Evaluation and blending of ATMS and AMSR2 snow water equivalent retrievals over the conterminous United States

Yanjun Gan[1,*], Yu Zhang[1,*], Cezar Kongoli[2], Christopher Grassotti[2], Yuqiong Liu[3], Yong-Keun Lee[2], and Dong-Jun Seo[1]

[1]Department of Civil Engineering, University of Texas at Arlington, Arlington, TX, USA.

[2]Earth System Science Interdisciplinary Center/Cooperative Institute for Satellite and Earth System Studies, University of Maryland at College Park, College Park, MD, USA.

[3]Office of Water Prediction, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, MD, USA.

Submitted to *Remote Sensing of Environment*

* Corresponding authors. E-mail addresses: yanjun.gan@uta.edu (Y. Gan), yu.zhang@uta.edu (Y. Zhang).

Abstract

This study first compares two different passive microwave snow water equivalent (SWE) retrievals, namely the retrieval from the Suomi National Polar-orbiting Partnership (S-NPP) Advanced Technology Microwave Sounder (ATMS) and that from the Global Change Observation Mission – Water (GCOM-W1) Advanced Microwave Scanning Radiometer 2 (AMSR2); it further creates an optimal blending mechanism that merges the two retrievals with in situ observations from the Snow Telemetry (SNOTEL) and Cooperative Observer Program (COOP) networks. The assessments of the two products are done over conterminous United States (CONUS) for the snow seasons (November–June) of the water years 2017–2019 using in situ data and the SNOw Data Assimilation System (SNODAS) SWE analysis. Both satellite products tend to underestimate SWE. Between the two, AMSR2 retrieval outperforms in terms of correlation with observations and depth of saturation, but it exhibits a distinctive, seasonally varying bias that is not seen in ATMS retrieval. The negative bias over the early snow season, as further analysis indicates, most likely stems from AMSR2 retrieval's use of a high frequency channel (i.e., 89 GHz) for shallow snow detection, while the impact of differing assumptions of snow density is marginal. The blending scheme, developed on the basis of the validation experiment, features a histogram-based bias correction as a supplement to optimal interpolation. Cross-validation suggests that interpolated station product without the satellite background broadly underperforms the blended in situ-satellite product, confirming the utility of the satellite retrievals. Furthermore, the a priori bias correction mechanism is shown to be effective in mitigating large fluctuations in bias. Finally,

the bias-corrected, blended in situ-satellite product performs comparably or even favorably against SNODAS over many parts of the CONUS, with important implications for joint use of satellite and in situ observations for hydrological monitoring and forecasting.

Keywords: ATMS, AMSR2, passive microwave, snow water equivalent, bias correction, optimal interpolation, weighted averaging.

1. Introduction

Snowpack plays an important role in modulating global climate and hydrologic cycle (Dong, 2018; Lettenmaier et al., 2015; Sturm, 2015). Accurate estimates of snowpack properties are of critical importance to a variety of hydrologic and climate-related applications (Chang et al., 2005; Dozier et al., 2016). Many gridded products have been created to provide long-term snow depth (SD) or snow water equivalent (SWE) estimates. Such products include land surface reanalysis (Dee et al., 2011; Gelaro et al., 2017; Rodell et al., 2004; Xia et al., 2012), snow model simulations (Brun et al., 2013), regional climate model simulations (Wrzesien et al., 2018), and ground-based interpolation data (Brown and Brasnett, 2010; Broxton et al., 2016a). Among these, the model simulations and reanalysis are subject to large uncertainties stemming from those in model structures, parameters, as well as forcing data (Mortimer et al., 2020; Mudryk et al., 2015). Meanwhile, the interpolation data are constrained by the density and locations of stations.

In recent decades, satellite retrievals are seeing increasing applications in snowpack monitoring and prediction, especially in regions with poor ground measurements (Frei et al., 2012; Nolin, 2010). In particular, passive microwave (PMW) SD/SWE retrievals have the advantage of being directly relevant to water balance calculation, available for both day and night-time conditions, and not subject to interference by clouds or atmospheric gases as are snow cover retrieved by optical sensors (Clifford, 2010; Lee et al., 2015). Currently, operational spaceborne PMW sensors that can retrieve SD/SWE include the Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave Imager/Sounder (SSMIS; Bommarito, 1993; Kunkee et al., 2008), the Global Change Observation Mission (GCOM) Advanced Microwave Scanning Radiometer (AMSR) series (Imaoka et al., 2002, 2010), and the Joint Polar Satellite System (JPSS) Advanced Technology Microwave Sounder (ATMS; Weng et al., 2012). Typically, the radiance observations from the PMW sensors are used to retrieve SD by exploiting empirical brightness temperature (Tb) – SD relationships (Kelly, 2009), and SD is then converted to SWE through empirical estimates of snow density (Sturm et al., 2010). The ATMS SWE retrieval algorithm is somewhat unique that it assimilates radiance observations from sounding channels into the Community Radiative Transfer Model (CRTM; Han et al., 2006) using the Microwave Integrated Retrieval System (MiRS; Boukabara et al., 2011). SWE is then retrieved by comparing the MiRS retrieved emissivity spectra with those from a precomputed catalog that relates surface emissivity to SWE to find the closest match. The catalog is generated from a dense medium radiative transfer snow emissivity model (Weng et al., 2001).

In spite of the many promising aspects of PWM SD/SWE retrievals, these products are limited in spatial resolution and are known to suffer from large errors (Dawson et al., 2018; De Lannoy et al., 2010; Frei et al., 2012). The errors may stem from sensor signal saturation, vegetation and terrain interference, snow wetness, and simplifying assumptions underpinning the retrieval algorithms (Dong et al., 2005; Liu et al., 2015). For example, Vuyovich et al. (2014) suggested that forest cover and deep snow have significant impact on AMSR – Earth Observing System (AMSR-E; Imaoka et al., 2002) and Special Sensor Microwave Imager (SSM/I; Hollinger, 1989) SWE estimates. Dai et al. (2017) showed that the mountainous topography and the coarse resolution of PMW sensor underlie the large disagreement between AMSR-E SD and in situ observations. Cho et al. (2020) illustrated that slope and surface heterogeneity impact the SWE difference between the SSMI/S (i.e., SSM/I and SSMIS) and gamma SWE. Tuttle et al. (2018) found that up to half of the error in AMSR-E SWE is potentially due to subpixel scale variability. While these studies advance our understanding of the error sources of PMW data, most of them fall short in proposing or establishing effective mechanisms for mitigating the errors. Furthermore, considering the number of sensors and retrieval products that are currently available, there is a clear, and heretofore unfulfilled demand for identifying and leveraging the complementary strengths of different PMW retrievals and in situ products, and thereby facilitating the application adoption of the retrievals.

One way to address the shortcomings of stand-alone PMW SD/SWE data is through the joint use of PMW radiometry and ground observations as information sources in the retrieval algorithms (Pulliainen, 2006). For example, the Global Snow Monitoring for Climate Research (GlobSnow) SWE product (Pulliainen et al., 2020; Takala et al., 2011) assimilates different sources of PMW Tb (from 18.7 and 36.5 GHz channels) and in situ SD into a semi-empirical snow emission model and provides 25-km daily SWE estimates from 1979 to 2018 over the Northern Hemisphere excluding alpine areas. This product, however, still exhibits large errors inherited from structural limitations of the Tb–SD relation and the snow emission models (Hancock et al., 2013; Larue et al., 2017; Mudryk et al., 2015). An alternative way is through the assimilation of the PMW observations into snow models (Dong et al., 2007; Dziubanski and Franz, 2016) or land surface models (Che et al., 2014; De Lannoy et al., 2012; Kwon et al., 2017; Liu et al., 2013). For example, Kumar et al. (2019) assimilated different sources of PMW SD into the Noah model (Ek et al., 2003) to improve SD estimates from 1979 to 2015 for the conterminous United States (CONUS). While these model-based products were demonstrated to be generally superior to stand-alone PMW SWE (Cho et al., 2020; Dawson et al., 2018; Mortimer et al., 2020), their creation entails high computational costs and their accuracy remains subject to questions in data scarce areas (Broxton et al., 2016b; Clark et al., 2011; Rutter et al., 2009).

The objectives of this study are twofold. The first is to assess the complementary skills of the two different PMW SWE retrievals, namely that from GCOM AMSR2 and the one based on JPSS ATMS, over the CONUS. The second is to develop a lightweight, computationally efficient blending algorithm that optimally combining the two satellite products and in situ observations. The blending scheme has the advantage of being simple and independent of any snow model, and its product can be assimilated to the latter to further improve the prediction of snow and other hydrologic variables (Kumar et al., 2015; Liu et al., 2015). We choose to focus on SWE rather than SD, as the former can be directly used in hydrologic analysis and predictions. The AMSR2 and ATMS SWE retrievals are selected for the following reasons. First, these products are based on two relatively new instruments that contrast sharply in scanning and channel configurations. AMSR2, like its predecessor AMSR-E, is a conical scanner measuring orthogonally polarized radiation (vertical and horizontal) at specific window frequencies (Imaoka et al., 2010), whereas ATMS is a cross-track scanner measuring radiation at all its channels at varying scan angles (Weng et al., 2012). The second reason is that the retrieval algorithms for the two products are quite different. While the AMSR2 retrieval algorithm has undergone extensive assessments (Lee et al., 2015; Wang et al., 2019; Zhang et al., 2017), neither the ATMS SWE retrieval nor the associated algorithm in MiRS has received much attention. This study is intended to fill the latter knowledge gap by gauging the relative strengths of ATMS retrieval against the more established AMSR2 counterpart.

The primary research questions of this study are as follows. First, how do the AMSR2 and ATMS SWE retrievals differ in their accuracy among different geographic regions, and what are the complementary strengths (if any) of the two products? Second, how the differences in skill can be attributed to differences in scanning patterns of each instrument and in the retrieval algorithms? Third, can the introduction of a priori bias correction improve upon the optimal interpolation-based blending scheme, which is typically employed in the field of snow analysis (Brasnett, 1999; Brown et al., 2003; Liu et al., 2015)?

The rest of this paper is structured as follows. Section 2 describes the study area, the two PWM SW retrievals, and the in situ observations for analysis and validation. Sections 3 offers an overview of the blending algorithm, and Section 4 presents and interprets key observations emerging from the evaluation. Section 5 summarizes the major findings and concludes the study.

2. Study area and data

Figure 1 shows the elevation and spatial distribution of Snow Telemetry (SNOTEL) and Cooperative Observer Program (COOP) stations in the 18 hydrologic units (HUs) over the CONUS.

The research domain extends from 25° to 53° north and 125° to 67° west with a 0.125° grid spacing, which is same as the North American Land Data Assimilation System (NLDAS; Mitchell et al., 2004). The elevation data were aggregated from the 30 arc second Global Multi-resolution Terrain Elevation Data (GMTED2010; Danielson and Gesch, 2011). We acquired daily in situ observations and satellite retrievals over the CONUS for the main snow seasons of the water years 2017–2019 (i.e., 11/01/2016–06/30/2017, 11/01/2017–06/30/2018, and 11/01/2018–06/30/2019).

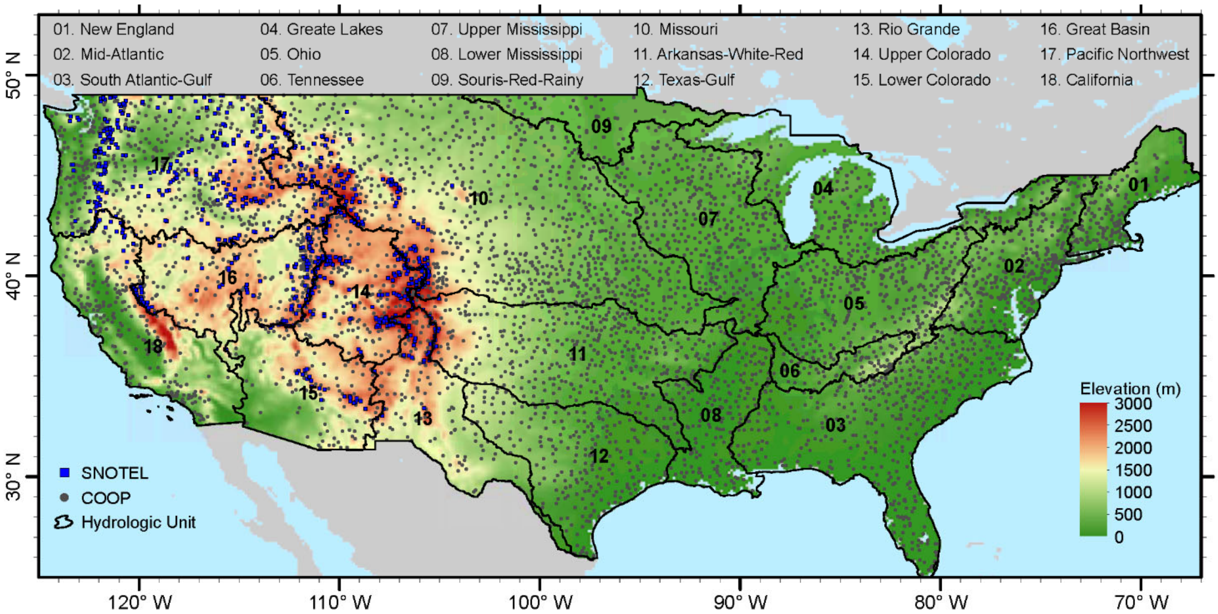

Figure 1. Elevation and spatial distribution of SNOTEL and COOP stations in the 18 hydrologic units over the CONUS.

2.1. SNOTEL SWE observations

SNOTEL is an automated network of stations that records snow and meteorological variables in the western US (Serreze et al., 1999). It provides reliable and quality-controlled data for over 800 high-elevation mountain stations in 12 states. At each station, a snow pillow measures how much water is in the snowpack by weighing the snow with a pressure transducer and the weight of snow is then converted into the SWE. We downloaded daily SWE data for 789 stations excluding those in Alaska from the Natural Resources Conservation Service (NRCS) (https://www.wcc.nrcs.usda.gov/snow).

2.2. COOP SWE observations

Because of the limited spatial coverage of SNOTEL SWE observations, additional observations of SD were collected from the National Weather Service (NWS) COOP network, which consists of more than 8000 active stations across the US. We acquired COOP SD data for 5073 stations from Iowa State University (https://mesonet.agron.iastate.edu/COOP). They were then converted to SWE using the snow bulk density method (Sturm et al., 2010) considering the effects of SD, snow aging, and snow cover classes as

$$\rho_{h_i,DOY_i} = (\rho_{max} - \rho_0)\left[1 - e^{(-k_1 \times h_i - k_2 \times DOY_i)}\right] + \rho_0 \quad (1)$$

where ρ_{max} is the maximum bulk density, ρ_0 is the initial density, and k_1 and k_2 are densification parameters for snow depth h_i and day of the year (DOY). DOY runs from −92 (1 October) to +181 (30 June) with no 0 value in the Northern Hemisphere. The parameter values of ρ_{max}, ρ_0, k_1, and k_2 for five snow cover classes as alpine, maritime, prairie, tundra, and taiga can be found in Table 4 of Sturm et al. (2010). The 0.5° × 0.5° global snow cover classes climatology map were acquired from the Arctic Data Center (https://arcticdata.io/catalog/view/doi:10.5065/D69G5JX5) and remapped to the research domain using nearest neighbor interpolation. This dataset divides the world's seasonal snowpack into six classes as tundra, taiga, maritime, ephemeral, prairie, and alpine snow, based on physical properties such as depth, density, thermal conductivity, number of layers, and degree of wetting (Sturm et al., 1995). We set the ephemeral snow to the prairie snow since there are no parameter values for this class of snow in Sturm et al. (2010).

2.3. ATMS SWE retrieval

ATMS was launched in October 2011 onboard the National Oceanic and Atmospheric Administration (NOAA) Suomi National Polar-orbiting Partnership (S-NPP) satellite. It is the next generation cross-track microwave sounder with 22 channels that spans the frequency range from 23.8 to 183 GHz. ATMS observes from 824 km above the Earth with the scan angular span of ±52.77° relative to nadir and obtains data over an approximately 2600 km swath. It provides two observations daily for a given location from a sun-synchronous orbit. Note that the resolution of the ATMS observations declines at larger scan angles due to the expanded field of view. A detailed description of the instrument characteristics of ATMS can be found in Weng et al. (2012).

The ATMS SWE product is retrieved using NOAA's MiRS, an operational microwave retrieval platform based on the one-dimensional variational (1DVAR) inversion algorithm (Boukabara et al., 2011). A detailed description of the MiRS is given by Liu et al. (2016). A flowchart of the ATMS SWE retrieval algorithm is presented in Figure 2. There are five primary steps to the retrieval algorithm:

Step 1. The raw sensor radiance observations are footprint matched using a footprint averaging/resampling algorithm to ensure that all channels for the retrieval view the same location on the Earth (Kongoli et al., 2011).

Step 2. The radiance observations are bias corrected to reduce the influences of cloud, precipitation and coastal contamination, by adjusting the histogram of the Tb difference between simulated and actual measurements to make it centered about zero (Liu et al., 2016).

Step 3. A first guess for the 1DVAR retrieval is generated, which comes either from a dynamic climatology varying with location, time of year, and time of day (temperature profile, water vapor profile, and skin temperature), or a combination of fixed climatology and regression (emissivity and hydrometeors).

Step 4. The 1DVAR is employed to assimilate the preprocessed radiance observations at multiple ATMS channels using the CRTM (Han et al., 2006) as the forward model to retrieve surface emissivity, along with all other key atmospheric and surface parameters simultaneously.

Step 5. A post-processing step is performed to compare the retrieved emissivity spectra with those from a precomputed catalog that relates surface emissivity to snowpack properties to find the closest match. The primary channels used in the catalog search are ATMS channels 1, 2, 3, and 16 (i.e., 23.8, 31.4, 50.3, and 88.2 GHz, respectively). The catalog is generated offline from a dense medium radiative transfer snow emissivity model which accounts for the dielectric properties of ice grains and assumes a constant snow density of 0.25 g/cm^3 (Weng et al., 2001). The catalog search of SWE subjects to additional constraints based on a temporally and spatially varying SWE climatology. A correction based on forest fraction is also applied to the emissivity spectrum prior to the catalog search.

We acquired ATMS level 2 SWE data from NOAA's Comprehensive Large Array-data Stewardship System (CLASS) (https://www.bou.class.noaa.gov). Level 2 data at the raw satellite observation field of views were binned into gridded fields of daily SWE on a resolution of 0.125° latitude/longitude as in our research domain.

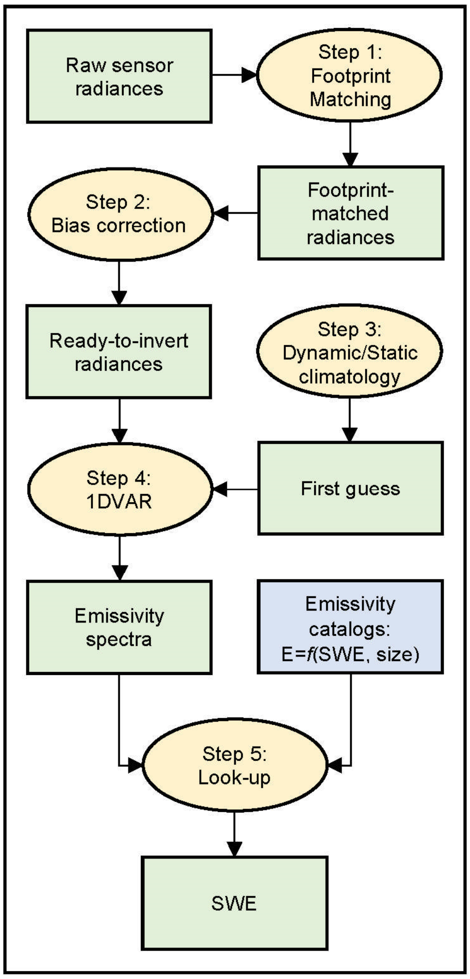

Figure 2. Flowchart of the ATMS SWE retrieval algorithm.

2.4. AMSR2 SWE retrieval

AMSR2 was launched in May 2012 onboard the Japanese Aerospace Exploration Agency (JAXA) GCOM – Water (GCOM-W1) satellite. It is a sensor to observe microwave radiation emitted naturally from the Earth surface and the atmosphere, using six different frequency bands ranging from 6.9 to 89 GHz. AMSR2 observes from about 700 km above the Earth with a nominal incidence angle of 55° and obtains data over a 1450 km swath. This conical scan mechanism enables AMSR2 to acquire a set of daytime and nighttime data at a nearly constant spatial resolution over more than 99% coverage of the Earth every two days. A detailed description of the instrument characteristics of AMSR2 can be found in Imaoka et al. (2010).

The retrieval algorithm of AMSR2 follows that of AMSR-E with higher spatial resolution. The SD algorithm separates out a retrieval for two classes as forest and non-forest and weights the summed estimate base on the fractional content of both classes in a grid cell (Kelly, 2009). The equation to calculate the SD is give as

$$SD = ff \times SD_f + (1 - ff) \times SD_{nf} \quad (2)$$

where SD_f is the SD for forested component, SD_{nf} is the SD for non-forested component, and ff is the forest fraction ranging from 0 to 0.75. For the forested component, the Tb difference between the 18.7 and 36.5 GHz channels forms the basis; while for the non-forested component, the Tb difference between the 10.7 and 36.5 GHz channels retrieves moderate snow and that between the 10.7 and 18.7 GHz channels retrieves deep snow. The equations to calculate the SD_f and SD_{nf} are given as

$$SD_f = p_1 \frac{Tb_{V18.7} - Tb_{V36.5}}{1 - 0.6 \times fd} \quad (3)$$

$$SD_{nf} = p_1(Tb_{V10.7} - Tb_{V36.5}) + p_2(Tb_{V10.7} - Tb_{V18.7}) \quad (4)$$

where

$$p_1 = \frac{1}{\log_{10}(Tb_{V36.5} - Tb_{H36.5})} \quad (5)$$

$$p_2 = \frac{1}{\log_{10}(Tb_{V18.7} - Tb_{H18.7})} \quad (6)$$

and fd is the forest density ranging from 0 to 1. Accordingly, subscript alphabets V and H denote vertical and horizontal polarization, respectively, and subscript numbers stand for corresponding frequencies. A nominal SD of 5 cm is assigned to shallow snow since it is challenging to get robust estimation of SD (Kelly, 2009). The shallow snow is detected by the following threshold

$$\begin{gathered} Tb_{V89} < Tb_{V23.8} \;\&\; Tb_{H89} < Tb_{H23.8} \\ Tb_{V89} < 255\ K \;\&\; Tb_{H89} < 255\ K \\ T_{ss} < 267\ K \end{gathered} \quad (7)$$

where T_{ss} is the snowpack surface temperature.

We acquired AMSR2 level-2 global SD data at the raw satellite observation field of views from the JAXA Global Portal System (G-Portal; https://gportal.jaxa.jp/gpr) and binned them into gridded fields of daily SD on a resolution of 0.125° latitude/longitude as in our research domain. The AMSR2 SD was then converted to SWE using the same snow bulk density method (Sturm et al., 2010) as in COOP SWE conversion.

2.5. IMS snow cover analysis

The Interactive Multisensor Snow and Ice Mapping System (IMS; Helfrich et al., 2007; Ramsay, 1998), provides snow and ice cover maps for the Northern Hemisphere from February 1997 to the present at 1 km, 4 km, and 14 km resolutions. This dataset is derived from a variety of data products including satellite imagery and in situ data. We downloaded the IMS daily 4 km resolution Northern Hemisphere snow and ice analysis dataset from the National Snow & Ice Data

Center (https://nsidc.org/data/g02156). We then remapped this dataset to our research domain using majority resampling. IMS was used to mask the PMW SWE, since it is more reliable for snow cover estimation (Brown et al., 2007; Chen et al., 2012).

2.6. SNODAS SWE analysis

The SNOw Data Assimilation System (SNODAS) is a modeling and data assimilation system developed by the National Operational Hydrologic Remote Sensing Center (NOHRSC) to provide daily estimates of SWE and other snow properties over the CONUS at a resolution of 1 km from 1 October 2003 to the present (https://nsidc.org/data/G02158). It integrates snow data from satellite, airborne platforms, and ground stations (including SNOTEL and COOP) with model estimates of snow cover and other snowpack properties (Carroll et al., 2001). On a regular basis, operators use the SNODAS platform to compute the differences between station observations and collocated model estimates. These differences undergo spatial interpolation and the interpolated fields are used to correct the model estimates. SNODAS is a widely used operational product and often serve as the reference for assessing other blended or model-simulated snowpack quantities (Tedesco and Narvekar, 2010). It is, however, worth noting that SNODAS product can be subject to errors in model physics as well as those in short-term forecasts and analysis that serve as forcing to the model (Lv et al., 2019). In this study, we remapped SNODAS dataset to our research domain by averaging the 1-km resolution values within each of the 0.125° grid cell. It was then used as a benchmark in examining the spatial distribution of ATMS, AMSR2 and blended SWE products.

3. Blending algorithm framework

The algorithm framework for blending satellite retrievals with in situ observations is shown in Figure 3. The first step is to mask the PMW SWE with the IMS snow cover maps. Namely, the PMW SWE retrievals are retained only over snow covered grids in order to remove possible false alarms; in addition, any grids that are seen to be snow covered in IMS but not detected by the PMW retrievals are filled with 5 mm of SWE. The outcome from the masking then undergoes the following processing steps: i) bias correction, ii) optimal interpolation, and iii) weighted averaging. We designed nine comparative experiments (Table 1) to investigate the effect of each step on the accuracy of the resulting product. Exp1 and Exp2 are experiments for the raw ATMS and AMSR2 SWE products, respectively, and Exp3 to Exp9 are different blending experiments. In each blending experiment, one, or a combination of processing steps is applied to zero to two satellite retrieval(s) in conjunction with the in situ observations.

Table 1. Design of comparative experiments.

No.	Background	Bias correction	Optimal interpolation	Weighted averaging
Exp1	ATMS	no	no	no
Exp2	AMSR2	no	no	no
Exp3	N/A	no	yes	no
Exp4	ATMS	no	yes	no
Exp5	AMSR2	no	yes	no
Exp6	ATMS	yes	yes	no
Exp7	AMSR2	yes	yes	no
Exp8	ATMS and AMSR2	no	yes	yes
Exp9	ATMS and AMSR2	yes	yes	yes

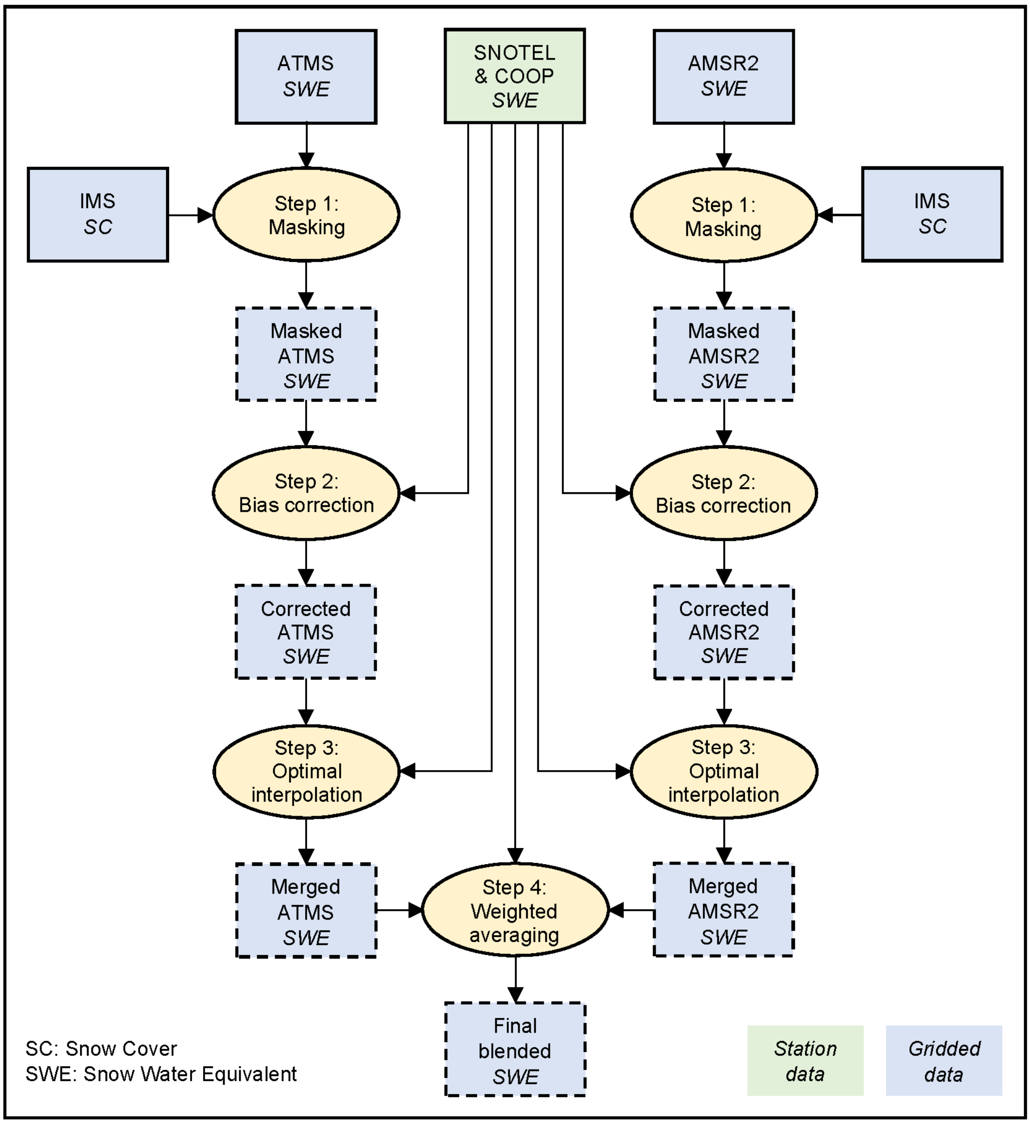

Figure 3. The algorithm framework for blending satellite retrievals with in situ observations.

3.1. Bias correction: CDF matching

In existing literature, bias in satellite SD/SWE products can be corrected through optimal interpolation, whereby the satellite estimate is the background (or the first guess) and subsequently blended with in situ data increments from surrounding stations (Kongoli et al., 2019; Liu et al., 2015). However, the robustness of optimal interpolation as a bias correction mechanism is debatable. A separate bias correction is often done prior to multisensory blending in precipitation analysis (Seo et al., 2000; Seo and Breidenbach, 2002; Xie and Xiong, 2011) and soil moisture analysis (Liu et al., 2011). In order to assess the potential impact of bias correction applied to satellite SWE prior to blending, we introduce a bias correction method based on cumulative distribution function (CDF) matching (Xie and Xiong, 2011). To this end, collocated pairs of in situ and satellite data are collected over a spatial domain of 120 km radius and a vertical distance of 800 m centered at the target grid cell and over a time period of 30 days, ending at the target date. The CDFs are then calculated for the satellite and in situ data, respectively. The initial data collection domain of 120 km radius is expanded when necessary until no less than 600 pairs of data (20 stations × 30 days) are collected to ensure stable CDFs. Under the assumption that the SWE at a percentage point in the CDF table for the satellite retrievals should be the same as that for the in situ observations, the bias of satellite retrievals at specific percentage point p for the target grid cell k can be corrected by

$$\hat{S}_{k,p} = S_{k,p} + \left(\tilde{O}_{k,p} - \tilde{S}_{k,p}\right) \tag{8}$$

where $\hat{S}_{k,p}$ and $S_{k,p}$ are the corrected and raw satellite retrievals, respectively; $\tilde{O}_{k,p}$ and $\tilde{S}_{k,p}$ are the SWE values at the percentage point p in the CDF table for in situ observations and satellite retrievals, respectively.

3.2. Optimal interpolation

Optimal interpolation, or OI (Gandin, 1965) was reported to provide the most accurate and stable analyses among several popular objective analysis methods (Chen et al., 2008). In our algorithm, bias-corrected satellite retrievals are merged with in situ observations using the optimal interpolation algorithm as in Brasnett (1999). The final analysis $\hat{S}_k$ at a target grid cell k is obtained by adjusting the background S_k using observations and backgrounds at the surrounding stations:

$$\hat{S}_k = S_k + \sum_{i=1}^{n} w_i (O_i - S_i) \tag{9}$$

where O_i is the observation at station i; S_i is the background at the grid cell where station i is located; n is the number of stations used for interpolation; and w_i is the optimum weight associated with station i for calculating the adjustment to the grid cell k. The weight vector **w** is calculated as follows:

$$\mathbf{w} = (\mathrm{P} + \mathrm{O})^{-1}\mathrm{q} \tag{10}$$

where **P** is the correlation coefficient matrix of background errors between all pairs of surrounding stations; **O** is the covariance matrix of observational errors normalized by the background error variance between all pairs of surrounding stations; and **q** is the correlation coefficient vector of background errors between the surrounding stations and the target grid cell.

The correlation coefficients of **P** and **q** are assumed to have the form

$$\mu_{ij} = \alpha(\Delta r_{ij})\beta(\Delta z_{ij}) \tag{11}$$

with the horizontal and vertical correlation functions calculated following Eq. (5) and (6), respectively

$$\alpha(\Delta r_{ij}) = (1 + c\Delta r_{ij})\exp(-c\Delta r_{ij}) \tag{12}$$

$$\beta(\Delta z_{ij}) = \exp\left[-(\Delta z_{ij}/h)^2\right] \tag{13}$$

where Δr_{ij} and Δz_{ij} denote the horizontal and vertical separations between points i and j, respectively, with $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$ for calculating the correlation coefficients in **P**, and $i = 1, 2, \ldots, n$ and $j = k$ for calculating the correlation coefficients in $\mathbf{q}$; c and h are two constants that prescribe the horizontal and vertical length scales, respectively. Here, we set c to 0.018 km^{-1} (corresponding to an e-folding distance of 120 km) and h to 800 m following Brasnett (1999).

To address the influence of terrain aspect (measured clockwise in degrees from north) on the snowpack, we derived the 0.125° terrain aspect map from the aggregated GMTED2010 elevation data and categorized the grid cells into north-facing (aspect ≤ 90° or aspect ≥270°) and south-facing (90° < aspect < 270°) slopes following Liu et al. (2015). Surrounding stations sharing the same slopes with the target grid cell are incorporated in optimal interpolation. This constraint condition of terrain aspect is applied only to grid cells that have an elevation higher than 900 m as identified in latter analysis that ATMS and AMSR2 have poor performance in these areas.

3.3. Weighted averaging

To further leverage the spatially varying skills of satellite-based merged products, the in situ-satellite merged SWE based on ATMS and AMSR2 retrievals are combined by weighted averaging

$$\bar{S}_k = \sum_{i=1}^{n} w_{k,i} \hat{S}_{k,i} \tag{14}$$

where $\hat{S}_{k,i}$ denotes the in situ-satellite merged SWE at the kth grid using the ith satellite product; $w_{k,i}$ is the optimal weight which can be determined by the reciprocal of mean squared error as

$$w_{k,i} = \frac{1}{\sigma_{k,i}^2} \Big/ \sum_{i=1}^{n} \frac{1}{\sigma_{k,i}^2} \tag{15}$$

where $\sigma_{k,i}^2$ is the mean squared error calculated based on the collocated pairs of in situ and satellite data over a spatial domain of 120 km radius and a vertical distance of 800 m centered at the target grid cell k and over a time period of 30 days, ending at the target date.

4. Results and discussion

In this section we first present the outcomes from the comparison of ATMS and AMSR2 SWE retrievals, then we describe the results of cross-validation experiments with a focus on the differential skills of products generated through each scheme. In the end we further compare the best product as determined from the cross-validation experiments with SNODAS analysis.

4.1. Evaluation of the PMW SWE retrievals

The relative accuracy and error sources of ATMS and AMSR2 SWE retrievals are examined in the following respects: a) geographic variations and associated seasonal contrasts; b) temporal dynamics of snowpack; c) influences of terrain, vegetation and SD; and d) potential sources of error. Results for each aspect are presented below.

4.1.1. Geographic and seasonal variation

Figure 4 shows the geographic distribution of multi-year (i.e., water years 2017–2019) mean daily SWE in different months for ATMS, AMSR2 and SNODAS. SWE retrievals from ATMS and AMSR2 agree well in the large-scale spatial patterns, but both largely underestimate SWE, as compared to SNODAS, in the north-central and northeastern US as well as in the Intermountain West. This is consistent with the findings of several other studies in which PMW retrievals were found lower than SNODAS analysis in the presence of complex terrain, high forest cover, deep snowpack, and snow ablation (Tedesco and Narvekar, 2010; Vuyovich et al., 2014). The maximum daily SWE for different months ranges from 72 to 221 mm for ATMS and from 208 to 279 mm for AMSR2, which are much lower than the range exhibited by SNODAS product (1539–2919 mm). This suggests that ATMS saturates around 220 mm SWE and AMSR2 saturates around 280 mm SWE, most likely because the Tb at one of the primary retrieval channels (31.4 GHz for ATMS and 36.5 GHz for AMSR2) is no longer sensitive to increasing SWE (Hancock et al., 2013). Although the maximum daily SWE of ATMS is lower than that of AMSR2 for all months, the mean daily SWE of ATMS is generally higher than that of AMSR2 over almost all regions in all months. The exceptions are regions of the Northern Plains in February and over the western Souris-Red-Rainy region in March. ATMS and AMSR2 capture the relatively high SWE from January to March over the Souris-Red-Rainy region, the Northern Plains, and the Rockies. Both PMW SWE products have better performance over the Northern Plains because this region is relatively flat and consists of mostly open prairie or farmland, where the snowpack has limited melt-refreeze effects (Josberger and Mognard, 2002). On the other hand, both products fail to track the shallow snow in November–December over the New England region, the Great Lakes region, the Rockies, the Cascade, and the Sierra Nevada, due to the weak scattering of shallow snow, which is difficult to be detected by PMW sensors (Foster et al., 2011). Meanwhile, they also fail to track the wet snow in March–June over these regions because meltwater in the snowpack significantly reduces the scattering signal compared with dry snow, resulting in a decrease of the high- and low-frequency Tb difference (Dawson et al., 2018; Foster et al., 2005).

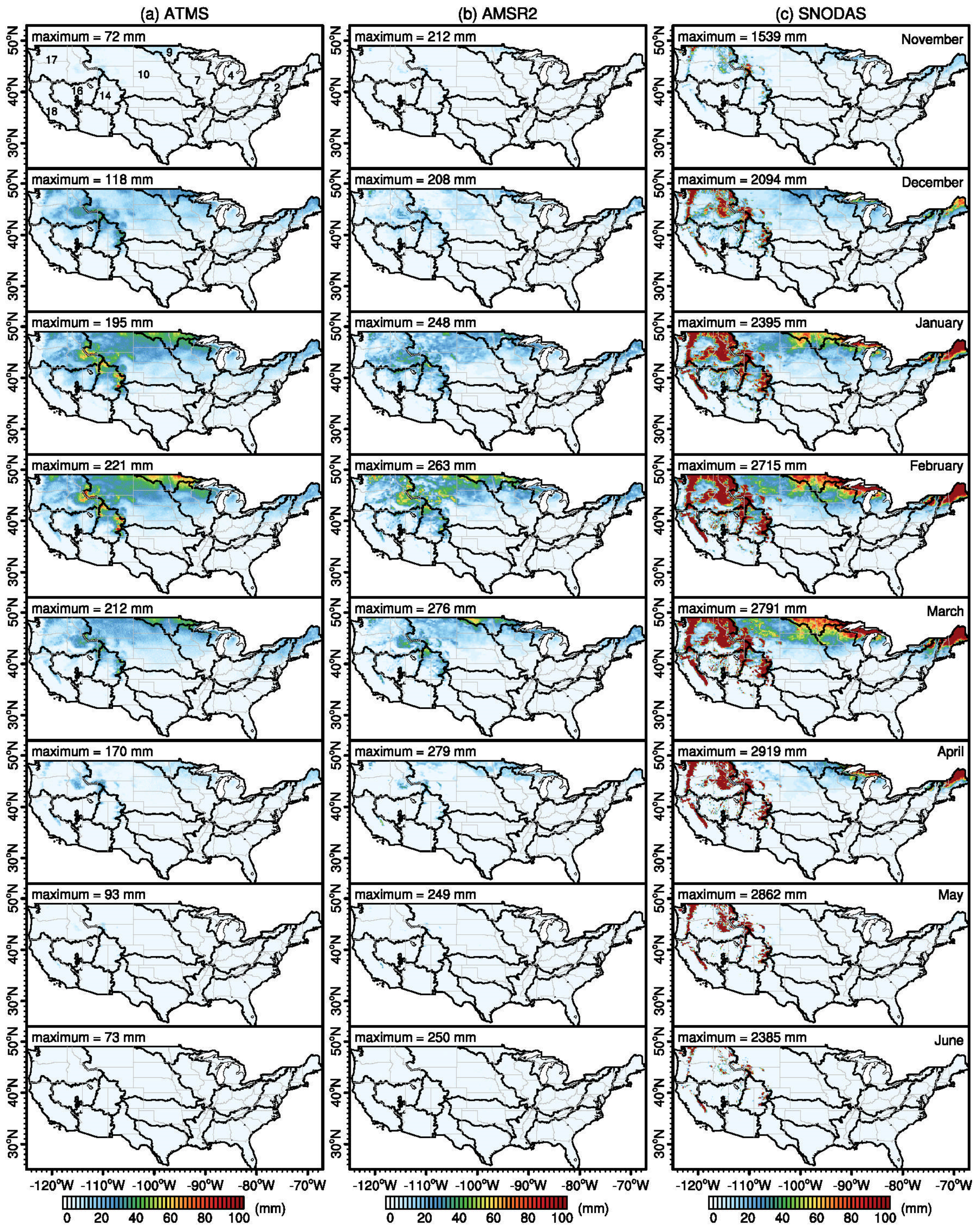

Figure 4. Geographic distribution of multi-year (i.e., water years 2017–2019) mean daily SWE in different months. Text at the top of each subfigure shows the maximum daily SWE in that month.

We also evaluated the ATMS, AMSR2, and SNODAS against the IMS snow cover analysis for the water years 2017–2019 to check their reliability in detecting snow cover. Grid cells were classified as snow or non-snow covered based on a SWE threshold of 1 mm as in Brown et al. (2007). Figure 5 shows the false alarm ratio (FAR) versus probability of detection (POD) in different months over ten snow covered HUs. Overall, both PMW products underperform SNODAS analysis in snow detection, which demonstrates the need for an accurate snow mask (such as the IMS snow cover analysis) before blending PMW products with the in situ observations. One caveat is that the high POD in SNODAS could be due to overestimation in some cases, as indicated by the high FAR for November and April–June as well as over parts of the Intermountain West (HUs 16–18) and the Great Lakes region (HU 4). The PMW products have clear issues in observing snow cover in April–June due to snow ablation (particularly increasing water content due to melting) and for almost all months over the complex terrain in the Sierra Nevada (HU 18). ATMS and AMSR2 clearly exhibit complementary snow cover detecting skills. In general, AMSR2 shows better detection skills over the New England region (HU 1) and the Great Lakes region (HU 4), but for some of the regions, including the Upper Mississippi (HU 7), the Souris-Red-Rainy (HU 9), the Missouri (HU 10), and the Pacific Northwest (HU 17), ATMS exhibits slightly higher detection skills. On the other hand, higher incidence of false alarms is seen in ATMS product for most of the study regions, and is particularly pronounced for April–May and over the east of the Rockies (HUs 1, 2, 4, 7, 9, and 10). These differences can be explained by the fact that ATMS SWE retrieval has on the average a lower resolution than the AMSR2 counterpart. As indicated earlier, the resolution of ATMS is lower than AMSR2 and it tends to decline at wider scanning angle. As the size of the grid pixel increases, it is more likely to encompass areas that are snow-free. This effect tends to be more pronounced in regions with ephemeral and patchy snow cover, and this is why the FAR for ATMS is particularly high in April and over the east of Rockies.

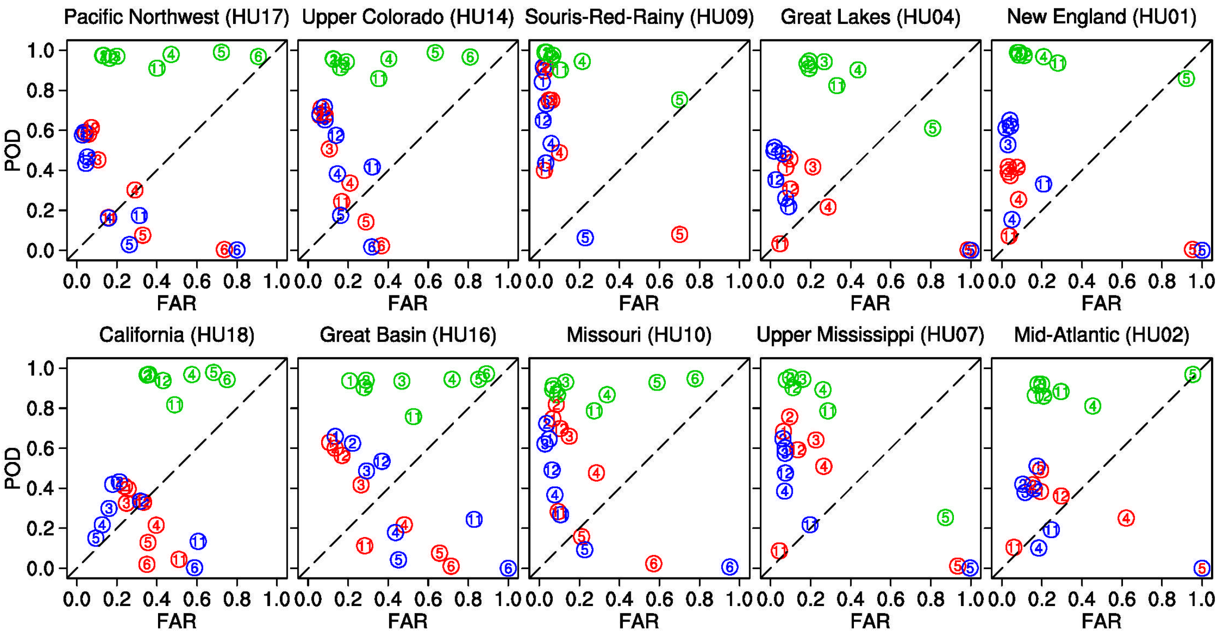

Figure 5. False alarm ratio (FAR) versus probability of detection (POD) for the snow cover of ATMS (red), AMSR2 (blue), and SNODAS (green) against IMS analysis in different months over ten snow covered HUs from the western to the eastern US. FAR and POD are calculated based on the daily snow cover of the water years 2017–2019. Numbers in the circles show the months. The closer a circle is to the upper left corner, the better it estimates.

4.1.2. Temporal dynamics

Figure 6 compares the time series of multi-station mean daily SWE for ATMS, AMSR2, SNODAS, and in situ observations over ten snow covered HUs. SNODAS provides relatively accurate depiction of SWE temporal variation in all HUs, with a correlation between 0.93 and 0.98 and a bias between −42 and −4 mm. At the same time, ATMS and AMSR2 perform poorly over the Mountainous West (HUs 10, 14 and 16–18); ATMS has a correlation ranging from 0.56 to 0.67

and a bias ranging from −153 to −36 mm and AMSR2 has a correlation ranging from 0.52 to 0.68 and a bias ranging from −155 to −39 mm. The initial snow accumulation phase and the abrupt end of season snow melt from the PMW SWE appear to track closely the in situ observations over the Northeast (HUs 1 and 2) and the Upper Midwest (HUs 4, 7, and 9), where the correlation is between 0.78 and 0.90 and the bias is between −31 and −7 mm for ATMS and the correlation is between 0.71 and 0.88 and the bias is between −36 and −8 mm for AMSR2. The PMW SWE peak slightly earlier than the in situ data, potentially a result of sensor saturation and liquid water effect, which reduces the scattering signal of the snowpack and thus limits the retrieval of SWE. Overall, ATMS outperforms AMSR2 for HUs 2, 4, 7, 9, and 16–18, whereas AMSR2 SWE is more closely correlated with in situ data yet it exhibits generally larger absolute bias than ATMS in HUs 10 and 14. These observations point to broad complementarity between ATMS and AMSR2 SWE products.

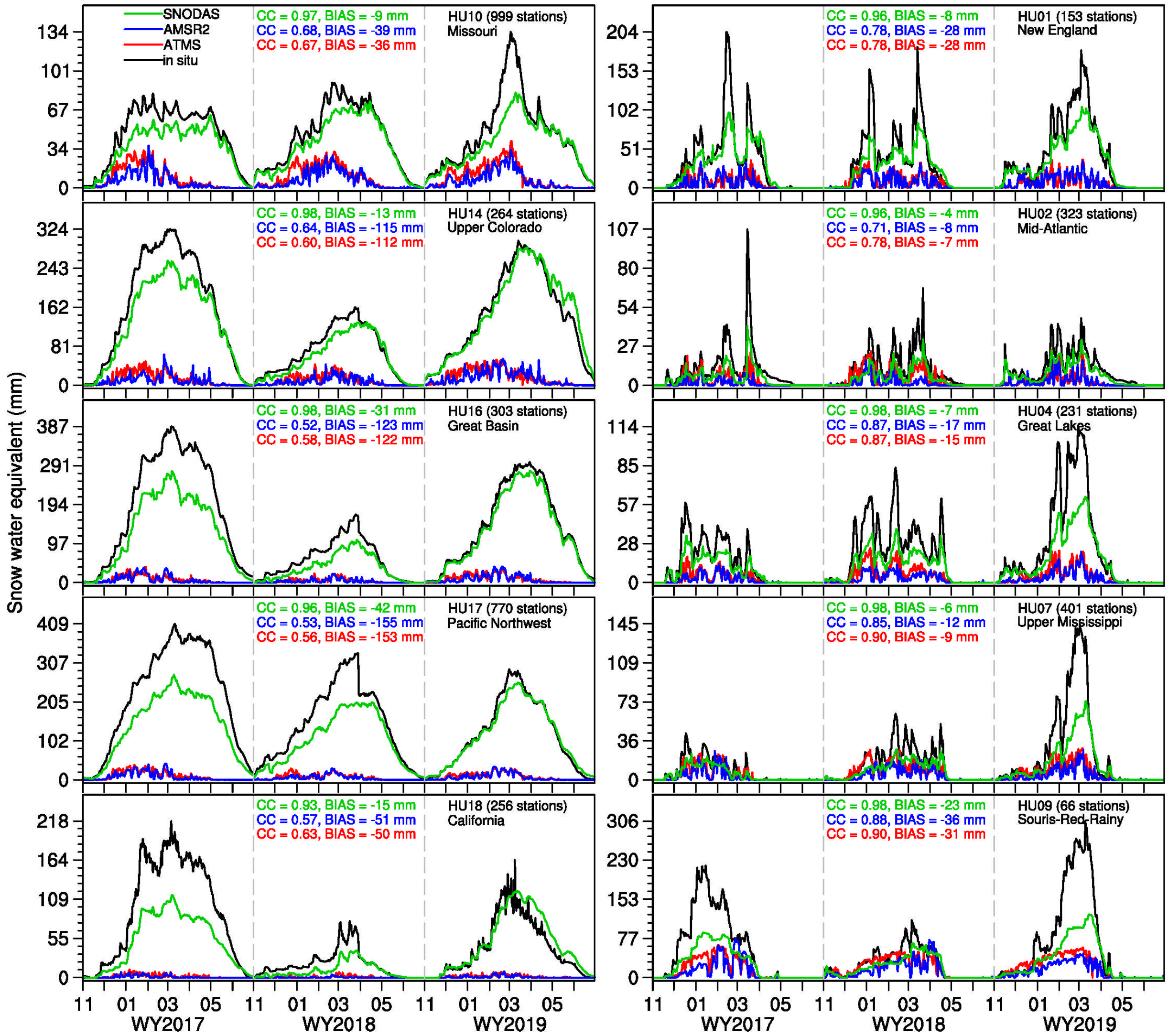

Figure 6. Time series of multi-station mean daily SWE for the main snow seasons of the water years 2017–2019 (different water years are separated by gray dashed lines) in ten snow covered HUs. SNODAS, AMSR2, ATMS, and in situ data are denoted by green, blue, red, and black colors, respectively. CC represents correlation and BIAS represents bias.

A distinctive feature in Figure 6 is that ATMS SWE is conspicuously higher than AMSR2 SWE in late December and early January in HUs 2, 4, 7, 9, and 10, especially for the water years 2017 and 2018 in HU 10. As the two retrievals rely on different assumptions of snow density, we performed a simple analysis to assess the specific role of the differing snow density in this early winter contrast. We applied a constant snow density of 0.25 g/cm^3 (as in ATMS) to the AMSR2

SD to derive an alternative AMSR2 SWE product. In addition, AMSR2 assigns a nominal SD of 5 cm to shallow snow, which is quite different from the ATMS retrieval algorithm. We categorized stations into two types as SD < 100 mm and SD ≥ 100 mm to contrast the early winter SWE difference. Figure 7 shows the multi-station mean daily SWE of AMSR2 with revised snow density against those of raw AMSR2 and ATMS products for HU 10. It is clear that the snow density assumption has marginal impact on the AMSR2 SWE series (i.e., slight difference in CC and BIAS for the two AMSR2 products), and thus it is unlikely to be the key cause of the difference between ATMS and AMSR2 SWE retrievals. The only time when this difference is relatively large is in the early winter and where snow is shallow (i.e., SD < 100 mm stations). It therefore appears that AMSR2's rather pronounced underestimation of SWE in late December and early January is more likely an outcome of inappropriate treatment of shallow snow in its algorithm. More specifically, AMSR2 uses the 89 GHz channel for detecting "shallow snow" as a separate class, to which a minimal SD value is assigned (Kelly, 2009). However, the 89 GHz channel is more sensitive to interference by atmospheric water vapor and precipitation, has lower snowpack penetration depth and is noisier due to snowpack grain size variations than the lower frequency channels. During the early phase of snow accumulation, this shallow snow switch is more likely to be activated, resulting in large negative bias.

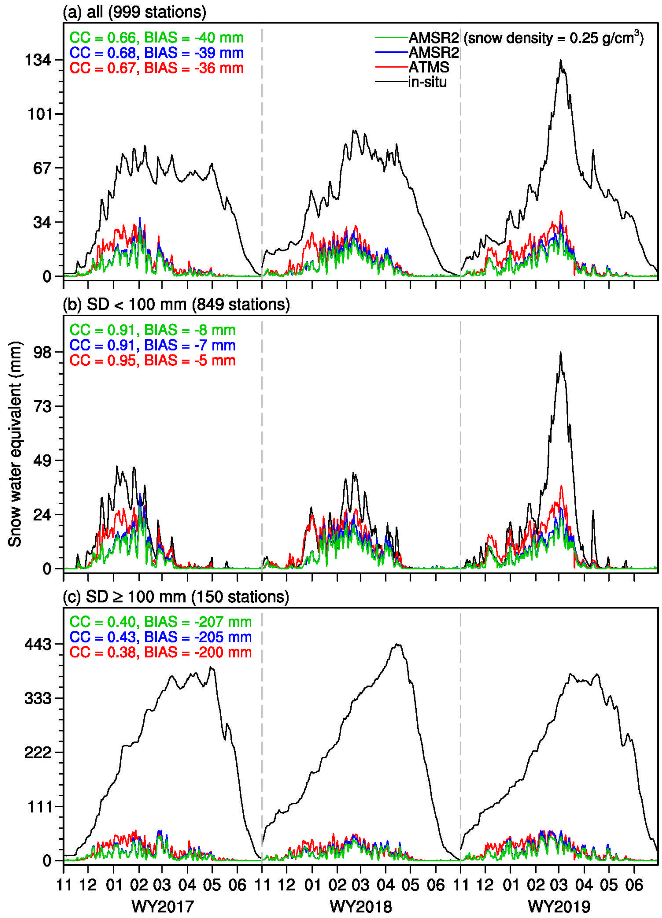

Figure 7. Time series of multi-station mean daily SWE for the main snow seasons of the water years 2017–2019 (different water years are separated by gray dashed lines) in HU 10. AMSR2 with constant snow density, AMSR2, ATMS, and in situ data are denoted by green, blue, red, and black colors, respectively. CC represents correlation and BIAS represents bias.

4.1.3. Effects of terrain, vegetation, and SD

The uncertainties in PMW SWE retrievals are attributable to many factors such as terrain, vegetation, and sensor saturation (Foster et al., 2005; Hancock et al., 2013), as well as snowpack properties such as snow grain size and volume fraction. We therefore compute a new set of validation statistics for ATMS, AMSR2 and SNODAS SWE products stratified by station elevation, grid elevation range, mean green vegetation fraction (GVF), and mean SD to determine the impact of each factor on SWE (Figure 8). The elevation and SD data were collected from the observation stations. The elevation range within each of the 0.125° grid cell was determined using the 30 arc second GMTED2010 elevation data. The GVF data were derived from the MODIS global Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation (FPAR) product (MOD15A2; Myneni et al., 2015) using Weather Research and Forecasting Preprocessing System (WPS).

As expected, SNODAS performs much better than ATMS and AMSR2 in both correlation and bias because it assimilates SNOTEL and COOP data. ATMS generally has lower correlation but smaller absolute bias than AMSR2, which demonstrates the complementary nature of, and thus the utility in blending both products. The performance of ATMS and AMSR2 declines with elevation, elevation range, GVF, and SD. Considering both correlation and bias, it is evident that PMW SWE is more reliable when elevation is below 900 m, elevation range is smaller than 300 m, GVF is less than 20%, and SD is lower than 200 mm. Snow properties exhibit wide spatial variations in high elevation areas as elevation ≥ 900 m due to complex terrain, which is difficult to be captured by the coarse spatial resolution of the available PMW sensors (Mätzler and Standley, 2000). The difference between PMW and in situ SWE widens with the elevation range as the surface heterogeneity becomes more significant (Cho et al., 2020). The masking effect of the forest canopy overwhelms the scattering signal from the snowpack when GVF ≥ 20%, which can lead to SWE underestimation (Foster et al., 2005). Additionally, there is a clear trend that the underestimation of SWE becomes increasingly severe at higher SD, especially when SD ≥ 200 mm; and this is reflecting the signal saturation in deep snow (Clifford, 2010).

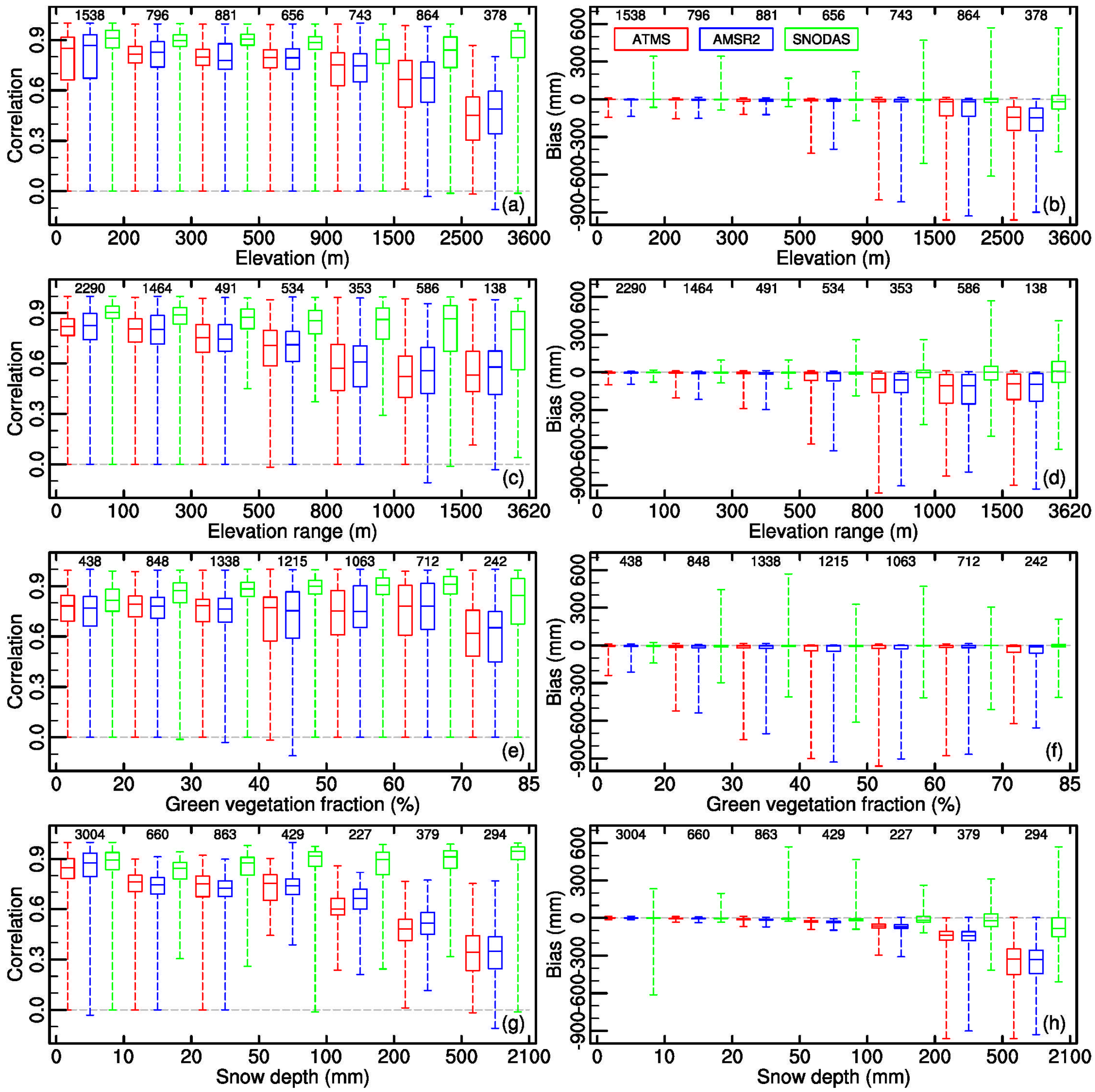

Figure 8. Box plots for correlation and bias of the daily SWE (November–June for the water years 2017–2019) stratified by station elevation, grid elevation range, mean green vegetation fraction, and mean snow depth. ATMS, AMSR2, and SNODAS are represented by red, blue, and green boxes, respectively. The number above the box is the total valid number of the statistic for each class.

4.1.4. Potential sources of error

The similarities and differences in the performance of ATMS and AMSR2 SWE retrievals are closely linked to those in instruments as well as in the retrieval algorithms. PMW SD retrieval algorithms rely on the microwave scattering signal produced by snow grains, which is manifested in a decrease of emitted radiation at higher frequencies relative to lower frequencies (Kelly, 2009). This decrease is a complex function of snowpack characteristics such as SD, grain size and snowpack density, the latter two of which can also vary within the snow column. AMSR2 retrieval algorithm, being an empirical mechanism, relates variation in Tb to SD but it does not account for grain size or snowpack density explicitly. At higher SD, Tb difference among channels becomes insensitive to changes in SD and this gives rise to signal saturation. Though ATMS retrieval algorithm employs a different approach (i.e., assimilation of radiance into the CRTM to retrieve surface emissivity spectra which in turn are interpreted through a simple snow emission model to retrieve SWE), its SWE retrieval remains subject to signal saturation that arises from reduced sensitivity of radiance to SD variations for deeper snowpack. In general, the relatively lower correlation, lower detection rate, and higher false alarm rate for ATMS SWE are indication that its retrieval mechanism, though physically based, is prone to errors arising from model parameterizations and underrepresentation of processes. Possible sources of error include misclassification of surface type for the a priori background spectrum in the snow emissivity, errors in atmospheric temperature and moisture profiles (Boukabara et al., 2013), and errors in the structure of the snow emissivity model and CRTM. The assumption of constant snow density most likely degrades the performance of ATMS SWE retrieval, as it ignores the variation as a result of compaction and snow metamorphism (Dawson et al., 2017). However, this impact appears limited as only small difference in statistics is observed in the AMSR2 SWE when a constant density is imposed.

The information gained can assist algorithm developers in anticipating error characteristics of other SWE products, e.g., the MiRS-based Advanced Microwave Sounding Unit-A (AMSUA)/Microwave Humidity Sounder (MHS) onboard the Meteorological Operational Satellite (Metop)-A, B, and C (Klaes et al., 2007), as well as help users establish the circumstances where each product, or a combination of the products can be effectively applied. The relative skills and error characteristics serve as the basis for the blending algorithm.

4.2. Cross-validation of different blended SWE products

We performed k-fold cross-validation to assess the performance of the blended in situ-satellite SWE. Such cross-validation is commonly used for validating model and observational analysis (Fushiki, 2011; Gan et al., 2015). Specifically, the in situ SWE observations from about $(k-1)/k$ of the 5862 stations (789 SNOTEL and 5073 COOP stations) were used in the blending process whereas the remainder $1/k$ were withheld for validation. This blending process was repeated k times so that the observation at each station was withdrawn once. The blended SWE analyses and corresponding statistics at the withdrawn stations were then calculated.

4.2.1. Comparison of different blending schemes

We first performed 10-fold cross-validation, which uses observations from 90% of stations in the blending process, to assess the performances of different blending schemes. The box-percentile plots for correlation, bias, root mean square error (RMSE), and unbiased RMSE of the different experiments in Table 1 and the SNODAS product are shown in Figure 9. Overall, all blended experiments (Exp3–Exp9) show much better performance than the raw ATMS (Exp1) and AMSR2 (Exp2). However, the optimal interpolation with the in situ observations alone experiment (Exp3) still exhibits large negative biases and performs much worse than the optimal interpolation with background experiments (Exp4 and Exp5). This confirms the utility of the satellite retrievals as background, despite their deficiencies relative to in situ observations. Exp4 and Exp5 improve the mean correlation from 0.72 (Exp1) to 0.82 (Exp4) for ATMS and from 0.72 (Exp2) to 0.83 (Exp5) for AMSR2; meanwhile, the mean bias is improved from −41 mm (Exp1) to −14 mm (Exp4) for ATMS and from −43 mm (Exp2) to −14 mm (Exp5) for AMSR2. Compared to the ATMS and AMSR2 SWE, the experiments with both bias correction and optimal interpolation (Exp6 and Exp7) show improved mean correlation, i.e., from 0.72 (Exp1) to 0.79 (Exp6) for ATMS and from 0.72 (Exp2) to 0.79 (Exp7) for AMSR2; meanwhile, the mean bias is improved from −41 mm (Exp1) to −10 mm (Exp6) for ATMS and from −43 mm (Exp2) to −10 mm (Exp7) for AMSR2. It is noted that the absolute bias would be reduced even more but the correlation would be improved relatively less when bias correction was applied before optimal interpolation. The weighted average of Exp4 and Exp5, i.e. Exp8, has higher correlation (with a mean value of 0.85) and smaller absolute bias (with a mean value of −12 mm) than both Exp4 and Exp5; while Exp9, which is the weighted average of Exp6 and Exp7, has much higher correlation (with a mean value of 0.84) but slightly larger absolute bias (with a mean value of −12 mm) than both Exp6 and Exp7. Compared to the ATMS and AMSR2 SWE, Exp8 increases the mean correlation by 18% and 17%, respectively, and reduces the mean bias by 70% and 71%, respectively; while Exp9 increases the mean correlation by 17% and 16%, respectively, and reduces the mean bias by 70% and 71%, respectively. Over 75% stations have a correlation higher than 0.80 and over 65% stations have an absolute bias smaller than 10 mm for Exp8 and Exp9, which are better than all the other experiments and are comparable to SNODAS SWE product.

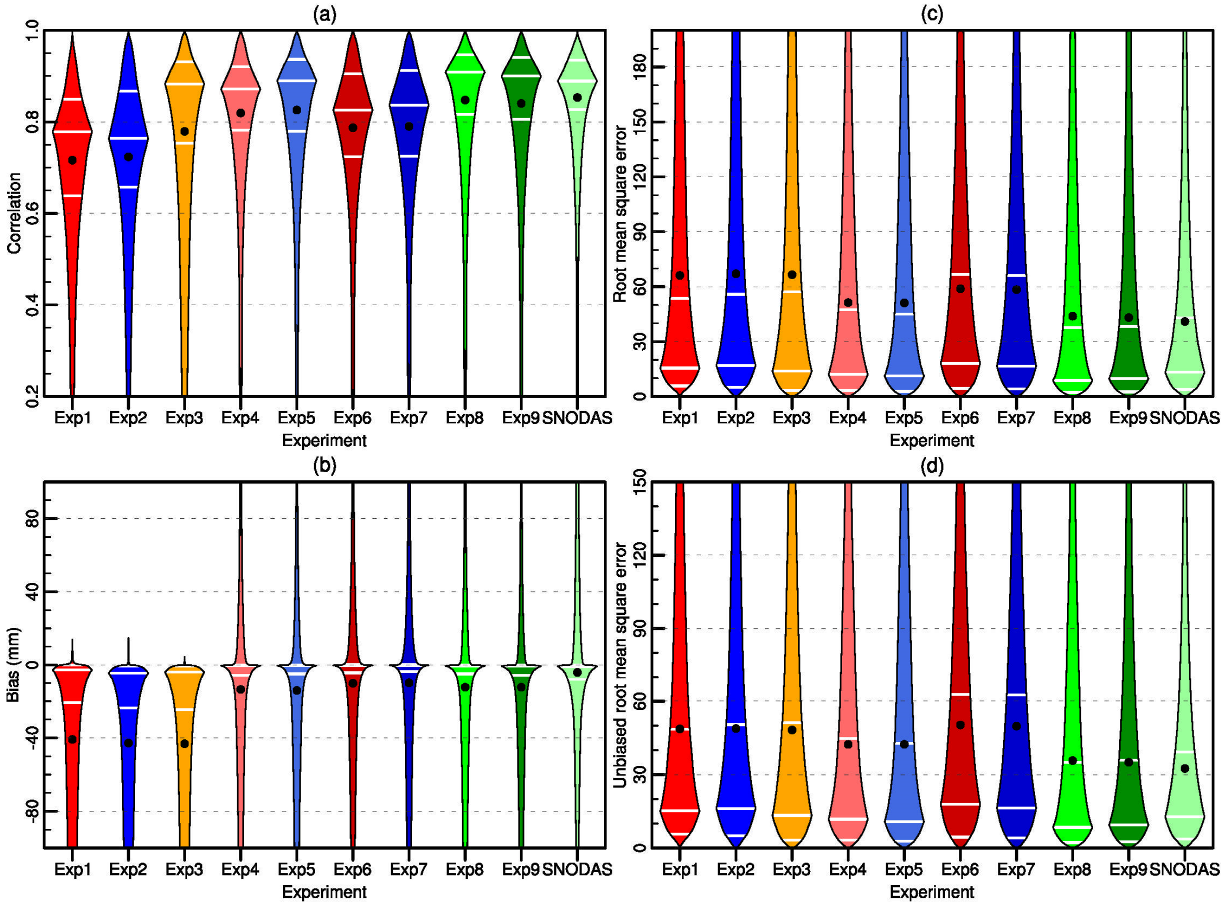

Figure 9. Box-percentile plots for (a) correlation, (b) bias, (c) RMSE, and (d) unbiased RMSE of the different experiments in Table 1 and the SNODAS product. Exp1 and Exp2 are experiments for the raw ATMS and AMSR2 SWE, respectively. Exp3 to Exp9 are different blending experiments with 10-fold cross-validation scheme. In each box, the black dot represents the mean value and the white lines from top to bottom represent 75%, 50%, and 25% percentiles, respectively. The width of the box shows the distribution of the data.

4.2.2. Impact of bias correction

We also repeated the comparisons for products Exp8 and Exp9 using 2- and 4-fold cross-validation schemes. Figure 10 shows the box-percentile plots for correlation, bias, RMSE, and unbiased RMSE of Exp8 and Exp9 as derived from 2-, 4-, and 10-fold cross-validations. As expected, the cross-validation schemes with more station records tend to perform better. While Exp8 and Exp9 perform similarly in the 10-fold cross-validation, all statistics except the bias of Exp8 are better than those of Exp9 in the 2- and 4-fold cross-validations. This means that the blending scheme with a prior bias correction (Exp9) does not always improve overall performance – other than bias – compared to that without bias correction (Exp8). Histogram matching intended to reduce the large negative bias of PMW SWE can degrade correlation to a certain extent, and the degradation can be alleviated when more stations are included in the blending process.

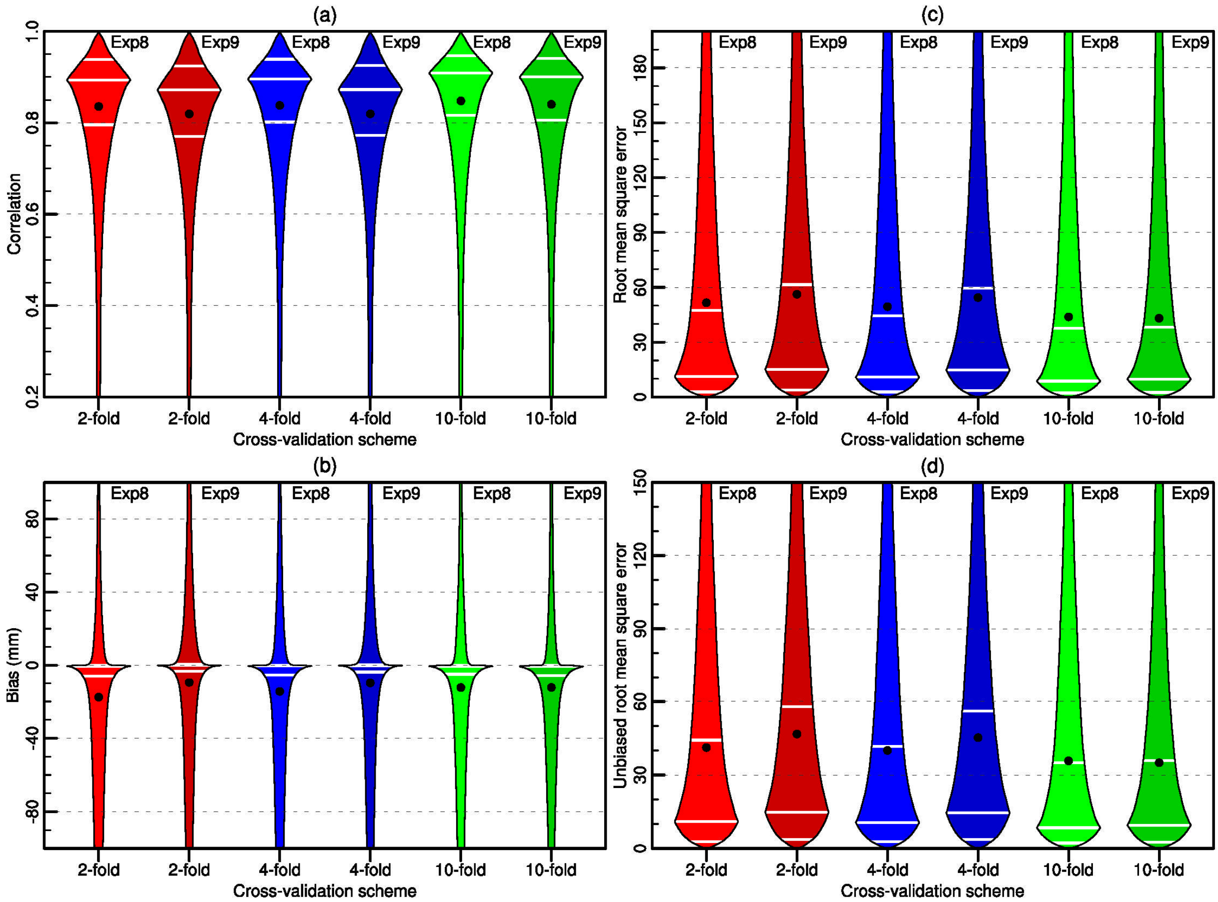

Figure 10. Box-percentile plots for (a) correlation, (b) bias, (c) RMSE, and (d) unbiased RMSE of Exp8 and Exp9 with different cross-validation schemes. In each box, the black dot represents the mean value and the white lines from top to bottom represent 75%, 50%, and 25% percentiles, respectively. The width of the box shows the distribution of the data.

In order to closely examine the role of bias correction, we compared the time series of Exp8 and Exp9 with 10-fold cross-validation scheme against in situ and SNODAS SWE on a multi-station mean basis for selected HUs in Figure 11. Exp8 and Exp9 outperform SNODAS in HUs 1, 4, 7, and 9, with correlation ranging from 0.98 to 0.99 and bias ranging from −15 to −2 mm. Meanwhile, Exp9 also outperforms SNODAS in HU 10, with a correlation of 0.99 and a bias of −7 mm. Nonetheless, Exp8 and Exp9 show relatively larger underestimation than SNODAS in HUs 14 and 16–18, with correlation ranging from 0.93 to 0.99 and bias ranging from −72 to −25 mm. Overall, the agreement between the blended products and the in situ observations is fairly well except Exp8 shows some spurious spikes in HUs 10, 14 and 16–18. This suggests that although optimal interpolation can generate realistic estimates of SWE at validation points, it may introduce spikes when the bias between the backgrounds and observations is large. These observations indicate that an independent bias correction method is warranted, at least for the Intermountain West where the in situ network is sparse and the spatial representativeness of in situ observations is limited by topography (Broxton et al., 2016a). Nevertheless, it is worth pointing out that bias correction does not always render the product bias free, or even improve the bias. For example, Exp9 exhibits slightly worse negative bias than Exp8 for HU9, and again this can be explained by a combination of a relatively sparse in situ network and severe bias that make the bias correction less effective.

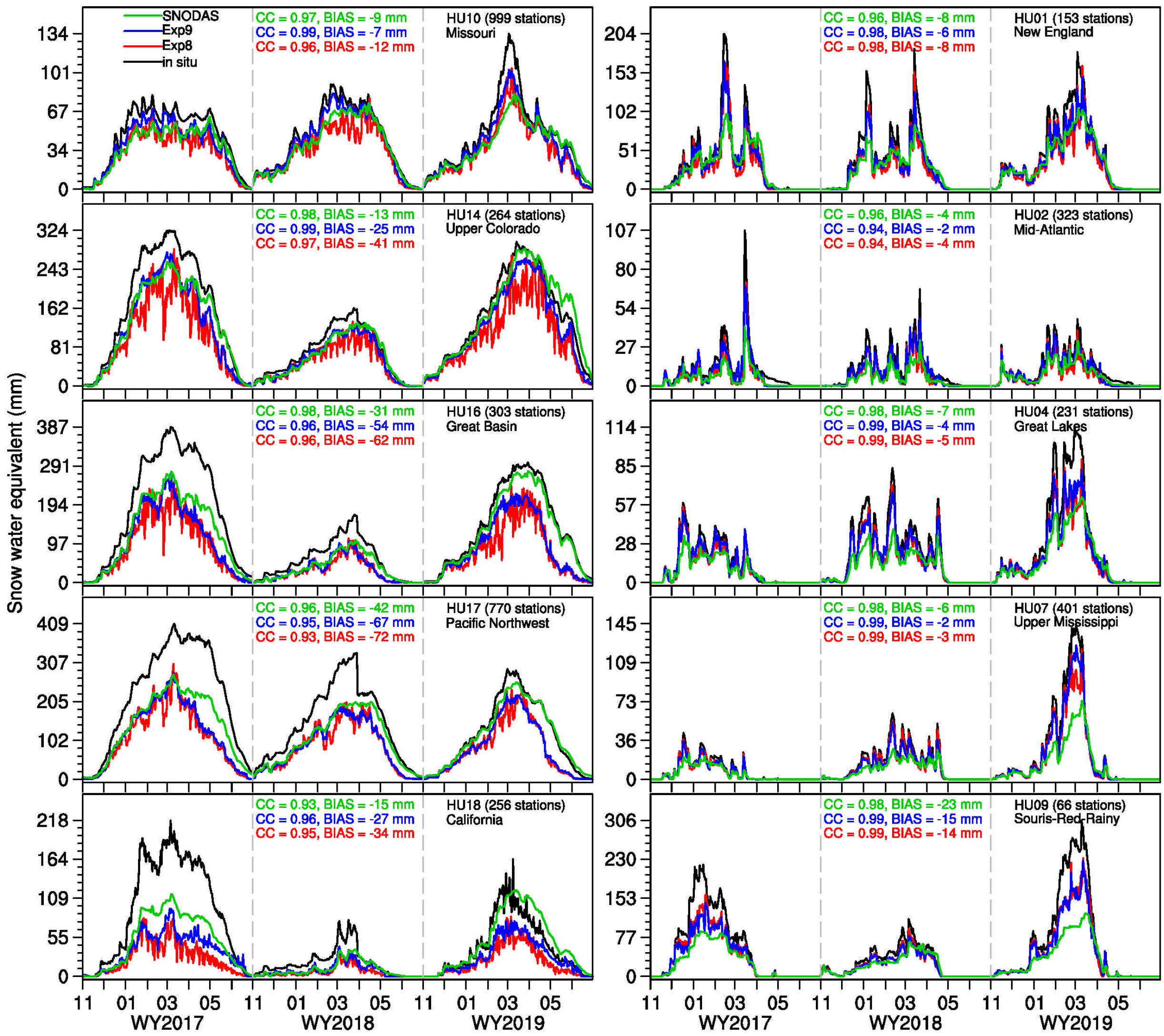

Figure 11. Time series of multi-station mean daily SWE for the main snow seasons of the water years 2017–2019 (different water years are separated by gray dashed lines) in ten snow covered HUs. Exp8 (red color) and Exp9 (blue color) are blending experiments as shown in Table 1 with 10-fold cross-validation. SNODAS and in situ data are denoted by green color and black color, respectively. CC represents correlation and BIAS represents bias.

4.3. Evaluation of the final blended SWE product

The earlier analysis suggests that Exp9, the blended product that underwent bias correction, broadly outperforms others. In this section we further explore the difference between this dataset and SNODAS SWE analysis to illustrate its potential practical utility over different geographic settings. Figure 12 shows the geographic distribution of the multi-year (i.e., water years 2017–2019) mean daily SWE for Exp9, SNODAS, and their differences across snow season. Generally, the final blended product (Exp9) agrees well with the SNODAS analysis in spatial pattern. The overestimation over the Rockies, the Northern Plains, and the Souris-Red-Rainy region from January to March is reasonable since SNODAS was demonstrated to underestimate the actual SWE in these regions as shown in Figure 11. SNODAS tends to underestimate snow density and thus SWE, because it assimilates SD and SWE observations across different scales and platforms without using snow density to constrain the assimilation (Dawson et al., 2017). Previous studies also found that SNODAS slightly underestimates SD in heavily forested regions (Anderson, 2011) and considerably underestimates SD in mountainous regions (Clow et al., 2012). On the other hand, the final blended product underestimates SWE over the Cascade and Sierra Nevada from December to June as well as the northern New England and northern Great Lakes regions from December to April. This discrepancy could be partly attributed to the fact that SNODAS tends to overestimate in these regions in April–June as indicated by the high FAR in Figure 5. It could also be attributed to the sparsity of observation stations in these regions (see Figure 1) to correct the large negative bias of the backgrounds (ATMS and AMSR2 SWE). Furthermore, the representativeness of the stations for their surrounding areas might be inadequate due to the spatial heterogeneity (Meromy et al., 2013), which also limits the accuracy of the blended in situ-satellite product. Interested readers can find additional information on the differential improvements of blended products for different land surface characteristics in the supplementary material.

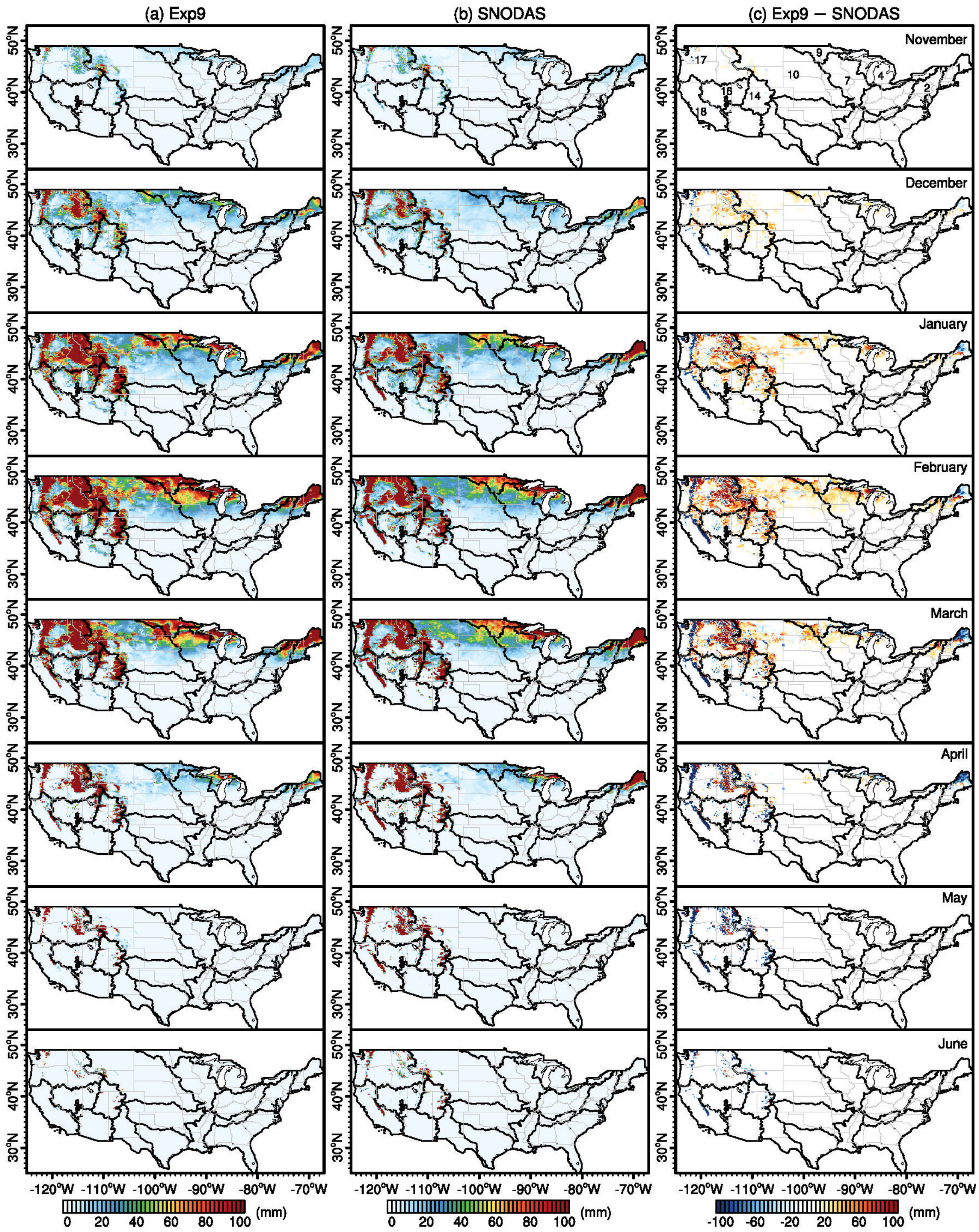

Figure 12. Geographic distribution of multi-year (i.e., water years 2017–2019) mean daily SWE in different months.

5. Summary and conclusions

This paper compares two different PMW SWE products, namely ATMS and AMSR2, with SNOTEL and COOP in situ observations and the SNODAS analysis over the CONUS. A blending algorithm is then designed to optimally combine the in situ and satellite SWE to obtain a reliable gridded product at a relatively high spatial resolution ($0.125° \times 0.125°$).

The comparison results indicate that the accuracy of ATMS and AMSR2 SWE, despite derived using different instruments and retrieval algorithms, have much in common in terms of geographic distribution of performance. Both products capture the temporal variability of in situ SWE well when elevation is below 900 m, elevation range is smaller than 300 m, GVF is less than 20%, and SD is lower than 200 mm. On the other hand, both products considerably underestimate SWE in the north-central and northeastern US as well as in the Intermountain West, possibly due to a combination of the influence of complex terrain, high forest cover, deep snowpack, and snow ablation. These similarities notwithstanding, there are notable differences in the performance of the two retrievals that point substantially to their complementarity as sources of information for SWE estimates. Relative to AMSR2, the signal saturation for ATMS appears to occur at a lower SWE (220 mm SWE vs 280mm SWE for AMSR2), and the ATMS tends to over-detect snow covered areas due to its larger field of views. On the other hand, ATMS does fare better in detecting snow cover over the early and middle of snow season (December to February) for regions spanning from the Pacific Northwest to the Upper Midwest (except for the Souris-Red-Rainy region), though it exhibits broadly higher false alarm rates for a majority of these regions. Further analysis also reveals a potential defect of AMSR2 retrieval – it tends to severely underestimate the SWE for early snow season over the Northern Plains, and this is related to its use of high frequency channel (i.e., 89 GHz) whose radiance observations tend to saturate at shallower SD.

Our blending algorithm provides a simple, yet effective way to produce reliable blended in situ-satellite SWE estimates by exploiting the complementary strengths of AMSR2 and ATMS SWE retrievals. The final, blended product outperforms the interpolated station-only product as well as the raw ATMS and AMSR2 SWE: for the latter, the mean correlation sees 17% and 16% increases, while the mean bias drops by 70% and 71%, respectively. In particular, our analysis shows that an independent bias correction is effective in improving upon the optimal interpolation-based blended product for much of the snow season. The only exception is over the Intermountain West, where the sparsity of in situ stations and their preferred topographic locations constrain the efficacy of the bias correction. Nonetheless, even in these regions, bias correction is still helpful as it reduces the wild fluctuations in SWE that itself is likely a consequence of data paucity. Additional research is warranted to identify the station density threshold and potential temporal smoothing approaches for damping out the oscillations.

The blended SWE can be used in a number of practical contexts, which include, but are not limited to assisting with situational hydrologic awareness for forecasters and water managers and serving as observation field assimilated to snow and land surface models. As demonstrated in the study, the blended product in fact slightly outperforms the SNODAS analysis in some of the regions over Northeast to the Upper Midwest where snowpack was shallow and ephemeral, though

SNODAS analysis remains a superior dataset over the Intermountain West where snowpack is much thicker than what satellites can sense. This suggests that there is additional room for improving the snow analysis by leveraging the strengths of blended products and model simulation through data assimilation. In addition, observations from emerging platforms, e.g., airborne lidar (Painter et al., 2016), GPS-reflectometry (Larson, 2016), Sentinel-1 C-band synthetic aperture radar (Lievens et al., 2019), and airborne gamma radiation detector (Cho et al., 2020) offer new opportunities to address the sparsity of in situ data over mountainous regions and adoption of these datasets in the blending framework will be explored in future studies.

CRediT authorship contribution statement

Yanjun Gan: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. **Yu Zhang:** Conceptualization, Methodology, Formal analysis, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Cezar Kongoli:** Data curation, Formal analysis, Writing - review & editing. **Christopher Grassotti:** Formal analysis, Writing - review & editing. **Yuqiong Liu:** Methodology, Writing - review & editing. **Yong-Keun Lee:** Data curation, Writing - review & editing. **Dong-Jun Seo:** Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Oceanic and Atmospheric Administration (grant #NA18OAR4590410) and this support is graciously acknowledged here. We would also like to thank Greg Fall and Brian Cosgrove at the Office of Water Prediction (OWP), and Hui Shao at the Joint Center for Satellite Data Assimilation (JCSDA) for their helpful suggestions. We appreciate the constructive comments made by the three anonymous reviewers. The daily 0.125° blended in situ-satellite SWE data are available at https://www.hydroshare.org/resource/7ac869e3be06411d8f43bf152b62a755/.

References

Anderson, B.T., 2011. Spatial Distribution and Evolution of a Seasonal Snowpack in Complex Terrain: An Evaluation of the SNODAS Modeling Product. Master of Science, Boise State University, Boise, Idaho.

Bommarito, J.J., 1993. DMSP special sensor microwave imager sounder (SSMIS), in: Proceedings of SPIE, Orlando, Florida, pp. 230–238. https://doi.org/10.1117/12.152601.

Boukabara, S.A., Garrett, K., Chen, W., Iturbide-Sanchez, F., Grassotti, C., Kongoli, C., Chen, R., Liu, Q., Yan, B., Weng, F., Ferraro, R., Kleespies, T.J., Meng, H., 2011. MiRS: An all-weather 1DVAR satellite data assimilation and retrieval system. IEEE T. Geosci. Remote 49 (9), 3249–3272. https://doi.org/10.1109/TGRS.2011.2158438.

Boukabara, S.A., Garrett, K., Grassotti, C., Iturbide Sanchez, F., Chen, W., Jiang, Z., Clough, S.A., Zhan, X., Liang, P., Liu, Q., 2013. A physical approach for a simultaneous retrieval of sounding, surface, hydrometeor, and cryospheric parameters from SNPP/ATMS. J. Geophys. Res. Atmos. 118 (22), 12,600–12,619. https://doi.org/10.1002/2013JD020448.

Brasnett, B., 1999. A global analysis of snow depth for numerical weather prediction. J. Appl. Meteor. 38 (6), 726–740. https://doi.org/10.1175/1520-0450(1999)038<0726:AGAOSD>2.0.CO;2.

Brown, R., Derksen, C., Wang, L., 2007. Assessment of spring snow cover duration variability over northern Canada from satellite datasets. Remote Sens. Environ. 111 (2), 367–381. https://doi.org/10.1016/j.rse.2006.09.035.

Brown, R.D., Brasnett, B., Robinson, D., 2003. Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. Atmos. Ocean 41 (1), 1–14. https://doi.org/10.3137/ao.410101.

Brown, R.D., Brasnett, B., 2010. Canadian Meteorological Centre (CMC) Daily Snow Depth Analysis Data. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA. https://doi.org/10.5067/W9FOYWH0EQZ3.

Broxton, P.D., Dawson, N., Zeng, X., 2016a. Linking snowfall and snow accumulation to generate spatial maps of SWE and snow depth. Earth Space Sci. 3 (6), 246–256. https://doi.org/10.1002/2016EA000174.

Broxton, P.D., Zeng, X., Dawson, N., 2016b. Why do global reanalyses and land data assimilation products underestimate snow water equivalent? J. Hydrometeorol. 17 (11), 2743–2761. https://doi.org/10.1175/JHM-D-16-0056.1.

Brun, E., Vionnet, V., Boone, A., Decharme, B., Peings, Y., Valette, R., Karbou, F., Morin, S., 2013. Simulation of Northern Eurasian local snow depth, mass, and density using a detailed snowpack model and meteorological reanalyses. J. Hydrometeorol. 14 (1), 203–219. https://doi.org/10.1175/JHM-D-12-012.1.

Carroll, T., Cline, D., Fall, G., Nilsson, A., Li, L., Rost, A., 2001. NOHRSC operations and the simulation of snow cover properties for the coterminous US, in: Proceedings of the 69th Annual Western Snow Conference, Sun Valley, Idaho, pp. 1–14.

Chang, A.T.C., Foster, J.L., Kelly, R.E.J., Josberger, E.G., Armstrong, R.L., Mognard, N.M., 2005. Analysis of ground-measured and passive-microwave-derived snow depth variations in midwinter across the northern Great Plains. J. Hydrometeorol. 6 (1), 20–33. https://doi.org/10.1175/JHM-405.1.

Che, T., Li, X., Jin, R., Huang, C., 2014. Assimilating passive microwave remote sensing data into a land surface model to improve the estimation of snow depth. Remote Sens. Environ. 143, 54–63. https://doi.org/10.1016/j.rse.2013.12.009.

Chen, C., Lakhankar, T., Romanov, P., Helfrich, S., Powell, A., Khanbilvardi, R., 2012. Validation of NOAA-Interactive Multisensor Snow and Ice Mapping System (IMS) by comparison with ground-based measurements over continental United States. Remote Sens. 4 (5), 1134–1145. https://doi.org/10.3390/rs4051134.

Chen, M., Shi, W., Xie, P., Silva, V.B.S., Kousky, V.E., Wayne Higgins, R., Janowiak, J.E., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res. Atmos. 113 (D4), D04110. https://doi.org/10.1029/2007JD009132.

Cho, E., Jacobs, J.M., Vuyovich, C.M., 2020. The value of long-term (40 years) airborne Gamma radiation SWE record for evaluating three observation-based gridded SWE data sets by seasonal snow and land cover classifications. Water Resour. Res. 56 (1), e2019WR025813. https://doi.org/10.1029/2019WR025813.

Clark, M.P., Hendrikx, J., Slater, A.G., Kavetski, D., Anderson, B., Cullen, N.J., Kerr, T., Örn Hreinsson, E., Woods, R.A., 2011. Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. Water Resour. Res. 47, W07539. https://doi.org/10.1029/2011WR010745.

Clifford, D., 2010. Global estimates of snow water equivalent from passive microwave instruments: history, challenges and future developments. Int. J. Remote Sens. 31 (14), 3707–3726. https://doi.org/10.1080/01431161.2010.483482.

Clow, D.W., Nanus, L., Verdin, K.L., Schmidt, J., 2012. Evaluation of SNODAS snow depth and snow water equivalent estimates for the Colorado Rocky Mountains, USA. Hydrol. Process. 26 (17), 2583–2591. https://doi.org/10.1002/hyp.9385.

Dai, L., Che, T., Ding, Y., Hao, X., 2017. Evaluation of snow cover and snow depth on the Qinghai–Tibetan Plateau derived from passive microwave remote sensing. The Cryosphere 11 (4), 1933–1948. https://doi.org/10.5194/tc-11-1933-2017.

Danielson, J.J., Gesch, D.B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010). U.S. Geological Survey, Reston, Virginia. https://doi.org/10.3133/ofr20111073.

Dawson, N., Broxton, P., Zeng, X., 2017. A new snow density parameterization for land data initialization. J. Hydrometeorol. 18 (1), 197–207. https://doi.org/10.1175/JHM-D-16-0166.1.

Dawson, N., Broxton, P., Zeng, X., 2018. Evaluation of remotely sensed snow water equivalent and snow cover extent over the Contiguous United States. J. Hydrometeorol. 19 (11), 1777–1791. https://doi.org/10.1175/JHM-D-18-0007.1.

De Lannoy, G.J.M., Reichle, R.H., Houser, P.R., Arsenault, K.R., Verhoest, N.E.C., Pauwels, V.R.N., 2010. Satellite-scale snow water equivalent assimilation into a high-resolution land surface model. J. Hydrometeorol. 11 (2), 352–369. https://doi.org/10.1175/2009JHM1192.1.

De Lannoy, G.J.M., Reichle, R.H., Arsenault, K.R., Houser, P.R., Kumar, S., Verhoest, N.E.C., Pauwels, V.R.N., 2012. Multiscale assimilation of Advanced Microwave Scanning Radiometer–EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado. Water Resour. Res. 48 (1), W01522. https://doi.org/10.1029/2011WR010588.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. Roy. Meteor. Soc. 137 (656), 553–597. https://doi.org/10.1002/qj.828.

Dong, C., 2018. Remote sensing, hydrological modeling and in situ observations in snow cover research: A review. J. Hydrol. 561, 573–583. https://doi.org/10.1016/j.jhydrol.2018.04.027.

Dong, J., Walker, J.P., Houser, P.R., 2005. Factors affecting remotely sensed snow water equivalent uncertainty. Remote Sens. Environ. 97 (1), 68–82. https://doi.org/10.1016/j.rse.2005.04.010.

Dong, J., Walker, J.P., Houser, P.R., Sun, C., 2007. Scanning multichannel microwave radiometer snow water equivalent assimilation. J. Geophys. Res. Atmos. 112 (D7), D07108. https://doi.org/10.1029/2006JD007209.

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

Dziubanski, D.J., Franz, K.J., 2016. Assimilation of AMSR-E snow water equivalent data in a spatially-lumped snow model. J. Hydrol. 540, 26–39. https://doi.org/10.1016/j.jhydrol.2016.05.046.

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for

Environmental Prediction operational mesoscale Eta model. J. Geophys. Res. Atmos. 108, 8851. https://doi.org/10.1029/2002JD003296.

Foster, J.L., Sun, C., Walker, J.P., Kelly, R., Chang, A., Dong, J., Powell, H., 2005. Quantifying the uncertainty in passive microwave snow water equivalent observations. Remote Sens. Environ. 94 (2), 187–203. https://doi.org/10.1016/j.rse.2004.09.012.

Foster, J.L., Hall, D.K., Eylander, J.B., Riggs, G.A., Nghiem, S.V., Tedesco, M., Kim, E., Montesano, P.M., Kelly, R.E.J., Casey, K.A., Choudhury, B., 2011. A blended global snow product using visible, passive microwave and scatterometer satellite data. Int. J. Remote Sens. 32 (5), 1371–1395. https://doi.org/10.1080/01431160903548013.

Frei, A., Tedesco, M., Lee, S., Foster, J., Hall, D.K., Kelly, R., Robinson, D.A., 2012. A review of global satellite-derived snow products. Adv. Space Res. 50 (8), 1007–1029. https://doi.org/10.1016/j.asr.2011.12.021.

Fushiki, T., 2011. Estimation of prediction error by using K-fold cross-validation. Stat. Comput. 21 (2), 137–146. https://doi.org/10.1007/s11222-009-9153-8.

Gan, Y., Liang, X.-Z., Duan, Q., Choi, H.I., Dai, Y., Wu, H., 2015. Stepwise sensitivity analysis from qualitative to quantitative: Application to the terrestrial hydrological modeling of a Conjunctive Surface-Subsurface Process (CSSP) land surface model. J. Adv. Model. Earth Sy. 7 (2), 648–669. https://doi.org/10.1002/2014MS000406.

Gandin, L.S., 1965. Objective Analysis of Meteorological Fields. Israel Program for Scientific Translations, Jerusalem.

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., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The Modern-Era Retrospective Analysis for research and applications, version 2 (MERRA-2). J. Climate 30 (14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1.

Han, Y., van Delst, P., Liu, Q., Weng, F., Yan, B., Treadon, R., Derber, J., 2006. JCSDA Community Radiative Transfer Model (CRTM)–Version 1. NOAA Technical Report NESDIS 122. NOAA, Silver Spring, Maryland.

Hancock, S., Baxter, R., Evans, J., Huntley, B., 2013. Evaluating global snow water equivalent products for testing land surface models. Remote Sens. Environ. 128, 107–117. https://doi.org/10.1016/j.rse.2012.10.004.

Helfrich, S.R., McNamara, D., Ramsay, B.H., Baldwin, T., Kasheta, T., 2007. Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). Hydrol. Process. 21 (12), 1576–1586. https://doi.org/10.1002/hyp.6720.

Hollinger, J.P., 1989. DMSP Special Sensor Microwave/Imager Calibration/Validation. Space Sensing Branch, Naval Research Laboratory, Washington D.C.

Imaoka, K., Sezai, T., Takeshima, T., Kawanishi, T., Shibata, A., 2002. Instrument characteristics and calibration of AMSR and AMSR-E, in: IEEE International Geoscience and Remote

Sensing Symposium, Toronto, Ontario, Canada, pp. 18–20. https://doi.org/10.1109/IGARSS.2002.1024927.

Imaoka, K., Kachi, M., Kasahara, M., Ito, N., Nakagawa, K., Oki, T., 2010. Instrument performance and calibration of AMSR-E and AMSR2. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 38 (8), 13–16.

Josberger, E.G., Mognard, N.M., 2002. A passive microwave snow depth algorithm with a proxy for snow metamorphism. Hydrol. Process. 16 (8), 1557–1568. https://doi.org/10.1002/hyp.1020.

Kelly, R., 2009. The AMSR-E snow depth algorithm: Description and initial results. J. Remote Sens. Soc. Japan 29 (1), 307–317. https://doi.org/10.11440/rssj.29.307.

Klaes, K.D., Cohen, M., Buhler, Y., Schlüssel, P., Munro, R., Luntama, J., von Engeln, A., Clérigh, E.Ó., Bonekamp, H., Ackermann, J., Schmetz, J., 2007. An introduction to the EUMETSAT polar system. B. Am. Meteorol. Soc. 88 (7), 1085–1096. https://doi.org/10.1175/BAMS-88-7-1085.

Kongoli, C., Boukabara, S., Yan, B., Weng, F., Ferraro, R., 2011. A new sea-ice concentration algorithm based on microwave surface emissivities—Application to AMSU measurements. IEEE T. Geosci. Remote 49 (1), 175–189. https://doi.org/10.1109/TGRS.2010.2052812.

Kongoli, C., Key, J., Smith, T.M., 2019. Mapping of snow depth by blending satellite and in-situ data using two-dimensional optimal interpolation—Application to AMSR2. Remote Sens. 11 (24), 3049. https://doi.org/10.3390/rs11243049.

Kumar, S.V., Peters-Lidard, C.D., Arsenault, K.R., Getirana, A., Mocko, D., Liu, Y., 2015. Quantifying the added value of snow cover area observations in passive microwave snow depth data assimilation. J. Hydrometeorol. 16 (4), 1736–1741. https://doi.org/10.1175/JHM-D-15-0021.1.

Kumar, S.V., Jasinski, M., Mocko, D.M., Rodell, M., Borak, J., Li, B., Beaudoing, H.K., Peters-Lidard, C.D., 2019. NCA-LDAS land analysis: Development and performance of a multisensor, multivariate land data assimilation system for the national climate assessment. J. Hydrometeorol. 20 (8), 1571–1593. https://doi.org/10.1175/JHM-D-17-0125.1.

Kunkee, D.B., Poe, G.A., Boucher, D.J., Swadley, S.D., Hong, Y., Wessel, J.E., Uliana, E.A., 2008. Design and evaluation of the first Special Sensor Microwave Imager/Sounder. IEEE T. Geosci. Remote 46 (4), 863–883. https://doi.org/10.1109/TGRS.2008.917980.

Kwon, Y., Yang, Z.-L., Hoar, T.J., Toure, A.M., 2017. Improving the radiance assimilation performance in estimating snow water storage across snow and land-cover types in North America. J. Hydrometeorol. 18 (3), 651–668. https://doi.org/10.1175/JHM-D-16-0102.1.

Larson, K.M., 2016. GPS interferometric reflectometry: applications to surface soil moisture, snow depth, and vegetation water content in the western United States. WIREs Water 3 (6), 775–787. https://doi.org/10.1002/wat2.1167.

Larue, F., Royer, A., De Sève, D., Langlois, A., Roy, A., Brucker, L., 2017. Validation of GlobSnow-2 snow water equivalent over Eastern Canada. Remote Sens. Environ. 194, 264–277. https://doi.org/10.1016/j.rse.2017.03.027.

Lee, Y., Kongoli, C., Key, J., 2015. An in-depth evaluation of heritage algorithms for snow cover and snow depth using AMSR-E and AMSR2 measurements. J. Atmos. Ocean. Tech. 32 (12), 2319–2336. https://doi.org/10.1175/JTECH-D-15-0100.1.

Lettenmaier, D.P., Alsdorf, D., Dozier, J., Huffman, G.J., Pan, M., Wood, E.F., 2015. Inroads of remote sensing into hydrologic science during the WRR era. Water Resour. Res. 51 (9), 7309–7342. https://doi.org/10.1002/2015WR017616.

Lievens, H., Demuzere, M., Marshall, H., 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., De Lannoy, G.J.M., 2019. Snow depth variability in the Northern Hemisphere mountains observed from space. Nat. Commun. 10 (1), 4629. https://doi.org/10.1038/s41467-019-12566-y.

Liu, Q., Grassotti, C., Garrett, K., Liu, S., Boukabara, S., 2016. Microwave Integrated Retrieval System (MIRS) User Manual. NOAA/NESDIS/STAR, College Park, Maryland.

Liu, Y., Peters-Lidard, C.D., Kumar, S., Foster, J.L., Shaw, M., Tian, Y., Fall, G.M., 2013. Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska. Adv. Water Resour. 54, 208–227. https://doi.org/10.1016/j.advwatres.2013.02.005.

Liu, Y., Peters-Lidard, C.D., Kumar, S.V., Arsenault, K.R., Mocko, D.M., 2015. Blending satellite-based snow depth products with in situ observations for streamflow predictions in the

Upper Colorado River Basin. Water Resour. Res. 51 (2), 1182–1202. https://doi.org/10.1002/2014WR016606.

Liu, Y.Y., Parinussa, R.M., Dorigo, W.A., De Jeu, R.A.M., Wagner, W., van Dijk, A.I.J.M., McCabe, M.F., Evans, J.P., 2011. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. Hydrol. Earth Syst. Sc. 15 (2), 425–436. https://doi.org/10.5194/hess-15-425-2011.

Lv, Z., Pomeroy, J.W., Fang, X., 2019. Evaluation of SNODAS snow water equivalent in western Canada and assimilation into a Cold Region Hydrological Model. Water Resour. Res. 55 (12), 11166–11187. https://doi.org/10.1029/2019WR025333.

Mätzler, C., Standley, A., 2000. Technical note: Relief effects for passive microwave remote sensing. Int. J. Remote Sens. 21 (12), 2403–2412. https://doi.org/10.1080/01431160050030538.

Meromy, L., Molotch, N.P., Link, T.E., Fassnacht, S.R., Rice, R., 2013. Subgrid variability of snow water equivalent at operational snow stations in the western USA. Hydrol. Process. 27 (17), 2383–2400. https://doi.org/10.1002/hyp.9355.

Mitchell, K.E., Lohmann, D., Houser, P.R., Wood, E.F., Schaake, J.C., Robock, A., Cosgrove, B.A., Sheffield, J., Duan, Q., Luo, L., Higgins, R.W., Pinker, R.T., Tarpley, J.D., Lettenmaier, D.P., Marshall, C.H., Entin, J.K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B.H., Bailey, A.A., 2004. The multi-institution North American Land Data Assimilation System (NLDAS):

Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. J. Geophys. Res. 109, D07S90. https://doi.org/10.1029/2003JD003823.

Mortimer, C., Mudryk, L., Derksen, C., Luojus, K., Brown, R., Kelly, R., Tedesco, M., 2020. Evaluation of long-term Northern Hemisphere snow water equivalent products. The Cryosphere 14 (5), 1579–1594. https://doi.org/10.5194/tc-14-1579-2020.

Mudryk, L.R., Derksen, C., Kushner, P.J., Brown, R., 2015. Characterization of Northern Hemisphere snow water equivalent datasets, 1981–2010. J. Climate 28 (20), 8037–8051. https://doi.org/10.1175/JCLI-D-15-0229.1.

Myneni, R., Knyazikhin, Y., Park, T., 2015. MOD15A2 MODIS/Terra Leaf Area Index/FPAR 8-Day L4 Global 1km SIN Grid. NASA LP DAAC. https://doi.org/10.5067/MODIS/MOD15A2.006.

Nolin, A.W., 2010. Recent advances in remote sensing of seasonal snow. J. Glaciol. 56 (200), 1141–1150. https://doi.org/10.3189/002214311796406077.

Painter, T.H., Berisford, D.F., Boardman, J.W., Bormann, K.J., Deems, J.S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S.M., Seidel, F.C., Winstral, A., 2016. The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. Remote Sens. Environ. 184, 139–152. https://doi.org/10.1016/j.rse.2016.06.018.

Pulliainen, J., 2006. Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations. Remote Sens. Environ. 101 (2), 257–269. https://doi.org/10.1016/j.rse.2006.01.002.

Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T., Norberg, J., 2020. Patterns and trends of Northern Hemisphere snow mass from 1980 to 2018. Nature 581 (7808), 294–298. https://doi.org/10.1038/s41586-020-2258-0.

Ramsay, B.H., 1998. The interactive multisensor snow and ice mapping system. Hydrol. Process. 12 (10–11), 1537–1546. https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11<1537::AID-HYP679>3.0.CO;2-A.

Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The global land data assimilation system. B. Am. Meteorol. Soc. 85 (3), 381–394. https://doi.org/10.1175/BAMS-85-3-381.

Rutter, N., Essery, R., Pomeroy, J., Altimir, N., Andreadis, K., Baker, I., Barr, A., Bartlett, P., Boone, A., Deng, H., Douville, H., Dutra, E., Elder, K., Ellis, C., Feng, X., Gelfan, A., Goodbody, A., Gusev, Y., Gustafsson, D., Hellström, R., Hirabayashi, Y., Hirota, T., Jonas, T., Koren, V., Kuragina, A., Lettenmaier, D., Li, W., Luce, C., Martin, E., Nasonova, O., Pumpanen, J., Pyles, R.D., Samuelsson, P., Sandells, M., Schädler, G., Shmakin, A., Smirnova, T.G., Stähli, M., Stöckli, R., Strasser, U., Su, H., Suzuki, K., Takata, K., Tanaka, K., Thompson, E., Vesala, T., Viterbo, P., Wiltshire, A., Xia, K., Xue, Y., Yamazaki, T., 2009. Evaluation of forest snow processes models (SnowMIP2). J. Geophys. Res. Atmos. 114 (D6), D06111. https://doi.org/10.1029/2008JD011063.

Seo, D.-J., Breidenbach, J., Fulton, R., Miller, D., O'Bannon, T., 2000. Real-time adjustment of range-dependent biases in WSR-88D rainfall estimates due to nonuniform vertical profile of reflectivity. J. Hydrometeorol. 1 (3), 222–240. https://doi.org/10.1175/1525-7541(2000)001<0222:RTAORD>2.0.CO;2.

Seo, D.-J., Breidenbach, J.P., 2002. Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. J. Hydrometeorol. 3 (2), 93–111. https://doi.org/10.1175/1525-7541(2002)003<0093:RTCOSN>2.0.CO;2.

Serreze, M.C., Clark, M.P., Armstrong, R.L., McGinnis, D.A., Pulwarty, R.S., 1999. Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. Water Resour. Res. 35 (7), 2145–2160. https://doi.org/10.1029/1999WR900090.

Sturm, M., Holmgren, J., Liston, G.E., 1995. A seasonal snow cover classification system for local to global applications. J. Climate 8 (5), 1261–1283. https://doi.org/10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2.

Sturm, M., Taras, B., Liston, G.E., Derksen, C., Jonas, T., Lea, J., 2010. Estimating snow water equivalent using snow depth data and climate classes. J. Hydrometeorol. 11 (6), 1380–1394. https://doi.org/10.1175/2010JHM1202.1.

Sturm, M., 2015. White water: Fifty years of snow research in WRR and the outlook for the future. Water Resour. Res. 51 (7), 4948–4965. https://doi.org/10.1002/2015WR017242.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J., Koskinen, J., Bojkov, B., 2011. Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. Remote Sens. Environ. 115 (12), 3517–3529. https://doi.org/10.1016/j.rse.2011.08.014.

Tedesco, M., Narvekar, P.S., 2010. Assessment of the NASA AMSR-E SWE product. IEEE J.-STARS 3 (1), 141–159. https://doi.org/10.1109/JSTARS.2010.2040462.

Tuttle, S.E., Jacobs, J.M., Vuyovich, C.M., Olheiser, C., Cho, E., 2018. Intercomparison of snow water equivalent observations in the Northern Great Plains. Hydrol. Process. 32 (6), 817–829. https://doi.org/10.1002/hyp.11459.

Vuyovich, C.M., Jacobs, J.M., Daly, S.F., 2014. Comparison of passive microwave and modeled estimates of total watershed SWE in the continental United States. Water Resour. Res. 50 (11), 9088–9102. https://doi.org/10.1002/2013WR014734.

Wang, Y., Huang, X., Wang, J., Zhou, M., Liang, T., 2019. AMSR2 snow depth downscaling algorithm based on a multifactor approach over the Tibetan Plateau, China. Remote Sens. Environ. 231, 111268. https://doi.org/10.1016/j.rse.2019.111268.

Weng, F., Yan, B., Grody, N.C., 2001. A microwave land emissivity model. J. Geophys. Res. Atmos. 106 (D17), 20,115–20,123. https://doi.org/10.1029/2001JD900019.

Weng, F., Zou, X., Wang, X., Yang, S., Goldberg, M.D., 2012. Introduction to Suomi national polar-orbiting partnership advanced technology microwave sounder for numerical weather prediction and tropical cyclone applications. J. Geophys. Res. Atmos. 117, D19112. https://doi.org/10.1029/2012JD018144.

Wrzesien, M.L., Durand, M.T., Pavelsky, T.M., Kapnick, S.B., Zhang, Y., Guo, J., Shum, C.K., 2018. A new estimate of North American mountain snow accumulation from regional climate model simulations. Geophys. Res. Lett. 45 (3), 1423–1432. https://doi.org/10.1002/2017GL076664.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., Mocko, D., 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. J. Geophys. Res. 117, D03109. https://doi.org/10.1029/2011JD016048.

Xie, P., Xiong, A., 2011. A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. J. Geophys. Res. Atmos. 116 (D21), D21106. https://doi.org/10.1029/2011JD016118.

Zhang, R., Liang, T., Feng, Q., Huang, X., Wang, W., Xie, H., Guo, J., 2017. Evaluation and adjustment of the AMSR2 snow depth algorithm for the Northern Xinjiang region, China. IEEE J.-STARS 10 (9), 3892–3903. https://doi.org/10.1109/JSTARS.2016.2620521.

List of Figure Captions

Figure 1. Elevation and spatial distribution of SNOTEL and COOP stations in the 18 hydrologic units over the CONUS.

Figure 2. Flowchart of the ATMS SWE retrieval algorithm.

Figure 3. The algorithm framework for blending satellite retrievals with in situ observations.

Figure 4. Geographic distribution of multi-year (i.e., water years 2017–2019) mean daily SWE in different months. Text at the top of each subfigure shows the maximum daily SWE in that month.

Figure 5. False alarm ratio (FAR) versus probability of detection (POD) for the snow cover of ATMS (red), AMSR2 (blue), and SNODAS (green) against IMS analysis in different months over ten snow covered HUs from the western to the eastern US. FAR and POD are calculated based on the daily snow cover of the water years 2017–2019. Numbers in the circles show the months. The closer a circle is to the upper left corner, the better it estimates.

Figure 6. Time series of multi-station mean daily SWE for the main snow seasons of the water years 2017–2019 (different water years are separated by gray dashed lines) in ten snow covered HUs. SNODAS, AMSR2, ATMS, and in situ data are denoted by green, blue, red, and black colors, respectively. CC represents correlation and BIAS represents bias.

Figure 7. Time series of multi-station mean daily SWE for the main snow seasons of the water years 2017–2019 (different water years are separated by gray dashed lines) in HU 10. AMSR2 with constant snow density, AMSR2, ATMS, and in situ data are denoted by green, blue, red, and black colors, respectively. CC represents correlation and BIAS represents bias.

Figure 8. Box plots for correlation and bias of the daily SWE (November–June for the water years 2017–2019) stratified by station elevation, grid elevation range, mean green vegetation fraction, and mean snow depth. ATMS, AMSR2, and SNODAS are represented by red, blue, and green boxes, respectively. The number above the box is the total valid number of the statistic for each class.

Figure 9. Box-percentile plots for (a) correlation, (b) bias, (c) RMSE, and (d) unbiased RMSE of the different experiments in Table 1 and the SNODAS product. Exp1 and Exp2 are experiments for the raw ATMS and AMSR2 SWE, respectively. Exp3 to Exp9 are different blending experiments with 10-fold cross-validation scheme. In each box, the black dot represents the mean

value and the white lines from top to bottom represent 75%, 50%, and 25% percentiles, respectively. The width of the box shows the distribution of the data.

Figure 10. Box-percentile plots for (a) correlation, (b) bias, (c) RMSE, and (d) unbiased RMSE of Exp8 and Exp9 with different cross-validation schemes. In each box, the black dot represents the mean value and the white lines from top to bottom represent 75%, 50%, and 25% percentiles, respectively. The width of the box shows the distribution of the data.

Figure 11. Time series of multi-station mean daily SWE for the main snow seasons of the water years 2017–2019 (different water years are separated by gray dashed lines) in ten snow covered HUs. Exp8 (red color) and Exp9 (blue color) are blending experiments as shown in Table 1 with 10-fold cross-validation. SNODAS and in situ data are denoted by green color and black color, respectively. CC represents correlation and BIAS represents bias.

Figure 12. Geographic distribution of multi-year (i.e., water years 2017–2019) mean daily SWE in different months.