Advancing snow data assimilation with a dynamic observation uncertainty

Devon Dunmire[1], Michel Bechtold[1], Lucas Boeykens[1, 2], and Gabriëlle J. M. De Lannoy[1]

[1]Department of Earth and Environmental Sciences, KU Leuven, Leuven, Belgium
[2]Department of Environment, Ghent University, Ghent, Belgium

Correspondence: Devon Dunmire (devon.dunmire@kuleuven.be)

Abstract. Seasonal snow is a critical resource for society by providing water for billions, supporting agriculture, clean energy, and tourism, and is an important element within the climate system by influencing the global energy balance. However, accurately quantifying snow mass, particularly in mountainous regions, remains a challenge due to substantial observational and modeling limitations. As such, data assimilation (DA) offers a powerful solution by integrating observations with physically-based models to improve estimates of the snowpack. Previous snow DA studies have employed an Ensemble Kalman Filter (EnKF) to assimilate Sentinel-1 satellite-based snow depth retrievals, demonstrating improved accuracy in modeled snow depth, mass, and streamflow when evaluated against in-situ measurements. In those studies, the uncertainty of the assimilated retrievals was assumed to be static in time and space, likely leading to a suboptimal use of the observational information. Here, we present several advances in snow DA. Using an EnKF, we assimilate novel snow depth retrievals derived from a machine learning product that leverages Sentinel-1 backscatter observations, land cover, and topographic information over the European Alps. We also incorporate a spatiotemporally dynamic observation error, whereby the uncertainty of the assimilated snow depth retrieval varies in space and time with snow depth. The machine learning snow depth retrieval product is assimilated into the Noah-MP land surface model over the entire European Alps at 1 km resolution for the years 2015-2023 and snow depth, snow water equivalent, and snow cover are evaluated against independent in-situ data and satellite observations. This work demonstrates the benefits of machine learning based snow depth retrievals and dynamic observation errors in EnKF-based snow DA.

1 Introduction

Snow is a valuable natural resource, integral for societal needs and in the climate system. The runoff from seasonal snow serves as a water source for billions of people (Barnett et al., 2005; Mankin et al., 2015), supports clean hydroelectric energy generation (Wasti et al., 2022), and sustains irrigated agriculture (Qin et al., 2020). Snow is also necessary for the multi-billion dollar winter tourism industry (Outdoor Industry Association, 2017; Parthum and Christensen, 2022; Steiger et al., 2019). The total economic value of snow is estimated to be in the trillions of dollars (Sturm et al., 2017).

~~Snow~~ Furthermore, snow has a high albedo and therefore plays an important role within the climate system by exerting a large-scale cooling effect. Variability in snow cover therefore impacts the Earth's surface energy balance and has been shown

to potentially affect Northern Hemisphere atmospheric circulation (Henderson et al., 2018). Significant changes including a decline in snow-covered area, particularly at low elevations (Bormann et al., 2018; Estilow et al., 2015), shifts in the timing of snow melt (Musselman et al., 2021; Vorkauf et al., 2021), and an increasing transition from snowfall to rainfall at lower elevations (Safeeq et al., 2016) have been observed in recent decades, with these changes projected to intensify throughout the 21st century (IPCC, 2021).

Despite the importance of snow within Earth's climate and as a natural resource, accurately quantifying snow mass (or snow water equivalent, SWE) in mountainous, complex terrain remains a challenge. Because SWE is difficult and costly to directly quantify (Dozier et al., 2016), measurements and retrieval algorithms more commonly focus on snow depth, which is related to SWE via snow density. In-situ observation stations provide point-based snow depth measurements with good temporal frequency, but fail to capture spatial snow variability, which can be great even in a small area (López-Moreno et al., 2015; Miller et al., 2022). Airborne surveys provide accurate snow depth maps at a fine spatial resolution (Deems et al., 2013), but their high costs and logistical constraints limit the frequency and spatial coverage of these measurements. Snow depth has also been retrieved using satellite observations, which have the benefit of providing frequent, global coverage (Lievens et al., 2019). ~~Passive~~ One approach estimates snow depth by comparing digital elevation models (DEMs) from snow-on and snow-off conditions. These DEMs can be generated from satellite laser altimetry such as ICESat-2 (Enderlin et al., 2022; Deschamps-Berger et al., 2023; Besso et al., 2024) or from very-high-resolution stereoscopic satellite imagery via photogrammetric methods (Marti et al., 2016; Shaw et al., 2020; Deschamps-Berger et al., 2020). Globally, passive microwave and synthetic aperture radar (SAR) observations are ~~currently widely~~ more commonly used to estimate snow depth. (Kelly et al., 2019; Luojus et al., 2021; Lievens et al., 2022)~~; however, these satellite-based estimates also have substantial limitations and uncertainties. For example~~. However, passive microwave imagery has a coarse spatial resolution (∼25 km) and saturates above 1 m snow depth (Tedesco and Narvekar, 2010; Vander Jagt et al., 2013), while SAR observations are challenged by wet snow, shallow snow, and forest cover (Broxton et al., 2024; Hoppinen et al., 2024; Lievens et al., 2022). Although recent work has utilized machine learning (ML) techniques to enhance SAR-based snow depth retrievals (Daudt et al., 2023; Broxton et al., 2024; Dunmire et al., 2024), there is still some way to go for accurate global SWE estimation.

Ultimately, complex feedbacks between changes in snow and other components of the global climate system are currently best studied using physics-based models (Girotto et al., 2020). Since in-situ SWE observations are far sparser than snow depth measurements (Dunmire et al., 2024), snow mass estimates also rely primarily on modeling approaches. However, these models are limited by uncertainties in mountain precipitation and low-quality forcing data (Günther et al., 2019; Raleigh et al., 2016; Terzago et al., 2020). In light of these observational and modeling challenges, data assimilation (DA) offers a way to overcome shortcomings of both the model and observations by integrating in-situ and remote satellite observations with physics-based models to improve modeled snow variables (Helmert et al., 2018; Smyth et al., 2020, 2022).

~~The~~ One method for assimilating observations into a physical model is via direct insertion, whereby the model's state variables are directly replaced with observations without any statistical blending or error weighting (Rodell and Houser, 2004; Toure et al., 2018). Increasing in sophistication, optimal interpolation methods, which consider model and observational uncertainty to blend the model and observations using statistically optimal weights (Liston and Hiemstra, 2008)

, are commonly used at operational centers (Helmert et al., 2018). Also common among operational centers (Helmert et al., 2018) , and one of the most used DA techniques within the land surface modeling community, is the Ensemble Kalman Filter (EnKF; Reichle et al. (2002))~~is a commonly utilized scheme for assimilating snow observations into land surface models via state updating~~. With an EnKF, the background-error covariance ~~in the Kalman Filter~~ is not explicitly computed, but instead estimated using an ensemble of model trajectories. While this ensemble approach is advantageous for high-dimensional, nonlinear systems where an exact computation of the background-error covariance is impractical, the assumption of unbiased, normally distributed model-state errors is often violated for cumulative state variables like snow depth. Despite its reliance on Gaussian assumptions, the EnKF has been extensively used in previous snow data assimilation work (Slater and Clark, 2006; Durand and Margulis, 2006; De Lannoy et al., 2012; Huang et al., 2017; Pflug et al., 2024). ~~In contrast~~An alternative solution that is commonly used in snow DA, particle batch filters and smoothers ~~provide as alternative~~ are capable of handling non-Gaussian noise and complex posterior distributions~~, though they can be more computationally expensive given the large number of particles required~~. In particular, particle batch smoothers have been commonly applied to create snow reconstructions (Margulis et al., 2015; Baldo and Margulis, 2018; Girotto et al., 2024).

Recent studies have used both particle batch smoothers and the EnKF to assimilate SAR-based snow depth retrievals from Sentinel-1 (S1), thereby improving modeled snow depth, SWE and streamflow compared to in-situ measurements (De Lannoy et al., 2024; Brangers et al., 2024; Girotto et al., 2024). However, these previous snow DA studies make the simplifying assumption that the observation uncertainty is constant in space and time, meaning that a 10 cm snowpack is assumed to have the same absolute uncertainty as a 400 cm snowpack, contributing to a suboptimal use of the observational information.

Here, we present several advances in snow DA. First, we assimilate snow depth retrievals from an ML product that uses ~~Sentinel-1 (~~S1 ~~)~~ observations, land cover, and topographic information to estimate snow depth in the European Alps (Dunmire et al., 2024). These ML-based snow depth retrievals have a higher accuracy and lower bias compared to previous S1-based retrievals from a conceptual model (Lievens et al., 2022), when validated against in-situ observations and airborne snow depth maps ~~.~~ from the European Alps. For instance, compared to 798 Alps-wide in-situ measurement sites, the ML model has an average site mean absolute error (MAE) of 0.18 m and an average site bias of -8 mm, compared to an MAE of 0.22 m and a bias of -99 mm for the conceptual model, respectively. We assimilate these ML-based snow depth retrievals within a land surface model over the entire European Alps, a domain much larger than most previous snow DA efforts which focus primarily on smaller, regional scales. Finally, we incorporate a dynamic observation error, whereby the uncertainty of the assimilated snow depth observation varies in space and time, reflecting the more realistic dynamics of uncertainty in snowpack observations. The primary goal of this work is to assess the utility of incorporating dynamic observation errors versus commonly used static observation errors in EnKF-based snow DA.

2 Materials and methodology

In this work, we utilized the NASA Land Information System (LIS; Kumar et al. (2006); Peters-Lidard et al. (2007)) version 7.5.0 to assimilate snow depth retrievals in the Noah-MP land surface model (Niu et al., 2011; Yang et al., 2011) version 4.0.1. The snow depth retrievals, land surface model, DA experiments, and evaluation data and methods are further described below.

2.1 Noah-MP land surface model

To simulate snow processes over the European Alps (3.9945°E–17.0175°E, 42.9945°N–48.6195°N), we ran Noah-MP on a regular latitude-longitude grid with a spatial resolution of 0.009°. The model was forced with atmospheric forcing from the ECMWF Reanalysis, version 5 (ERA5; Hersbach et al. (2020)). The ERA5 data were downscaled from their native resolution (31 km) to the domain grid through bilinear spatial interpolation and by applying a topographic lapse-rate correction to correct the air-temperature forcing. ERA5 has previously been used as atmospheric forcing in other snow DA studies (Pflug et al., 2024; De Lannoy et al., 2024), and Brangers et al. (2024) additionally demonstrated that ERA5 forcing leads to superior modeled snow depth, compared with simulations forced with The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. (2017)), and MERRA-2 gauge-corrected precipitation (M2CORR; Reichle et al. (2017)). From Figure 10 of Brangers et al. (2024), the ERA5, MERRA-2, and M2CORR atmospheric forcing led to average modeled snow depth MAEs of 0.367 m, 0.404 m, and 0.434 m, and average snow depth biases of -0.07 m, +0.138 m, and -0.363 m, respectively, compared to in-situ measurement stations in the Western European Alps.

In Noah-MP, snow is simulated in up to 3 layers, depending on the total snow depth. Snow processes and properties such as melt metamorphism, canopy interception, and snow cover fraction are represented by detailed physically-based parameterizations (Niu et al., 2011). For snow albedo, we used the Canadian Land Surface Scheme (CLASS; Verseghy (1991)). For other parameterization options, we followed Brangers et al. (2024).

Before beginning our DA experiments, we performed a 15-year model spin-up (2000-2015). The experiments were conducted over the period spanning October 1, 2015 – April 30, 2023 (8 snow seasons). Noah-MP was run with a 15 minute model time step and daily averages of state variables were written to output.

2.2 Machine learning snow depth retrieval

Previous work has assimilated snow depths retrieved from the S1 satellite constellation (SD_{S1}; Brangers et al. (2024); De Lannoy et al. (2024)). Here, we assimilated snow depth estimates from Dunmire et al. (2024) (SD_{ML}), which uses machine learning to enhance S1-based snow depth retrievals. Dunmire et al. (2024) use an eXtreme Gradient Boosting (XGBoost) model that incorporates 12 input features (elevation, slope, aspect angle, topographical position index, snow class, forest cover fraction, day of snow season, snow cover fraction, cumulative snow cover fraction, local incidence angle of the S1 observation, S1 VV backscatter, and S1 cross-polarization ratio) to estimate snow depth across the European Alps at 100 m resolution. When compared to in-situ snow depth stations and airborne photogrammetry snow depth maps, SD_{ML} is shown to ~~be more~~

Table 1. Perturbation parameters applied for the OL and DA runs. * We perturb the total snow depth and propagate these perturbations into the different snow layers.

Variable	Perturbation type	Standard deviation	Cross-correlation			
Forcing variables			SW	LW	P	T
SW: Incident shortwave ($\mathrm{W\,m^{-2}}$)	multiplicative	0.6	1	-0.5	-0.5	0.3
LW: Incident longwave ($\mathrm{W\,m^{-2}}$)	additive	50.0	0.5	1	0.5	0.6
P: Precipitation ($\mathrm{kg\,m^{-2}\,s^{-2}}$)	multiplicative	0.5	-0.5	0.5	1	-0.1
T: 2 m air temperature (K)	additive	1.0	0.3	0.6	-0.1	1
Forecast variable*						
Snow depth (m)	multiplicative	0.0005				

~~accurate than~~ reduce MAE and improve bias compared to SD_{S1} (MAE reduction from 0.22 m for SD_{S1} to 0.18 m for SD_{ML}, bias improvement from -99 mm for SD_{S1} to -8 mm for SD_{ML}) (Dunmire et al., 2024; Lievens et al., 2022).

We spatially averaged the SD_{ML} retrievals to the 0.009° model resolution and masked pixels with a glacier fraction above 50%, according to version 7 of the Randolph Glacier Inventory (Pfeffer et al., 2014; RGI 7.0 Consortium, 2023). We also temporally averaged the SD_{ML} retrievals every 7 days and assimilated these estimates weekly, in the center of the 7-day averaging window. This step was taken to avoid assimilating outlier snow depths (the SD_{ML} retrievals can be noisy in time) and to avoid negative consequences (e.g. spurious temporal trends) associated with a changing assimilation frequency (Dee, 2005).

2.3 Data assimilation approach and experiments

We conducted 3 different experiments: (1) an open loop, model-only experiment (OL) which serves as a benchmark to evaluate the added value of assimilating SD_{ML} retrievals, (2) a DA experiment with an assumed constant observation error ($\mathrm{DA_{const}}$), and (3) a DA experiment with a dynamic observation error that varies spatially and temporally ($\mathrm{DA_{var}}$). For all experiments, we utilized 12 ensemble members, created by perturbing forcing variables (precipitation, 2 m air temperature, and incident longwave and shortwave radiation) and the total forecasted snow depth (with the total snow depth perturbations distributed over the snow layers). Although a larger ensemble size is more optimal, our choice of 12 ensembles is reasonable as the control vector used in the assimilation consists of just total snow depth (Pflug et al., 2024). The perturbation parameters are summarized in Table 1 and follow Modanesi et al. (2022), Bechtold et al. (2023), and Pflug et al. (2024).

For the DA experiments, we used a one-dimensional EnKF to assimilate the SD_{ML} retrievals into Noah-MP. The Kalman gain matrix determines the strength of the model corrections at each location (x) and timestep (t), and is given by Equation 1 below:

$$K(x,t) = \frac{\sigma_f^2(x,t)}{\sigma_f^2(x,t) + \sigma_{obs}^2} \tag{1}$$

where σ_f is the standard deviation of the forecast error and represents the uncertainty in the forecast's total snow, and σ_{obs} is the standard deviation of the observation error and represents the uncertainty in the observations. The EnKF extends a traditional Kalman Filter by estimating σ_f using forecast ensembles, while σ_{obs} is a user-defined parameter. Here, we tested two different approaches for σ_{obs}, one that is constant ($\mathrm{DA_{const}}$) and one that varies in space and time ($\mathrm{DA_{var}}$).

As per De Lannoy et al. (2024), the $\mathrm{DA_{const}}$ experiment assumes a constant value of ~~$\sigma^{obs} = 0.3$~~$\sigma_{obs} = 0.3$ m. The multiplicative factor for the snow depth state perturbations (Table 1) was determined experimentally through trial and error, with the optimal value selected based on its performance compared to in-situ snow depth observations over a subset region (Brangers et al. (2024), personal communication, Isis Brangers).

The $\mathrm{DA_{var}}$ experiment expands upon $\mathrm{DA_{const}}$ by varying σ_{obs} throughout space (x) and time (t) following Equation 2 below:

$$\sigma^{obs}_{obs}(x,t) = \begin{cases} 0.05, & SD_{ML}(x,t) \leq 0.167, \\ m * SD_{ML}(x,t), & 0.167 < SD_{ML}(x,t) < 3.5, \\ 1.05, & SD_{ML}(x,t) \geq 3.5. \end{cases} \quad (2)$$

where $SD_{ML}(x,t)$ is the assimilated observation at location x and time t and m is a user-defined multiplier. We ~~defined~~ calibrated m experimentally by selecting the optimal value when comparing modeled snow depth with in-situ observations in a subset region (6-8 ~~°E~~°E, 45-46 ~~°N~~°N). Here, we used ~~$m = 0.3$. To avoid issues when $SD_{ML}(i,t) = 0$~~$m = 0.3$. Equation 2 assumes that σ_{obs} varies linearly as a function of assimilated snow depth. Supplemental Figure 1 demonstrates that this assumption is valid at independent in-situ measurement sites. For SD_{ML} below 0.25 m, ~~we defined a minimum value of~~ the average error of the SD_{ML} product compared to in-situ measurements is 0.05 ~~m~~ (Supplemental Figure 1), and as such we chose this as a minimum threshold value for σ_{obs} ~~. We also defined a maximum value~~ (Equation 2). Setting this minimum threshold also avoids issues when SD_{ML}(i,t) = 0 m. We can see from Supplemental Figure 1 that there are no assimilated snow depths above 3 m at these in-situ measurement sites, making it difficult to characterize the observation error for deeper assimilated snow depths. As such, we also defined an upper threshold for σ_{obs} of 1.05 m~~. Figure~~ , corresponding to an assimilated snow depth of 3.5 m (Equation 2). This value was also chosen as an upper threshold because we observed that σ_f, which represents the uncertainty in the model-only (OL) simulated snow depth, given by the standard deviation of the model ensembles, levels off above 3.5 m snow depth (Supplemental Figure 2). We chose to reflect this feature of the forecast error in our characterization of the observation error. Figure 1 compares σ_{obs} from $\mathrm{DA_{const}}$ and $\mathrm{DA_{var}}$ as a function of the assimilated snow depth observation (SD_{ML}).

For both DA experiments, the snow updates were applied following the methodology of Brangers et al. (2024), whereby the increments applied to the total forecasted snow depth ~~and SWE~~ are divided over the different snow layers, proportionate to each layer's forecasted share of the total snowpack, and SWE is updated accordingly assuming snow density remains unchanged during each update. The compaction and redistribution of snow layers is done during the model propagation. This approach circumvents the need to compute dynamic error covariances between total snow depth and a varying number of snow state

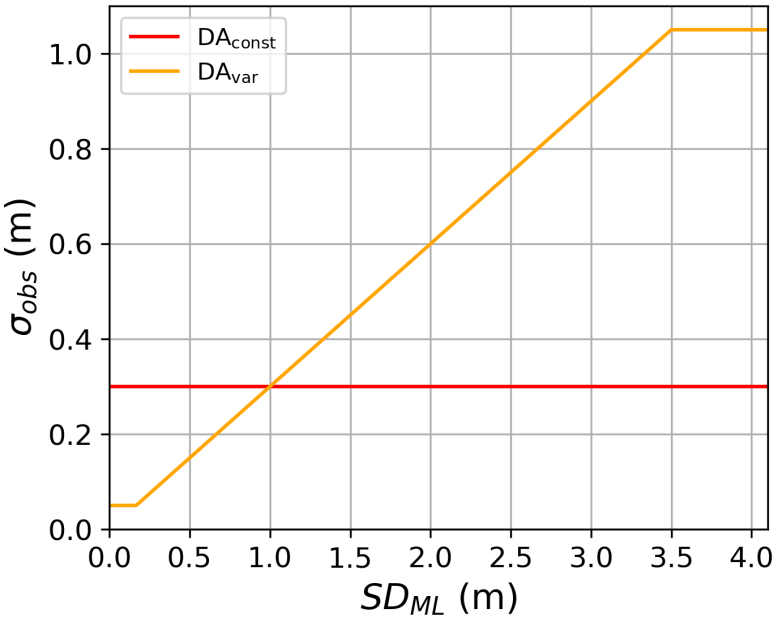

Figure 1. Assumed observation error standard deviation (σ_{obs}) as a function of the assimilated snow depth (SD_{ML}) for the two DA experiments.

variables in varying numbers of layers. We assimilated SD_{ML} estimates weekly each year from September 1 through March 31, ~~except over forested areas and glaciers, or~~ excluding assimilation further into the ablation period when wet snow complicates the S1 signal. Due to limitations of using S1 observations to estimate snow depth in forested terrain, and the unsuitability of the ML SD retrieval over glaciated terrain, we do not assimilate over forested areas or glaciers. Following De Lannoy et al. (2024), we also do not assimilate when the soil or vegetation temperature is above 5°C~~(De Lannoy et al., 2024)~~.

2.4 Evaluation

~~To evaluate the snow depth~~For each of our three experiments (OL, DA$_{const}$, DA$_{var}$), we utilized a variety of in-situ and satellite-based products to evaluate 1) snow depth, 2) SWE, and 3) snow cover fraction (SCF) and snow disappearance date (SDD). We also compared our results with those from De Lannoy et al. (2024), in which the SD_{S1} retrieval was assimilated with a static observation uncertainty.

2.4.1 Snow depth evaluation

Snow depth estimates from each experiment (OL, DA$_{const}$, DA$_{var}$) ~~, we first~~ were compared with in-situ snow depth observations from ~~a variety of snow monitoring networks~~across the European Alps. For comparing the performance of our DA experiments against the OL experiment, we utilized independent in-situ observations that were not included in the training for

the ML model from Dunmire et al. (2024), and sites located in places where the SD_{ML} retrievals were assimilated (i.e. not in dense forest, over glaciers). In total, we utilized snow depth data from 588 measurement sites, which report for varying parts of the 8 year study period. We obtained these point-scale snow depth measurements from the WSL – Institute for Snow and Avalanche Research SLF (Switzerland, 220 sites), Météo-France (France, 57 sites), GeoSphere Austria (Austria, 108 sites), the International Center for Environmental Monitoring CIMA Research Foundation (Italy, 10 sites), Provincia autonoma di Trento (Italy, 48 sites), Provincia autonoma di Bolzano - Alto Adige (Italy, 19 sites), Valle d'Aosta (Italy, 27 sites), the Agenzia Regionale per la Protezione Ambientale - Piemonte (Italy, 28 sites), the European Centre for Medium-Range Weather Forecasts' SYNOP snow depth measurement network (Global, 35 sites; de Rosnay et al. (2015)), and Global Historical Climatology Network (Global, 36 sites). For each experiment, we computed the mean absolute error (MAE), bias, and Pearson correlation coefficient (R) of the modeled snow depth compared with the in-situ observations obtained at these sites. To investigate how well the model captures spatial and temporal anomalies in snow depth patterns, we also computed spatial and temporal anomaly correlation coefficients (ACC). The spatial ACC was computed for each day throughout the snow season with more than 10 in-situ snow depth measurements available. Spatial anomalies were computed for each site by subtracting the spatial mean snow depth recorded across all measurement sites on that day. The temporal ACC was computed for each measurement site with 5 or more years of in-situ observations. Temporal anomalies were calculated at each site by subtracting the site's multi-year climatology (2015-2023) with a 10-day moving mean smoothing function applied. In order to utilize more sites with a longer time series of observations, we also included sites that were used in the ML training. Thus, for this metric, we only compared the two DA experiments, which both assimilated the same SD_{ML} retrievals.

2.4.2 SWE evaluation

Next, we evaluated modeled SWE, with in-situ measurements of SWE located (1) in places where DA was applied, and (2) not on a glacier, according to the Noah-MP glacier land cover class and the Randolph Glacier Inventory (Pfeffer et al., 2014; RGI 7.0 Consortium, 2023). We consolidated 8211 manual SWE measurements from the Bundesministerium für Land- und Forstwirtschaft, Regionen und Wasserwirtschaft (Austria, 676 measurements), The Climate Data Center of the German Weather Service (Germany, 2311 measurements), the WSL – Institute for Snow and Avalanche Research SLF (Switzerland, 1546 measurements), Provincia autonoma di Trento (Italy, 944 measurements), and Valle d'Aosta (Italy, 2793 measurements). As with snow depth, we compared MAE, bias, and R for the different experiments.

~~Finally, we~~

2.4.3 SCF and SDD evaluation

We further evaluated the impact of the DA on the timing of snow disappearance and modeled ~~snow cover fraction (SCF)~~SCF. We first compared the ~~snow disappearance date (SDD)~~ SDD of the model experiments at the in-situ snow measurement sites. We ~~computed~~ defined the SDD as the first day of five consecutive days with less than 0.1 mm snow depth, following the date of peak snow. For in-situ SDD, the day of peak snow was computed using the in-situ snow depth and for model SDD, the day of peak snow was computed using snow depth output from the appropriate model experiment. We also, in the same manner,

computed SDD using the Interactive Multisensor Snow and Ice Mapping System (IMS) product. IMS is a 1 km horizontal resolution binary snow cover dataset that is derived from a variety of satellite and in-situ data.

We also compared SCF and total snow covered area from our three model experiments with both the IMS product and the Copernicus Fractional Snow Cover product. The Copernicus product is available at a 20 m spatial resolution and is computed from Sentinel-2 ~~Level1-C~~ Level-1C imagery. The product is not gap-filled, thus data gaps exist when clouds are present. We regridded both snow cover products to our model domain grid using nearest neighbor interpolation for IMS, and averaging for the Copernicus product. For comparison with the IMS product, we converted modeled SCF to a binary value: $\mathrm{SCF} < 50\% = 0$, $\mathrm{SCF} \geq 50\% = 1$. For comparison with the Copernicus product, we ignored areas with data gaps.

2.4.4 Comparison to SD_{S1} DA

To compare with previous work that assimilates snow depth retrievals from the S1 change detection algorithm (SD_{S1}; Lievens et al. (2022)), we compared output from our two DA experiments with DA output from De Lannoy et al. (2024) (experiment $\mathrm{DA_{S1}}$). This $\mathrm{DA_{S1}}$ experiment utilized the same DA setup as in $\mathrm{DA_{const}}$, with a static observation uncertainty ($\sigma_{obs} = 0.3\ m$), but assimilates SD_{S1} retrievals instead of SD_{ML}. Here, we utilized 4548 manual SWE measurements collected within the Po River basin (the study domain of De Lannoy et al. (2024)) to compare SWE MAE between the $\mathrm{DA_{const}}$, $\mathrm{DA_{var}}$, and $\mathrm{DA_{S1}}$ experiments.

3 Results

3.1 Snow depth~~evaluation~~

The practical impact of the $\mathrm{DA_{var}}$ and $\mathrm{DA_{const}}$ experiments on snow depth estimates is illustrated in Figure 2. When the assimilated snow depth retrieval is 1 m, the observation uncertainty is equivalent for both experiments (Fig. 1). The variable observation uncertainty approach in $\mathrm{DA_{var}}$ dynamically adapts to assimilated snow depth, resulting in stronger corrections for shallow depths while $\mathrm{DA_{const}}$ provides stronger corrections at higher depths (Fig. 2). For assimilated snow depths below 1 m, the observation uncertainty is smaller in $\mathrm{DA_{var}}$ than in $\mathrm{DA_{const}}$, resulting in a lower observation error covariance (σ_{obs}) in the EnKF (Equation 1) and stronger corrections of the posterior state toward the observations in $\mathrm{DA_{var}}$ (Fig. 2a). In contrast, the constant σ_{obs} of 0.3 m in $\mathrm{DA_{const}}$ is relatively large for shallow snow depths and results in minimal corrections of the posterior state.

A measurement site with assimilated snow depths substantially greater than 1 m is demonstrated in Figure 2b. In this case, the observation uncertainty is smaller for $\mathrm{DA_{const}}$ than for $\mathrm{DA_{var}}$, resulting in stronger posterior state adjustments in $\mathrm{DA_{const}}$. At this measurement site we see that the OL experiment is closer to the in-situ snow depth than the assimilated observations, leading to a deterioration in model performance when the DA is applied (both with $\mathrm{DA_{var}}$ and $\mathrm{DA_{const}}$). For $\mathrm{DA_{var}}$, this phenomenon occurs at ~16% of all measurement site (Fig. 3a), with only 1% experiencing a deterioration in SD MAE greater than 125 mm.

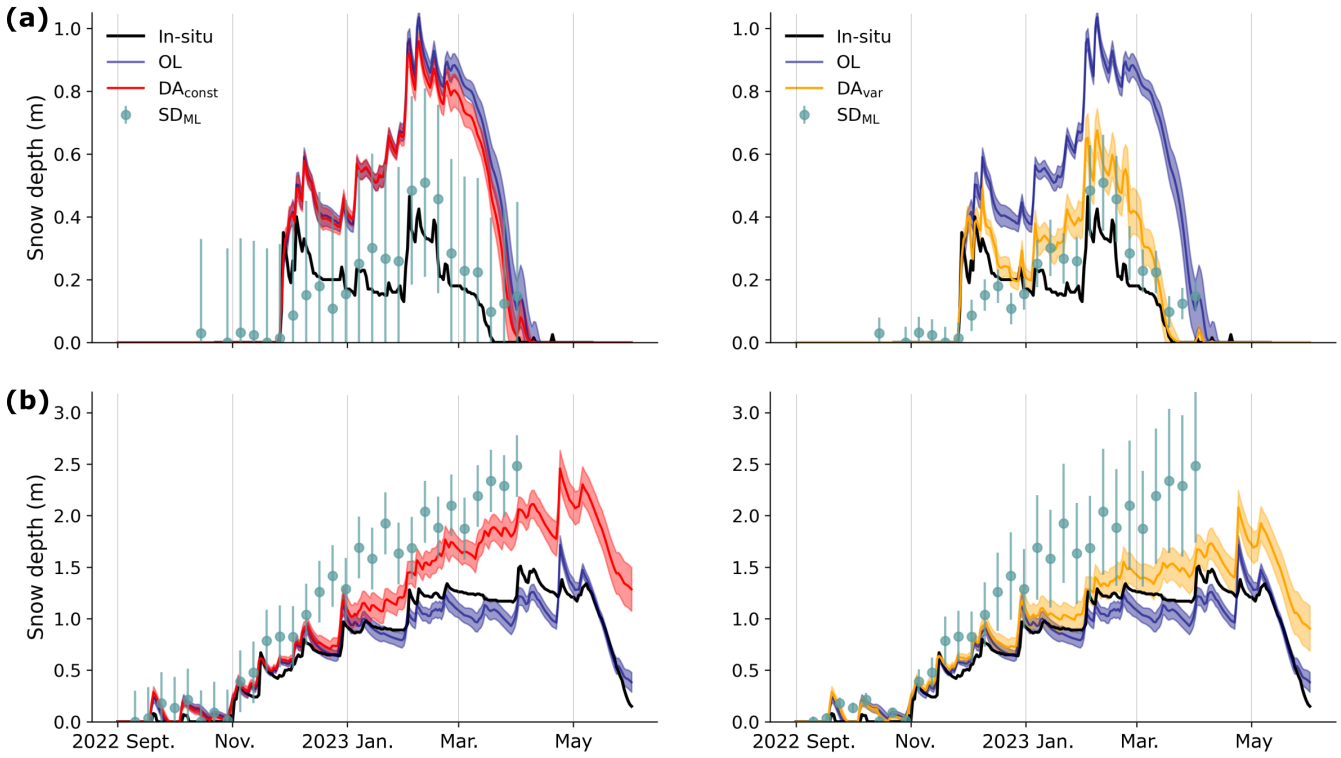

Figure 2. Snow depth estimates and independent in-situ measurements at two example sites. (a) Snow depth from $\mathrm{DA_{const}}$ (red, left) and $\mathrm{DA_{var}}$ (orange, right) compared with the OL (navy) from a measurement station in Austria (∼13.6228 °E, 47.0944 °N, 1050 m elevation). The shading represents ±1 standard deviation in the model ensembles. The sage green dots represent the assimilated SD_{ML} retrievals, with error bars for the assumed observation error standard deviation (σ_{obs}, Equation 2). (b) Same as (a), but for a different measurement station in Switzerland (∼7.7836 °E, 45.9872 °N, 2948 m elevation). These two sites were chosen due to a lack of gaps in the in-situ measurements and their general representativeness of locations where the DA removes and adds snow.

Across the 588 in-situ snow depth measurement sites used for evaluation, the corrections applied in $\mathrm{DA_{var}}$ result in snow depth estimates that align more closely with in-situ observations (Fig. 3). The OL experiment yields a site-average MAE of 0.244 m, a RMSE of of 0.300 m, a bias of 0.113 m and a Pearson correlation coefficient of 0.75. Both the $\mathrm{DA_{const}}$ and $\mathrm{DA_{var}}$ experiments show improved performance, with site-average MAE values of 0.237 m and 0.215 m, ~~and~~ RMSE values of 0.292 m and 0.268 m, and biases of 0.106 m and 0.055 m, respectively. These improvements are illustrated in Figure 3, which compares MAE from the $\mathrm{DA_{var}}$ experiment with the OL experiment (Fig. 3a) and $\mathrm{DA_{const}}$ (Fig. 3b). Relative to the OL, MAE is reduced in $\mathrm{DA_{var}}$ by more than 25 mm at 245 sites (42%), while 92 sites (16%) have an MAE increase exceeding 25 mm. Comparing $\mathrm{DA_{var}}$ to $\mathrm{DA_{const}}$, we find that MAE is reduced in $\mathrm{DA_{var}}$ by more than 15 mm at 297 sites (51%), while 71 sites (12%) experience a deterioration greater than 15 mm. While improvement in MAE from the OL experiment is

not significant for DA_{const}, the MAE improvement is significant for DA_{var} (p=0.001). The site-average Pearson correlation coefficient slightly deteriorated for DA_{const} and improved for DA_{var} to 0.75 and 0.76, respectively.

While the OL experiment already does a good job at representing spatial snow depth patterns (spatial ACC = 0.71), Figure 3c highlights that, for most of the snow season, the DA_{var} experiment ~~provides the best representation of spatial snow depth~~ offers slight improvements in the representation of these spatial patterns. Averaged across the entire year, the spatial ACC ~~is~~ increases from 0.71 for the OL experiment ~~,~~ to 0.72 for DA_{const} ~~, and~~ and to 0.73 for DA_{var}. The greatest improvement in spatial ACC for DA_{var} occurs during the early snow season (November), with values exceeding those of the OL and DA_{const} experiments by 0.058 and 0.047, respectively. From December through April, the spatial ACC for DA_{var} remains approximately 0.021 greater than that of the OL experiment. By mid-April, all three experiments exhibit similar performance in capturing spatial snow depth patterns. Additionally, ~~DA_{var} significantly outperforms~~ both DA_{const} ~~($p < 0.001$) in capturing~~ and DA_{var} well-capture temporal snow depth patterns~~(~~, with average temporal ACC values of 0.68 and 0.72, respectively. The improvement in temporal ACC for DA_{var} from both the OL and DA_{const} is statistically significant ($p < 0.01$, Fig. 3d). Across the 948 sites evaluated, 491 sites (52%) have an improved temporal ACC in DA_{var} ($>$ +0.02 compared to ~~DA_{const}~~DA_{const}), while only 103 sites (11%) experience a deterioration in temporal ACC ($<$ -0.02 compared to DA_{const}).

The OL experiment has an elevation-dependent snow depth bias, characterized by an overestimation of snow depth at lower elevations and early in the snow season, and an underestimation at higher elevations during peak snow accumulation (Fig. 4a). Both of these issues are mitigated in the DA_{var} experiment, which brings seasonal biases closer to zero across all elevation bands (Fig. 4c). In contrast, the DA_{const} experiment minimally corrects snow depth overestimation in the early season and at low elevations, due to the relatively higher assumed observation uncertainty for shallow snow (e.g., Fig. 2a). From September 1 through January 31, the DA_{var} experiment reduces the average bias across all sites by 46%, while the DA_{const} experiment achieves only a 10% reduction over the same period. These improvements are particularly notable at mid-elevations (1000–2000 m), where DA_{var} reduces model bias by 54% throughout the season, compared to a 13% reduction in model bias at these same sites in DA_{const}.

The MAE is also reduced by DA_{var} across most elevation bands and throughout much of the season. The difference in MAE between the OL and DA_{var} experiments (Fig. 4~~d~~e) indicate that the largest MAE improvements occur from early winter through peak accumulation. However, an increase in MAE at high elevations during the melt season (March onwards) suggests a tendency for the DA experiments to retain snow for too long, which could be due to limitations in the modeled melt processes or biases introduced by the assimilated observations at higher elevations (e.g., Figure 4d).

3.2 SWE~~evaluation~~

Compared with 8,211 manual SWE measurements from 231 different measurement sites across the Alps, the DA_{var} experiment also significantly improves the MAE of SWE estimates compared to both the OL and DA_{const} experiments ($p<<0.001$). Relative to the OL, DA_{var} reduces SWE MAE by at least 15 mm at a majority of these sites (57%), while only 23% of sites experience a deterioration in SWE MAE of more than 15 mm (Fig. 5a). Similar improvements are observed when comparing DA_{var} to DA_{const}, with DA_{var} outperforming DA_{const} at 56% of measurement sites (Fig. 5b).

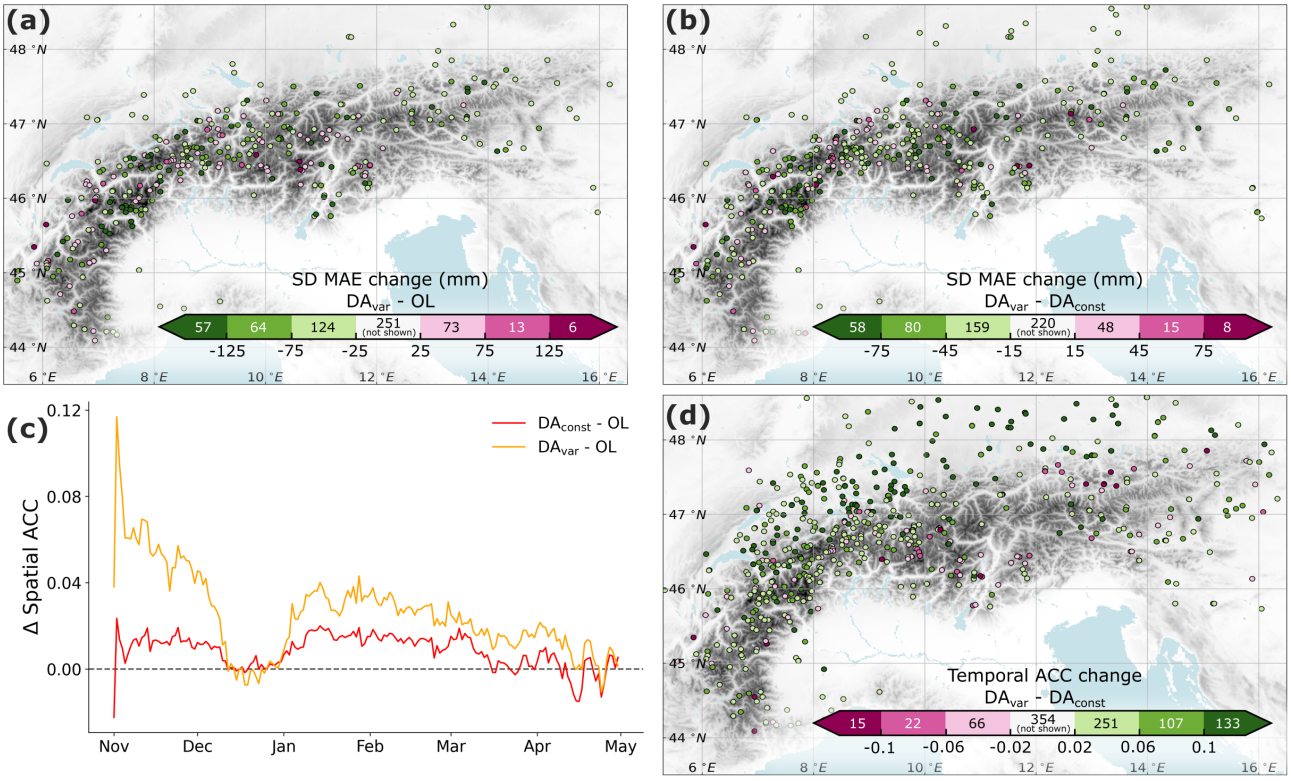

Figure 3. Experiment evaluation at in-situ snow depth measurement sites. (a) Change in MAE at each measurement site from the OL experiment to $\mathrm{DA_{var}}$. Green colors indicate an improvement in MAE in the $\mathrm{DA_{var}}$ experiment. On the color bar, the number of sites that fall within each color range is indicated and points within the white color are not plotted on the map. (b) Change in MAE at each measurement site from the $\mathrm{DA_{const}}$ experiment to $\mathrm{DA_{var}}$. (c) Change in the spatial anomaly correlation coefficient (ACC) for each DA experiment from the OL experiment. The spatial ACC is averaged over all snow seasons (2015/16 - 2022/23). (d) Change in the temporal ACC from the $\mathrm{DA_{const}}$ experiment to $\mathrm{DA_{var}}$. Green colors indicate an improvement in temporal ACC in the $\mathrm{DA_{var}}$ experiment.

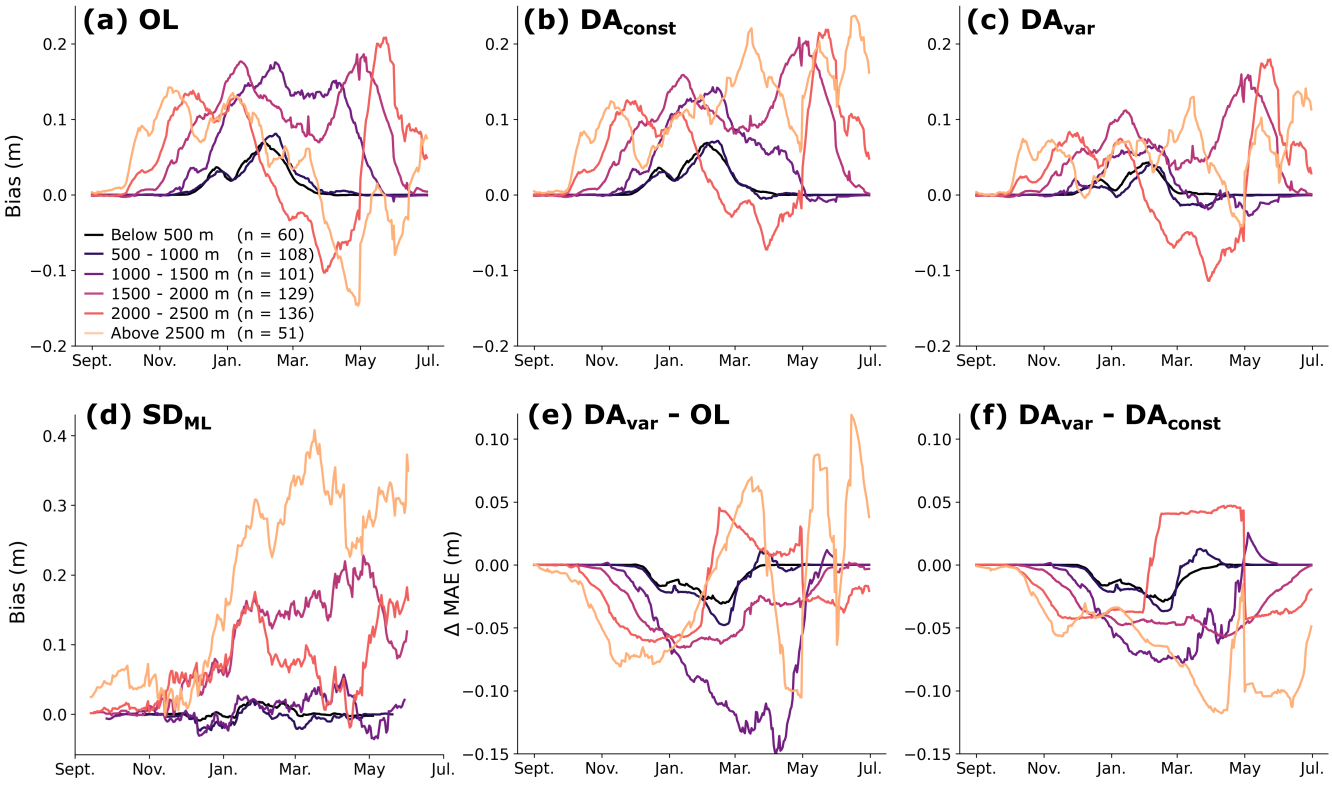

Figure 4. Seasonal evolution of bias and mean absolute error (MAE) stratified by elevation. Panels (a)-(~~c~~d) ~~snow~~ show the seasonal snow depth bias for the (a) OL, (b) DA_{const}, and (c) DA_{var} experiments~~.~~, and for *(d) the assimilated observations (SD_{ML}).* Bias is computed relative to in-situ snow depth measurements and is grouped by elevation bands (indicated by different colors). Panels (~~d~~e)–(~~e~~f) show the change in MAE between the OL and DA_{var} experiments (~~d~~e) and between the DA_{const} and DA_{var} experiments (~~e~~f). Negative values in (~~d~~e)–(~~e~~f) indicate improved performance (decreased MAE). Statistics are computed for each day, averaged over the entire ~~8 year~~ 8-year period (~~2015–2023~~2015–2023). A 14-day smoothing is applied to each timeseries and the number of in-situ measurement sites (n) within each elevation band is provided in the legend.

In the OL experiment, we observe a positive bias for low observed SWE and a negative bias for high observed SWE (Fig. 5c), similar to the bias patterns seen for snow depth. The DA_{var} experiment effectively reduces both biases, with the most substantial improvement occurring for low observed SWE values. As a result, the overall average SWE bias decreases from +81 mm in the OL to +18 mm in DA_{var}. This bias reduction is significantly greater than that for DA_{const} (+76 mm bias), which only marginally corrects the ~~high~~ positive bias for low observed SWE, due to minimal model adjustments for shallow assimilated snow depths (e.g., Fig. 2a). Both DA_{const} and DA_{var} also substantially improve the Pearson correlation coefficient ~~,~~(R = 0.60 for OL, R = 0.72 for DA_{const}, R = 0.71 for DA_{var}), indicating a stronger correlation with measured SWE.

Across all experiments, SWE typically peaks during the first week of March (March 1–7). The 2016/2017 snow season recorded the lowest modeled SWE in our OL experiment, and correspondingly saw the largest SWE increases in DA_{var} prior to early March, particularly in the Central Alps and Austrian Alps (Fig. 6a). However, DA_{var} SWE improvements were mixed during this year. Of the 41 manual measurements taken between March 1 and March 7, 2017, only 24% demonstrated improved SWE of more than 15 mm in DA_{var}. While the data assimilation led to more accurately estimated SWE at some sites (e.g., Supplemental Fig. S1b,d), it resulted in an overestimation of SWE at others (e.g., Supplemental Fig. S1c,e,f).

The most significant reductions of the positive SWE bias from the OL to the DA_{var} experiment occurred during the 2017/2018 snow season, particularly in the Bavarian Alps, Swiss Alps, and French Alps (Fig. 6b). In general, the reduced SWE in DA_{var} aligns more closely with in-situ observations (e.g., Supplemental Fig. S2), improving SWE error by more than 15 mm in 59% of the 68 manual measurements taken between March 1 and March 7, 2018.

The 2020/2021 snow season also experienced substantial SWE reductions, especially in the Swiss Alps and Eastern Dolomites. A lack of in-situ observations in the Dolomites region makes it difficult to assess whether these reductions are realistic; however, limited observation sites along the Italy-Austria border suggest that the SWE reductions may be too strong (e.g., Supplemental Fig. S3d).

3.3 Snow cover ~~evaluation~~and snow disappearance

The DA also affects snow cover estimates, contributing to a decrease in total snow-covered area leading up to peak snow accumulation in early March, and a slight increase in snow-covered area later in the season (April-May), compared with the OL experiment (Fig. 7a). During peak snow accumulation in early March (March 1–7), the DA_{var} experiment reduces total snow-covered area by 6,077 km^2 compared to the OL, averaged across the 2016-2023 period. Total snow-covered area during this same period in the DA_{const} experiment is comparatively reduced by only 1,409 km^2. The relative difference in snow-covered area between DA_{var} and the OL fluctuates more than for DA_{const} (Fig. 7a), primarily due to the shallower early-season and low-elevation snowpacks in DA_{var} which melt out more quickly.

The reduction in snow cover primarily occurs in low-elevation areas along the northern Alps (Fig. 7b), and aligns more closely with observed snow cover estimates from the IMS and Copernicus snow cover products, both of which indicate substantially less snow-covered area than any of our model simulations. For example, on March 1, 2021, the OL and DA_{var} experiments have, respectively, 79,345 km^2 and 58,091 km^2 more snow-covered area than the Copernicus fractional snow

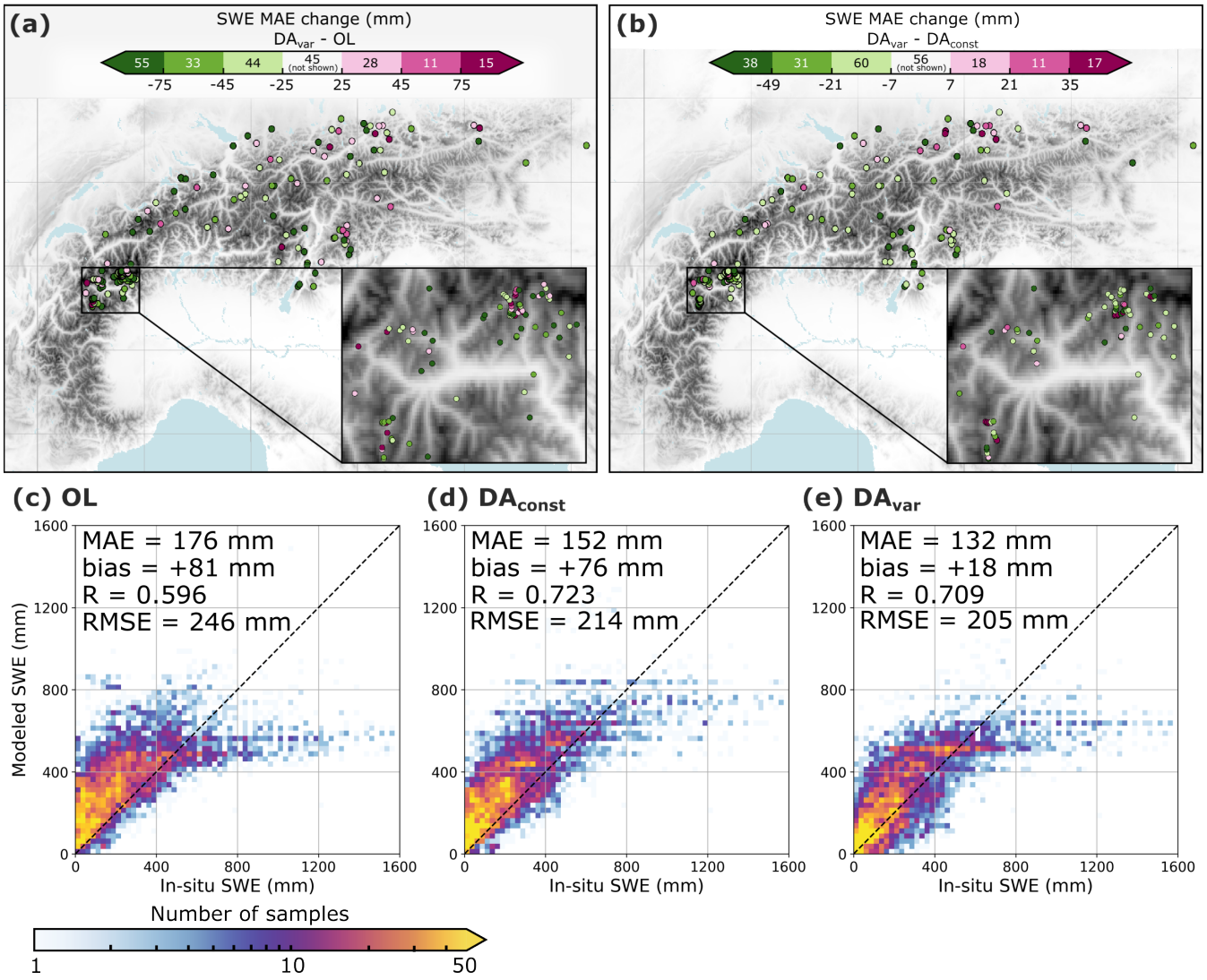

Figure 5. Evaluation of SWE from the OL and DA experiments against in-situ measurements. (a-b) Change in SWE MAE (a) $\mathrm{DA_{var}}$ relative to the OL experiment, and (b) $\mathrm{DA_{var}}$ relative to $\mathrm{DA_{const}}$, where green indicates error reduction and magenta indicates a deterioration in performance. Measurements from within the same 1 km model grid cell are averaged for visualization purposes. On the color bar, the number of sites that fall within each color range is indicated and points within the white color are not shown on the map. (c-e) 2D histograms comparing modeled SWE to in-situ SWE observations for (c) OL, (d) $\mathrm{DA_{const}}$, and (e) $\mathrm{DA_{var}}$. All non-zero SWE measurements are included and the spatiotemporal MAE, bias, ~~and~~ R, and RMSE are provided for each approach.

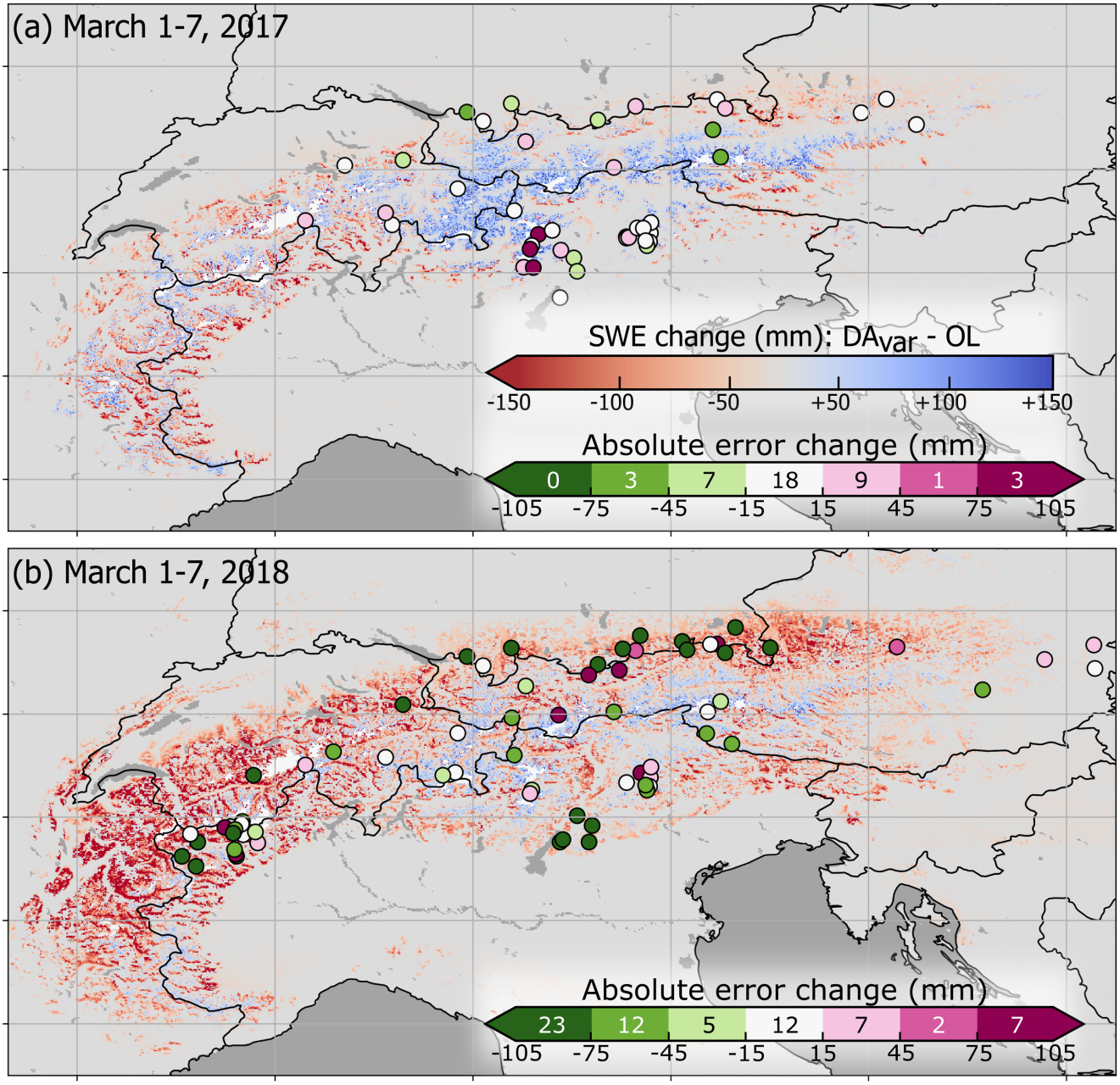

Figure 6. Change in SWE during the period March 1-7, between the $\mathrm{DA_{var}}$ and OL experiments for (a) 2017 and (b) 2018. Manual SWE measurements taken during this period are plotted as dots, colored according to the change in absolute error between the $\mathrm{DA_{var}}$ and OL experiments. On the error change color bar, the number of sites that fall within each color range is indicated.

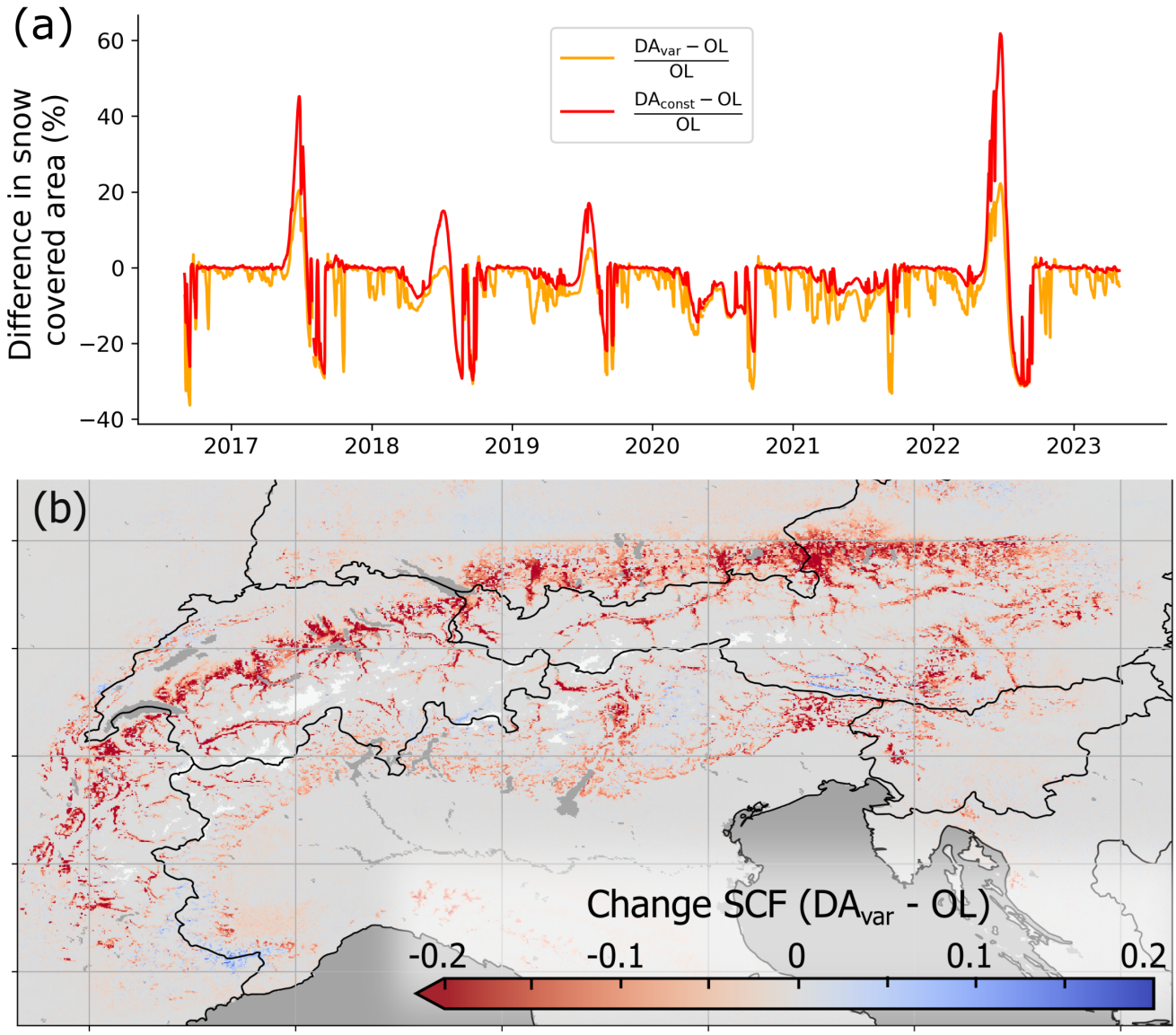

Figure 7. Difference in estimated snow covered area. (a) Timeseries of the percent difference in total snow covered area between $\mathrm{DA_{var}}$ and OL (orange), and $\mathrm{DA_{const}}$ and OL (red). (b) Average change in snow cover fraction (SCF) between the $\mathrm{DA_{var}}$ and OL experiments during the period March 1-7 (all years). White indicates glaciers.

cover product, and 55,578 km^2 and 32,526 km^2 more than the IMS snow cover product (Supplemental Fig. S4). These discrepancies will be discussed further in Section 4.

At the majority of in-situ snow depth measurement sites, the estimated snow persists for too long compared to in-situ observations. Figure 8 presents cumulative distribution functions (CDFs), which show the cumulative number of sites with

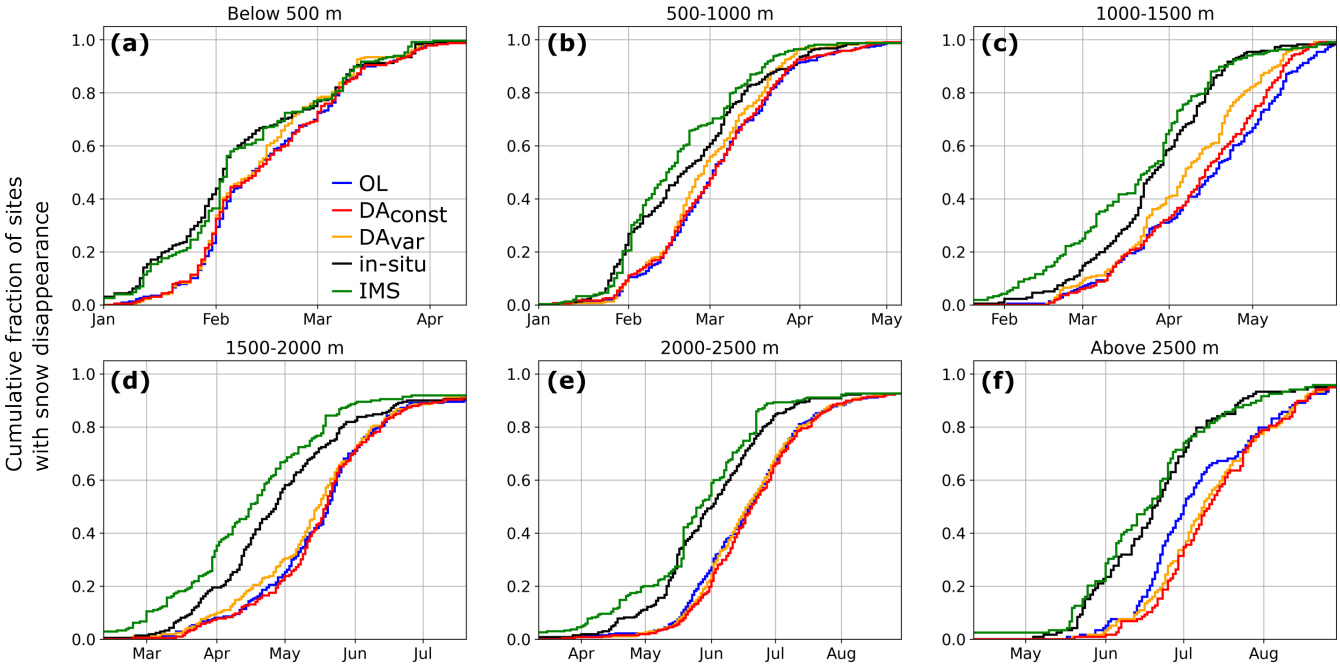

Figure 8. Cumulative ~~number~~ fraction of measurement sites with snow disappearance (following peak snow depth) at in-situ measurement sites, stratified by elevation: (a) below 500 m, (b) 500-1000 m, (c) 1000-1500 m, (d) 1500-2000 m, (e) 2000-2500 m, (f) above 2500 m.

snow-free conditions after peak snow, stratified by elevation band. In all three model experiments, the snow disappearance date (SDD) occurs later than observed, indicating an overestimation of snow persistence across all elevation bands.

In $\mathrm{DA_{var}}$, the SDD timing is improved at a majority of the observation sites located below 2000 m, with 51% of sites experiencing a SDD closer to in-situ observations, 22% experiencing a SDD farther from in-situ observations, and 27% remaining unchanged. The improvement is less pronounced for $\mathrm{DA_{const}}$, in which 40% of sites show better agreement with observations, 24% show worse agreement, and 36% remain unchanged. The reduced SWE at lower elevations in $\mathrm{DA_{var}}$ (see Section 3.2) likely results in more realistic timing for snow-free conditions at these sites. As we only assimilate observations through March, thus limiting assimilation during times of ablation, changes in SDD are mainly a result in changes of peak SWE. In general, the IMS observations underestimate snow persistence (Fig. 8), leading to an earlier SDD compared to in-situ observations, which may result from the binary (as opposed to fractional) nature of the IMS observations.

3.4 DA increments and spread

In $\mathrm{DA_{const}}$, model updates predominantly occur later in the accumulation season, with positive average increments above 2500 m and negative average increments below 1500 m (Fig. 9a). In contrast, $\mathrm{DA_{var}}$ exhibits stronger negative increments earlier in the snow season, and at lower elevations (Fig. 9b), suggesting that assimilated observations influence the entire accumulation period rather than just times near peak SWE. Additionally, the magnitude of positive increments in $\mathrm{DA_{var}}$ is

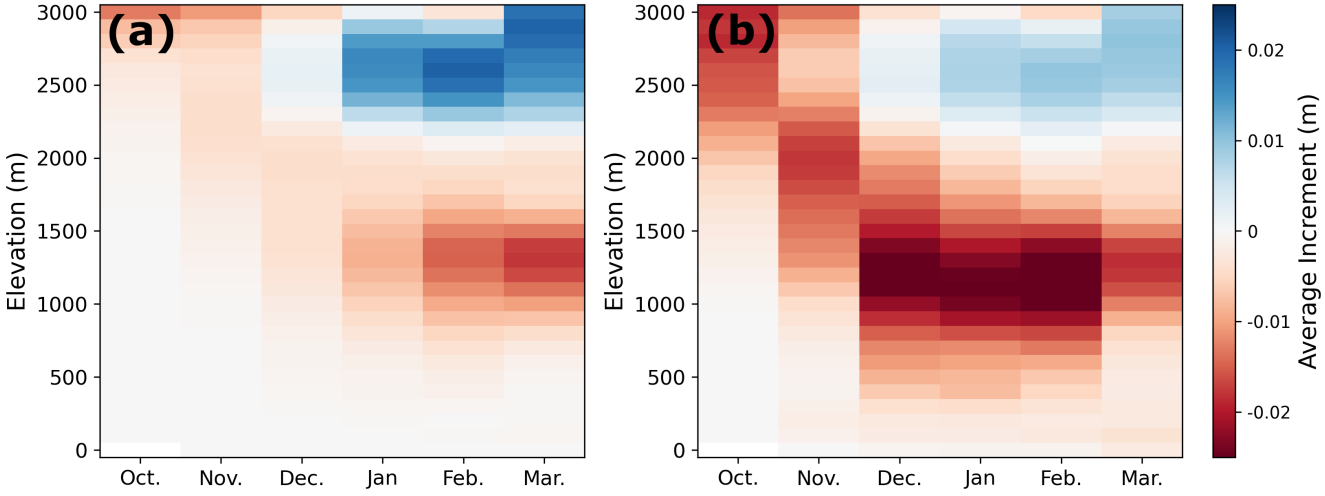

Figure 9. Average total snow depth increments (m) over the accumulation season (x-axis), stratified by elevation (y-axis) for (a) $\mathrm{DA_{const}}$ and (b) $\mathrm{DA_{var}}$. Increments are averaged over all years (2015-2023)

reduced, meaning that less snow is added at higher elevations in $\mathrm{DA_{var}}$. While the OL has a negative snow bias in these higher elevation areas, the weaker positive increments in $\mathrm{DA_{var}}$ may be more realistic, given that Figure 4b indicates a strong positive snow depth bias for sites about 2500 m in $\mathrm{DA_{const}}$, and a reduced positive bias at these same sites in $\mathrm{DA_{var}}$.

The change in observation uncertainty also has an impact on the analysis ensemble spread, with primarily decreased ensemble spread in $\mathrm{DA_{var}}$, compared to $\mathrm{DA_{const}}$, especially in lower elevation regions (Supplemental Fig. S5a). Changes in analysis spread are related to changes in the observation uncertainty, with decreases in spread corresponding to decreases in average observation uncertainty (Supplemental Fig. S5b). For example, for all model grid cells where σ_{obs} decreases, on average, from $\mathrm{DA_{const}}$ to $\mathrm{DA_{var}}$, 83% indicate a corresponding decrease in the snow depth analysis ensemble spread. In contrast, for all grid cells where σ_{obs} increases, 65% have a corresponding increase in analysis ensemble spread. The reason for this decrease in ensemble spread is likely two-fold. This overall decrease in ensemble spread is likely driven by two factors: (1) lower observation uncertainty in many regions, and (2) reduced snow depth, which results in smaller multiplicative perturbations to the forecast state.

4 Discussion

This work enhances snow DA by incorporating an ML-based snow depth retrieval product using spatio-temporally dynamic error estimates into the assimilation scheme. The ML snow depth model integrates multiple sources of information, including S1 backscatter observations, fractional snow cover from optical imagery, and land cover information to estimate snow depth. Future work could experiment with integrating additional satellite-based information into the assimilated ML product (e.g., passive microwave, X-band, lidar data). The snow depth estimated from this ML model has been shown to possess superior

accuracy compared to prior S1 snow depth retrieval work by Lievens et al. (2022) (SD_{S1}) (Dunmire et al., 2024), which has previously been assimilated into the Noah-MP land surface model using an Ensemble Kalman Filter (De Lannoy et al., 2024; Brangers et al., 2024). ~~Future work could experiment with integrating additional satellite-based information into the assimilated ML product (e.g., passive microwave, X-band, ICESat-2 data).~~

~~To compare with previous work that assimilates snow depth retrievals from the S1 change detection algorithm (SD_{S1}; Lievens et al. (2022)), we compared output from our two DA experiments with DA output from De Lannoy et al. (2024) (DA_{S1}). The DA_{S1} experiment utilizes the same DA setup as DA_{const}, with a static observation uncertainty ($\sigma obs = 0.3\ m$), but assimilates SD_{S1} retrievals instead of SD_{ML}.~~ In the OL, we see an overestimation of SWE at measurement sites with low recorded SWE, and an underestimation of SWE at measurement sites with high recorded SWE (Fig. 5c). Previous work has demonstrated that forcing bias is the dominant source of uncertainty in snow modeling (Raleigh et al., 2015). Here, we use ERA5 atmospheric forcing, which has a relatively coarse spatial resolution (31 km). While we apply a standard lapse-rate correction to downscale the near-surface air temperature forcing, precipitation is not downscaled, and therefore is unable to resolve orographic precipitation, resulting in relatively low precipitation and SWE spatial variability, and an underestimation of high SWE values. Furthermore, the SD_{ML} product has also been demonstrated to underestimate deep snow, likely due to these measurements being underrepresented in the ML training (Dunmire et al., 2024). As such, the assimilation of this product is unable to fully correct the negative SWE bias for measured SWE $>$ ~800 mm, as can be seen in Figure 5d/e.

Here, we also highlight the implications of accounting for dynamic estimates of the observation uncertainty and demonstrate that this system results in a more realistic modeled snow state. The EnKF depends on accurate uncertainty estimates for both the model and observations, using these to weigh the information and obtain an optimal state. With this in mind, Dee (1995) argues that proper characterization of both model and observation uncertainties is necessary for successful implementation of the EnKF. While the specification of observation uncertainty substantially influences DA performance, in snow DA systems, this uncertainty is often prescribed as a constant value (Helmert et al., 2018). Some previous studies have incorporated dynamic observations errors (e.g., Magnusson et al. (2017); Oberrauch et al. (2024)); however, the utility of dynamic observation errors, relative to an assumed static observation error, in snow DA has not yet been explored prior to this work. Moreover, most operational land data assimilation systems (e.g., NASA Land Data Assimilation Systems, ECMWF Land Data Assimilation System) and recent studies that assimilates SAR-based snow depth retrievals assume a static observation error. For instance, Brangers et al. (2024) assumed $\sigma_{obs} = 0.36\ m$, and Girotto et al. (2024)) and De Lannoy et al. (2024) both assume $\sigma_{obs} = 0.30\ m$ (applied here in DA_{const}).

We find that ~~SWE MAE is significantly lower in DA_{var} compared to DA_{S1} ($p \ll 0.001$), whereas~~ assimilating SD_{ML} with a dynamic observation error (DA_{var}) offers a significant improvement to SWE MAE ($p \ll 0.001$) compared to assimilating SD_{S1} with a static observation error (DA_{S1}, Supplemental Fig. S6). Meanwhile, DA_{const} does not demonstrate ~~a meaningful improvement~~ any significant improvements to SWE MAE (Supplemental Fig. S6). Using 4548 manual SWE measurements collected within the Po River basin~~(the study domain of De Lannoy et al. (2024))~~, we find an MAE of 225 mm from the OL experiment, while the MAE for DA_{S1}, DA_{const}, and DA_{var} is 193, 195, and 177 mm, respectively. ~~The~~ Generally, the SD_{ML} retrievals are more accurate than SD_{S1} for in-situ snow depths below 2.5 m, while for snow depth exceeding 3 m,

SD_{S1} performs better (see Figure 3a from Dunmire et al. (2024)). This suggests that assimilating SD_{ML} should provide improvements particularly for shallower snow. However, in the $\mathrm{DA_{const}}$ experiment, the use of a static observation uncertainty, where relatively large errors are assumed for shallow snow observations, limits these potential improvements (e.g. Fig. 2a) and results in an overall performance of $\mathrm{DA_{const}}$ that is similar to $\mathrm{DA_{S1}}$. This analysis highlights that the treatment of the observation uncertainty is as critical as the observations themselves. A poorly parameterized observation uncertainty can restrict the benefits of DA, underscoring the need for options in DA systems to dynamically vary the observation error.

~~The EnKF depends on accurate uncertainty estimates for both the model and observations, using these to weigh the information and obtain an optimal state. With this in mind, Dee (1995) argues that proper characterization of both model and observation uncertainties is necessary for successful implementation of the EnKF. While the specification of observation uncertainty substantially influences DA performance, in snow DA systems, this uncertainty is often prescribed as a constant value (Helmert et al., 2018) , and the use of dynamic observation error estimation in snow DA had not yet been explored prior to this work. Moreover, most operational land data assimilation systems (e.g., NASA Land Data Assimilation Systems, ECMWF Land Data Assimilation System) and recent studies that have applied the EnKF to assimilate~~ Implementing the dynamic observation error generally improves performance in both places the DA adds and removes snow. In the OL experiment, snow depth has a positive bias at low elevations and a negative bias at high elevations (Fig. 4a). The $\mathrm{DA_{const}}$ experiment applies a static observation error that is relatively too large for shallow assimilated snow depths (e.g. Fig. 2a), limiting snow removal at lower elevations and leading to a still large positive bias at these locations. At higher elevations (above $\sim$1500 m), the assimilated observations exhibit a strong positive bias (Fig. 4d). The relatively small static observation error for deeper assimilated snow depths (e.g. Fig. 2b) leads to too much added snow in some cases, particularly above 2500 m (Fig. 4b). In contrast, in $\mathrm{DA_{var}}$, less snow is added at high elevations (Fig. 9), resulting in improvements where snow needs to be added as well (Fig. 4f). However, we see that the $\mathrm{DA_{var}}$ experiment performs worse than $\mathrm{DA_{const}}$ between February and May within the 2000-2500 m elevation band (Fig. 4f). In this range, the OL experiment has a positive snow depth bias until approximately February, followed by a negative snow depth bias until May (Fig. 4a). $\mathrm{DA_{var}}$ more effectively reduces this early season positive bias, resulting in lower mean snow depths later in the season, and poorer performance during the period when the OL is negatively biased. This suggests that a lack of early-season corrections in $\mathrm{DA_{const}}$ can, in some cases, propagate to more accurate late-season snow depths, although this effect is likely limited to locations where the snow depth is not consistently positively or negatively biased throughout the season.

While $\mathrm{DA_{var}}$ improves performance at most snow depth and SWE measurement sites, some locations see little benefit, or even a deterioration in performance (approximately 12% of snow depth sites and 20% of SWE sites). These degradations are more likely to occur where the SD_{ML} product is less accurate than the OL experiment. To account for known limitations of SAR-based snow depth retrievals~~assume a static observation error~~, we did not assimilate the SD_{ML} product over dense forests or glaciers, and after March 31. Nevertheless, SD_{ML} remains inaccurate in some places, leading to localized deterioration when these observations are assimilated. Locations with minimal differences between $\mathrm{DA_{const}}$ and $\mathrm{DA_{var}}$ typically occur where the observations already agree well with the OL, or where $\sigma_{obs} >> \sigma_f$, thus the DA increments are small, and the model receives limited benefit from the observational information. Despite these spatial inconsistencies, $\mathrm{DA_{var}}$ nearly doubles the

improvement in absolute SWE error compared to DA_{const}. For instance, ~~Brangers et al. (2024) assumed $\sigma obs = 0.36\ m$, and~~ ~~De Lannoy et al. (2024) assume $\sigma obs = 0.30\ m$ (applied here in~~ the SWE MAE decreases from 152 mm in DA_{const} to 132 mm in DA_{var} (-13.2%), while the overall impact of DA_{const} relative to the OL is a 13.6% reduction (176 mm in the OL to 152 mm in DA_{const}). Importantly, previous studies have demonstrated that even modest improvements in snow depth or SWE from DA propagate to further improvements in streamflow (Brangers et al., 2024; De Lannoy et al., 2024).

~~Here, we highlight the implications of accounting for dynamical estimates of observation uncertainty,~~ Snow cover fraction affects the energy balance, and consequently, has implications for numerical weather prediction. While the DA experiments generally reduce the snow-covered area by largely removing snow at lower elevation regions, all three experiments still exhibit a substantial overestimation of total snow-covered area compared with both Copernicus and IMS snow cover products. Several factors may contribute to this discrepancy. First, a positive bias in snowfall forcing data at low elevations will result in unrealistically large snow-covered area. Second, the higher-resolution Copernicus product (20 m) inherently captures finer-scale variation between snow-covered and ~~demonstrate that this system results in a more realistic modeled snow state. In~~ snow-free conditions, often resulting in lower overall snow cover estimates compared to coarser-resolution products. Third, inaccuracies in the parameterization of snow cover fraction within Noah-MP may also play a role. In Noah-MP, the snow cover fraction is parameterized as a function of snow depth, density, and ground roughness length. (Niu et al., 2011; Lee et al., 2024). It should be investigated whether the current parameterizations in Noah-MP remain appropriate for regions with complex terrain, where subgrid variation in topography can significantly influence fractional snow cover. Finally, uncertainty in the Copernicus and IMS snow cover, for example due to cloud and forest cover, contribute to errors in these validation data sets and potentially influence the perceived model biases.

Finally, in DA_{var}, Equation 2 (~~$\sigma^{obs} = m * SD_{ML}$~~ $\sigma_{obs} = m * SD_{ML}$, $m = 0.3$) is used to adapt the standard deviation of the observation error in space and time based on the assimilated snow depth. This relationship is a first-order approximation that assumes that the observation error increases linearly with the observation magnitude; however, σ_{obs} could be defined to vary in more complex ways. Future work could explore applying relationships where σ_{obs} varies non-linearly with the assimilated snow depth observation, or statistical parameterizations of σ_{obs} depending on other conditions such as elevation, or forest cover. Furthermore, σ_{obs} could be directly linked to the SD_{ML} retrieval quality which could be obtained, for example, through error propagation. The effectiveness of a ~~variable~~ dynamic observation error also depends on the magnitude of the forecast error, as the Kalman gain matrix, which determines the strength of the corrections, depends on both forecast and observation error. To maximize benefits, the observation error, whether static or dynamic, should be properly tuned in relation to forecast error. While most operational systems do not currently include options to dynamically vary the observation error, this functionality is not complicated to incorporate, and the snow-specific MuSA (Multiple Snow Data Assimilation System) system does already provide an option for a user-defined observation error that can vary dynamically (Alonso-González et al., 2022).

4.1 Limitations of bias-blind DA systems

The EnKF is widely used in snow DA systems due to its efficiency; however, a key assumption is that both the observations and model are unbiased. We see from Figure 4 that this assumption is not satisfied by neither the observations nor the model. Here,

we implement a bias-blind system by not bias-correcting either the observations or the model, thereby violating this assumption. Bias-aware systems which a priori correct the model bias to align with the observation climatology assume that the assimilated observations are more realistic than the model. While this assumption may be realistic in many situations, satellite-based snow retrievals also exhibit substantial bias. Since snow is a cumulative variable, biases in either the observations or the model typically persist throughout the snow season. While in-situ measurement stations can help quantify these biases, they are often inconsistent spatially and on an interannual basis (i.e. Supplemental Fig. S7), which provides a challenge for correcting them a priori.

Two major issues exist with bias-blind systems: (1) model drift towards its original state, leading to a sawtooth-like pattern that can result in unrealistic fluxes in other variables, and (2) unrealistic model trends in DA output due to changes in assimilated observation frequency (Dee, 2005). For snow, model biases primarily stem from errors in precipitation forcing data. Consequently, we do not expect model drift to occur as observed in De Lannoy et al. (2007); Mocko et al. (2021); Scherrer et al. (2023), unless there is an instantaneous precipitation forcing error. We also assimilate observations weekly throughout the study period, thereby mitigating the potential effects of assimilation frequency in bias-blind DA. Scherrer et al. (2023) further compare bias-blind and bias-aware assimilation of leaf area index - a cumulative variable - using the EnKF. Their results show that the bias-blind DA more effectively updates the model state variable, and leads to larger improvements in water balance components such as evapotranspiration and runoff. In contrast, while the bias-aware approach yields smaller improvements in state variables, it improves temporal anomalies and internal DA diagnostics indicate a more optimal DA system performance. Given our focus on improving the modeled snow state rather than snow anomalies, along with the inherent challenges of a priori bias correcting the observations and model, we opt for a bias-blind approach, recognizing that this may lead to suboptimal DA performance (i.e. temporally correlated residuals).

~~Snow cover fraction affects the energy balance, and consequently, has implications for numerical weather prediction. While the DA experiments generally reduce the snow-covered areaby largely removing snow at lower elevation regions, all three experiments still exhibit a substantial overestimation of total snow-covered area compared with both Copernicus and IMS snow cover products. Several factors may contribute to this discrepancy. First, a high bias in snowfall forcing data at low elevations will result in unrealistically large snow-covered area. Second,~~

4.2 Limitations of site evaluation representativeness

Previous studies have shown that mountain snow is highly variable, and point-scale measurements don't necessarily well-represent the surrounding area, even at spatial scales as fine as 10 m (López-Moreno et al., 2011; Fassnacht et al., 2018). Meromy et al. (2013) found that approximately half of the SNOTEL sites they analyzed where representative of the surrounding 1 km area, defining "representative" as snow station biases within 10% of the surrounding mean observed depth. More recently, Herbert et al. (2024) reported that roughly one-third of 476 paired lidar–station data observations were representative at the ~~higher-resolution Copernicus product (20 m) inherently captures finer-scale variation between snow-covered and snow-free conditions, often resulting in lower overall snow cover estimates compared to coarser-resolution products. Third, inaccuracies in the parameterization of snow cover fraction within Noah-MP may also play a role. In Noah-MP, the snow cover fraction is parameterized as a~~

~~function of snow depth, density, and ground roughness length. (Niu et al., 2011; Lee et al., 2024). It should be investigated whether the current parameterizations in Noah-MP remain appropriate for regions with complex terrain, where subgrid variation in topography can significantly influence fractional snow cover. Finally, uncertainty in the Copernicus and IMS snow cover, for example due to cloud and forest cover, contribute to errors in these validation data sets and potentially influence the perceived model biases~~1 km scale, with representativeness defined as in-situ measured snow within $\pm$10 cm of the lidar-mean snow depth at that scale. However, they also showed little change between the 500 m and 1 km scales, with 35% of stations considered as representative at 500 m. Generally, in-situ snow stations exhibit a positive bias as these sites are often located in flat terrain that preferentially accumulates snow (Grünewald and Lehning, 2011).

In this study, we use in-situ snow depth and manual SWE measurements as the best-available reference in the European Alps that cover a range of terrain conditions and spans many years. Unlike in the western United States, where high-resolution spatial snow depth products from the Airborne Snow Observatory and NASA SnowEx missions are available, such publicly available products are extremely limited in the European Alps. As such, it is not feasible to assess the representativeness of all 588 snow depth measurement sites and 8211 manual SWE measurements at the 1 km scale, and these point-scale measurements provide the best available Alps-wide, multi-year data available for evaluation. Nevertheless, by leveraging a large network of sites that span a range of elevations and terrain types, we can reduce sampling-related limitations by increased coverage of terrain diversity, although this does not mitigate the general positive bias noted above.

5 Conclusions

In this manuscript, we demonstrate the utility of incorporating a dynamic observation uncertainty into a snow depth data assimilation scheme. For the first time, we assimilate satellite-based snow depth estimates from a novel machine learning model into the Noah-MP land surface model using the EnKF to update snow depth and SWE. We compare two data assimilation experiments: one with a static observation error ($\mathrm{DA_{const}}$), and one with an observation error that is dynamic in space and time ($\mathrm{DA_{var}}$).The performance of these DA experiments is evaluated against the open-loop experiment (OL, model-only) using in-situ snow depth observations, manual SWE measurements, and two different snow cover products. We show that the dynamic observation error makes better use of the assimilated observations, thereby leading to stronger model corrections, particularly at times when the assimilated snow depth observation is much shallower than the model forecast (e.g., early in the accumulation period or at lower elevations). The DA experiment that incorporates this dynamic observation error more effectively corrects biases introduced by errors in the forcing data, and improves SWE estimates by 25% and 13% compared to the OL and $\mathrm{DA_{const}}$ experiments, respectively. While snow cover is overestimated in all three model experiments, $\mathrm{DA_{var}}$ also leads to stronger reductions in snow cover than $\mathrm{DA_{const}}$, better aligning with existing snow cover products. As most snow DA work and operational snow DA systems assume that the observational uncertainty is constant in space and time, this work highlights the impact of a better constrained observational error and the importance of these considerations when designing a DA system.

Code and data availability. The ML-based snow depth retrieval product is publicly available at (https://doi.org/10.5281/zenodo.13342108). The NASA LIS software is available at https://github.com/NASA-LIS/LISF. Publicly available in-situ snow depth and SWE data used for evaluation can be accessed at:

- https://www.doi.org/10.16904/15 (Switzerland)
- https://www.doi.org/10.16904/envidat.380 (Switzerland)
- https://www.doi.org/10.16904/envidat.590 (Switzerland)
- https://www.doi.org/10.16904/envidat.406 (Switzerland)
- https://www.arpa.piemonte.it/rischi_naturali/snippets_arpa_graphs/map_meteoweb/?rete=stazione_meteorologica (Italy)
- https://www.meteotrentino.it/index.html#!/home (Italy)
- https://data.civis.bz.it/de/dataset/p-bz-southtyrolean-weatherservice-weatherstations/resource/ef2f6f24-cffd-4993-8699-5023696a49b5 (Italy)
- https://dataset.api.hub.geosphere.at/app/frontend/station/historical/klima-v2-1d (Austria)
- https://donneespubliques.meteofrance.fr/?fond=recherche (France)
- https://cdc.dwd.de/portal/ (Germany)

Additional snow depth and SWE data were obtained from the Italian Department of Civil Protection and processed by the Centro Internazionale in Monitoraggio Ambientale (CIMA).

The configuration files used for the modeling experiments and code used for the analysis and creation of figures will be made freely available after review.

Author contributions. DD and GDL conceived and designed the study. DD conducted the model experiments, with help from MB. Material preparation and data collection were performed by DD and LB. Formal analysis, visualization, and original draft preparation was done by DD and all authors contributed to the manuscript reviewing and editing.

Competing interests. The authors declare that there are no competing interests.

Acknowledgements. This work was funded by the project C14/21/057 of KU Leuven and SNOWTRANE (SR/00/407) of the Belgian Science Policy (Belspo). The computer resources and services were provided by the High Performance Computing system of the Vlaams Supercomputer Center, funded by FWO and the Flemish Government (incl. Storage4Climate collaborative grant). The authors acknowledge CIMA for providing reference data. During the preparation of this work the authors used ChatGPT in order to improve readability, flow, and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Alonso-González, E., Aalstad, K., Baba, M. W., Revuelto, J., López-Moreno, J. I., Fiddes, J., Essery, R., and Gascoin, S.: The Multiple Snow Data Assimilation System (MuSA v1.0), Geoscientific Model Development, 15, 9127–9155, https://doi.org/10.5194/gmd-15-9127-2022, 2022.

Baldo, E. and Margulis, S. A.: Assessment of a multiresolution snow reanalysis framework: a multidecadal reanalysis case over the upper Yampa River basin, Colorado, Hydrology and Earth System Sciences, 22, 3575–3587, https://doi.org/10.5194/hess-22-3575-2018, 2018.

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, 303–309, https://doi.org/10.1038/nature04141, 2005.

Bechtold, M., Modanesi, S., Lievens, H., Baguis, P., Brangers, I., Carrassi, A., Getirana, A., Gruber, A., Heyvaert, Z., Massari, C., Scherrer, S., Vannitsem, S., and De Lannoy, G.: Assimilation of Sentinel-1 Backscatter into a Land Surface Model with River Routing and Its Impact on Streamflow Simulations in Two Belgian Catchments, Journal of Hydrometeorology, 24, 2389–2408, https://doi.org/10.1175/JHM-D-22-0198.1, 2023.

Besso, H., Shean, D., and Lundquist, J. D.: Mountain snow depth retrievals from customized processing of ICESat-2 satellite laser altimetry, Remote Sensing of Environment, 300, 113 843, https://doi.org/10.1016/j.rse.2023.113843, 2024.

Bormann, K. J., Brown, R. D., Derksen, C., and Painter, T. H.: Estimating snow-cover trends from space, Nature Climate Change, 8, 924–928, https://doi.org/10.1038/s41558-018-0318-3, 2018.

Brangers, I., Lievens, H., Getirana, A., and De Lannoy, G. J. M.: Sentinel-1 Snow Depth Assimilation to Improve River Discharge Estimates in the Western European Alps, Water Resources Research, 60, https://doi.org/10.1029/2023WR035019, 2024.

Broxton, P., Ehsani, M. R., and Behrangi, A.: Improving Mountain Snowpack Estimation Using Machine Learning With Sentinel-1, the Airborne Snow Observatory, and University of Arizona Snowpack Data, Earth and Space Science, 11, https://doi.org/10.1029/2023EA002964, 2024.

Daudt, R. C., Wulf, H., Hafner, E. D., Bühler, Y., Schindler, K., and Wegner, J. D.: Snow depth estimation at country-scale with high spatial and temporal resolution, ISPRS Journal of Photogrammetry and Remote Sensing, 197, 105–121, https://doi.org/10.1016/j.isprsjprs.2023.01.017, 2023.

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Pauwels, V. R. N., and Verhoest, N. E. C.: Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter, Water Resources Research, 43, https://doi.org/10.1029/2006WR005449, 2007.

De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C., and Pauwels, V. R. N.: Multiscale assimilation of Advanced Microwave Scanning Radiometer–EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, Water Resources Research, 48, https://doi.org/10.1029/2011WR010588, 2012.

De Lannoy, G. J. M., Bechtold, M., Busschaert, L., Heyvaert, Z., Modanesi, S., Dunmire, D., Lievens, H., Getirana, A., and Massari, C.: Contributions of Irrigation Modeling, Soil Moisture and Snow Data Assimilation to High-Resolution Water Budget Estimates Over the Po Basin: Progress Towards Digital Replicas, Journal of Advances in Modeling Earth Systems, 16, https://doi.org/10.1029/2024MS004433, 2024.

de Rosnay, P., Isaksen, L., and Dahoui, M.: Snow data assimilation at ECMWF, Tech. rep., ECMWF Newsletter, 2015.

Dee, D. P.: On-line Estimation of Error Covariance Parameters for Atmospheric Data Assimilation, Monthly Weather Review, 123, 1128–1145, https://doi.org/10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2, 1995.

Dee, D. P.: Bias and data assimilation, Quarterly Journal of the Royal Meteorological Society, 131, 3323–3343, https://doi.org/10.1256/qj.05.137, 2005.

Deems, J. S., Painter, T. H., and Finnegan, D. C.: Lidar measurement of snow depth: a review, Journal of Glaciology, 59, 467–479, https://doi.org/10.3189/2013JoG12J154, 2013.

Deschamps-Berger, C., Gascoin, S., Berthier, E., Deems, J., Gutmann, E., Dehecq, A., Shean, D., and Dumont, M.: Snow depth mapping from stereo satellite imagery in mountainous terrain: evaluation using airborne laser-scanning data, The Cryosphere, 14, 2925–2940, https://doi.org/10.5194/tc-14-2925-2020, 2020.

Deschamps-Berger, C., Gascoin, S., Shean, D., Besso, H., Guiot, A., and López-Moreno, J. I.: Evaluation of snow depth retrievals from ICESat-2 using airborne laser-scanning data, The Cryosphere, 17, 2779–2792, https://doi.org/10.5194/tc-17-2779-2023, 2023.

Dozier, J., Bair, E. H., and Davis, R. E.: Estimating the spatial distribution of snow water equivalent in the world's mountains, WIREs Water, 3, 461–474, https://doi.org/10.1002/wat2.1140, 2016.

Dunmire, D., Lievens, H., Boeykens, L., and De Lannoy, G. J.: A machine learning approach for estimating snow depth across the European Alps from Sentinel-1 imagery, Remote Sensing of Environment, 314, 114 369, https://doi.org/10.1016/j.rse.2024.114369, 2024.

Durand, M. and Margulis, S. A.: Feasibility Test of Multifrequency Radiometric Data Assimilation to Estimate Snow Water Equivalent, Journal of Hydrometeorology, 7, 443–457, https://doi.org/10.1175/JHM502.1, 2006.

Enderlin, E. M., Elkin, C. M., Gendreau, M., Marshall, H., O'Neel, S., McNeil, C., Florentine, C., and Sass, L.: Uncertainty of ICESat-2 ATL06- and ATL08-derived snow depths for glacierized and vegetated mountain regions, Remote Sensing of Environment, 283, 113 307, https://doi.org/10.1016/j.rse.2022.113307, 2022.

Estilow, T. W., Young, A. H., and Robinson, D. A.: A long-term Northern Hemisphere snow cover extent data record for climate studies and monitoring, Earth System Science Data, 7, 137–142, https://doi.org/10.5194/essd-7-137-2015, 2015.

Fassnacht, S. R., Brown, K. S. J., Blumberg, E. J., López Moreno, J. I., Covino, T. P., Kappas, M., Huang, Y., Leone, V., and Kashipazha, A. H.: Distribution of snow depth variability, Frontiers of Earth Science, 12, 683–692, https://doi.org/10.1007/s11707-018-0714-z, 2018.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Girotto, M., Musselman, K. N., and Essery, R. L. H.: Data Assimilation Improves Estimates of Climate-Sensitive Seasonal Snow, Current Climate Change Reports, 6, 81–94, https://doi.org/10.1007/s40641-020-00159-7, 2020.

Girotto, M., Formetta, G., Azimi, S., Bachand, C., Cowherd, M., De Lannoy, G., Lievens, H., Modanesi, S., Raleigh, M. S., Rigon, R., and Massari, C.: Identifying snowfall elevation patterns by assimilating satellite-based snow depth retrievals, Science of The Total Environment, 906, 167 312, https://doi.org/10.1016/j.scitotenv.2023.167312, 2024.

Grünewald, T. and Lehning, M.: Altitudinal dependency of snow amounts in two small alpine catchments: can catchment-wide snow amounts be estimated via single snow or precipitation stations?, Annals of Glaciology, 52, 153–158, https://doi.org/10.3189/172756411797252248, 2011.

Günther, D., Marke, T., Essery, R., and Strasser, U.: 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, 2779–2800, https://doi.org/10.1029/2018WR023403, 2019.

Helmert, J., Şensoy Şorman, A., Alvarado Montero, R., De Michele, C., de Rosnay, P., Dumont, M., Finger, D., Lange, M., Picard, G., Potopová, V., Pullen, S., Vikhamar-Schuler, D., and Arslan, A.: Review of Snow Data Assimilation Methods for Hydrological, Land Surface, Meteorological and Climate Models: Results from a COST HarmoSnow Survey, Geosciences, 8, 489, https://doi.org/10.3390/geosciences8120489, 2018.

Henderson, G. R., Peings, Y., Furtado, J. C., and Kushner, P. J.: Snow–atmosphere coupling in the Northern Hemisphere, Nature Climate Change, 8, 954–963, https://doi.org/10.1038/s41558-018-0295-6, 2018.

Herbert, J. N., Raleigh, M. S., and Small, E. E.: Reanalyzing the spatial representativeness of snow depth at automated monitoring stations using airborne lidar data, The Cryosphere, 18, 3495–3512, https://doi.org/10.5194/tc-18-3495-2024, 2024.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hoppinen, Z., Palomaki, R. T., Brencher, G., Dunmire, D., Gagliano, E., Marziliano, A., Tarricone, J., and Marshall, H.-P.: Evaluating snow depth retrievals from Sentinel-1 volume scattering over NASA SnowEx sites, The Cryosphere, 18, 5407–5430, https://doi.org/10.5194/tc-18-5407-2024, 2024.

Huang, C., Newman, A. J., Clark, M. P., Wood, A. W., and Zheng, X.: Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States, Hydrology and Earth System Sciences, 21, 635–650, https://doi.org/10.5194/hess-21-635-2017, 2017.

IPCC: IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Tech. rep., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021.

Kelly, R., Li, Q., and Saberi, N.: 'The AMSR2 Satellite-Based Microwave Snow Algorithm (SMSA): A New Algorithm for Estimating Global Snow Accumulation, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 5606–5609, IEEE, ISBN 978-1-5386-9154-0, https://doi.org/10.1109/IGARSS.2019.8898525, 2019.

Kumar, S., Peterslidard, C., Tian, Y., Houser, P., Geiger, J., Olden, S., Lighty, L., Eastman, J., Doty, B., and Dirmeyer, P.: Land information system: An interoperable framework for high resolution land surface modeling, Environmental Modelling & Software, 21, 1402–1415, https://doi.org/10.1016/j.envsoft.2005.07.004, 2006.

Lee, W. Y., Gim, H.-J., and Park, S. K.: Parameterizations of Snow Cover, Snow Albedo and Snow Density in Land Surface Models: A Comparative Review, Asia-Pacific Journal of Atmospheric Sciences, 60, 185–210, https://doi.org/10.1007/s13143-023-00344-2, 2024.

Lievens, H., Demuzere, M., Marshall, H.-P., Reichle, R. H., Brucker, L., Brangers, I., de Rosnay, P., Dumont, M., Girotto, M., Immerzeel, W. W., Jonas, T., Kim, E. J., Koch, I., Marty, C., Saloranta, T., Schöber, J., and De Lannoy, G. J. M.: Snow depth variability in the Northern Hemisphere mountains observed from space, Nature Communications, 10, 4629, https://doi.org/10.1038/s41467-019-12566-y, 2019.

Lievens, H., Brangers, I., Marshall, H.-P., Jonas, T., Olefs, M., and De Lannoy, G.: Sentinel-1 snow depth retrieval at sub-kilometer resolution over the European Alps, The Cryosphere, 16, 159–177, https://doi.org/10.5194/tc-16-159-2022, 2022.

Liston, G. E. and Hiemstra, C. A.: A Simple Data Assimilation System for Complex Snow Distributions (SnowAssim), Journal of Hydrometeorology, 9, 989–1004, https://doi.org/10.1175/2008JHM871.1, 2008.

López-Moreno, J. I., Fassnacht, S. R., Beguería, S., and Latron, J. B. P.: Variability of snow depth at the plot scale: implications for mean depth estimation and sampling strategies, The Cryosphere, 5, 617–629, https://doi.org/10.5194/tc-5-617-2011, 2011.

López-Moreno, J. I., Revuelto, J., Fassnacht, S. R., Azorín-Molina, C., Vicente-Serrano, S. M., Morán-Tejeda, E., and Sexstone, G. A.: Snowpack variability across various spatio-temporal resolutions, Hydrological Processes, 29, 1213–1224, https://doi.org/10.1002/hyp.10245, 2015.

Luojus, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Mortimer, C., Derksen, C., Mudryk, L., Moisander, M., Hiltunen, M., Smolander, T., Ikonen, J., Cohen, J., Salminen, M., Norberg, J., Veijola, K., and Venäläinen, P.: GlobSnow v3.0 Northern Hemisphere snow water equivalent dataset, Scientific Data, 8, 163, https://doi.org/10.1038/s41597-021-00939-2, 2021.

Magnusson, J., Winstral, A., Stordal, A. S., Essery, R., and Jonas, T.: Improving physically based snow simulations by assimilating snow depths using the particle filter, Water Resources Research, 53, 1125–1143, https://doi.org/10.1002/2016WR019092, 2017.

Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., and Diffenbaugh, N. S.: The potential for snow to supply human water demand in the present and future, Environmental Research Letters, 10, 114 016, https://doi.org/10.1088/1748-9326/10/11/114016, 2015.

Margulis, S. A., Girotto, M., Cortés, G., and Durand, M.: A Particle Batch Smoother Approach to Snow Water Equivalent Estimation, Journal of Hydrometeorology, 16, 1752–1772, https://doi.org/10.1175/JHM-D-14-0177.1, 2015.

Marti, R., Gascoin, S., Berthier, E., de Pinel, M., Houet, T., and Laffly, D.: Mapping snow depth in open alpine terrain from stereo satellite imagery, The Cryosphere, 10, 1361–1380, https://doi.org/10.5194/tc-10-1361-2016, 2016.

Meromy, L., Molotch, N. P., Link, T. E., Fassnacht, S. R., and Rice, R.: Subgrid variability of snow water equivalent at operational snow stations in the western USA, Hydrological Processes, 27, 2383–2400, https://doi.org/10.1002/hyp.9355, 2013.

Miller, Z. S., Peitzsch, E. H., Sproles, E. A., Birkeland, K. W., and Palomaki, R. T.: Assessing the seasonal evolution of snow depth spatial variability and scaling in complex mountain terrain, The Cryosphere, 16, 4907–4930, https://doi.org/10.5194/tc-16-4907-2022, 2022.

Mocko, D. M., Kumar, S. V., Peters-Lidard, C. D., and Wang, S.: Assimilation of Vegetation Conditions Improves the Representation of Drought over Agricultural Areas, Journal of Hydrometeorology, 22, 1085–1098, https://doi.org/10.1175/JHM-D-20-0065.1, 2021.

Modanesi, S., Massari, C., Bechtold, M., Lievens, H., Tarpanelli, A., Brocca, L., Zappa, L., and De Lannoy, G. J. M.: Challenges and benefits of quantifying irrigation through the assimilation of Sentinel-1 backscatter observations into Noah-MP, Hydrology and Earth System Sciences, 26, 4685–4706, https://doi.org/10.5194/hess-26-4685-2022, 2022.

Musselman, K. N., Addor, N., Vano, J. A., and Molotch, N. P.: Winter melt trends portend widespread declines in snow water resources, Nature Climate Change, 11, 418–424, https://doi.org/10.1038/s41558-021-01014-9, 2021.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research, 116, D12 109, https://doi.org/10.1029/2010JD015139, 2011.

Oberrauch, M., Cluzet, B., Magnusson, J., and Jonas, T.: Improving Fully Distributed Snowpack Simulations by Mapping Perturbations of Meteorological Forcings Inferred From Particle Filter Assimilation of Snow Monitoring Data, Water Resources Research, 60, https://doi.org/10.1029/2023WR036994, 2024.

Outdoor Industry Association: The outdoor recreation economy, Tech. rep., Boulder, CO, USA, 2017.

Parthum, B. and Christensen, P.: A market for snow: Modeling winter recreation patterns under current and future climate, Journal of Environmental Economics and Management, 113, 102 637, https://doi.org/10.1016/j.jeem.2022.102637, 2022.

Peters-Lidard, C. D., Houser, P. R., Tian, Y., Kumar, S. V., Geiger, J., Olden, S., Lighty, L., Doty, B., Dirmeyer, P., Adams, J., Mitchell, K., Wood, E. F., and Sheffield, J.: High-performance Earth system modeling with NASA/GSFC's Land Information System, Innovations in Systems and Software Engineering, 3, 157–165, https://doi.org/10.1007/s11334-007-0028-x, 2007.

Pfeffer, W. T., Arendt, A. A., Bliss, A., Bolch, T., Cogley, J. G., Gardner, A. S., Hagen, J.-O., Hock, R., Kaser, G., Kienholz, C., Miles, E. S., Moholdt, G., Mölg, N., Paul, F., Radić, V., Rastner, P., Raup, B. H., Rich, J., and Sharp, M. J.: The Randolph Glacier Inventory: a globally complete inventory of glaciers, Journal of Glaciology, 60, 537–552, https://doi.org/10.3189/2014JoG13J176, 2014.

Pflug, J. M., Wrzesien, M. L., Kumar, S. V., Cho, E., Arsenault, K. R., Houser, P. R., and Vuyovich, C. M.: Extending the utility of space-borne snow water equivalent observations over vegetated areas with data assimilation, Hydrology and Earth System Sciences, 28, 631–648, https://doi.org/10.5194/hess-28-631-2024, 2024.

Qin, Y., Abatzoglou, J. T., Siebert, S., Huning, L. S., AghaKouchak, A., Mankin, J. S., Hong, C., Tong, D., Davis, S. J., and Mueller, N. D.: Agricultural risks from changing snowmelt, Nature Climate Change, 10, 459–465, https://doi.org/10.1038/s41558-020-0746-8, 2020.

Raleigh, M. S., Lundquist, J. D., and Clark, M. P.: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, Hydrology and Earth System Sciences, 19, 3153–3179, https://doi.org/10.5194/hess-19-3153-2015, 2015.

Raleigh, M. S., Livneh, B., Lapo, K., and Lundquist, J. D.: How Does Availability of Meteorological Forcing Data Impact Physically Based Snowpack Simulations?*, Journal of Hydrometeorology, 17, 99–120, https://doi.org/10.1175/JHM-D-14-0235.1, 2016.

Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic Data Assimilation with the Ensemble Kalman Filter, Monthly Weather Review, 130, 103–114, https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2, 2002.

Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P. P., and Partyka, G. S.: Land Surface Precipitation in MERRA-2, Journal of Climate, 30, 1643–1664, https://doi.org/10.1175/JCLI-D-16-0570.1, 2017.

RGI 7.0 Consortium: Randolph Glacier Inventory - A Dataset of Global Glacier Outlines, Version 7.0, 2023.

Rodell, M. and Houser, P. R.: Updating a Land Surface Model with MODIS-Derived Snow Cover, Journal of Hydrometeorology, 5, 1064–1075, https://doi.org/10.1175/JHM-395.1, 2004.

Safeeq, M., Shukla, S., Arismendi, I., Grant, G. E., Lewis, S. L., and Nolin, A.: Influence of winter season climate variability on snow–precipitation ratio in the western United States, International Journal of Climatology, 36, 3175–3190, https://doi.org/10.1002/joc.4545, 2016.

Scherrer, S., De Lannoy, G., Heyvaert, Z., Bechtold, M., Albergel, C., El-Madany, T. S., and Dorigo, W.: Bias-blind and bias-aware assimilation of leaf area index into the Noah-MP land surface model over Europe, Hydrology and Earth System Sciences, 27, 4087–4114, https://doi.org/10.5194/hess-27-4087-2023, 2023.

Shaw, T. E., Gascoin, S., Mendoza, P. A., Pellicciotti, F., and McPhee, J.: Snow Depth Patterns in a High Mountain Andean Catchment from Satellite Optical Tristereoscopic Remote Sensing, Water Resources Research, 56, https://doi.org/10.1029/2019WR024880, 2020.

Slater, A. G. and Clark, M. P.: Snow Data Assimilation via an Ensemble Kalman Filter, Journal of Hydrometeorology, 7, 478–493, https://doi.org/10.1175/JHM505.1, 2006.

Smyth, E. J., Raleigh, M. S., and Small, E. E.: Improving SWE Estimation With Data Assimilation: The Influence of Snow Depth Observation Timing and Uncertainty, Water Resources Research, 56, https://doi.org/10.1029/2019WR026853, 2020.

Smyth, E. J., Raleigh, M. S., and Small, E. E.: The Challenges of Simulating SWE Beneath Forest Canopies are Reduced by Data Assimilation of Snow Depth, Water Resources Research, 58, https://doi.org/10.1029/2021WR030563, 2022.

Steiger, R., Scott, D., Abegg, B., Pons, M., and Aall, C.: A critical review of climate change risk for ski tourism, Current Issues in Tourism, 22, 1343–1379, https://doi.org/10.1080/13683500.2017.1410110, 2019.

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

Tedesco, M. and Narvekar, P. S.: Assessment of the NASA AMSR-E SWE Product, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3, 141–159, https://doi.org/10.1109/JSTARS.2010.2040462, 2010.

Terzago, S., Andreoli, V., Arduini, G., Balsamo, G., Campo, L., Cassardo, C., Cremonese, E., Dolia, D., Gabellani, S., von Hardenberg, J., Morra 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, Hydrology and Earth System Sciences, 24, 4061–4090, https://doi.org/10.5194/hess-24-4061-2020, 2020.

Toure, A. M., Reichle, R. H., Forman, B. A., Getirana, A., and De Lannoy, G. J. M.: Assimilation of MODIS Snow Cover Fraction Observations into the NASA Catchment Land Surface Model, Remote Sensing, 10, 316, https://doi.org/10.3390/rs10020316, 2018.

Vander Jagt, B. J., Durand, M. T., Margulis, S. A., Kim, E. J., and Molotch, N. P.: The effect of spatial variability on the sensitivity of passive microwave measurements to snow water equivalent, Remote Sensing of Environment, 136, 163–179, https://doi.org/10.1016/j.rse.2013.05.002, 2013.

Verseghy, D. L.: Class—A Canadian land surface scheme for GCMS. I. Soil model, International Journal of Climatology, 11, 111–133, https://doi.org/10.1002/joc.3370110202, 1991.

Vorkauf, M., Marty, C., Kahmen, A., and Hiltbrunner, E.: Past and future snowmelt trends in the Swiss Alps: the role of temperature and snowpack, Climatic Change, 165, 44, https://doi.org/10.1007/s10584-021-03027-x, 2021.

Wasti, A., Ray, P., Wi, S., Folch, C., Ubierna, M., and Karki, P.: Climate change and the hydropower sector: A global review, WIREs Climate Change, 13, https://doi.org/10.1002/wcc.757, 2022.

Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins, Journal of Geophysical Research, 116, D12 110, https://doi.org/10.1029/2010JD015140, 2011.