https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

GEMS v1.0: Generalizable empirical model of snow accumulation and melt based on daily snow mass changes in response to climate and topographic drivers

Atabek Umirbekov[1,2,5], Richard Essery[3], Daniel Müller[1,2,4]

[1] Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Theodor-Lieser-Str. 2, 06120 Halle (Saale), Germany
[2] Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[3] School of Geosciences, University of Edinburgh, EH9 3JW, Edinburgh, United Kingdom
[4] Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), Humboldt Universität-zu-Berlin, Berlin, Germany
[5] Tashkent Institute of Irrigation and Agricultural Mechanization Engineers (TIIAME), 39 Kari Niyazov Str., Tashkent, 100000, Uzbekistan

Correspondence to: Atabek Umirbekov (umirbekov@iamo.de)

Abstract: Snow modeling is often hampered by the availability of input and calibration data, which can affect the choice of model complexity and its transferability. To address the trade-off between model parsimony and transferability, we present the Generalizable Empirical Model of Snow Accumulation and Melt (GEMS), a machine learning-based model that requires only daily precipitation, temperature or its daily diurnal cycle, and basic topographic features to simulate snow water equivalent. The model embeds a Support Vector Regression pretrained on a large dataset of daily observations from a diverse set of the Snowpack Telemetry Network (SNOTEL) stations in the United States. GEMS does not require any user calibration, except for the option to adjust the temperature threshold for rain-snow partitioning, though the model achieves robust simulation results with the default value. We validated the model with long term daily observations from numerous independent SNOTEL stations not included in the training and with data from reference stations of the Earth System Model-Snow Model Intercomparison Project. We demonstrate how the model advances large scale SWE modelling in regions with complex terrain that lack in-situ snow mass observations for calibration, such as the Pamir and Andes, by assessing the model`s ability to reproduce daily snow cover dynamics. Future model development should consider the effects of vegetation, improve simulation accuracy for shallow snow in warm locations at lower elevations and address wind-induced snow redistribution. Overall, GEMS provides a new approach for snow modeling that can be useful for hydro-climatic research and operational monitoring in regions where in-situ snow observations are scarce.

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

1. Introduction

Snow is a vital component of the global climate system and plays a key role in regulating the temperature of the Earth's surface and in governing the hydrologic cycle on both global and regional scales (Zhang, 2005; Sturm et al., 2017). Furthermore, snow plays an important role as a natural means of water storage and supply for human activities (Barnett et al., 2005), with a substantial share of the world`s population relying on snowmelt to provide water for agriculture and domestic needs (Mankin et al., 2015; Kraaijenbrink et al., 2021). Snowmelt is particularly crucial for densely populated downstream areas, where the timing and quantity of snow accumulation and melting in mountainous regions determine the availability of water (Armstrong et al., 2019; Immerzeel et al., 2020). Accurate estimation of snow mass accumulation and melt is therefore essential for water resource management as well as for early warning of droughts and floods.

Energy-balance and temperature-index snow models are the two main types of models to simulate snow accumulation and melting. Energy-balance snow models, also referred to as physics-based models, calculate the amount of snow mass based on the balance between the energy input to the snowpack and the energy output from the snowpack (Essery, 2006). These models consider multiple factors such as incoming solar radiation, air temperature, humidity, precipitation, and wind speed, as well as the physical properties of the snowpack, such as snow density and surface albedo. Due to high input data requirement of energy-balance models, which are often lacking especially in countries of the Global South, researchers often opt for relatively simpler conceptual temperature-index models, which rely on temperature and precipitation data. These models estimate the amount of snowmelt by determining empirical relationship between temperature and amount of snowmelt.

Despite the differences in the number of internal processes represented and the corresponding data requirements, both types of models produce similar results when calibrated and applied to the same spatial domain (Kumar et al., 2013). The growing number of the intercomparison studies conclude that model complexity does not determine performance (Essery et al., 2013; Magnusson et al., 2015; Menard et al., 2021), and simpler models may perform equally well or even outperform more sophisticated snow models in some cases, e.g. when input data is of low quality (Terzago et al., 2020). However, physics-based snow models are known to show better temporal and spatial transferability than temperature-index models (Magnusson et al., 2015), since they are able to capture the dynamic physical processes that govern formation, accumulation, and melting of snow, which allows them to simulate snow under a wide range of climate conditions. The generalizability and transferability of snow models are important considerations in their development and deployment, especially for applications over geographical domains where in-situ snow-measurements are non-existent or scarce.

In recent years, the research community saw an emergence of so-called data-driven approaches for snow modeling, which usually employ machine learning techniques on many snow observations and predictor variables. The studies employ machine learning techniques with a combination of onsite measurement and airborne or satellite data to estimate instant properties of snow pack or interpolate the spatial distribution of snow mass (e.g. Broxton, van Leeuwen and Biederman, 2019; Mital *et al.*, 2022; Santi *et al.*, 2022), rather than explicitly modeling snow mass

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

accumulation and melt dynamics. In many instances, these approaches also rely on is-situ observations, e.g. of snow depth or solar radiation, which again restricts their wider applicability due to unavailability of such data in many instances. The ability of pretrained machine learning models to generalize to new geographic and climatic domains remains another challenge; machine learning models often perform less well outside the data distribution used to train them (Chase et al., 2022; Hernanz et al., 2022).

We address these challenges with the Generalizable Empirical Model of Snow accumulation and melt (GEMS) that, by leveraging the power of machine learning to learn from a large number of diverse experiments, generates accurate estimates of snow water equivalent from a limited range of input data. Instead of modeling snow as a dynamic system, the GEMS employs assimilated statistical relationship between changes in snow mass in response to climate variables while accounting for topographic features. By incorporating diverse climate and topographic observations into the model training, we demonstrate how it simulates snow water equivalent with acceptable accuracy even in distant out-of-sample geographical locations.

2. Model description

GEMS is an empirical model based on statistical learning of daily changes in snow water equivalent in response to precipitation, temperature, and topography. It incorporates Support Vector Regression (SVR) that was trained using more than 28,000 observations of daily snow accumulation and melt from 94 stations of the Snowpack Telemetry Network (SNOTEL) in the United States. The model has only one adjustable parameter, a temperature threshold that specifies when 100% of precipitation falls as snow, which is used to confine the SVR simulations during the rain-to-snow transition and snow accumulation phases. **Figure 1** depicts the model's workflow and its primary components, which are described in greater detail in the following sections.

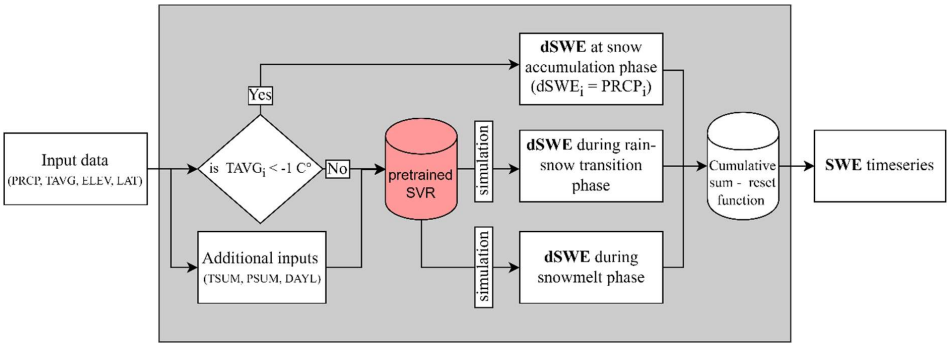


Figure 1. GEMS workflow. Model elements and abbreviations are described in the sub-sections that follow

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

The GEMS v1.0 model is implemented in the R programming environment (R Core Team, 2020), and is available as a setof functions that take input data on daily time steps, calculates additional predictors (described in the "Data for training support vector regression" section), and generates corresponding estimates of snow water equivalent. It can be applied for both single-point and spatially distributed simulations by feeding input data in tabular form or raster files, respectively.

The model is available in four variations of the required input data (**Table 1**), with the simplest one, GEMS-4P (the "P" suffix specifies the number of required inputs), requires four predictors, such as daily precipitation, average temperature, latitude, and elevation. Three other modifications, GEMS-5P, GEMS-6P and GEMS-7P, require additional predictors, such as daily diurnal temperature range (daily maximum and minimum temperatures) and a location-specific heat-insolation index, which can be retrieved through the Google Earth Engine.

Table 1. Required forcing data for GEMS (highlighted with grey per each model version)

Input data	GEMS-4P	GEMS-5P	GEMS-6P	GEMS-7P
Precipitation (mm)	✓	✓	✓	✓
Mean daily temperature (°C)	✓	✓	✓	✓
Maximum daily temperature (°C)			✓	✓
Minimum daily temperature (°C)			✓	✓
Latitude (decimal degrees)	✓	✓	✓	✓
Elevation (meters)	✓	✓	✓	✓
Heat-insolation index		✓		✓

2.1 Support vector regression

In its core embedding, GEMS is built on a pretrained SVR that estimates daily accumulation and melt of SWE given the meteorological conditions and terrain features. SVR is a supervised machine learning algorithm that projects data into a higher dimensional space, then minimizes error by generating a set of hyperplanes that explain as many observations as possible (Awad and Khanna, 2015; VAPNIK and N., 1995). SVR utilizes radial basis function kernels (Schölkopf et al., 2004) and is calibrated for optimal cost and sigma hyperparameters, which govern training errors and degree of influence of a single training point. The SVR can be expressed as:

$$SVR(X) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)\, K(x_i, x) + b \qquad (1)$$

where, α_i, α_i^* *are Langrage multipliers*, such that $\alpha_i \geq 0$ *and* $\alpha_i^* \leq 0$, and the radial basis function kernel is

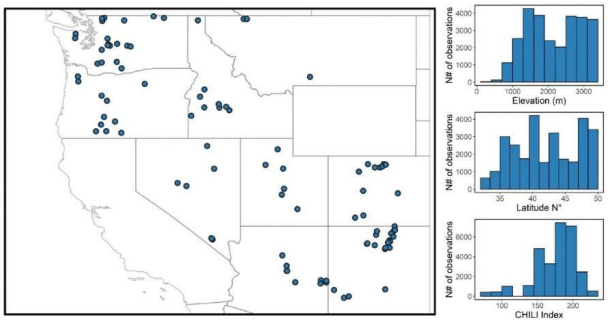
$$K(x_i, x_j) = \exp\left[-\frac{||x_i - x_j||^2}{2\sigma^2}\right]$$

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

105 **2.2 Model inputs and data for training support vector regression**

For training the SVR, we used the SNOTEL data, the largest network of automated weather stations that collect data on snow water equivalent, precipitation, temperature, and other climatic variables. We used daily observations from 94 SNOTEL stations located in the contiguous United States (**Figure 2**) for two hydrological years, 2017 and 2018.
110 As it was noted above the SVR model has two tunable parameters: cost and gamma, which can be optimized to achieve the best prediction performance, but overfitting can occur if these are overtuned. To avoid overfitting, we trained the model using data from 2017 and 2018 and fine-tuned the hyperparameters so that the model produces similar levels of accuracy when applied to observations from the same stations for 2019 and 2020. For parsimony we limit the climate inputs to precipitation and temperature variables to develop the empirical model (**Table 1**).



115

Figure 2. Location of SNOTEL stations used for training the SVR and their density distributions in terms of latitude, elevation and heat-insolation index.

In the 1990s, the temperature observations from SNOTEL showed anomalous trends (Pepin et al., 2005), which were eventually attributed to a new temperature sensor (Oyler et al., 2015), installed with an incorrect equation algorithm.
120 To correct for this bias, we applied a debiasing equation on SNOTEL temperature data proposed by Brown *et al.* (2019) and using metadata of affected stations (Air Temperature Bias Correction).

The input data includes a heat-insolation index to account for the influence of topographic shading, which may result in a significant variability of surface energy balance and therefore in snowmelt rate, particularly in complex terrain. We used the Continuous Heat-Insolation Load Index (CHILI), which approximates effects of insolation and
125 topographic shading on evapotranspiration and is determined by estimating insolation in the early afternoon at equinox sun height (Theobald et al., 2015). The Google Earth Engine provides access to CHILI data on a global scale with a horizontal resolution of 90 m. Since CHILI is a location-specific static characteristic, we also augmented the forcing data with daylength, which is a time-varying variable estimated using latitude of a location and day of a year.

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

Table 2. Climate and topographic data used to train the model

Variable	Abbreviation	Source/reference
Daily change of Snow water equivalent (mm)	dSWE	SNOTEL
Precipitation (mm)	PRCP	SNOTEL
Mean daily temperature (C°)	TAVG	SNOTEL
Maximum daily temperature (C°)	TMAX	SNOTEL
Minimum daily temperature (C°)	TMIN	SNOTEL
Rolling sum of temperature over preceding three days (C°)	TSUM	Calculated using TAVG
Cumulative sum of precipitation over preceding three days (mm)	PSUM	Calculated using PRCP
Daylength (hours)	DAYL	Calculated as a function of latitude and day of a year (Forsythe et al., 1995)
Elevation (meter)	ELEV	SNOTEL
Heat-insolation index	CHILI	Global Continuous Heat-Insolation Load Index (Theobald et al., 2015)

2.3 Temperature threshold constraint and model-wrapper function

Due to instabilities of daily changes in SWE (dSWE) estimated by the SVR during rain-snow transition phases (described in the Model validation section below), simulated dSWE values are constrained as follows:

$$\mathrm{dSWE_t} = \begin{cases} \mathrm{SVR}(X_i), & \text{if } \mathrm{TAVG_i} \geq \mathrm{T_S} \\ \mathrm{PRCP_i}, & \text{if } \mathrm{TAVG_i} < \mathrm{T_S} \end{cases} \quad (2)$$

where,

T_S is a 100% rain-snow temperature threshold, with default value of -1℃

The dSWE estimates are then aggregated into daily SWE timeseries using the cumulative sum-reset function:

$$\mathrm{SWE_t} = \begin{cases} 0, & \text{if } t = 0 \\ \max(\mathrm{dSWE}_i + \mathrm{SWE_{i-1}}, 0), & \text{if } t > 0 \end{cases} \quad (3)$$

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions
Open Access
EGU

3. Model validation

145

The evaluation of the model performance followed a three-tiered structure. First, we assessed the model performance using observations from SNOTEL stations that were not included in the training. The selection of stations for validation followed two main criteria: First. we excluded stations that exhibit precipitation undercatch, which we formulate as when SWE accumulated by March is greater than the accumulated precipitation during October to March.
150 Out of the filtered stations we selected only stations that have complete daily observations for at least five water years, defined as October of the preceding year to September next year for any year from 2011 to 2022. The selection algorithm filtered 520 stations from a total of approximately 703 contiguous US SNOTEL stations that had not been used for model training.

Second, we evaluated the model performance using snow and meteorological data from seven reference stations,
155 which were used in the Earth System Model-Snow Model Intercomparison Project (SnowMIP), hereinafter referred to as SnowMIP reference stations. **Table 3** below provides descriptions of these sites.

Table 3. Geographic and climate characteristics of the SnowMIP reference stations

Site	Abbreviation	Latitude (ºN)	Elevation (m)	Snow cover classification	Köppen climate classification
Col de Porte, France	CDP	45.3	1,325	Alpine	Warm-summer humid continental climate
Reynolds Mountain, East Idaho, USA	RME	43.19	2,060	Alpine	Warm-summer humid continental climate
Sapporo, Japan	SAP	43.08	15	Maritime	Hot summer continental climates
Senator Bec, Colorado, USA	SNB	37.91	3,714	Alpine	Polar and alpine (montane) climates
Swamp Angel, Colorado, USA	SWA	37.91	3,371	Alpine	Subarctic climate
Sodankylä, Finland	SOD	67.37	179	Taiga	Subarctic climate
Weissfluhjoch, Switzerland	WFJ	46.83	2,536	Alpine	Polar and alpine (montane) climates

Source: Ménard *et al.*, 2019

160 The evaluation metrics for single point simulations across SNOTEL and SnowMIP reference sites consist of the Nash–Sutcliffe Efficiency (NSE) coefficient, mean absolute percentage error of peak SWE (maxSWE MAPE), bias of the simulated peak SWE (maxSWE BIAS), and difference in snow melt-out dates:

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

$$NSE\left(SWE, \widehat{SWE}\right) = 1 - \frac{\sum_{i=1}^{N(days)} (SWE_i - \widehat{SWE_i})^2}{\sum_{i=1}^{N(days)} \left(SWE_i - mean(SWE_i)\right)^2}$$

where,

SWE_i – observed daily SWE;

$\widehat{SWE_i}$ – simulated daily SWE

$$maxSWE\ MAPE(y, \hat{y}) = \frac{100\%}{N\ (years)} \sum_{w=1}^{N\ (years)} \frac{|y_w - \hat{y}_w|}{y_w}$$

$$maxSWE\ BIAS(y, \hat{y}) = \frac{100\%}{N(years)} \sum_{w=1}^{N(years)} \frac{y_w - \hat{y}_w}{y_w}$$

where,

y_w– observed peak SWE in wth hydrological year;

$\hat{y}_w$ – simulated peak SWE in wth hydrological year

$$Snow\ melt\ out\ date\ error = \frac{1}{N(years)} \sum_{w=1}^{N(years)} mdate_w - \widehat{mdate}_w$$

where,

$mdate_w$– actual date of snow disappearance in wth hydrological year;

$\widehat{mdate}_w$– date of the snow disappearance according to model simulations

Finally, we assessed the performance of the model using distributed large-scale climate data over two regions with complex terrain (**Figure 3**) by comparing observed and simulated snow cover. We used temperature and precipitation data at 1 km resolution from CHELSA-W5E5 dataset (Karger et al., 2022a) to force the model, and compared the extent of SWE simulated during the two consecutive snow seasons between 2014 and 2016 with MODIS-derived snow cover retrievals using the cloud-gap filled MOD10A1F product images (Riggs et al., 2019).

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development
Discussions
Open Access
EGU

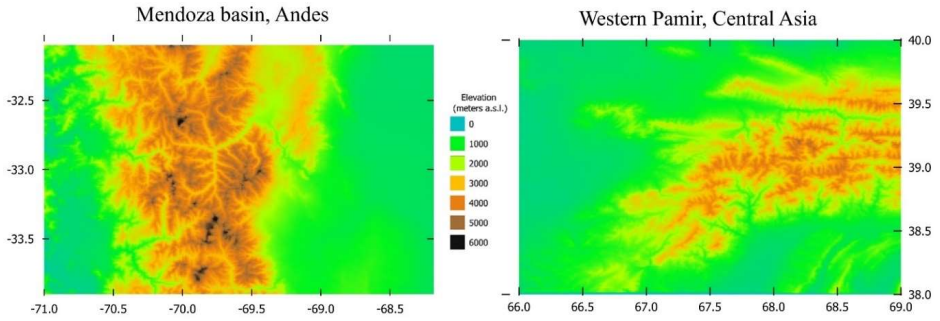


Figure 3. Selected regions for distributed snow modelling

All simulations for the validation are implemented with the GEMS-7P version of the model that uses seven predictors (**Table 1**). The section 4.3 "Performance of GEMS model under different input requirements" compares the overall performance of the model's four different versions (GEMS-7P, GEMS-6P, GEMS-5P, and GEMS-4P).

3.1 Observed and modeled daily changes in SWE across training and validation SNOTEL stations

The pretrained SVR yields plausible estimates of the dSWE, albeit the variance is greater at higher melt rates (**Figure 4**). There is a greater variance between simulated and observed values in the validation dataset, although it should be noted that the validation dataset has a much larger number of observations compared to the training and calibration datasets (1.36 million, 28,600, and 32,600 observations, respectively), which results in more outliers. In each of the three instances, the slope of the robust linear regression between the observed and simulated values ranges between 0.98 to 1.02.

The validation dataset simulations exhibited a bigger proportion of outliers in the upper tile corresponding to snow accumulation phase (dSWE>0). To determine the accuracy of the SVR's performance for this phase, we compared model simulations using a sample of the validation dataset that includes observations with incremental changes of SWE at the beginning of the snow season. **Figure 5** depicts the rain-to-snow transition modelled using the metadata of the 520 validation SNOTEL stations. We conclude that average daily temperatures at which the model predicts precipitation to fall partially as snow may range from -5 to more than 5 °C and have a relatively higher association with maximum temperature and elevation. The comparison also reveals that the simulations tend to underestimate snow accumulation, since in some cases the solid component of precipitation in simulations does not exceed 100% even at temperatures below -5 °C. In this regard, we have introduced a constraint (specified above in the section 3.4 "Temperature threshold constraint and model-wrapper function" section), which imposes that any daily precipitation after a certain temperature threshold (T_S) is considered to fall as 100% snow. We set the default value of T_S as -1 C°,

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

which the simulations revealed to be the optimal common threshold based on observations from the validation dataset. Consequently, GEMS estimates a snow fraction of precipitation using the assimilated statistical relationships until the average temperature falls below -1 C°, at which point all precipitation is considered to be snowfall.

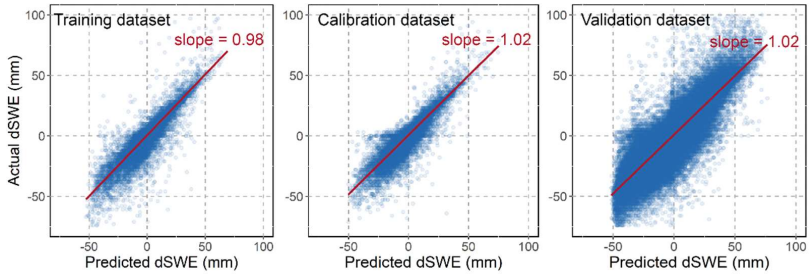


Figure 4. Predicted and observed dSWE values for training, calibration, and validation datasets. The red line represents the slope of the robust linear regression run on observed and predicted dSWE values

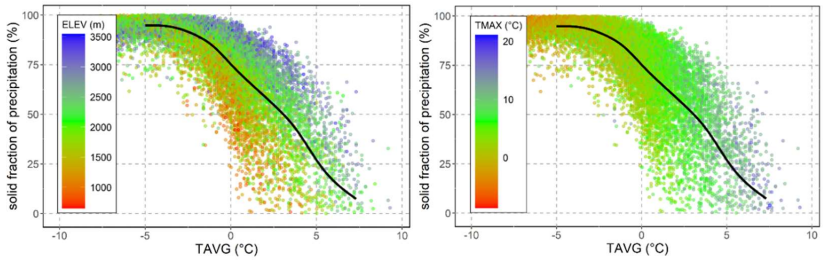


Figure 5. The rain-snow transition simulated by the pretrained SVR using metadata from the SNOTEL validation data. The two graphs illustrate the same simulations and highlight distributions of elevation **(left)** and maximum temperature **(right)**. The black line is the median of the resulting solid fraction of precipitation across the simulations.

3.2 Model evaluation in simulation of SWE timeseries for the independent SNOTEL stations

The validation on multi-annual data from the 520 independent SNOTEL stations showed that the model produces accurate simulations of SWE timeseries in most cases (**Figure 6**). The median NSE for simulations across all the stations is 0.91, and for 84% of the stations the model achieved NSE of greater than or equal to 0.8. For 80% of all stations, the maxSWE erorr (maxSWE MAPE) of the simulations is less than 20%, with the median value for all stations being 14%. The median error of the snow meltout date was four days and did not exceed ten days in 74% of instances. We found no spatial associations for NSE values and maxSWE errors, while bias for maxSWE and snow

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

meltout date error tend to be larger in the western part of the study domain (in the vicinity of the Cascade mountains, Oregon state). Here the simulations overestimate maxSWE and snow meltout date by a larger margin. Another concentration of overestimation of simulated snow melt-out date occurs in stations located in the Sierra-Nevada mountains. In contrast, the model systematically underestimates maxSWE in some stations in the north-eastern part (Montana and Wyoming), where it consequently simulates earlier snow disappearance.

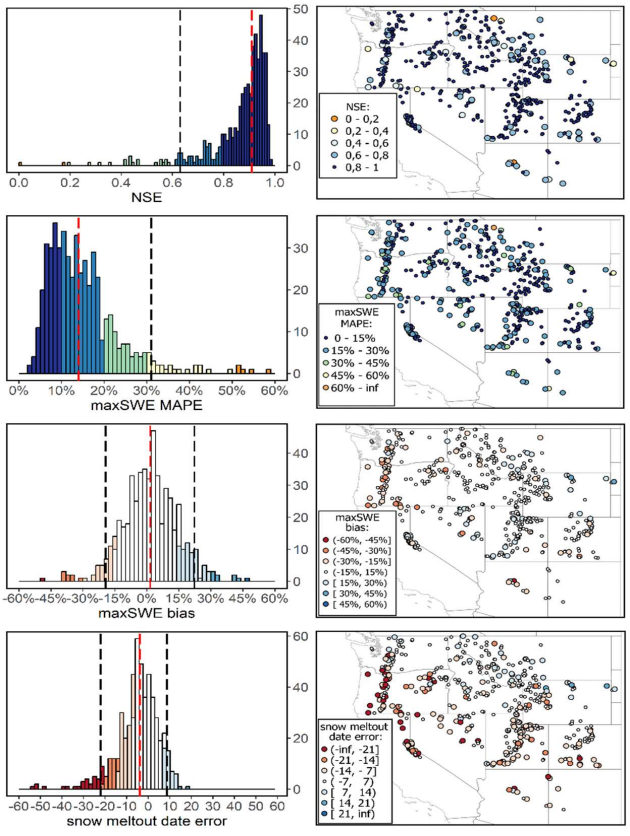


Figure 6. GEMS performance metrics for independent SNOTEL stations: histograms (left) and spatial distributions (right) of the resultant NSE, maxSWE MAPE, maxSWE bias and snow meltout date error. Vertical red dashed lines on the histograms denote the median across all stations; the vertical black dashed line correspond to the 5th percentile (and 95th percentile in case of two-tailed distributions).

In addition to simulations generated with the default T_S value, we also examined the model's accuracy using Ts values calibrated to each SNOTEL station (**Figure 7**). We bounded the range of calibrated T_S to -5 to +5 °C. Accordingly, the station-adjusted modeling incrementally improves all evaluation metrics of the simulations result, though with a lesser impact on mean maxSWE error. Adjusted T_S values tends to be negative across the stations on mountain ranges,

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

particularly across the Cascade and Sierra-Nevada and Rocky Mountains. A cluster of a few stations with positive T_S appear in the northeastern portion of the study region.

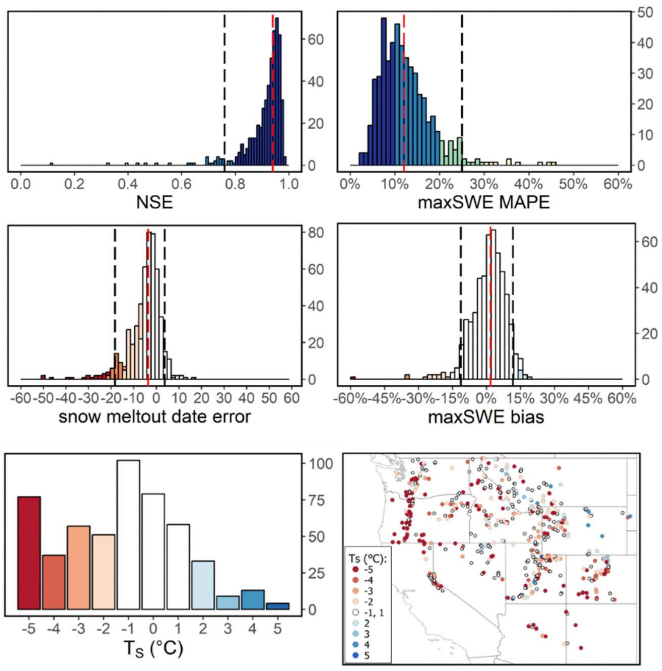


Figure 7. GEMS performance metrics for the independent SNOTEL stations with station-adjusted T_S threshold. The **bottom** histogram and map show density and spatial distribution of the adjusted T_S values.

While the median of the adjusted T_S values for all stations agrees with its default threshold (-1 °C), the density distribution also shows a high frequency of calibrated Ts resulting at the lowest bound of -5 °C. Here we should mention that the selection algorithm, designed to filter out stations with systematic precipitation undercatch also effectively excluded stations where discrepancies between SWE and accumulated precipitation are due to snow drifts into to snow pillows. However, the algorithm does not prevent inclusion of stations where snow drifts out of the station's snowpillow, which may be a frequent phenomenon across the SNOTEL network (Meyer et al., 2012). Thus in many cases where calibrated Ts values are close to the lowest boundary, the calibration likely aligned model simulations to "undercatched" snow observations.

3.3 Model evaluation using SnowMIP reference station data

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

The model performance at SnowMIP reference sites was robust for the majority of stations with the T_S threshold set to -1 C° by default (**Table 4**; **Figure 8**). In general, simulated SWE was more accurate for the stations located at higher elevations and characterized by higher snow accumulation rates (RME, CDP, WFJ, and SWA), except for SNB, which had the lowest NSE value (0.34) and the highest maxSWE error (17%) among all SnowMIP stations. The poor performance of the model for the SNB station is attributed to prevalence of wind-induced snow redistribution, which can reportedly reduce peak SWE on the site by up to 40% (Landry et al., 2014). For the same reason, the largest SWE errors were recorded for the SNB site by the majority of models that participated in SnowMIP (Menard et al., 2021).

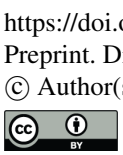
SWE simulations for SOD and SAP stations, while inferior to those of other sites, are acceptable with NSE values of around 0.7 and maxSWE MAPE errors of 8% and 18%, respectively. It is important to note that in terms of latitude and thus the range of daylengths, the SOD station is situated much beyond the range of the data utilized to pre-train the GEMS model. In addition, since the Global Continuous Heat-Insolation Load Index (CHILI) does not extend beyond the arctic circle, we estimated it for SOD based on the nearest known value and assuming flat terrain. Regarding the SAP station, GEMS' performance may be affected by the site's anomalous precipitation phase partitioning, in which precipitation reportedly can fall as rain at low temperatures and as snow at temperatures over 5C° (Ménard et al., 2019).

Table 4 also contains an evaluation of model runs using site-adjusted T_S thresholds, which improved SWE simulations across all stations, except CDP and SAP sites where the default T_S threshold remained optimal. The incremental improvements in terms of NSE and maxSWE errors are particularly noticeable for RME and SOD sites. The SNB site's NSE also increased significantly when T_S was set to -5 C°, though the simulations for the nearby SWA site show the best performance with a much lower maxSWE error when T_S is set to 1 C°. This contradiction, coupled with the fact that the maxSWE error does not change for updated SNB simulations, may suggest that the new T_S threshold rather compensates for wind-blown snow than improves SWE simulations per se.

Table 4. GEMS performance metrics for the SnowMIP reference stations.

	Performance under default T_S (Ts = -1 C°)			**Performance under adjusted T_S thresholds**			
	NSE	maxSWE MAPE (%)	maxSWE bias (%):	NSE	maxSWE MAPE (%)	maxSWE bias (%):	Ts threshold (C°):
RME	0.80	13	-13	0.88	9	-6	-5.0
CDP	0.84	14	0	0.84	14	0	default
WFJ	0.85	14	4	0.87	15	5	-1.5
SAP	0.72	17	3	0.72	17	3	default
SWA	0.85	15	14	0.86	8	8	1
SNB	0.34	17	-13	0.56	17	-4	-5.0
SOD	0.68	8	6	0.9	4	-4	3

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023



Geoscientific Model Development Discussions
Open Access
EGU

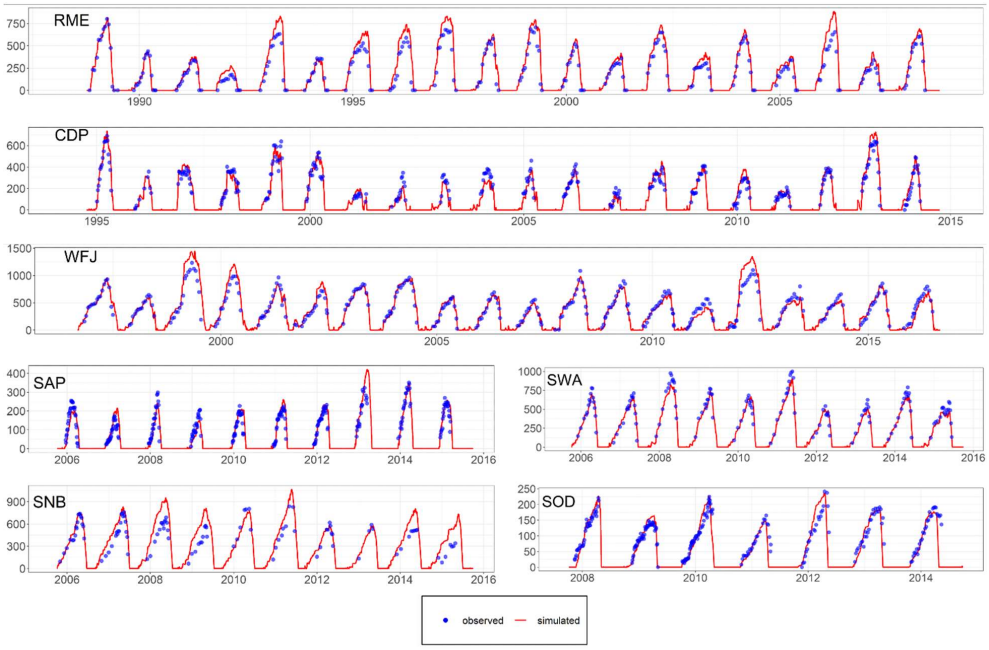


Figure 8. Observed and modelled SWE at the SnowMIP reference stations (with the default T_S threshold)

3.4 Evaluation of the model for large-scale simulations

GEMS accurately reproduced seasonal cycles and interannual variations of snow cover in the Western Pamir and Mendoza-Andes region, which have distinctive seasonal patterns (**Figure 9**). The simulations capture short-term spikes in the snow cover extent in the middle of the snow seasons over the Pamir. Overall pixel-wise accuracy of snow/no-snow detection for both regions was 92%, while the class-balanced accuracy was 87% on average.

All validation sites used previously are in the northern hemisphere because we were unable to locate representative station-based snow and climate forcing data for validation in the southern hemisphere. The validation of the model in the Mendoza-Andes region implies that the model may have comparable performance for locations in the southern hemisphere.

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

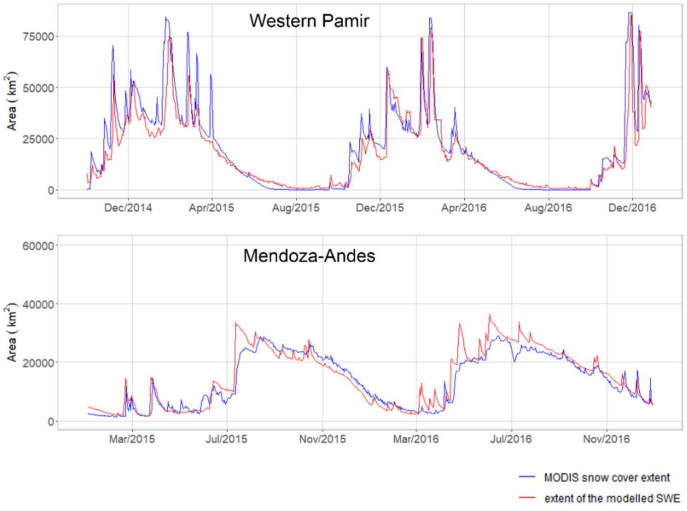


Figure 9. Observed and simulated snow cover area for Western Pamir and Mendoza-Andes region

300

4. Model sensitivity and uncertainty assessment

4.1 Relative importance of climate and topographic variables

305 We conducted a permutation-based feature importance analysis to determine how individual input variables affect dSWE simulation. The method randomly shuffles input data and compares the model's baseline performance on the original dataset to performance after permuting a feature's values. We applied the permutation-based feature importance analysis on the entire training dataset of the independent SNOTEL stations as well as its subsamples containing only snow accumulation or melt observations.

310 The results unequivocally identified precipitation and average temperature followed by daylength as the most significant variables, but they also demonstrate that their importance varies considerably depending on the phase considered (**Error! Reference source not found.**). In terms of snow accumulation, precipitation is by far the most obvious and significant variable, followed by a wide margin by maximum temperature. In contrast, the model relies heavily on average temperature and daylength to predict snow melt, followed by precipitation and other remaining
315 variables, again by a wide margin. At first glance, the results suggest that topographic variables are among the least influential, but it should be noted that this is a relative comparison, and the effect of elevation is evident from the simulation of the precipitation partitioning phase. Furthermore, climate variables can be highly variable, whereas

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

topographic features are constant per each location, which contributes to a wider gap between their relative importance.

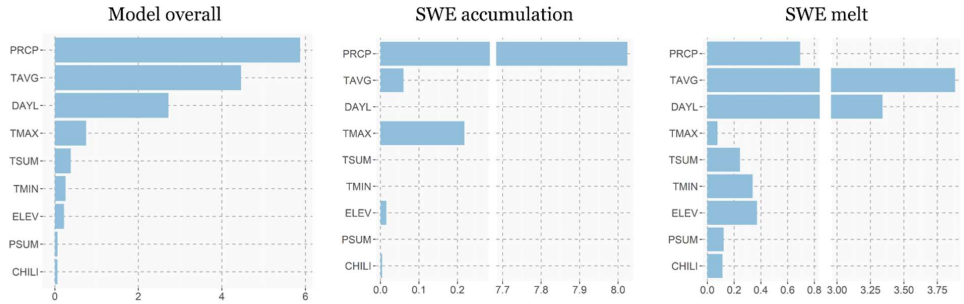


Figure 10. Relative importance of model inputs during SWE accumulation and melt phases.

4.2 Climatic and topographic attributes of locations where the model exhibited lower accuracy

In order to evaluate model uncertainty, the simulations for the SNOTEL stations with NSE values below 0.70 were separated from those with NSE values above that level. We then calculated probability distribution densities for several climatic and topographic characteristics (**Figure 11**) for each group to compare how stations with relatively poorer model performance differ from those with good model performance.

Accordingly, there is a higher likelihood that a station where the GEMS shows relatively inferior performance typically yields lower seasonal snowpack and has higher average seasonal temperatures. In addition, stations with poor model performance tend to have slightly higher diurnal fluctuations during the snow season. We have not detected any significant differences between two groups of stations in terms of average elevation or distribution of heat-insolation indices. Despite not using station latitude in the model as a direct input (it is required to estimate daylength for the location), the comparison suggests that there was a relatively higher proportion of poorly performing stations at lower latitudes.

These distinguishing characteristics of poorly performing stations are not mutually exclusive. E.g. locations with higher seasonal temperatures usually tend to have lower seasonal SWE peaks under identical conditions. Similarly, lower latitudes in the western US have generally greater diurnal air temperature variations. We hypothesize that the performance of the model under such climatic conditions could be enhanced by incorporating more respective observations into the training dataset, which apparently included fewer SNOTEL stations from the southern part of the training domain (see **Figure 2**).

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

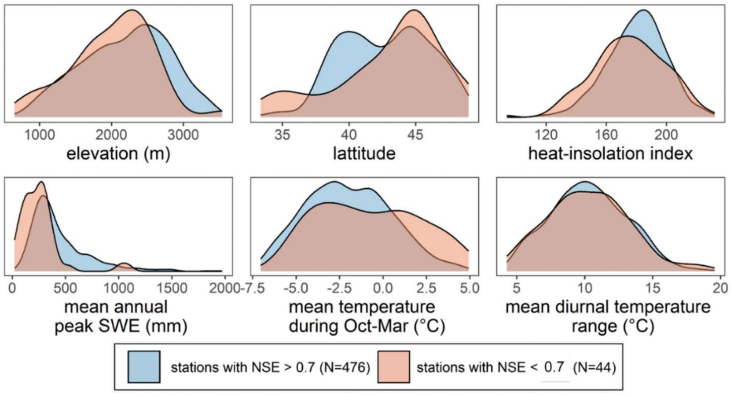


Figure 11. Density distributions of topographic and climatic characteristics of the SNOTEL stations where the model shows high and low performance.

345

4.3 Performance of GEMS model under different input requirements

To evaluate the performance of various input ranges, we compared simulations of the three versions of GEMS using the SNOTEL validation dataset. Overall, the incorporation of diurnal temperature range and heat-insolation index
350 enhances simulation accuracy as measured by a smaller interquartile range of NSE and maxSWE error (Figure 12). Compared to utilizing only the maximum and minimum temperatures, the heat-insolation index is a predictor that appears to modestly improve model accuracy. This improvement is evident, as compared to GEMS-6P, GEMS-5P exhibits somewhat better performance across the four metrics used.

Besides, GEMS-7P and GEMS-5P have a tighter range between the minimum and maximum NSE and maxSWE error
355 when outliers are controlled for. However, there is no discernible difference in the snow meltout date and maximum SWE bias across the three model versions. Although GEMS-4P has a slightly lower NSE and maxSWE error accuracy, its overall performance is still robust and it has the benefit of requiring less inputs (only precipitation, average temperature, elevation, and latitude).

Running any of the three versions of the model on a desktop computer using single CPU core (Intel i7) took less than
360 6 seconds for 20-year long Weissfluhjoch station data, which approximates to 0.3 seconds per site-year.

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

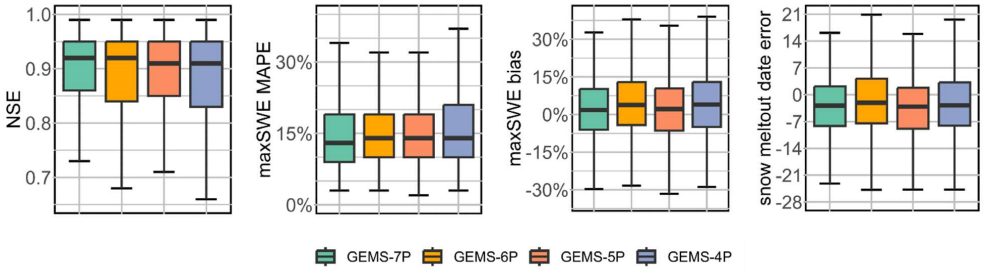


Figure 12. Comparison of performance across three GEMS models. The boxplots` minimum and maximum limits correspond to 1st Q-1.5*IQR and 3rd Q+1.5*IQR respectively.

5. Summary

We present a computationally efficient model that emulates snow mass accumulation and melting using only a few climate and topographic inputs. The absence of the explicit need for calibration is the most distinctive aspect of the GEMS model, with 100% rain-snow transition temperature threshold (T_S) being the only parameter that can be modified (in most validation cases, robust simulations were obtained just using the default T_S value). Despite its parsimony and no extensive calibration options, the model achieves robust transferability across a variety of climate and geographic conditions. In addition to avoiding computationally demanding calibration, GEMS also addresses the equifinality issue that is pertinent to hydrological and snow modelling.

The emulator was developed by training a machine learning model on daily changes in snow water equivalent as a response to daily climate inputs and diverse topographic features. Despite the dynamic nature of snow processes, our simplified "static" approach effectively captured the impact of precipitation, temperature, and topography on snow melt, as indicated by the validation results. This corroborates the conclusion of several intercomparison studies that model complexity is not necessarily a predeterminant of its performance (Essery et al., 2013; Magnusson et al., 2015; Menard et al., 2021).

The main motivation behind the development of GEMS was to balance the trade-off between complexity, data requirement, and transferability, which can be instrumental for operational monitoring and hydrological modelling in data scarce domains. The derived empirical relationships may also proof useful for the advancement of snow

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Geoscientific Model Development Discussions

modelling. GEMS can, for instance, provide information for the parameterization of physics-based models, e.g. precipitation phase partitioning and its elevational dependence.

5.1 Model limitations

Instances of less consistent simulations generated by the model can arise from various sources of uncertainty, including internal uncertainty within the model, as well as uncertainty in input data and unaccounted external factors.

One of the major limitations of the model is that it does not account for vegetation, which is known to have a complex and divergent effect on snow accumulation and melt under different climate conditions (Dickerson-Lange et al., 2021; Sun et al., 2022). Because most SNOTEL stations are situated in forest clearings or open bushes, we initially assumed the training sample locations to be free of canopy obstruction. A visual check using Google Earth of the stations used in both the training and validation samples reveals, however, that some sites can be intercepted by the tree canopies in their surroundings. In addition, we have detected the shadowing of some snow pillows by the dense forests that surround them. Both phenomena are possible sources of model uncertainty, and future model development should try to incorporate vegetation effects.

Based on a comparison of high performing and low performing site simulations, the model may be less accurate at simulating shallow snow in warm sites at lower elevations. When these factors combine with large diurnal temperature fluctuations, model simulations may even become more distorted. These issues could be resolved with a more sophisticated sampling strategy and by incorporating additional observations into the training of the model. It is questionable, however, whether an improved sampling strategy could also better approximate rain-on-snow effects, as these are driven by dynamic processes of energy exchanges across snow layers that the model does not capture for.

The model's parsimonious design, which relies only on precipitation and temperature variables as climate data inputs, also precludes the incorporation of wind- or gravity-induced snow redistribution. While this may compromise the accuracy of single-point simulations for wind-exposed sites, it may be a minor concern for large-scale distributed applications of the model.

Code and data availability: The current version of GEMS model is available from the project website: https://github.com/iamo-lsg/GEMS . The exact version of the model used to produce the results used in this paper is available on https://doi.org/10.5281/zenodo.7929181 under the Creative Commons Attribution 4.0 International license. The model repository (https://doi.org/10.5281/zenodo.7929181) also contains files with a validation set of temperature-bias-corrected SNOTEL data, as well as SnowMIP stations data aggregated to daily time scales. Additional code and supplementary files to produce results and graphs presented in this paper are available at https://zenodo.org/record/8171491. The original SNOTEL data is accessible via https://wcc.sc.egov.usda.gov/reportGenerator (USDA, 2022). The original SnowMIP reference station data is

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

accessible at https://doi.pangaea.de/10.1594/PANGAEA.897575 (Menard and Essery, 2019). CHELSA-W5E5 v1.0 data is accessible at https://data.isimip.org/10.48364/ISIMIP.836809.3 (Karger et al., 2022b)

Author contributions: AU and DM designed the study, AU performed computations, RE provided feedback on model evaluation. All authors contributed to writing and review of the manuscript. DM supervised the project.

Competing interests: The authors declare that they have no conflict of interest.

Acknowledgments: This research has been supported by the Volkswagen Foundation within the 'Structured doctoral programme on Sustainable Agricultural Development in Central Asia' (SUSADICA) project, Grant Number 96 264. We would like to acknowledge the use of IAMO`s computing facilities in this research.

References

Armstrong, R. L., Rittger, K., Brodzik, M. J., Racoviteanu, A., Barrett, A. P., Khalsa, S.-J. S., Raup, B., Hill, A. F., Khan, A. L., Wilson, A. M., Kayastha, R. B., Fetterer, F., and Armstrong, B.: Runoff from glacier ice and seasonal snow in High Asia: separating melt water sources in river flow, Reg. Environ. Chang., 19, 1249–1261, https://doi.org/10.1007/s10113-018-1429-0, 2019.

Awad, M. and Khanna, R.: Support Vector Regression, in: Efficient Learning Machines, Apress, Berkeley, CA, 67–80, https://doi.org/10.1007/978-1-4302-5990-9_4, 2015.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions., Nature, 438, 303–9, https://doi.org/10.1038/nature04141, 2005.

Brown, C. R., Domonkos, B., Brosten, T., DeMarco, T., and Rebentisch, A.: Transformation of the SNOTEL Temperature Record – Methodology and Implications, 2019.

Broxton, P. D., van Leeuwen, W. J. D., and Biederman, J. A.: Improving Snow Water Equivalent Maps With Machine Learning of Snow Survey and Lidar Measurements, Water Resour. Res., 55, 3739–3757, https://doi.org/https://doi.org/10.1029/2018WR024146, 2019.

Chase, R. J., Harrison, D. R., Burke, A., Lackmann, G. M., and McGovern, A.: A Machine Learning Tutorial for Operational Meteorology. Part I: Traditional Machine Learning, Weather Forecast., 37, 1509–1529, https://doi.org/https://doi.org/10.1175/WAF-D-22-0070.1, 2022.

Dickerson-Lange, S. E., Vano, J. A., Gersonde, R., and Lundquist, J. D.: Ranking Forest Effects on Snow Storage: A Decision Tool for Forest Management, Water Resour. Res., 57, e2020WR027926, https://doi.org/https://doi.org/10.1029/2020WR027926, 2021.

Essery, R.: Snow Modelling, https://www.ecmwf.int/sites/default/files/elibrary/2007/9315-snow-modelling.pdf,

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

2006.

Essery, R., Morin, S., Lejeune, Y., and B Ménard, C.: A comparison of 1701 snow models using observations from an alpine site, Adv. Water Resour., 55, 131–148, https://doi.org/https://doi.org/10.1016/j.advwatres.2012.07.013, 2013.

Forsythe, W. C., Rykiel, E. J., Stahl, R. S., Wu, H., and Schoolfield, R. M.: A model comparison for daylength as a function of latitude and day of year, Ecol. Modell., 80, 87–95, https://doi.org/https://doi.org/10.1016/0304-3800(94)00034-F, 1995.

Hernanz, A., García-Valero, J. A., Domínguez, M., and Rodríguez-Camino, E.: A critical view on the suitability of machine learning techniques to downscale climate change projections: Illustration for temperature with a toy experiment, Atmos. Sci. Lett., 23, e1087, https://doi.org/https://doi.org/10.1002/asl.1087, 2022.

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B. J., Elmore, A. C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J. S., Koppes, M., Kraaijenbrink, P. D. A., Kulkarni, A. V., Mayewski, P. A., Nepal, S., Pacheco, P., Painter, T. H., Pellicciotti, F., Rajaram, H., Rupper, S., Sinisalo, A., Shrestha, A. B., Viviroli, D., Wada, Y., Xiao, C., Yao, T., and Baillie, J. E. M.: Importance and vulnerability of the world's water towers, Nature, 577, 364–369, https://doi.org/10.1038/s41586-019-1822-y, 2020.

Karger, D. N., Lange, S., Hari, C., Reyer, C. P. O., Conrad, O., Zimmermann, N. E., and Frieler, K.: CHELSA-W5E5: Daily 1\,km meteorological forcing data for climate impact studies, Earth Syst. Sci. Data Discuss., 2022, 1–28, https://doi.org/10.5194/essd-2022-367, 2022a.

Karger, D. N., Lange, S., Hari, C., Reyer, C. P. O., and Zimmermann, N. E.: CHELSA-W5E5 v1.0: W5E5 v1.0 downscaled with CHELSA v2.0, https://doi.org/10.48364/ISIMIP.836809.3, 2022b.

Kraaijenbrink, P. D. A., Stigter, E. E., Yao, T., and Immerzeel, W. W.: Climate change decisive for Asia's snow meltwater supply, Nat. Clim. Chang., 11, 591–597, https://doi.org/10.1038/s41558-021-01074-x, 2021.

Kumar, M., Marks, D., Dozier, J., Reba, M., and Winstral, A.: Evaluation of distributed hydrologic impacts of temperature-index and energy-based snow models, Adv. Water Resour., 56, 77–89, https://doi.org/https://doi.org/10.1016/j.advwatres.2013.03.006, 2013.

Landry, C. C., Buck, K. A., Raleigh, M. S., and Clark, M. P.: Mountain system monitoring at Senator Beck Basin, San Juan Mountains, Colorado: A new integrative data source to develop and evaluate models of snow and hydrologic processes, Water Resour. Res., 50, 1773–1788, https://doi.org/https://doi.org/10.1002/2013WR013711, 2014.

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A., and Jonas, T.: Evaluating snow models with varying process representations for hydrological applications, Water Resour. Res., 51, 2707–2723, https://doi.org/https://doi.org/10.1002/2014WR016498, 2015.

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

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, Environ. Res. Lett., 10, 114016, https://doi.org/10.1088/1748-9326/10/11/114016, 2015.

Menard, C. and Essery, R.: ESM-SnowMIP meteorological and evaluation datasets at ten reference sites (in situ and bias corrected reanalysis data), https://doi.org/10.1594/PANGAEA.897575, 2019.

Menard, C. B., Essery, R., Krinner, G., Arduini, G., Bartlett, P., Boone, A., Brutel-Vuilmet, C., Burke, E., Cuntz, M., Dai, Y., Decharme, B., Dutra, E., Fang, X., Fierz, C., Gusev, Y., Hagemann, S., Haverd, V., Kim, H., Lafaysse, M., Marke, T., Nasonova, O., Nitta, T., Niwano, M., Pomeroy, J., Schädler, G., Semenov, V. A., Smirnova, T., Strasser, U., Swenson, S., Turkov, D., Wever, N., and Yuan, H.: Scientific and Human Errors in a Snow Model Intercomparison, Bull. Am. Meteorol. Soc., 102, E61–E79, https://doi.org/10.1175/BAMS-D-19-0329.1, 2021.

Ménard, C. B., Essery, R., Barr, A., Bartlett, P., Derry, J., Dumont, M., Fierz, C., Kim, H., Kontu, A., Lejeune, Y., Marks, D., Niwano, M., Raleigh, M., Wang, L., and Wever, N.: Meteorological and evaluation datasets for snow modelling at 10 reference sites: description of in situ and bias-corrected reanalysis data, Earth Syst. Sci. Data, 11, 865–880, https://doi.org/10.5194/essd-11-865-2019, 2019.

Meyer, J. D. D., Jin, J., and Wang, S.-Y.: Systematic Patterns of the Inconsistency between Snow Water Equivalent and Accumulated Precipitation as Reported by the Snowpack Telemetry Network, J. Hydrometeorol., 13, 1970–1976, https://doi.org/https://doi.org/10.1175/JHM-D-12-066.1, 2012.

Mital, U., Dwivedi, D., Özgen-Xian, I., Brown, J. B., and Steefel, C. I.: Modeling Spatial Distribution of Snow Water Equivalent by Combining Meteorological and Satellite Data with Lidar Maps, Artif. Intell. Earth Syst., 1, e220010, https://doi.org/10.1175/AIES-D-22-0010.1, 2022.

Oyler, J. W., Dobrowski, S. Z., Ballantyne, A. P., Klene, A. E., and Running, S. W.: Artificial amplification of warming trends across the mountains of the western United States, Geophys. Res. Lett., 42, 153–161, https://doi.org/https://doi.org/10.1002/2014GL062803, 2015.

Pepin, N. C., Losleben, M., Hartman, M., and Chowanski, K.: A Comparison of SNOTEL and GHCN/CRU Surface Temperatures with Free-Air Temperatures at High Elevations in the Western United States: Data Compatibility and Trends, J. Clim., 18, 1967–1985, https://doi.org/10.1175/JCLI3375.1, 2005.

R Core Team: R: A language and environment for statistical computing., https://www.r-project.org/, 2020.

Riggs, G., Hall, D., and Salomonson, V.: MODIS snow products user guide to collection 6.1, 2019.

Santi, E., De Gregorio, L., Pettinato, S., Cuozzo, G., Jacob, A., Notarnicola, C., Günther, D., Strasser, U., Cigna, F., Tapete, D., and Paloscia, S.: On the Use of COSMO-SkyMed X-Band SAR for Estimating Snow Water Equivalent in Alpine Areas: A Retrieval Approach Based on Machine Learning and Snow Models, IEEE Trans. Geosci. Remote Sens., 60, 1–19, https://doi.org/10.1109/TGRS.2022.3191409, 2022.

https://doi.org/10.5194/gmd-2023-103
Preprint. Discussion started: 1 August 2023

Schölkopf, B., Tsuda, K., and Vert, J.-P.: A primer on kernel methods, in: Kernel Methods in Computational Biology, MIT Press, 2004.

Sturm, M., Goldstein, M. A., and Parr, C.: Water and life from snow: A trillion dollar science question, Water Resour. Res., 53, 3534–3544, https://doi.org/https://doi.org/10.1002/2017WR020840, 2017.

Sun, N., Yan, H., Wigmosta, M. S., Lundquist, J., Dickerson-Lange, S., and Zhou, T.: Forest Canopy Density Effects on Snowpack Across the Climate Gradients of the Western United States Mountain Ranges, Water Resour. Res., 58, e2020WR029194, https://doi.org/https://doi.org/10.1029/2020WR029194, 2022.

Terzago, S., Andreoli, V., Arduini, G., Balsamo, G., Campo, L., Cassardo, C., Cremonese, E., Dolia, D., Gabellani, S., von Hardenberg, J., di Cella, U., Palazzi, E., Piazzi, G., Pogliotti, P., and Provenzale, A.: Sensitivity of snow models to the accuracy of meteorological forcings in mountain environments, Hydrol. Earth Syst. Sci., 24, 4061–4090, https://doi.org/10.5194/hess-24-4061-2020, 2020.

Theobald, D. M., Harrison-Atlas, D., Monahan, W. B., and Albano, C. M.: Ecologically-Relevant Maps of Landforms and Physiographic Diversity for Climate Adaptation Planning, PLoS One, 10, 1–17, https://doi.org/10.1371/journal.pone.0143619, 2015.

Air Temperature Bias Correction: https://www.nrcs.usda.gov/wps/portal/wcc/home/snowClimateMonitoring/temperature/temperatureBiasCorrection/.

USDA: Report Generator 2.0 for the SNOwpack TELemetry Network (SNOTEL) database, 2022.

VAPNIK and N., V.: The Nature of Statistical Learning, II., 334 pp., 1995.

Zhang, T.: Influence of the seasonal snow cover on the ground thermal regime: An overview, Rev. Geophys., 43, https://doi.org/https://doi.org/10.1029/2004RG000157, 2005.