete historical snow seasons help SWECast handle variability in SWE evolution across locations, while dynamic weighting of SWE and meteorological inputs enables the model to respond to abrupt weather changes.

To address the challenges illustrated in Figure 1, we select meteorological variables and construct an input configuration that combines the current season history, complete historical snow seasons, and meteorological inputs over the forecast horizon. SWECast combines a Prior Conditioner (PC) with stacked Adaptive Periodicity Blocks (APBs). The PC matches the current season against complete historical records to derive priors about SWE evolution at a given location, while the APBs model the current SWE state and its interactions with meteorological inputs in the frequency domain. The priors from the PC modulate the representations learned by the APBs. Experimental results show that SWECast consistently outperforms representative baselines at stations excluded from model training.

Our contributions can be summarized as follows:

- We design an effective input configuration and propose SWECast to forecast SWE while accounting for the variability in SWE evolution across locations and abrupt weather changes.
- We introduce a prior modulation mechanism that integrates matched historical SWE patterns with frequency-domain modeling of both the current SWE state and its interactions with meteorological inputs. This enables SWECast to preserve SWE evolution at each station while responding to rapid weather changes.
- Experiments show that SWECast consistently outperforms the compared models across all forecast horizons and evaluation metrics when forecasting SWE at stations excluded from model training.

2 Related Work

2.1 Process-Based Snow Models

Process based snow models simulate snow accumulation and ablation using degree day or energy balance formulations. SNOW-17 (Anderson, 1973), a representative degree day model, estimates snow accumulation and melt from precipitation and air temperature, but its calibrated melt factors depend on local topographic and climatic conditions, limiting transfer across catchments and changing climates (Hock, 2003). Energy balance models, including the Utah Energy Balance model (Tarboton & Luce, 1996), Crocus (Brun et al., 1989; Vionnet et al., 2012), and SNOWPACK (Bartelt

& Lehning, 2002; Lehning et al., 2002), represent snow processes in greater detail but require extensive meteorological forcing and parameterization, whose errors can reduce forecast reliability. SWECast learns SWE evolution directly from observations without simulating snow processes or calibrating physical parameters for individual locations.

2.2 Data-Driven Models

Early statistical and LSTM models operate at individual stations and do not directly support forecasting at unseen stations (Sarhadi et al., 2014; Song et al., 2024). GEMS improves generalization by training across stations (Umirbekov et al., 2024), while later models use spatiotemporal attention and station background information (Thapa et al., 2024; 2026). Unlike these approaches, SWECast derives historical priors from matched observations in past snow seasons and dynamically weights SWE and meteorological inputs without using station background information or correlations among stations.

Alongside fully data-driven models, physics-informed approaches either use outputs from process-based snow models as auxiliary predictors or additional training data (Pomarol Moya et al., 2026), or encode physical constraints within neural models (Charbonneau et al., 2025; Maharjan et al., 2025). Unlike these approaches, SWECast obtains prior guidance directly from observed past snow seasons without physical simulations or predefined constraints.

3 Task and Input Design

3.1 Problem Definition

Given SWE observations and meteorological information up to time t, together with meteorological inputs for the H future time steps, the task is to forecast the complete SWE time series over the same H time steps. Let the forecasting target be $\mathbf{y} = [y_{t+1}, y_{t+2}, \ldots, y_{t+H}]$, where y_{t+h} denotes the SWE value at the h-th forecast step. The model f with trainable parameters θ learns the following mapping:

$$f_\theta : \mathcal{X} \rightarrow \mathbb{R}^H, \tag{1}$$

where $\mathcal{X}$ denotes all information provided to the model. Given this information, the model produces:

$$\hat{\mathbf{y}} = f_\theta(\mathcal{X}) = [\hat{y}_{t+1}, \hat{y}_{t+2}, \ldots, \hat{y}_{t+H}], \tag{2}$$

where $\hat{y}_{t+h}$ denotes the SWE forecast at the h-th forecast step.

3.2 Input Selection and Configuration

SWE changes are driven by snow accumulation and snowmelt, which are in turn governed by temperature, precipitation, and other meteorological factors such as solar radiation, humidity, and wind speed. To determine which combination of meteorological variables yields the most effective forecasting, we evaluate the performance of SWE only, SWE with individual meteorological variables, and SWE with two or more meteorological variables. Experimental analysis shows that SWE, temperature (Tm), and total precipitation (TP) together yield the best overall forecasting performance among all evaluated combinations. Detailed results are presented in the Input Configuration Analysis section. We therefore select Tm and TP as the meteorological inputs.

Based on this selection, we draw the inputs from three temporal periods: complete historical snow seasons, the current season history, and the forecast horizon. Records covering complete historical snow seasons and the current season history provide complementary historical information. Within the complete historical snow seasons, segments similar to the current season history are identified through matching. These matched segments, together with the SWE evolution following them, provide references that help the model account for differences in SWE evolution across locations. The current season history serves two roles. First, the current SWE state and its recent evolution form the most direct basis for the forecasted values. Second, by providing SWE observations together with concurrent meteorological conditions, it enables the model to learn how SWE responds to meteorological forcing. Finally, meteorological inputs over the forecast horizon provide information about future weather conditions when forecasting SWE.

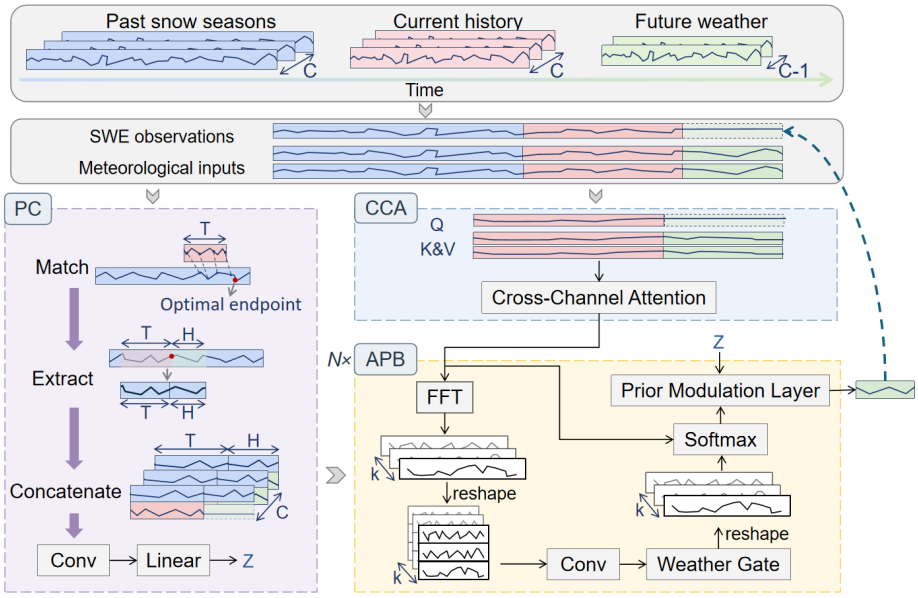


Figure 2: Overview of the SWECast architecture. Historical and future inputs are processed through the Prior Conditioner (PC), Cross-Channel Attention (CCA), and stacked Adaptive Periodicity Blocks (APBs) to produce SWE forecasts. $\mathbf{Z}$ denotes the Prior Modulation Parameters.

Formally, the model receives three inputs, each with $C = 3$ channels corresponding to SWE, temperature (Tm), and total precipitation (TP). $\mathbf{X}_{\text{prev}} \in \mathbb{R}^{S \times C}$ contains complete historical snow seasons, comprising S time steps. $\mathbf{X}_{\text{hist}} \in \mathbb{R}^{T \times C}$ contains a historical window of length T from the current season. $\mathbf{X}_{\text{fut}} \in \mathbb{R}^{H \times C}$ follows the same channel layout and contains Tm and TP forcing over the H-day forecast horizon. Because SWE is unavailable at future time steps, its channel in $\mathbf{X}_{\text{fut}}$ is assigned a sentinel value, and a binary observation mask distinguishes observed SWE from future placeholders.

4 Method

This section presents SWECast, a model designed for SWE forecasting under the variability in SWE evolution across locations and abrupt weather changes. As illustrated in Figure 2, SWECast comprises three main components: (1) a Prior Conditioner (PC), which extracts the segment most similar to the current season from complete historical snow season records and encodes it into Prior Modulation Parameters; (2) a Cross-Channel Attention (CCA) module, which learns how SWE responds to meteorological inputs from the current season history to initialize their weights over the forecast horizon; and (3) a stack of Adaptive Periodicity Blocks (APBs), which forms the main forecasting pathway. Each APB further refines the relations between SWE and meteorological inputs in the frequency domain, incorporates meteorological information over the forecast horizon, and uses the Prior Modulation Parameters to adapt the modeled patterns to SWE evolution at a given location.

4.1 Prior Conditioner

Similar current season histories at different locations may be followed by different future SWE trajectories. The Prior Conditioner addresses this variability by extracting relevant information from complete historical snow season records at the forecast location. Specifically, it matches the current season history against historical records to find the most similar segment, extracts that segment together with the following H time steps as a reference, and concatenates the reference with the current season time series to generate the Prior Modulation Parameters, which modulate the representations learned by the APBs.

To match the current season history, we first select all time steps in the historical records with at least H subsequent time steps as candidate endpoints. For each candidate endpoint q, we construct a candidate window $\overline{\mathbf{X}}_{\text{prev}}^{(q)} \in \mathbb{R}^{T\times C}$ from the T time steps ending at q, zero-padding on the left when fewer than T preceding time steps are available. We then determine the optimal endpoint by minimizing the Soft-DTW discrepancy (Cuturi & Blondel, 2017) between the current season history $\mathbf{X}_{\text{hist}}$ and each candidate window:

$$q^* = \arg\min_{q\in\mathcal{Q}} \ \text{SoftDTW}_\gamma \left(\mathbf{X}_{\text{hist}}, \overline{\mathbf{X}}_{\text{prev}}^{(q)}\right), \tag{3}$$

where $\mathcal{Q}$ denotes the set of candidate endpoints that leave at least H subsequent time steps. Soft-DTW flexibly aligns two time series by accumulating pointwise differences along an optimal alignment path, allowing similar SWE evolution patterns to be matched even when their timing differs across years. We compute the distance jointly over SWE, Tm, and TP so that the match reflects both the snow state and the driving meteorological conditions. The detailed formulation of Soft-DTW is provided in the Appendix.

After determining q^*, we extract the T time steps ending at q^* as the matched historical window and the following H time steps as its continuation, yielding a reference time series $\mathbf{X}_{\text{match}}^{+} \in \mathbb{R}^{(T+H)\times C}$.

We extend the current season historical window to the forecast horizon using $\mathbf{X}_{\text{fut}}$ defined in Section 3, yielding $\mathbf{X}_{\text{hist}}^{+}$. We concatenate $\mathbf{X}_{\text{hist}}^{+}$ and $\mathbf{X}_{\text{match}}^{+}$ along the channel dimension and process them with a convolutional encoder. The encoder compresses their joint information into a compact representation that captures how the current season relates to the matched historical time series and its continuation:

$$\mathbf{Z}' = \text{ConvEncoder}\left(\text{Concat}_c\left(\mathbf{X}_{\text{hist}}^{+}, \mathbf{X}_{\text{match}}^{+}\right)\right). \tag{4}$$

A linear projection then maps this representation to the Prior Modulation Parameters:

$$\mathbf{Z} = \text{Linear}\left(\mathbf{Z}'\right), \qquad \mathbf{Z} \in \mathbb{R}^{d_{\text{model}}}. \tag{5}$$

By conditioning the forecast on $\mathbf{Z}$, the model gains access to location-specific priors about SWE evolution without relying on static geographic attributes or cross-station correlations.

4.2 Cross-Channel Attention

The Cross-Channel Attention module produces an initial characterization of how SWE responds to meteorological conditions. At each time step, SWE queries the Tm and TP features via multi-head attention, producing attention weights that reflect the relation of each meteorological variable to SWE at that moment. A binary observation mask is appended to both $\mathbf{X}_{\text{hist}}$ and $\mathbf{X}_{\text{fut}}$ before they are concatenated along the temporal dimension, so that the module can distinguish observed SWE values from sentinel placeholders at each time step.

The module operates independently at each time step. At step t, the SWE value s_t and the observation mask value m_t are jointly projected to form the query, whereas the meteorological variables $\mathbf{x}_{\text{met},t} \in \mathbb{R}^{C-1}$ (Tm and TP at step t) provide the keys and values. The query and each meteorological token are mapped to d_{model}-dimensional representations through learned linear projections, and a shared projection produces the keys and values for all meteorological channels. Multi-head attention is then applied, with the SWE mask query attending to the meteorological tokens:

$$\mathbf{Q} = \text{Linear}_q([s_t, m_t]), \quad \mathbf{K}, \mathbf{V} = \text{Linear}_{kv}(\mathbf{x}_{\text{met},t}). \tag{6}$$

The outputs of all attention heads are concatenated and passed through a linear projection, followed by GELU activation and dropout.

4.3 Adaptive Periodicity Block

The APB serves as the core forecasting component of SWECast. Each APB consists of a Periodicity Modeling Module (PMM) and a Prior Modulation Layer (PML). PMM takes the feature map produced by CCA, identifies dominant periods in the frequency domain, and applies a weather gate that dynamically weights the feature channels using the meteorological variables at each time step. The PMM therefore combines periodicity modeling with dynamic conditioning on meteorological inputs.

4.3.1 PERIODICITY MODELING MODULE

SWE evolution exhibits recurring patterns at multiple temporal scales. To capture these, the PMM operates in the frequency domain. Inspired by TimesNet (Wu et al., 2023), we fold the 1D input time series into a 2D representation, aligning positions at the same phase across cycles for joint modeling through 2D convolution. Dominant periods are detected from the input feature map $\mathbf{F} \in \mathbb{R}^{(T+H)\times d_{\text{model}}}$ via FFT; the k frequency components with the largest amplitudes are selected as candidate cycle lengths. For each candidate period P, the time series is folded into a 2D grid and processed by a 2D convolution.

After the 2D convolution, a weather gate is applied. At each time step t, we project the meteorological variables through a linear layer followed by a sigmoid function to produce a gate vector $\mathbf{g}_t \in \mathbb{R}^{d_{\text{model}}}$, and broadcast it to all positions associated with t in the 2D feature map. This gate allows the model to emphasize or suppress feature channels according to the meteorological conditions at time step t.

After the weather gate, each period specific stream is flattened and truncated to the original length $T + H$. The outputs are combined through a weighted sum, with weights determined by the corresponding FFT amplitudes. A residual connection adds the block input (i.e., the CCA feature map) to the aggregated output, allowing the cross-channel relationships captured by CCA to be retained across the block.

4.3.2 PRIOR MODULATION LAYER

The Prior Modulation Layer recalibrates the current season representation using the cross year context encoded in $\mathbf{Z}$ via a FiLM style affine transformation (Perez et al., 2018):

$$\text{PriorMod}(\mathbf{x}, \mathbf{Z}) = \boldsymbol{\alpha} \odot \text{LN}(\mathbf{x}) + \boldsymbol{\beta}. \tag{7}$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^{d_{\text{model}}}$ are generated from $\mathbf{Z}$ by two independent linear projections. This separates the main forecasting pathway from auxiliary cross year correction: CCA and PMM learn temporal patterns and how SWE responds to meteorology from the current season, while previous year information affects the forecast only through the FiLM modulation. All stacked blocks receive the same $\mathbf{Z}$ but use independent modulation parameters.

After the final APB, a linear projection maps the d_{model}-dimensional representation at each time step to a scalar SWE estimate. The outputs at the final H time steps form the forecast $\hat{\mathbf{y}} \in \mathbb{R}^H$.

5 EXPERIMENTS

5.1 DATASETS

Daily snow water equivalent observations are obtained from the Snow Telemetry (SNOTEL) network (Fleming et al., 2023), with the data downloaded from the Natural Resources Conservation Service (NRCS) (USDA Natural Resources Conservation Service, National Water and Climate Center, 2026). We spatially and temporally match each SNOTEL station with variables from ERA5-Land (Muñoz-Sabater et al., 2021). We use its 2m temperature and total precipitation as the primary meteorological forcing variables, denoted Tm and TP respectively. The remaining ERA5-Land variables are grouped into snow related, surface, atmospheric, and vegetation categories (see the Appendix for the full name mapping of all abbreviated variables). These variables are treated as candidate auxiliary inputs in the input configuration analysis.

The initial dataset comprises 29 years of observations from 1993 to 2021 across 832 stations. We remove daily records with missing SWE, Tm, or TP values and retain only stations containing at least one consecutive three-year period in which the missing-data rate is below 20% for each year. After quality control, 736 stations remain.

We split the dataset by station, assigning 70% of the stations to the training set, 10% to the test set, and 20% to the validation set. Within each station and hydrological year, samples are generated using a sliding window with a stride of 14 days.

Table 1: SWE forecasting performance using the preceding 60 days of observations from the current snow season. Results are averaged over three independent runs with different random seeds. MAE and RMSE are reported as mean ± standard deviation across runs, and NSE as the mean. ↓ indicates lower is better; ↑ indicates higher is better. Best results are shown in **bold**, and second-best results are underlined. †SNOW-17 calibrated using all available historical records at each station.

Forecast Horizon	7 days			14 days		
Metric	MAE (mm) ↓	RMSE (mm) ↓	NSE ↑	MAE (mm) ↓	RMSE (mm) ↓	NSE ↑
Random Forest (Breiman, 2001)	11.88±0.16	19.69±0.51	0.968	18.87±0.25	32.57±0.32	0.922
XGBoost (Chen & Guestrin, 2016)	10.97±0.27	17.68±0.71	0.975	17.06±0.40	28.31±0.53	0.942
DLinear (Zeng et al., 2023)	43.64±2.13	57.89±0.96	0.739	54.82±1.81	74.59±0.71	0.601
PatchTST (Nie et al., 2023)	26.02±2.79	35.45±3.28	0.896	34.30±2.27	51.00±2.52	0.809
iTransformer (Liu et al., 2024)	14.03±0.47	21.64±0.78	0.961	20.19±0.42	32.37±0.35	0.917
TimesNet (Wu et al., 2023)	13.70±0.47	21.04±1.26	0.963	19.77±0.61	30.97±0.45	0.923
SNOW-17 (Anderson, 1973)	75.38±1.26	105.32±2.45	0.377	76.38±2.44	109.38±3.10	0.308
SNOW-17†	58.74±1.97	81.75±4.13	0.632	60.40±1.44	84.20±2.09	0.568
SWECast (Ours)	**10.10±0.27**	**16.10±0.76**	**0.978**	**14.82±0.37**	**24.20±1.04**	**0.954**

5.2 Experiment Setup

SWECast is trained for 50 epochs with a batch size of 256 using the MAE loss and the AdamW optimizer. We selected the learning rate and hidden dimension via grid search on the validation set, resulting in a learning rate of 1×10^{-3} and a hidden dimension of 256; the weight decay is set to 1×10^{-4}. The grid search ranges and parameter sensitivity analyses are provided in the Appendix. The forecasted SWE values are transformed back to their original scale before evaluation. We use MAE, RMSE, and the Nash–Sutcliffe efficiency coefficient (NSE) as evaluation metrics. The metrics are calculated independently for each test station, and their median across stations is reported. All experiments are conducted on an NVIDIA GeForce RTX 3090 GPU (24 GB).

5.3 Main Results

We compare SWECast against seven forecasting models including tree models (Random Forest (Breiman, 2001), XGBoost (Chen & Guestrin, 2016)), a linear model (DLinear (Zeng et al., 2023)), a CNN model that captures temporal variation through periodicity modeling (TimesNet (Wu et al., 2023)), Transformer architectures (PatchTST (Nie et al., 2023), iTransformer (Liu et al., 2024)), and process-based snow modeling (SNOW-17 (Anderson, 1973)).

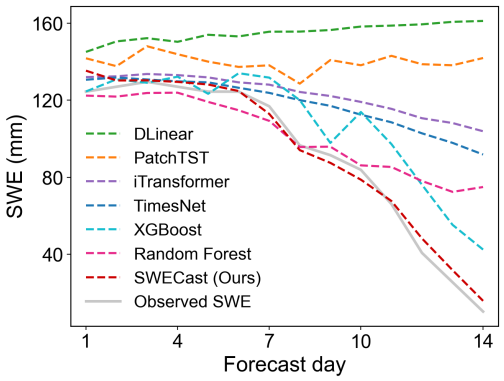


Figure 3: Comparison of SWE forecast time series produced by SWECast and baseline models.

For a fair comparison, all data-driven baselines use the same input information and station split as SWECast, with their input formats adapted to the corresponding architectures. Each model is independently tuned on the validation set, and the selected hyperparameters are provided in the Appendix. For SNOW-17, model parameters are calibrated separately at each station using the preceding complete snow season. The calibrated model is subsequently driven by current-season meteorological forcing to generate daily SWE forecasts.

Table 1 summarizes the SWE forecasting performance of all methods. SWECast achieves the best results across both forecast horizons and all evaluation metrics, with improvements that are large relative to the run to run variation. For the 7-day forecasts, SWECast reduces MAE and RMSE by 7.9% and 8.9%, respectively, relative to the strongest baseline. For the 14-day forecasts, the corresponding improvements increase to 13.1% and 14.5%. SWECast also shows the smallest degradation as

the horizon extends from 7 to 14 days: its MAE increases by only 4.72 mm, compared with 6.09 mm for the strongest baseline.

SNOW-17 performs substantially worse than the data-driven models, even when calibrated using all available historical records at each station. This result is consistent with the limitations discussed in the introduction: SNOW-17 relies on station specific parameter calibration and continuous meteorological forcing, making it sensitive to parameter uncertainty and gaps in the forcing records, whereas the data-driven methods can learn jointly from observations across stations.

As shown in Figure 3, DLinear, PatchTST, iTransformer, and TimesNet produce relatively smooth forecasts that fluctuate within a narrow range around their respective mean levels and do not capture the substantial SWE change in this example. In contrast, Random Forest and XGBoost exhibit larger but irregular fluctuations. Such behavior is consistent with the piecewise constant predictions of tree based models and their lack of explicit modeling of temporal continuity across forecast days. SWECast more closely follows the observed SWE time series in this example, including both the overall direction and short term variations in SWE evolution.

5.4 Ablation Studies

To evaluate the contribution of each component in SWECast, we construct five ablation variants, all evaluated at a forecast horizon of 14 days using otherwise identical settings:

- **w/o PC** replaces the Prior Conditioner with direct flattening followed by a linear projection.
- **w/o Soft-DTW** removes Soft-DTW matching and summarizes the previous season time series through mean pooling and a linear projection.
- **w/o CCA** replaces Cross-Channel Attention with max pooling and a linear projection.
- **w/o PML** replaces the Prior Modulation Layer with layer normalization, thereby removing the modulation of current-season representations by the cross-year prior.
- **w/o PMM** removes the Periodicity Modeling Module and skips frequency domain processing, operating directly on the 1D input sequence.

Table 2 reports the ablation results. Removing PMM causes the largest degradation (MAE increases by 10.10 mm), confirming that frequency domain modeling is essential for capturing the recurring temporal patterns in SWE evolution. Removing PML follows closely (MAE increases by 7.47 mm), showing that the cross year prior must actively recalibrate the current season representations rather than merely be encoded as auxiliary information. Removing the PC also leads to a substantial decline, with RMSE increasing by 5.50 mm, demonstrating that effectively conditioning on matched historical references is critical for accurate forecasting. The remaining variants, w/o Soft-DTW and w/o CCA, each degrade performance by smaller but consistent margins, confirming that flexible temporal alignment and cross-channel integration both contribute to the full performance of SWECast.

Table 2: Ablation results for 14-day SWE forecasting using 60 days of current-season observations. MAE and RMSE are in mm; best results are in **bold**.

Method	MAE ↓	RMSE ↓	NSE ↑
SWECast	**14.46**	**23.83**	**0.958**
w/o PC	17.54	29.33	0.934
w/o Soft-DTW	16.59	26.56	0.945
w/o CCA	16.69	27.67	0.942
w/o PML	21.93	34.25	0.909
w/o PMM	24.56	40.34	0.886

5.5 Input Configuration Analysis

We examine how forecasting performance changes with the choice of meteorological variables, the number of complete historical snow seasons, and the current season historical window length. All experiments in this section use a forecast horizon of 14 days and are based on a single run.

Table 3 shows that adding either Tm or TP to SWE improves performance, while combining both provides a substantially greater improvement. This result indicates that temperature and precipitation provide complementary information for forecasting future SWE evolution. Adding snow,

surface, atmospheric, or vegetation variables does not further improve performance, as all four augmented configurations produce higher MAE and RMSE.

We further investigate the number of complete historical snow seasons required to construct the historical prior. As shown in Figure 4(a), incorporating complete historical snow seasons substantially improves forecasting performance compared with using no cross-year information. Increasing the number of complete historical snow seasons beyond one, however, produces only minor fluctuations rather than consistent improvements, suggesting that the primary benefit comes from the most recent complete snow season.

Table 3: Effects of input configurations on 14-day SWE forecasting using 60 days of current-season observations. MAE and RMSE are in mm; best results are in **bold**. Tm: temperature; TP: total precipitation.

Input	MAE ↓	RMSE ↓	NSE ↑
SWE only	24.00	41.64	0.881
SWE + Tm	23.18	34.49	0.914
SWE + TP	19.19	33.57	0.908
SWE + Tm + TP (Base)	**14.46**	**23.83**	**0.958**
Base + Snow	14.95	24.67	0.956
Base + Surface	14.87	24.05	0.954
Base + Atmospheric	15.81	25.54	0.954
Base + Vegetation	15.03	24.59	0.954

We also examine the effect of the current season historical window length. Figure 4(b) shows that extending the window does not lead to a monotonic improvement in forecasting performance. Performance improves up to 60 days and then degrades, indicating that recent observations provide sufficient information about the current snow state while earlier ones contribute little to short-term SWE evolution. We therefore use a 60-day window, which achieves the lowest MAE.

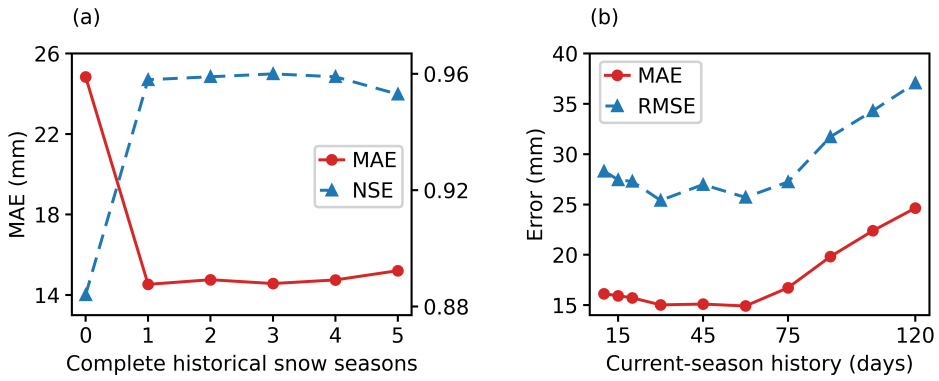


Figure 4: Input-history sensitivity of SWECast. (a) Effect of the number of complete historical snow seasons. (b) Effect of the current-season history length. Both panels report a 14-day forecast horizon from a single run.

6 Conclusion

We proposed SWECast, which addresses variability in SWE evolution across locations using historical priors derived from matched segments in complete snow seasons, and responds to abrupt changes caused by weather events by dynamically weighting SWE and meteorological inputs in the frequency domain. Designed for the one-to-two-week forecast window, SWECast provides actionable lead time for reservoir operations and flood preparedness. Experiments on SNOTEL stations demonstrate that SWECast consistently outperforms the compared baselines across forecast horizons and evaluation metrics. However, SWECast still requires a complete historical snow season at each target station, and the present evaluation uses meteorological variables from ERA5-Land over the forecast horizon. Future work will explore transferable priors for locations with shorter or unavailable observation records and examine the sensitivity of SWECast to uncertainty in meteorological forcing. Extending the approach to ensemble forecasting could further support decision making that accounts for uncertainty in water management.

AI USE STATEMENT

Generative AI tools were used as auxiliary tools in this work. They assisted with language refinement and translation, literature retrieval, code implementation, and data cleaning. The authors checked the original sources of all cited literature.

Generative AI tools were not used to determine the research objectives, formulate the core methodology, design the experiments, interpret the results, or draw scientific conclusions. These research decisions and analyses were made by the authors. The authors reviewed AI-assisted code and data processing before using their outputs in the experiments.

The authors take full responsibility for the manuscript, code, data, experimental results, and conclusions presented in this work.

ETHICS STATEMENT

This work investigates snow water equivalent (SWE) forecasting, which may support research and planning related to water resources. Forecast errors could nevertheless lead to inappropriate decisions if predictions are used without considering their uncertainty and the conditions under which the model was evaluated. We therefore present the results as research findings rather than guarantees of performance in operational settings. The authors are responsible for accurately reporting the data, methods, results, and limitations of this work, and for using the underlying datasets in accordance with their applicable terms.

REPRODUCIBILITY STATEMENT

The task definition, input construction, model components, datasets, station split, and evaluation protocol are described in Sections 3, 4, and 5. Appendix A.1 provides the complete Soft-DTW matching procedure. Appendix B.1 lists the auxiliary input variables, Appendix B.2 reports the hyperparameter search spaces and selected configurations, and the computing environment, evaluation metrics, and random seeds are documented in Appendices B.3–B.5. Additional hyperparameter and subgroup analyses appear in Appendices C.1 and C.2.

REFERENCES

Eric A. Anderson. National Weather Service River Forecast System: snow accumulation and ablation model. NOAA Technical Memorandum NWS HYDRO-17, National Weather Service, Office of Hydrology, Silver Spring, Maryland, November 1973. URL https://repository.library.noaa.gov/view/noaa/13507.

Tim P Barnett, Jennifer C Adam, and Dennis P Lettenmaier. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438(7066):303–309, 2005.

Perry Bartelt and Michael Lehning. A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model. *Cold Regions Science and Technology*, 35(3):123–145, 2002.

Martin Beniston, Daniel Farinotti, Markus Stoffel, Liss M Andreassen, Erika Coppola, Nicolas Eckert, Adriano Fantini, Florie Giacona, Christian Hauck, Matthias Huss, et al. The European mountain cryosphere: a review of its current state, trends, and future challenges. *The Cryosphere*, 12(2):759–794, 2018.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

E Brun, E Martin, V Simon, C Gendre, and C Coleou. An energy and mass model of snow cover suitable for operational avalanche forecasting. *Journal of glaciology*, 35(121):333–342, 1989.

Andrew Charbonneau, Katherine Deck, and Tapio Schneider. A physics-constrained neural differential equation framework for data-driven snowpack simulation. *Artificial Intelligence for the Earth Systems*, 4(3):240040, 2025.

Tianqi Chen and Carlos Guestrin. XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

Marianne Cowherd, Utkarsh Mital, Stefan Rahimi, Manuela Girotto, Andrew Schwartz, and Daniel Feldman. Climate change-resilient snowpack estimation in the Western United States. *Communications Earth & Environment*, 5(1):337, 2024.

Guotao Cui, Michael Anderson, and Roger Bales. Mapping of snow water equivalent by a deep-learning model assimilating snow observations. *Journal of Hydrology*, 616:128835, 2023.

Marco Cuturi and Mathieu Blondel. Soft-DTW: a differentiable loss function for time-series. In *International conference on machine learning*, pp. 894–903. PMLR, 2017.

Chris J. Delaney, Robert K. Hartman, John Mendoza, Michael Dettinger, Luca Delle Monache, Jay Jasperse, F. Martin Ralph, Cary Talbot, James Brown, David Reynolds, and Simone Evett. Forecast informed reservoir operations using ensemble streamflow predictions for a multipurpose reservoir in Northern California. *Water Resources Research*, 56(9):e2019WR026604, 2020. doi: 10.1029/2019WR026604.

Sean W. Fleming, Lucas Zukiewicz, Michael L. Strobel, Heather Hofman, and Adam G. Goodbody. SNOTEL, the soil climate analysis network, and water supply forecasting at the Natural Resources Conservation Service: past, present, and future. *JAWRA Journal of the American Water Resources Association*, 59(4):585–599, 2023. doi: 10.1111/1752-1688.13104.

Daniel Günther, Thomas Marke, Richard Essery, and Ulrich Strasser. Uncertainties in snowpack simulations – assessing the impact of model structure, parameter choice, and forcing data error on point-scale energy balance snow model performance. *Water Resources Research*, 55(4):2779–2800, 2019.

Regine Hock. Temperature index melt modelling in mountain areas. *Journal of hydrology*, 282(1-4): 104–115, 2003.

W. W. Immerzeel, Arthur F Lutz, Marcos Andrade, A Bahl, Henk Biemans, Tobias Bolch, Sam Hyde, S Brumby, BJ Davies, AC Elmore, et al. Importance and vulnerability of the world's water towers. *Nature*, 577(7790):364–369, 2020.

Tobias Jonas, Christoph Marty, and Jan Magnusson. Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. *Journal of Hydrology*, 378(1-2):161–167, 2009.

Matthieu Lafaysse, Bertrand Cluzet, Marie Dumont, Yves Lejeune, Vincent Vionnet, and Samuel Morin. A multiphysical ensemble system of numerical snow modelling. *The Cryosphere*, 11: 1173–1198, 2017. doi: 10.5194/tc-11-1173-2017.

Michael Lehning, Perry Bartelt, Bob Brown, Charles Fierz, and Pramod Satyawali. A physical SNOWPACK model for the Swiss avalanche warning: Part II. snow microstructure. *Cold regions science and technology*, 35(3):147–167, 2002.

Jiangtao Liu, Chaopeng Shen, Fearghal O'Donncha, Yalan Song, Wei Zhi, Hylke E. Beck, Tadd Bindas, Nicholas Kraabel, and Kathryn Lawson. From RNNs to transformers: benchmarking deep learning architectures for hydrologic prediction. *Hydrology and Earth System Sciences*, 29: 6811–6828, 2025. doi: 10.5194/hess-29-6811-2025.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=JePfAI8fah`.

Surendra Maharjan, Wenzhao Li, Rejoice Thomas, Shahryar Fazli, Arshad Ansari, Hesham Morgan, Ali Elgendy, Mohamed Allali, and Hesham El-Askary. Physics-informed deep learning reveals climate-driven snowpack decline and threatens ecological water availability in a Californian snow-fed catchment. *Ecological Informatics*, 92:103526, 2025. doi: 10.1016/j.ecoinf.2025.103526.

Miika Malin, Jarkko Okkonen, and Jaakko Suutala. Snow water equivalent forecasting in sub-arctic and arctic regions: efficient recurrent neural networks approach. *Environmental Modelling & Software*, 194:106695, 2025. doi: 10.1016/j.envsoft.2025.106695.

Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9):4349–4383, 2021. doi: 10.5194/essd-13-4349-2021.

Keith N Musselman, Martyn P Clark, Changhai Liu, Kyoko Ikeda, and Roy Rasmussen. Slower snowmelt in a warmer world. *Nature Climate Change*, 7(3):214–219, 2017.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023.

Thomas H. Painter et al. The Airborne Snow Observatory: fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, 184:139–152, 2016. doi: 10.1016/j.rse.2016.06.018.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. FiLM: visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Oriol Pomarol Moya, Madlene Nussbaum, Siamak Mehrkanoon, Philip DA Kraaijenbrink, Isabelle Gouttevin, Derek Karssenberg, and Walter W Immerzeel. Improving snow water equivalent modelling: a comparative study of hybrid machine learning techniques. *The Cryosphere*, 20(2):1427–1444, 2026.

Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R. Andersson, et al. Probabilistic weather forecasting with machine learning. *Nature*, 637:84–90, 2025. doi: 10.1038/s41586-024-08252-9.

Yue Qin, John T Abatzoglou, Stefan Siebert, Laurie S Huning, Amir AghaKouchak, Justin S Mankin, Chaopeng Hong, Dan Tong, Steven J Davis, and Nathaniel D Mueller. Agricultural risks from changing snowmelt. *Nature Climate Change*, 10(5):459–465, 2020.

Mark S Raleigh, Jessica D Lundquist, and Martyn P Clark. Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrology and Earth System Sciences*, 19(7):3153–3179, 2015.

Ali Sarhadi, Richard Kelly, and Reza Modarres. Snow water equivalent time-series forecasting in Ontario, Canada, in link to large atmospheric circulations. *Hydrological Processes*, 28(16): 4640–4653, 2014.

Yalan Song, Wen-Ping Tsai, Jonah Gluck, Alan Rhoades, Colin Zarzycki, Rachel McCrary, Kathryn Lawson, and Chaopeng Shen. LSTM-based data integration to improve snow water equivalent prediction and diagnose error sources. *Journal of Hydrometeorology*, 25(1):223–237, 2024.

David G Tarboton and Charles H Luce. *Utah Energy Balance snow accumulation and melt model (UEB)*. Utah Water Research Laboratory, 1996.

Krishu K Thapa, Bhupinderjeet Singh, Supriya Savalkar, Alan Fern, Kirti Rajagopalan, and Ananth Kalyanaraman. Attention-based models for snow-water equivalent prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 22969–22975, 2024.

Krishu K. Thapa, Supriya Savalkar, Bhupinderjeet Singh, Trong Nghia Hoang, Kirti Rajagopalan, and Ananth Kalyanaraman. ForeSWE: forecasting snow-water equivalent with an uncertainty-aware attention model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 39348–39357, 2026. doi: 10.1609/aaai.v40i46.41284.

Leung Tsang, Michael Durand, Chris Derksen, Ana Barros, Do-Hyuk Kang, Hans Lievens, Hans-Peter Marshall, Jiyue Zhu, Joel Johnson, Joshua King, et al. Global monitoring of snow water equivalent using high-frequency radar remote sensing. *The Cryosphere*, 16(9):3531–3573, 2022.

Atabek Umirbekov, Richard Essery, and Daniel Müller. 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(2):911–929, 2024.

USDA Natural Resources Conservation Service, National Water and Climate Center. Air and water database. Water and Climate Information System, 2026. URL https://nwcc-apps.sc.egov.usda.gov/. Accessed: 2026-07-27.

Vincent Vionnet, Eric Brun, S Morin, A Boone, Stéphanie Faroux, Patrick Le Moigne, Eric Martin, and J-M Willemet. The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2. *Geoscientific model development*, 5(3):773–791, 2012.

William R Wieder, Daniel Kennedy, Flavio Lehner, Keith N Musselman, Keith B Rodgers, Nan Rosenbloom, Isla R Simpson, and Ryohei Yamaguchi. Pervasive alterations to snow-dominated ecosystem functions under climate change. *Proceedings of the National Academy of Sciences*, 119(30):e2202393119, 2022.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: temporal 2D-variation modeling for general time series analysis. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=ju_Uqw384Oq.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11121–11128, 2023.

Appendix

A Method Details

A.1 Soft-DTW Matching Details

The Prior Conditioner retrieves a historical reference trajectory whose observed evolution resembles that of the current snow season. Specifically, it searches a complete historical snow season $\mathbf{X}_{\text{prev}} \in \mathbb{R}^{S\times C}$ using the current season historical window $\mathbf{X}_{\text{hist}} \in \mathbb{R}^{T\times C}$. In SWECast, the C channels correspond to SWE, Tm, and TP, and all channels are jointly considered during matching.

Let $\mathcal{Q}$ denote the set of candidate endpoints. A valid endpoint must leave at least H subsequent time steps in the same historical snow season. We index the time steps in the previous snow season from 1 to S. To ensure that each matched segment is followed by H available time steps, the candidate endpoints are defined as

$$\mathcal{Q} = \{1, \ldots, S - H\}. \tag{8}$$

For each $q \in \mathcal{Q}$, we extract the T time steps ending at q. To describe the construction explicitly, let

$$\begin{aligned} a_q &= \max(1, q - T + 1), \\ p_q &= \max(0, T - q), \end{aligned} \tag{9}$$

where a_q is the first available historical index and p_q is the required padding length. The candidate window is then

$$\overline{\mathbf{X}}^{(q)}_{\text{prev}} = \text{Concat}_t \left(\mathbf{0}_{p_q \times C}, \mathbf{X}_{\text{prev}}[a_q : q, :] \right), \tag{10}$$

where the endpoint q is included in the indexed segment. Thus, $\overline{\mathbf{X}}^{(q)}_{\text{prev}} \in \mathbb{R}^{T\times C}$ for every candidate. When $q \geq T$, the candidate contains the complete T preceding time steps and $p_q = 0$. When $q < T$, the unavailable steps at the beginning of the season are filled with zeros.

DTW aligns two time series using a path through a $T \times T$ cost matrix, where each position (i, ℓ) represents an alignment between time step i in one sequence and time step ℓ in the other. The path starts at $(1, 1)$ and ends at (T, T), moving only forward. From position (i, ℓ), the path can proceed to $(i + 1, \ell + 1)$, $(i + 1, \ell)$, or $(i, \ell + 1)$. The first move advances both sequences simultaneously, whereas the other two advance only one sequence. This allows one time step in either sequence to be aligned with nearby time steps in the other sequence. Consequently, similar accumulation or melt patterns can be matched even when they occur at different times across snow seasons.

For candidate q, the local cost matrix $\mathbf{D}^{(q)} \in \mathbb{R}^{T\times T}$ contains the pairwise dissimilarity between the two multivariate sequences. For $i, \ell \in \{1, \ldots, T\}$, each entry is

$$\begin{aligned} D^{(q)}_{i,\ell} &= \left\| \mathbf{X}_{\text{hist}}[i, :] - \overline{\mathbf{X}}^{(q)}_{\text{prev}}[\ell, :] \right\|_2^2 \\ &= \sum_{c=1}^{C} \left(X_{\text{hist}}[i, c] - \overline{X}^{(q)}_{\text{prev}}[\ell, c] \right)^2. \end{aligned} \tag{11}$$

Therefore, each local cost jointly measures the differences in SWE, Tm, and TP rather than matching the three channels independently.

Let Π_T denote the set of admissible alignment paths between two sequences of length T. The cost of a path $\pi \in \Pi_T$ is the sum of the local costs visited by that path:

$$C_q(\pi) = \sum_{(i,\ell)\in\pi} D^{(q)}_{i,\ell}. \tag{12}$$

Standard DTW uses the minimum path cost. Soft-DTW replaces this hard minimum with a differentiable soft minimum:

$$\begin{aligned} \text{SoftDTW}^{(q)}_{\gamma} &= \text{softmin}_{\gamma} \left\{ C_q(\pi) : \pi \in \Pi_T \right\} \\ &= -\gamma \log \sum_{\pi\in\Pi_T} \exp\left(-\frac{C_q(\pi)}{\gamma} \right), \end{aligned} \tag{13}$$

where $\gamma > 0$ is the smoothing parameter. This formulation assigns larger contributions to paths with lower alignment costs while retaining the contributions of other admissible paths.

The Soft-DTW score can be computed efficiently through dynamic programming. Let $\mathbf{R}^{(q)}$ denote the accumulated alignment-cost matrix. We initialize $R^{(q)}_{0,0} = 0$, $R^{(q)}_{i,0} = +\infty$, and $R^{(q)}_{0,\ell} = +\infty$ for $i, \ell \in \{1, \ldots, T\}$. The remaining entries are computed as

$$R^{(q)}_{i,\ell} = D^{(q)}_{i,\ell} + \mathrm{softmin}_\gamma \left(R^{(q)}_{i-1,\ell-1}, R^{(q)}_{i-1,\ell}, R^{(q)}_{i,\ell-1} \right), \tag{14}$$

where

$$\mathrm{softmin}_\gamma(a_1, \ldots, a_k) = -\gamma \log \sum_{j=1}^{k} \exp\left(-\frac{a_j}{\gamma} \right). \tag{15}$$

In implementation, the soft minimum is evaluated using the stabilized log-sum-exp operation to avoid numerical overflow or underflow. As γ approaches zero, the soft minimum approaches the standard minimum and the Soft-DTW score approaches the DTW score.

The matching score for candidate endpoint q is the accumulated cost at the final matrix entry:

$$s_q = \mathrm{SoftDTW}_\gamma \left(\mathbf{X}_{\mathrm{hist}}, \overline{\mathbf{X}}^{(q)}_{\mathrm{prev}} \right) = R^{(q)}_{T,T}. \tag{16}$$

Because all candidates have the same length and use the same local cost function, their scores can be compared directly. A lower score indicates a lower smoothed alignment cost. The matched endpoint is therefore selected as

$$q^* = \arg\min_{q \in \mathcal{Q}} s_q. \tag{17}$$

After selecting q^*, we retrieve the following H time steps from the same historical snow season:

$$\mathbf{X}^{(q^*)}_{\mathrm{future}} = \begin{bmatrix} \mathbf{X}_{\mathrm{prev}}[q^* + 1, :] \\ \vdots \\ \mathbf{X}_{\mathrm{prev}}[q^* + H, :] \end{bmatrix} \in \mathbb{R}^{H \times C}. \tag{18}$$

The matched historical window and its subsequent evolution are concatenated along the temporal dimension:

$$\mathbf{X}^{+}_{\mathrm{match}} = \mathrm{Concat}_t \left(\overline{\mathbf{X}}^{(q^*)}_{\mathrm{prev}}, \mathbf{X}^{(q^*)}_{\mathrm{future}} \right) \in \mathbb{R}^{(T+H) \times C}. \tag{19}$$

Only the first T time steps are used to determine q^*; the following H steps are retrieved only after matching. Therefore, the future portion of the historical reference does not influence the selection of the matched segment. The resulting sequence $\mathbf{X}^{+}_{\mathrm{match}}$ provides both the matched historical conditions and their subsequent evolution to the Prior Conditioner.

B Experimental Details

B.1 ERA5-Land Auxiliary Variable Groups

In the input configuration analysis, we organize the ERA5-Land variables beyond Tm and TP into four auxiliary groups. Each group is added separately to the base configuration consisting of SWE, Tm, and TP. Table 4 lists the abbreviations and full names of the variables in each group.

Variable group	Abbreviation	ERA5-Land variable
Snow	sf	Snowfall
	sc	Snow cover
	sd	Snow depth
	sal	Snow albedo
	snd	Snow density
Surface	stl	Snow surface temperature
	stl1	Soil temperature level 1
	skt	Skin temperature
	fal	Forecast albedo
Atmospheric	u10	10 metre U wind component
	v10	10 metre V wind component
	sp	Surface pressure
Vegetation	lai_hv	Leaf area index, high vegetation
	lai_lv	Leaf area index, low vegetation

Table 4: ERA5-Land auxiliary variable groups evaluated in the input configuration analysis. Each group is added separately to the base configuration consisting of SWE, Tm, and TP.

B.2 Hyperparameter Search Details

Table 5 reports the search spaces, fixed settings, and selected hyperparameters for SWECast and all baseline models. We tune the listed hyperparameters through grid search on the validation set and select the configuration with the best validation performance. All neural models use the common grid $\mathcal{B}$ over the learning rate and hidden dimension. SWECast and DLinear evaluate all 16 configurations in $\mathcal{B}$. For TimesNet, iTransformer, and PatchTST, we first select the best configuration from $\mathcal{B}$ and then, with the learning rate and hidden dimension fixed, evaluate three values of one model-specific hyperparameter. For Random Forest and XGBoost, we tune only the maximum tree depth and fix the remaining hyperparameters as listed in the table.

Model	Search space and fixed settings	Selected
SWECast	$(\mathrm{lr}, d_{\mathrm{model}}) \in \mathcal{B}$	$\mathrm{lr} = 1 \times 10^{-3}$, $d_{\mathrm{model}} = 256$
TimesNet	$(\mathrm{lr}, d_{\mathrm{model}}) \in \mathcal{B}$, $\mathtt{top_k} \in \{1, 3, 5\}$	$\mathrm{lr} = 1 \times 10^{-3}$, $d_{\mathrm{model}} = 64$, $\mathtt{top_k} = 5$
iTransformer	$(\mathrm{lr}, d_{\mathrm{model}}) \in \mathcal{B}$, $\mathtt{e_layers} \in \{1, 2, 3\}$	$\mathrm{lr} = 5 \times 10^{-4}$, $d_{\mathrm{model}} = 128$, $\mathtt{e_layers} = 3$
PatchTST	$(\mathrm{lr}, d_{\mathrm{model}}) \in \mathcal{B}$, $\mathtt{patch_len} \in \{8, 16, 24\}$	$\mathrm{lr} = 5 \times 10^{-3}$, $d_{\mathrm{model}} = 64$, $\mathtt{patch_len} = 24$
DLinear	$(\mathrm{lr}, d_{\mathrm{model}}) \in \mathcal{B}$	$\mathrm{lr} = 1 \times 10^{-4}$, $d_{\mathrm{model}} = 256$
Random Forest	$\mathtt{max_depth} \in \{3, 6, 9, 12\}$; fixed: $\mathtt{n_estimators} = 200$, $\mathtt{min_samples_leaf} = 5$	$\mathtt{max_depth} = 12$
XGBoost	$\mathtt{max_depth} \in \{3, 6, 9, 12\}$; fixed: $\mathtt{n_estimators} = 200$, $\mathrm{lr} = 0.1$, $\mathtt{subsample} = 0.8$	$\mathtt{max_depth} = 6$

Table 5: Hyperparameter search spaces, fixed settings, and selected values. The common grid over $(\mathrm{lr}, d_{\mathrm{model}})$ is $\mathcal{B} = \{1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}\} \times \{64, 128, 256, 512\}$.

B.3 Computing Environment

All experiments are conducted in the following computing environment:

- **Operating system**: Ubuntu 22.04 (x86_64)
- **GPU**: NVIDIA GeForce RTX 3090 (24 GB)
- **CPU**: Intel Xeon Gold 6133 @ 2.50 GHz

- **System memory**: 125 GB RAM
- **Python**: 3.11.15
- **PyTorch**: 2.5.1 with CUDA 12.1
- **Key libraries**: NumPy 2.1.2, pandas 2.3.3, Matplotlib 3.10.8, scikit-learn 1.7.2, XGBoost 3.2.0, Optuna 4.8.0, and tqdm 4.66.5

B.4 Evaluation Metrics

We evaluate forecasting performance using mean absolute error (MAE), root mean square error (RMSE), and Nash–Sutcliffe efficiency (NSE). For a given test station, let y_i and $\hat{y}_i$ denote the observed and forecast SWE values, respectively, let $\bar{y}$ denote the mean of the observations, and let N denote the number of paired observations and forecasts.

Mean Absolute Error (MAE):

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|\,. \tag{20}$$

MAE measures the average magnitude of forecast errors and provides a direct assessment of overall forecast accuracy.

Root Mean Square Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}. \tag{21}$$

RMSE assigns greater weight to large errors through squared differences and therefore emphasizes substantial deviations between forecasts and observations.

Nash–Sutcliffe Efficiency (NSE):

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}. \tag{22}$$

NSE evaluates forecast skill relative to using the mean of the observations and reflects how well the forecasts reproduce variations in observed SWE.

Lower MAE and RMSE values and higher NSE values indicate better forecasting performance. NSE ranges from $-\infty$ to 1: a value of 1 indicates a perfect forecast, a value of 0 indicates performance equivalent to using the mean of the observations, and a negative value indicates poorer performance than this reference. MAE and RMSE are reported in millimeters (mm) of SWE. Each metric is computed separately for every test station, and the median across all test stations is reported.

B.5 Random Seeds and Reproducibility

We set random seeds for Python's `random` module, NumPy, and PyTorch in all experiments. For the main comparison in Table 1 of the main paper, each model is independently trained and evaluated using three seeds: 42, 222, and 789. We report the mean and standard deviation across these three runs. All remaining experiments, including the input configuration analysis and ablation studies, use a single run with seed 42. We also enable `torch.backends.cudnn.deterministic` to reduce nondeterminism in GPU-accelerated cuDNN operations.

C Additional Experiments

C.1 Hyperparameter Sensitivity Analysis

Figure 5 examines the sensitivity of SWECast to the number k of dominant frequency components retained by the Periodicity Modeling Module (PMM), the hidden dimension, and the number of Adaptive Periodicity Blocks. As shown in Figure 5(a), retaining only one frequency component

results in substantially higher MAE and RMSE. Once k reaches 3, both metrics remain relatively stable, indicating that several dominant frequency components are necessary, whereas retaining additional components provides no consistent improvement.

Figures 5(b) and 5(c) show the joint effects of the hidden dimension and the number of Adaptive Periodicity Blocks. A single block consistently yields the largest errors across all hidden dimensions. Increasing the model depth and hidden dimension generally improves forecasting accuracy, but the gains diminish beyond a hidden dimension of 256, and adding more blocks does not lead to monotonic improvements. Among the evaluated configurations, five blocks with a hidden dimension of 256 achieve the lowest MAE of 14.77 mm while maintaining a competitive RMSE of 25.70 mm. Overall, SWECast remains stable across a reasonable range of these hyperparameters once sufficient model capacity is provided.

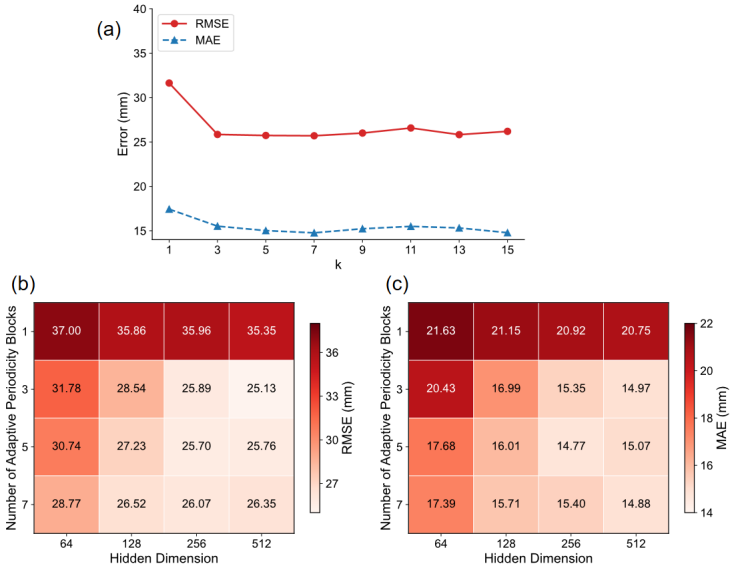


Figure 5: Hyperparameter sensitivity of SWECast. (a) MAE and RMSE for different numbers k of dominant frequency components retained by the PMM. (b) RMSE and (c) MAE for different combinations of hidden dimension and number of Adaptive Periodicity Blocks. Lower values indicate better performance.

C.2 Performance Across Months and SWE Levels

To examine how forecasting performance varies over the snow season and across different snow amounts, we analyze the forecasts by calendar month and station-level mean observed SWE. All results in this subsection are based on the 14-day forecast horizon.

Performance Across Months. Figure 6(a) compares the monthly distributions of observed and forecast SWE. SWECast closely reproduces the seasonal increase in SWE from December to April and its subsequent decline toward June, with similar medians and distributions between the forecasts and observations in each month. Figure 6(b) presents the distribution of station-level MAE and the median station-level NSE for each month. MAE generally increases during the snowmelt period, particularly from April to June, while NSE remains above 0.8 throughout the evaluated period but declines after March. These results indicate that SWECast captures the overall seasonal evolution of SWE, although forecasting SWE changes during the snowmelt period remains more challenging.

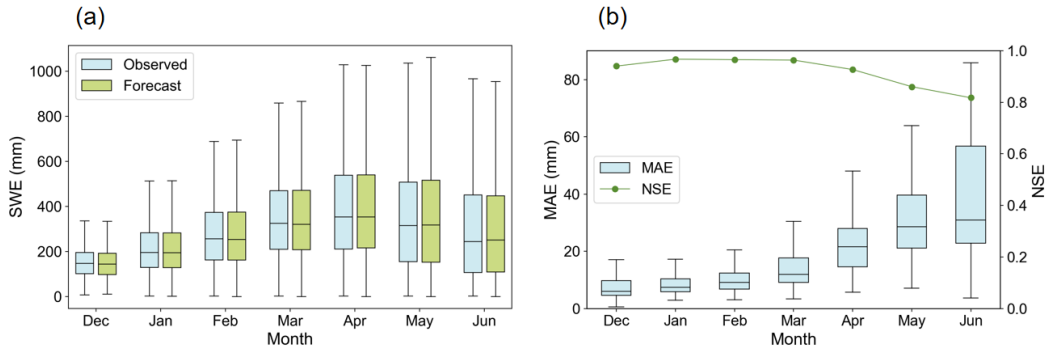


Figure 6: Forecasting performance across months for the 14-day forecast horizon. (a) Monthly distributions of observed and forecast SWE. (b) Station-level MAE distributions and NSE for each month.

Performance Across SWE Levels. Figure 7 evaluates performance across four groups defined by station-level mean observed SWE. As shown in Figure 7(a), the median percentage bias remains close to zero across all SWE groups, indicating no consistent tendency toward overestimation or underestimation. However, the lowest-SWE group exhibits greater variability in percentage bias. Figure 7(b) shows that NSE generally increases and becomes less variable as mean SWE increases. The point-wise comparison in Figure 7(c) further shows that the forecasts closely follow the one-to-one line over the full range of SWE values. Overall, SWECast maintains reliable performance across different snow amounts, although relative forecast errors are more variable at stations with low SWE.

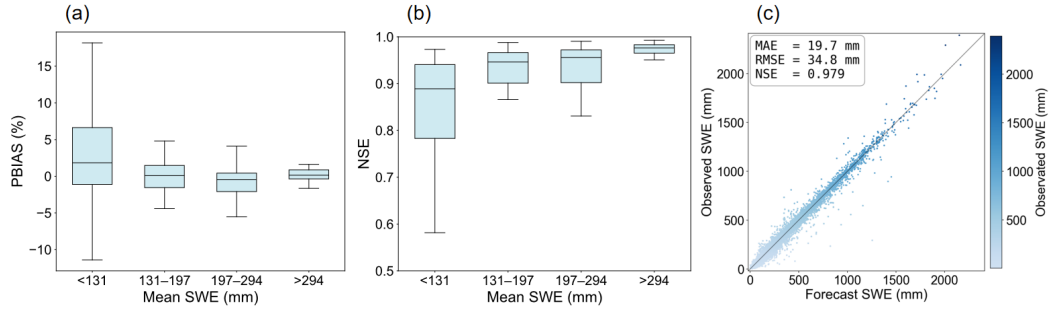


Figure 7: Forecasting performance across station-level mean observed SWE groups for the 14-day forecast horizon. Test stations are divided into four groups using the quartiles of their mean observed SWE, with the resulting SWE ranges shown on the x-axis. (a) Percentage bias (PBIAS) and (b) NSE across the four groups. (c) Observed versus forecast SWE for all test samples; the gray diagonal denotes one-to-one agreement.