Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR023400

Special Section:
Advances in remote sensing, measurement, and simulation of seasonal snow

Particle Filter Data Assimilation of Monthly Snow Depth Observations Improves Estimation of Snow Density and SWE

Eric J. Smyth[1], Mark S. Raleigh[1,2,3], and Eric E. Small[1]

[1]Department of Geological Sciences, University of Colorado Boulder, Boulder, CO, USA, [2]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, Boulder, CO, USA, [3]National Snow and Ice Data Center (NSIDC), University of Colorado Boulder, Boulder, CO, USA

Key Points:
- Particle filter assimilation of monthly snow depth observations, possible with many techniques, improved modeled snow density and SWE
- The particle filter yielded greater gains when using nonlocal (NLDAS) meteorological forcing than with high-quality station data
- Greatest improvements occurred when ensemble generation was designed to address sources of open loop model errors

Supporting Information:
- Supporting Information S1

Correspondence to:
E. J. Smyth,
eric.smyth@colorado.edu

Citation:
Smyth, E. J., Raleigh, M. S., & Small, E. E. (2019). Particle filter data assimilation of monthly snow depth observations improves estimation of snow density and SWE. *Water Resources Research*, *55*. https://doi.org/10.1029/2018WR023400

Received 1 JUN 2018
Accepted 4 JAN 2019
Accepted article online 17 JAN 2019

Abstract Snow depth observations and modeled snow density can be combined to calculate snow water equivalent (SWE). In this approach, SWE uncertainty is dominated by snow density uncertainty, which depends on meteorological data quality and process representation (e.g., compaction) in models. We test whether assimilating snow depth observations with the particle filter can improve modeled snow density, thus improving SWE estimated from intermittent depth observations. We model snowpack at Mammoth Mountain (California) over water years 2013–2016, assuming monthly snow depth data (e.g., sampling intervals relevant to lidar or manual surveys) for assimilation, and validate against observed SWE and density. The particle filter reduced density and SWE root-mean-square error by 27% and 28% relative to open loop simulations when using high-quality, point location forcing. Assimilation gains were greater (35% and 51% reduction in density and SWE root-mean-square error) when using coarse-resolution North American Land Data Assimilation System phase 2 meteorology. Ensembles created with both meteorological and compaction perturbations led to the greatest model improvements. Because modeled depth and density were both generally lower than observations, assimilation favored particles with higher precipitation and thus more overburden compaction. This moved depth and density (therefore SWE) closer to observations. In contrast, ensemble generation that varied only compaction parameters degraded performance. These results were supported by synthetic experiments with prescribed error sources. Thus, assimilation of snow depth data from lidar or other techniques can likely improve snow density and SWE derived at the basin scale. However, supplementary in situ observations are valuable to identify primary error sources in simulated snow depth and density.

1. Introduction

Seasonal snowpack serves as an important control on irrigation, drinking water, hydropower, ecological activity, weather, and regional and global climate (Bales et al., 2006; Serreze et al., 1999). Despite this, the spatial and temporal distribution of snow water equivalent (SWE) in mountain basins remains undermonitored with ground observations (Bales et al., 2006). In addition, no satellite remote sensing platform resolves local scale SWE variations across all snow-covered environments, ranging from mountains to polar regions (Dozier, 2011; Nolin, 2010; Sturm, 2015). A path forward is to calculate SWE by measuring spatial snow depth with remote sensing (e.g., airborne lidar, synthetic aperture radar, or photogrammetry) and estimating bulk snow density with models (Figures 1a–1c). This paper describes the use of data assimilation with snow depth observations to improve modeled bulk density and thus SWE.

The recent National Aeronautics and Space Administration Airborne Snow Observatory (ASO) missions have demonstrated that airborne lidar can yield near-real time monitoring of basin-wide snow depth at the meter scale for operational and research applications (Painter et al., 2016). Despite the promise of these high-resolution data, nonnegligible uncertainty in SWE maps persists because (1) there is no parallel advance in remote sensing or ground measurement of snow density across basins, (2) observed density can vary widely across an individual basin (e.g., from 200 to 500 kg/m^3 in Wetlaufer et al., 2016), (3) estimated density can vary over a wide range (e.g., ±30%) due to model selection alone and is further enhanced by uncertainties in meteorology and model parameters (Feng et al., 2008), and (4) uncertainty in modeled snow density often contributes more to the total SWE error than uncertainties in lidar snow depth

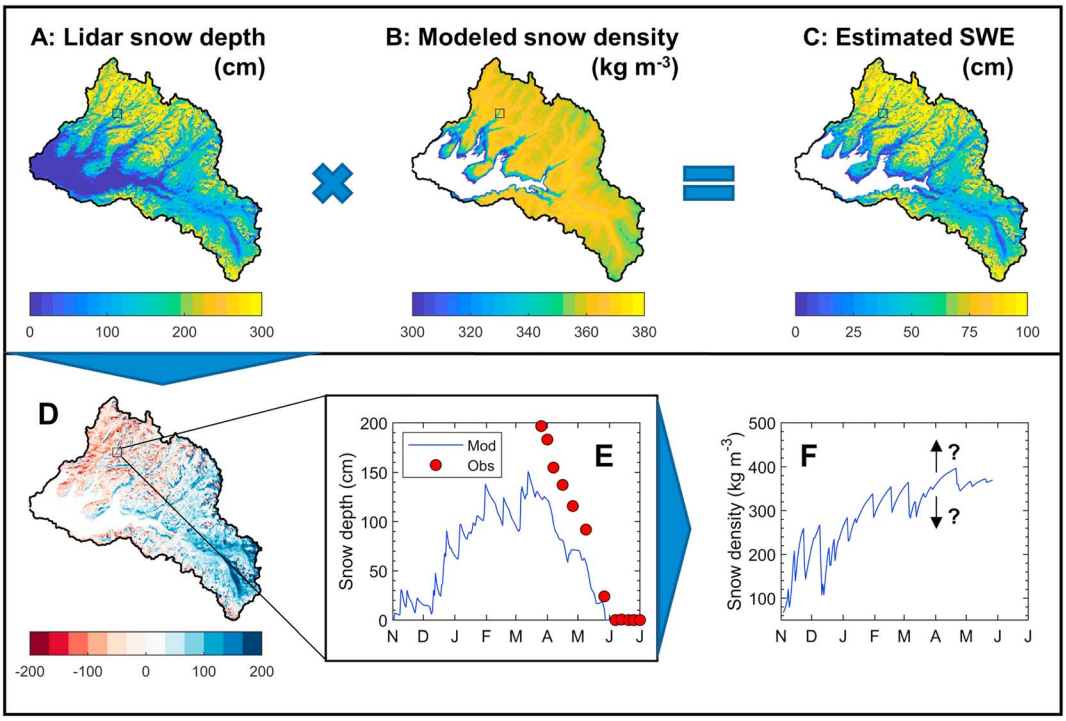

Figure 1. Conceptual diagram demonstrating how (a) lidar snow depth (ASO data from the Tuolumne Basin) can be combined with (b) modeled density to map (c) SWE. Differences between modeled and observed depth in (d) space and (e) time can be used to adjust (f) modeled density (e.g., through assimilation) and therefore SWE. ASO = Airborne Snow Observatory; SWE = snow water equivalent.

(Raleigh & Small, 2017). Given this, we must advance snow model approaches to accurately represent snow density variations in space and time.

One approach to improve density simulated by models is to calibrate parameters via comparison to field observations, such as in situ bulk density data from snow pits or collocated depth and snow pillow sensors. This approach is problematic for several reasons. First, data collection is labor intensive, and thus, observations are sparse in space and time (Elder et al., 2009). Second, snow stations are preferentially located at mid-elevations in areas with low slopes in clearings (e.g., Gleason et al., 2017; Wetlaufer et al., 2016). Third, errors can be hydrologically significant—due to method-to-method differences (on the order of 5–10%), person-to-person differences in skill and procedures, and environmental errors (e.g., snow bridging at snow pillow sites) (Conger & McClung, 2009; Proksch et al., 2016). Fourth, a model tuned for one basin may not produce accurate estimates in another basin, other years, or in real time (e.g., Rutter et al., 2009). In addition, calibrating a density model to combine with snow depth does not allow for quantification of density uncertainty and, by extension, total SWE uncertainty (e.g., Painter et al., 2016).

An alternative approach to improve modeled snow density is to incorporate snow depth observations via data assimilation, which constrains the evolution of model state variables over time. Data assimilation methods explicitly recognize that the true state vector of a system is unobservable but estimate the probability density of the state vector by generating an ensemble of model states. In doing so, assimilation can quantify model uncertainties that arise from various sources, such as model forcing data, parameter choices, and model decisions. Depending on the assimilation scheme, these estimates are updated or weighted by available observations to produce a posterior probability density for state variables of the system. Prior studies have improved snow simulations through data assimilation using observations of snow-covered area (Andreadis & Lettenmaier, 2006; De Lannoy et al., 2012; Girotto et al., 2014; Girotto, Margulis, et al., 2014; Kumar et al., 2015, 2008; Liu et al., 2013; Margulis et al., 2015; Rodell et al., 2004), optical reflectance (Charrois et al., 2016), passive microwave and broadband albedo (Durand & Margulis, 2006, 2007), SWE (De Lannoy et al., 2012; Kumar et al., 2008; Slater & Clark, 2006; Sun et al., 2004), and snow depth (Charrois et al., 2016; Kumar et al., 2015, 2014; Liston et al., 1999; Liu et al., 2013; Magnusson et al., 2017). Some

applications, such as the Snow Data Assimilation System, assimilate multiple variables like snow-covered area, SWE, and snow depth observations (Carroll et al., 2001). Our objective differs from these prior studies in that we focus on improvements in snow density that may be gained via assimilation of the more readily observed snow depth.

There are several ways to assimilate snow depth observations into a model, but the particle filter (PF) has several key advantages over Ensemble Kalman Filter, Kalman Smoother, and direct insertion methods for this application. First, Kalman updates are only appropriate for Gaussian inputs and linear models—while the PF can be used with nonlinear systems (van Leeuwen, 2009). Direct insertion and Kalman updating both change the assimilated state variable directly, which may violate model physics and conservation of mass (Magnusson et al., 2017). PF approaches give weights to individual particles but leave model states untouched. A direct insertion framework also assumes that observations are error free, leading to more severe model corrections than the PF (Magnusson et al., 2017). Lastly, the PF can be more computationally efficient than the Ensemble Kalman Filter and Smoother (Margulis et al., 2015)—an important consideration when designing a data assimilation scheme for all points across a basin (Figure 1) in near real time. We use the PF in this paper to assess improvements in modeled density and SWE (objectives and questions outlined in section 2).

2. Objectives and Approach

Over the past decade, there has been a technological revolution in remotely sensed snow depth measurement, including airborne and terrestrial lidar (Grünewald et al., 2010; Painter et al., 2016), spaceborne and drone-based photogrammetry (Marti et al., 2016; Vander Jagt et al., 2015), and airborne radar (Moller et al., 2017). These data share several common traits: they are spatially extensive, discontinuous through time, and have errors on the order of ~10 cm. Regardless of the measurement technique, model-based estimates of bulk snow density are required to estimate SWE. The overarching question addressed in this paper is: can assimilating intermittent snow depth observations with the PF improve modeled density and therefore SWE? Our focus is on measurements made approximately monthly, similar to the time between ASO lidar surveys in California, thus making the results relevant to the ongoing ASO campaign. Using a single snow study site, we resample hourly observations of snow depth to monthly resolution to match the sampling interval achievable with remote sensing or manual snow surveys.

Snow depth observations can be used to identify errors in simulated snow depth (Figures 1d and 1e). Snow depth and density are linked through a suite of physical processes (e.g., overburden compaction), and hence, snow depth errors can provide useful information on model density errors. Based on the magnitude and direction of depth errors, data assimilation could improve modeled snow density (Figure 1f). When these improved density estimates are combined with high-quality depth observations, the resulting estimates of SWE will similarly be improved. Magnusson et al. (2017) used daily depth data to weight model simulations to improve forecasted SWE and snowmelt. However, they did not explicitly consider the influence of assimilation on simulated density nor did they consider data with a longer time between successive observations, typical of current remote sensing platforms (e.g., every month). Hedrick et al. (2018) used a direct insertion assimilation technique with ASO lidar snow depth but did not examine whether modeled density was improved with assimilation.

The plan for this paper is as follows. In section 3, we describe the study site and data. Then, in the methods section (section 4), we describe the snow model and data assimilation approach. We complete a series of experiments and describe the results in section 5. First, we use local meteorological station forcing and compare the density and SWE improvements resulting from different ensemble generation methods: meteorological forcing perturbations only, compaction parameter perturbations only, and a combination of both. This analysis quantifies the gains associated with different ensemble generation decisions. Second, we repeat these experiments but use North American Land Data Assimilation System (NLDAS) forcing. This step more closely replicates how remotely sensed depth observations would be used in practice, as meteorological forcing is not available throughout a basin. Third, we complete a series of synthetic experiments with prescribed error sources. These experiments help explain the success and failure of different possible ensemble generation choices. Although we do not use spatially distributed data (e.g., lidar) in this paper,

in the discussion (section 6) we describe how our results can be used to improve estimates of SWE by combining spatially extensive snow depth observations with data assimilation and modeled snow density.

3. Study Site and Data

We conduct experiments at the Cold Regions Research and Engineering Laboratory and University of California—Santa Barbara Energy Site (CUES, 37.643 N, 119.029 W, 2,940 m). CUES data are publicly available and described by Bair et al. (2018). CUES sits on a small plateau halfway up Mammoth Mountain, CA.

The data set includes hourly meteorological and energy balance measurements for 2011 to 2017. The California Department of Water Resources installed a snow pillow in September 2012, which provides hourly SWE data for validation. For assimilation, we use snow depth measured hourly with an ultrasonic sensor located directly above the snow pillow but resampled at a 1-month interval (1 November to 1 May) as a proxy for typical intermittent remotely sensed depth data (described further in section 4.3). Hourly bulk snow density is derived by dividing the snow pillow SWE by the ultrasonic snow depth (and then normalized by liquid water density). Precipitation measurements are unreliable, as the CUES site is typically too windy for satisfactory catch efficiencies (Bair et al., 2018). We use hourly precipitation from the nearby (3.5 km) Mammoth Pass site (CDEC code MHP, 2,835 m) which is less wind affected and receives representative precipitation amounts for the area.

Extremely high precipitation buried the depth sensor at MHP from mid-February to May 2017, so we restrict our modeling to water years 2013–2016. The 2015 water year was historically dry and produced the lowest recorded peak snow depth at CUES. Years 2013 and 2014 were also below average in terms of snow accumulation, and 2016 was more representative of the area's historical precipitation. For details about measurement methods, data control, and sensor specifications, see Bair et al. (2018).

In typical spatial applications, the high-quality meteorological forcing data measured at CUES are not available throughout a basin. To assess the feasibility of using a spatially distributed data set, we also force our model with hourly 1/8° data from the NLDAS-2 (hereafter called NLDAS) grid cell that contains CUES (Xia et al., 2012). These inputs more closely replicate a current paradigm in estimating SWE: combining remotely sensed with modeled density using nonlocal meteorological forcing. We did not conduct any downscaling operations (see section 6). For reference, the elevation of the coincident NLDAS grid cell is approximately 200 m lower than the CUES site.

4. Methods and Data Assimilation Experiments

4.1. Snow Model Description

For this study we use the physical snowmelt model Snobal (Marks et al., 1992; Marks & Dozier, 1992). Snobal is a point model that solves the energy and mass budget for a two-layer snowpack, using input data on meteorological forcing, initial snow conditions, and measurement heights. The model has been applied in numerous studies (e.g., Painter et al., 2007, 2016; Pomeroy et al., 2016; Reba et al., 2014; Winstral & Marks, 2014). We chose Snobal because it was developed in the Sierra Nevada and is used by ASO to estimate snow density for lidar-based SWE in this region. The performance of other models in a data assimilation framework is a topic for future work.

In response to ASO's need for robust snow density modeling, the density routine within Snobal was recently updated to include the effects of destructive metamorphism and overburden compaction, based on equations from Anderson (1976; D. Marks, personal communication, December 2017). The model indexes the partition between rain and snow and the density of new snow from a lookup table, based on the precipitation (dew point) temperature. Bulk density is also adjusted for liquid water present in the snowpack at each time step. Like most snow models, Snobal does not account for wind compaction effects on snow density. Model inputs were prepared using the standard Image Processing Workbench tools.

4.2. The PF

A complete description of a state vector of interest (e.g., snow depth) is given by its probability density $p(x_k)$, where x_k is the "true" state vector at time step k. However, in practice, x_k is unobservable—we may only have an imperfect observation (z_k) and/or model estimate (x^i_k). The goal of the PF is to estimate the probability

density of the true state vector over time by generating an ensemble of model states, represented by a collection of individual model estimates ("particles") and constrained by observations.

Model runs are varied to reflect the model's main uncertainty sources. For example, each particle may be generated with different meteorological forcing. These particles are advanced in parallel between time steps $k-1$ and k. We assume that the ith particle (x^i_k) is generated from the probability distribution of x_k, and all the particles will define the shape of the distribution. In other words, the particles create a histogram, with particles falling into bins in the x_k dimension. More particles in a bin indicate a higher probability within the distribution. If a particle is far from an observation at time step k, it should have minimal impact on the probability distribution. This concept is formalized by calculating the particle's weight

$$w^i{}_k = w^i{}_{k-1}\, p\left(z_k|x^i{}_k\right) \tag{1}$$

where the weight on particle i at time step k is equal to its weight at the previous time step, multiplied by a likelihood function, $p\left(z_k|x^i{}_k\right)$. The likelihood describes the probability of observing z_k given the particle estimate $x^i{}_k$ (lower if z_k and $x^i{}_k$ are far apart). The function is assumed to be normally distributed

$$p\left(z_k|x^i{}_k\right) = \frac{1}{\sqrt{(2\pi)^N|C_v|}} e^{\left[-0.5\left(z_k-x^i_k\right)C_v^{-1}\left(z_k-x^i_k\right)\right]} \tag{2}$$

where N is the number of observations and C_v is the measurement error covariance. Weights are normalized to sum to 1. The collection of particles, now weighted by the observation at time step k, approximates the probability density of the true state vector

$$p\left(x_k|z_{1:k}\right) = \sum_{i=1}^{N} w^i_k \delta\left(x_k - x^i_k\right) \tag{3}$$

where δ is the Dirac delta function—which returns zero for all arguments, except for when its argument is zero. Functionally, it creates a histogram for the particle weights with infinitely small bins.

Particle weights refer to weights from previous time steps (equation (1)). Therefore, particles that diverge from several consecutive observations soon develop very low weights, and the statistical information contained in equation (3) becomes less meaningful as the probability distribution is defined by fewer particles. Resampling is one way to overcome this filter degeneracy (van Leeuwen, 2009).

With resampling, particles with very low weights are eliminated, while high weight particles are duplicated and advanced to the next time step. The higher the particle weight, the more duplicates it creates. After resampling, the total number of particles is kept constant at N and weights are reset to $1/N$. Importantly, the PF does not adjust the state variables of individual particles at any point—it only assigns weights (e.g., Magnusson et al., 2017). The particles diverge between time steps (even duplicates) because each particle gets different meteorological forcing, parameter values, etc.

Approaches have varied in defining the PF's single "best estimate" of the state vector over the assimilation period. For example, Magnusson et al. (2017) calculate a weighted average of ensemble members using particle weights, Dong et al. (2015) use the mean of the distribution, and Margulis et al. (2015) use the ensemble median.

For a more comprehensive description of the PF, including its Bayesian motivation and formula derivations, see Arulampalam et al. (2007), Dong et al. (2015), Magnusson et al. (2017), and van Leeuwen (2009). Selected example calculations are in the supporting information.

4.3. Data Assimilation Setup

We propagated 100 particles between assimilation intervals, as Magnusson et al. (2017) demonstrated that fewer than 100 particles are needed to reduce root-mean-square error (RMSE) and improve correlation with the PF. We also generated one "open loop" (OL, i.e., no assimilation) model run for comparison, using the default parameters in Snobal.

Although CUES measures snow depth hourly, we resampled the snow depth data set to monthly to make the assimilation results relevant to a sampling interval possible with remote sensing (e.g., ASO) and field

surveys. A well-instrumented site like CUES has the detailed meteorological and snow data required for evaluation of the PF technique that would otherwise not be available at a location with just snow depth observations. The monthly snow depth data here could represent any type of snow depth data (e.g., stereo imaging and structure-from-motion) that is not continuous through time.

The PF requires an assumption of depth measurement uncertainty. Raleigh and Small (2017) review previously reported lidar snow depth uncertainties, with a median of 8 cm. We set C_v equal to 5 cm (equation (2)). We note that the results are not sensitive to this parameter: tests with C_v values of 2 and 10 cm produced nearly identical results.

To avoid filter degeneracy, we resampled particles at every assimilation time step using a stochastic universal sampling algorithm (supporting information), which Kitagawa (1996) showed to have the lowest sampling noise among several methods. To confirm, we ran the PF over the 2016 water year with four resampling formulas (stochastic universal, probabilistic, residual, and Monte Carlo Metropolis-Hastings algorithms; van Leeuwen, 2009) and did not find substantial differences in the resulting weighted average modeled snow density (no figures shown).

4.3.1. Ensemble Generation and Uncertainties

To represent possible sources of model uncertainty, we generated ensembles of particles by running the model with different precipitation, radiation, compaction parameters, and combinations of the three.

Meteorological measurements are subject to stochastic and systematic errors. To capture potential stochastic precipitation errors, we perturbed each particle's hourly precipitation with additive random noise, from a normal distribution that had a mean of zero and was bound between ±25% of the OL precipitation value. Raleigh et al. (2015) showed that many snow model outputs are more responsive to precipitation biases than random errors. Therefore, we introduced systematic biases through different multiplicative snowfall correction factors (SCF). At every assimilation time step, each SCF was drawn randomly from a uniform distribution between −75% and 300%, following Raleigh et al. (2015). The particles did not retain "memory" of their SCF from previous intervals. This two-step approach to generate a precipitation ensemble is similar to Magnusson et al. (2017).

Radiation is a control on snow temperature, which is a factor that drives destructive metamorphism. It also drives snowmelt, and melt water retained in the snowpack can increase bulk density. Similar to precipitation, we perturbed the hourly radiation input for each particle with additive stochastic noise from a normal distribution—longwave noise bound between ±80 W/m^2 and shortwave noise between ±160 W/m^2. To capture systematic errors, we also applied an additive bias to each particle's full set of radiation inputs: longwave bias bound between ±25 W/m^2 and shortwave between ±100 W/m^2, following Raleigh et al. (2015).

Observations from Kojima (1967) indicate that overburden compaction is a function of the weight of overlying snow and the snowpack's viscosity coefficient. The coefficient can be calculated with a viscosity parameter, which according to Anderson (1976) is determined from observed data. Sturm and Holmgren (1998) report a range of parameter values for different classes of snow, including significant variations within individual classes (e.g., range of 13 cm^3/g for "maritime to alpine" snow). Given this uncertainty, we varied the parameter around Kojima's default value of 21 cm^3/g in a uniform range (±6 cm^3/g) to represent uncertain model parameterization. The snow viscosity parameter has different notation in the literature; here we follow the updated Snobal code notation "C5" (formerly called "C2," D. Marks, personal communication, December 2017).

We chose to focus on overburden compaction (C5) to investigate uncertainty in model density parameterization. Although we did not conduct a full sensitivity analysis with Snobal, offline tests with the model's density equations showed that the C5 parameter explains 90% of variance in snow density simulations. With Snobal, we found minimal model sensitivity to changes in new snowfall density and destructive metamorphism parameters (no figures shown). In the context of how Snobal computes density, focusing on overburden compaction is justified because this process dominates throughout the entire snow year. In contrast, new snowfall events are intermittent and destructive metamorphism acts only at low-to-moderate bulk density that exists early in the snow year.

4.3.2. SWE Calculation, Notation, and Validation

We define the PF best estimate of snowpack density as the weighted average of all particles over time, with weights generated by the PF (e.g., Magnusson et al., 2017). We calculated SWE by multiplying this density

estimate by observed snow depth—thereby minimizing one source of uncertainty (snow depth) and focusing our analysis on improving modeled density. Hereafter this quantity is called SWE_PF. To identify skill improvement, we generated an OL estimate of SWE by multiplying the OL modeled density by the same observed snow depth (SWE_OL). Because both SWE_PF and SWE_OL were based on observed snow depth, any improvements in SWE_PF relative to SWE_OL are derived from improvements in modeled density.

To compare the PF and OL, we calculated RMSE, mean bias error, and the mean absolute error (Willmott, 1982). Although the assimilated snow depth observations were treated as periodic, we calculated these objective functions through all hourly time steps, because these data were readily available and we wanted to characterize performance of the models between observations.

4.4. Experiments

4.4.1. Experiments Using Observed Forcings and Reanalysis Data

In the first set of experiments, we generated the OL model run using the default Snobal parameters and observed forcings at CUES (section 5.1). Then, we compared the performance of the PF with ensembles generated as described in section 4.2. Next, we generated an OL model run using default Snobal parameters and NLDAS forcing and again implemented the PF (section 5.2).

4.4.2. Synthetic Demonstration Scenarios

In reality, forcing data, parameterizations, and observations are imperfect. To better demonstrate the effect of the PF on modeled snow density, we ran a series of synthetic experiments with specific, prescribed errors:

Scenario 1) Here we assumed the model was perfect, and meteorological data were the only sources of uncertainty.
   a. First, we ran the model once for 2013 using observed precipitation and radiation and treated the resulting snow depth, SWE, and density outputs as "truth." Note that this is similar to the OL runs in previous sections, but here we assume that the model outputs are perfect.
   b. Next, we generated a new, synthetic OL run—this time with precipitation scaled by a multiplicative factor across a range of magnitudes (i.e., we introduced a known error).
   c. Finally, we implemented the PF by assimilating snow depth from the truth run (a) every month, with ensembles generated as in section 4.2, to determine whether the PF could adjust for meteorological errors.

Scenario 2) Here we assumed that the forcing inputs were perfect, and the model parameterization was the only source of uncertainty. As in step (a) above, we treated an initial model run as truth. Then we generated a synthetic OL run—with the only difference being a range of different values for the C5 parameter in the model's overburden compaction routine. We again implemented the PF by assimilating snow depth from the truth model run and generated particles as in scenario 4.2.

5. Results

5.1. Results With Model Forcings Observed at CUES

First, we describe PF results with the ensemble of particles constructed by varying meteorological forcings only (section 5.1.1). We then compare these results to those from an ensemble constructed with compaction variations alone (5.1.2) and a combination of the two (5.1.3).

5.1.1. Ensembles With Meteorological Perturbations

To start, we examined the performance of the PF when our ensemble was constructed by perturbing precipitation alone (Figure 2). Year 2013 was a representative water year for our results, when OL depth and density were generally lower than observations throughout the year, and modeled snow melted out too slowly in spring/summer. This is similar to 2014 and 2016 (supporting information). In 2015, OL depth was too low, but density fluctuated below and above observations (supporting information).

Data assimilation improved estimation of snow depth overall, lowering RMSE from 0.36 to 0.27 m (Figure 2a). The depth from the PF closely matched observations through the accumulation season (before peak SWE) and the beginning of the ablation season (post peak SWE). However, modeled snow melted out more slowly than observed during the latter half of the ablation season. With little precipitation during the ablation season, there were few opportunities for the SCF to act upon simulated depth, and the PF ensemble

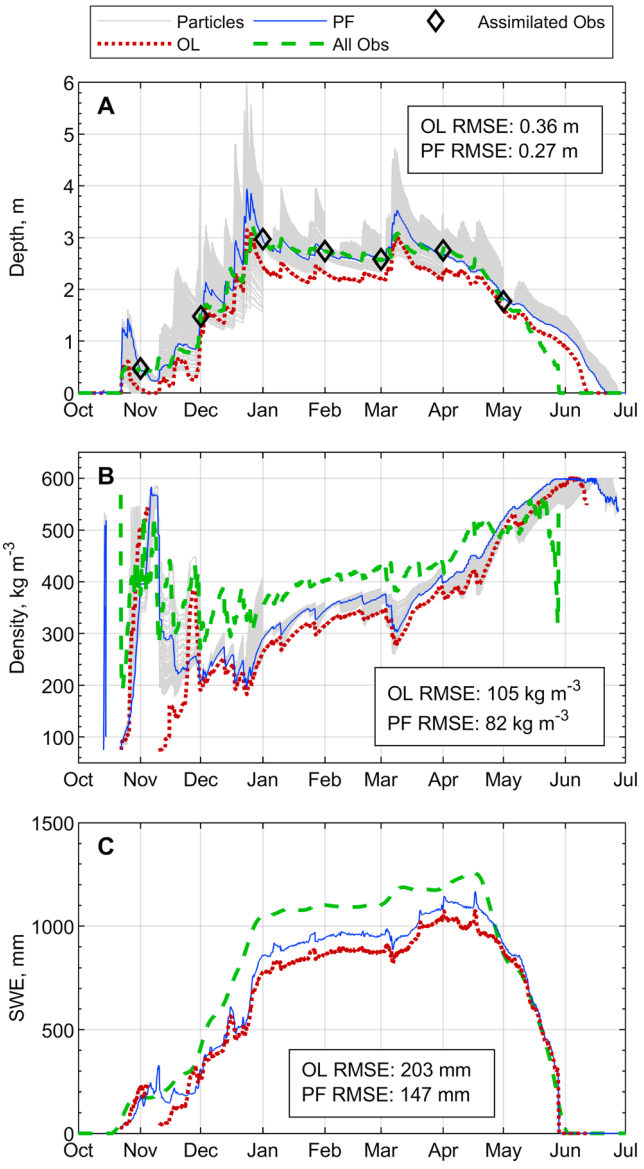

Figure 2. Performance of the PF in WY 2013, where particles are created by perturbing only precipitation forcing. (a) Snow depth, (b), density, and (c) estimated SWE, calculated as modeled density from (b) multiplied by observed (green) snow depth from (a). PF = particle filter; WY = water year; OL = open loop; RMSE = root-mean-square error; SWE = snow water equivalent.

was not wide enough to bring its estimate down (note that gray area does not encompass observations at the end of the season). During this interval, the PF performed worse than the OL in simulating snow depth. Through most of the year, the PF improved depth by favoring particles with higher precipitation. These particles also had higher modeled density, through enhanced overburden compaction. The outcome was lower errors in modeled density—reducing RMSE over 20% from 105 to 82 kg/m^3 (Figure 2b). The greatest improvements occurred during the middle of the water year (January through April).

Although particles did not retain their SCF between assimilation intervals, the highest weighted particles tended to have SCF values of approximately 1.5 (data not shown) throughout the season. As a test, we allowed resampled particles to hold their SCF values across points of assimilation (e.g., Magnusson et al., 2017) but found that the resulting ensemble became too narrow and did not improve modeled density (no figures shown).

OL density in 2013 was generally low compared to observations (Figure 2b), and the PF improved density by increasing it. However, OL density was not too low across all years. In Figure 3, we show the difference between PF and OL density across all years, when OL density was below or above observations. The bars show that when OL density was too low, the PF increased it, and vice versa. Although 2014 and 2016 were similar to 2013, this bidirectionality was important in 2015, when OL density was above and below observations at different parts of the year (supporting information). Figure 3 shows that the PF improved modeled density the most (y axis) and in the right direction when the OL errors were greatest (bins on x axis).

Improvements in modeled density led to better SWE estimates (Figure 2c). SWE RMSE was lowered 28% from 203 to 147 mm in 2013. Similar to density, SWE estimates were improved the most during the middle of the water year (January through April) and otherwise mostly unchanged from the OL. Figure 4b and Table 1 compare RMSE between SWE_OL and SWE_PF for different ensembles and water years. In many cases, improvements in SWE RMSE were greater than improvements in density RMSE on a percentage basis (e.g., 2013). This occurred because the PF improved estimates of density when snow depth was greatest—for example, during February and March of 2013 (Figure 2c).

When constructing an ensemble of particles by perturbing radiation inputs and precipitation, the PF estimates of snow depth and density were very similar in 2013, with RMSE of 0.27 m and 81 kg/m^3, respectively (Figure S2). Overall, the effect of precipitation dominated radiation—the ensembles constructed through variations in both forcings produced estimates that were close to those when precipitation was varied alone (see Figure 4 and Table 1). However, overall SWE RMSE was slightly improved (see Table 1) when perturbing both inputs. Therefore, when we perturb meteorological forcing in all subsequent sections, we vary both precipitation and radiation.

5.1.2. Ensembles With Compaction Perturbations

Next, we constructed our ensemble by varying the C5 compaction parameter alone (i.e., no perturbations to meteorological inputs). Although PF assimilation improved the snow depth simulation (RMSE reduced from 0.36 to 0.24 m), it did so in this case by placing higher weights on particles with lower compaction, which made density estimates worse in 2013 (Figure 5). As a result, estimated SWE had lower errors in the OL (RMSE = 203 mm) than in the PF (RMSE = 304 mm; Figure 5c). The results were similar in other years—Figure 4 shows that when C5 was varied alone, both density and SWE_PF RMSE increased in

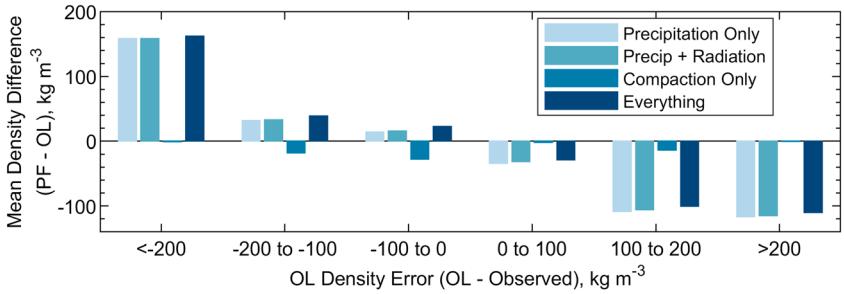

Figure 3. Average difference between PF and OL modeled density for 100 kg/m^3 wide "bins" of OL density error (OL—observed density). Colors indicate different ensemble generation choices. PF = particle filter; OL = open loop.

every year but 2015. On average, the compaction-only ensemble increased SWE RMSE by 36% (Table 1). Figure 3 shows that varying compaction alone often moved density in the wrong direction, for cases with both positive and negative OL density errors.

5.1.3. Ensembles With Meteorological and Compaction Perturbations

In the next experiments, we constructed our ensemble by varying precipitation, radiation, and the C5 compaction parameter together (Figure S6). Again, the PF improved snow depth estimates, lowering RMSE from 0.36 to 0.27 m. The PF increased depth by weighing particles with higher precipitation inputs (which acted to increase snow density), along with particles with lower rates of compaction (which acted to decrease density). The effect of precipitation again dominated, producing results similar to the precipitation-only ensemble (Table 1).

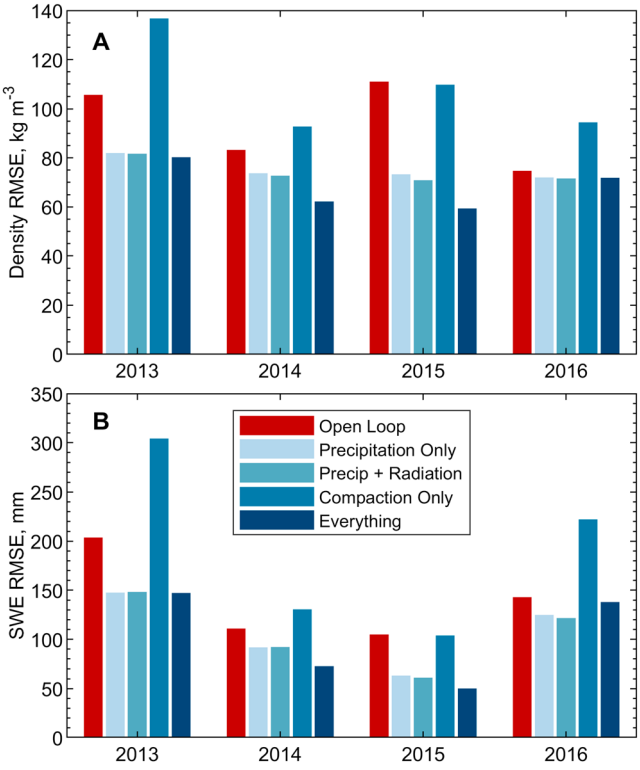

Figure 4. (a) Density and (b) SWE_PF RMSE for different PF ensemble generation choices, compared to the OL. The two charts differ because improvements in density have greater effects on SWE when snow depth is higher. PF = particle filter; OL = open loop; RMSE = root-mean-square error; SWE = snow water equivalent.

In Figure 4b, ensembles where meteorological forcing and compaction were perturbed together had lower errors than the OL in every year and always performed better than the compaction-only ensembles in depth, density, and SWE (Table 1). The compaction and meteorological forcing ensemble moved modeled density in the correct direction for each given OL density error (Figure 3).

5.2. NLDAS Results

To quantify model performance with a widely used, gridded forcing data set, we implemented the PF with NLDAS forcing (no downscaling) and CUES depth observations. NLDAS precipitation was lower than that measured at MHP, and as a result OL depth and density tended to be biased low relative to observations (supporting information; Table 1). As in section 5.1, we conducted experiments with different PF ensembles and computed summary statistics relative to the OL averaged across all water years (bottom of Table 1).

Assimilation with an ensemble constructed by perturbing compaction alone did not improve estimates of SWE: overall SWE RMSE increased by 28% compared to the OL (Table 1). Perturbing the meteorological forcing data alone produced the best results—that ensemble reduced overall SWE RMSE by 51% compared to the OL and brought each objective function in-line with the best results when using higher-quality CUES forcing data. The ensemble with meteorological forcing and compaction that varied together had similar improvements to perturbing meteorological forcing alone—reducing SWE RMSE by 38%.

The meteorological data measured at CUES were better than the NLDAS meteorological forcing. This is shown by the respective OL performances relative to observations: RMSE for SWE_OL was 140 mm when using the CUES data set, whereas RMSE was 179 mm with NLDAS forcing (Table 1). However, Table 1 also shows that the PF was able to

Table 1
Comparison of RMSE, MBE, and MAE for Different Ensembles

Ensemble description	SWE RMSE (mm)			Full season MBE			Full season MAE		
	Full season	Accumulation season	Ablation season	Depth (m)	Density (kg/m^3)	SWE (mm)	Depth (m)	Density (kg/m^3)	SWE (mm)
CUES open loop	140	149	106	−0.36	−25	−71	0.40	81	122
Precipitation only	106	112	82	**0.00**	−31	−62	0.16	62	92
Radiation only	135	144	102	−0.28	−33	−80	0.31	75	118
Precipitation and radiation	105	110	82	**0.00**	−29	−59	**0.15**	61	91
Compaction only	190	203	137	−0.23	−44	−114	0.30	97	164
Meteorology and compaction	**101**	**108**	**73**	0.02	**−23**	**−51**	0.17	**55**	**83**
NLDAS open loop	179	180	157	−0.98	−44	−105	0.98	89	149
Precipitation and radiation	**88**	**92**	**67**	**−0.02**	**−29**	**−53**	**0.18**	**52**	**75**
Compaction only	229	235	183	−0.90	−63	−149	0.90	106	192
Meteorology and compaction	111	114	80	0.03	−41	−74	0.19	64	97

Note. Each metric is averaged across all water years (2013–2016), and SWE RMSE is further broken down to the accumulation and ablation seasons. Input data are from CUES (top) and NLDAS (bottom). The best ensemble for each metric/season is shown in bold emphasis. RMSE = root-mean-square error; MBE = mean bias error; MAE = mean absolute error; CUES = University of California—Santa Barbara Energy Site; SWE = snow water equivalent; NLDAS = North American Land Data Assimilation System.

overcome the relatively poor NLDAS forcing and improve estimates to the extent that error statistics were similar to (and sometimes better than) PF estimates using CUES forcing. Therefore, relative to its OL, there were greater improvements when using NLDAS data. Downscaling NLDAS would have likely improved its OL performance but that was not attempted here.

5.3. Synthetic Demonstration Results

5.3.1. Scenario 1—Meteorological Forcing Uncertainty

The synthetic experiments were designed to demonstrate effects of different ensemble creation methods in systematic, controlled tests. In scenario 1, we assumed a perfect model and all uncertainty was attributed to meteorology. Figure 6 shows a synthetic truth generated with observed precipitation in 2013, and a synthetic OL with precipitation inputs scaled down uniformly by 30%. As a result, OL depth and density were both generally lower than the truth. With an ensemble generated by meteorological input perturbations (Figures 6a and 6c), the PF reduced depth RMSE from 0.56 to 0.04 m by placing higher weights on particles with higher precipitation inputs. These particles also had higher density, and therefore, the PF estimate of snow density increased, reducing RMSE from 22 to 4 kg/m^3.

Using the same synthetic truth and OL, we then generated our ensemble by perturbing the C5 parameter in a uniform range around Snobal's default value of 21 cm^3/g (Figures 6b and 6d). The PF again improved depth estimates (reducing depth RMSE from 0.56 to 0.37 m) but did so by favoring particles with lower compaction. Therefore, these particles estimated even lower snow density, which increased density RMSE from 22 to 61 kg/m^3. SWE RMSE was similarly degraded.

Figure 7a shows overall density RMSE for the OL simulation and three different ensemble creation methods (meteorological perturbations, compaction perturbations, and both) in 2013, when the synthetic OL precipitation inputs were scaled by different amounts. Ensembles with meteorological variations generally reduced density RMSE. In contrast, the ensembles generated with variations in compaction parameters increased density RMSE, with errors increasing with the magnitude of the precipitation error. This demonstrates that the PF cannot compensate for meteorological forcing errors by generating ensembles only with compaction parameters.

5.3.2. Scenario 2—Model Parameterization Uncertainty

In scenario 2, we assumed perfect meteorological data and uncertainty only in model parameterization. In Figure 8, we generated a synthetic truth using the model's default parameter values (e.g., C5 equal to 21 cm^3/g) and a synthetic OL run where C5 was biased low at 17 cm^3/g. As a result, OL depth was lower than truth, and density was higher—the synthetic OL compacted the snowpack too much. With an ensemble constructed by varying the compaction parameter (Figures 8b and 8d), the PF improved depth by placing higher

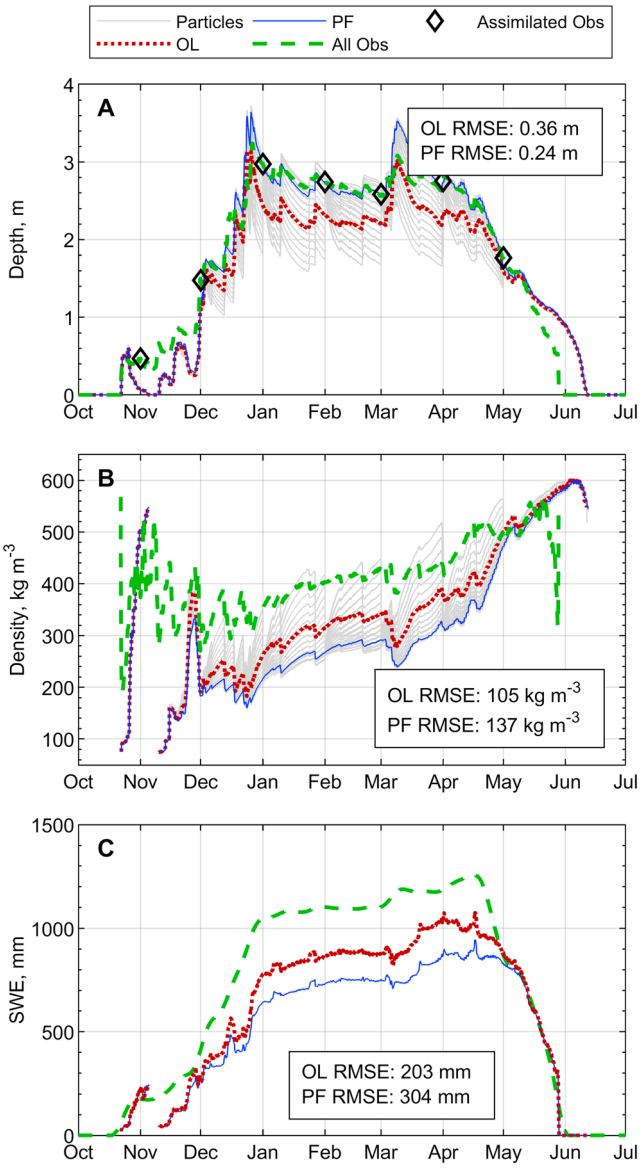

Figure 5. Performance of the PF in WY 2013, where particles are created by perturbing only compaction. (a) Snow depth, (b), density, and (c) estimated SWE, calculated as modeled density from (b) multiplied by observed (green) snow depth from (a). PF = particle filter; WY = water year; OL = open loop; RMSE = root-mean-square error; SWE = snow water equivalent.

weights on particles with lower compaction, reducing RMSE from 0.29 to 0.02 m. These particles also had lower density, which better matched the truth—lowering RMSE from 52 to 5 kg/m^3.

Using the same synthetic truth and OL, we then generated our ensemble by perturbing meteorological inputs (Figures 8a and 8c). The PF improved snow depth estimates by favoring particles with higher precipitation inputs. These particles had higher density estimates, which increased overall density RMSE from 52 to 79 kg/m^3.

We compared density RMSE for the OL and the same ensemble creation methods (Figure 7b), when the synthetic OL C5 value was changed by different amounts (synthetic experiment 2). In all cases, ensembles with compaction variations alone reduced the density RMSE, while ensembles generated with variations in meteorological forcing increased RMSE. The PF could not compensate for compaction parameter errors by changing meteorological inputs.

In both synthetic experiments, the greatest improvement relative to OL was for simulations with the greatest prescribed errors, either positive or negative. There was little to no improvement when prescribed precipitation errors were less than 10% and when C5 values were close to 21 cm^3/g.

6. Discussion

6.1. Ensemble Generation With the PF

The results demonstrate that it is possible to improve simulation of snowpack density and therefore SWE by assimilating snow depth with the PF. However, there is an inherent challenge when assimilating snow depth to improve snowpack density simulation: the PF more strongly weighs ensemble members with depth closest to observations but has no knowledge of density or reasons for depth differences (Figures 1d and 1e). Depending on how the ensemble is constructed, and the reason for differences in observed and modeled snow depth, it is possible for the filter to degrade snow density estimation when assimilating depth. For example, OL modeled depth and density were low relative to CUES observations in 2013. The PF favored particles with higher snow depth estimates—but when the ensemble of particles was generated by varying compaction parameters alone, the PF weighed particles that compacted less, resulting in an even lower (worse) density estimate (Figure 5). In contrast, when the particles were generated with different precipitation inputs, the PF favored those with higher precipitation which, through overburden compaction, also had higher (improved) density.

Based on the synthetic experiments, we conceptualize two cases when assimilating snow depth with the goal of improving density (Figure 9a):

Case 1) When depth and density errors are typically in the same direction: both either too low or too high relative to observations. In this case, creating an ensemble with different precipitation inputs will improve both model states (Figure 9a, white boxes).

Case 2) When depth and density errors are typically in opposite directions. Here creating an ensemble of model runs with different density compaction parameters will improve both model states (Figure 9a, gray boxes).

The 2-D histograms in Figures 9b and 9c show a depth ratio (OL modeled/observed depth for each hour, all years), plotted against a density ratio (OL modeled/observed density) for CUES. The four quadrants correspond to Figure 9a, and the colors indicate the number of data points in each histogram grid cell. In 2013,

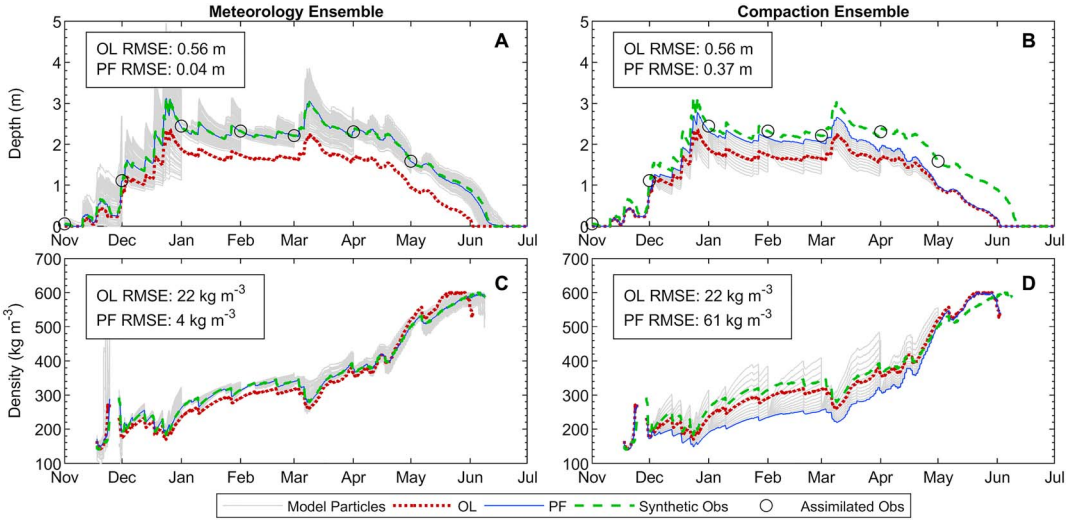

Figure 6. Synthetic experiment, scenario 1, with precipitation scaled down by 30% in the OL in 2013. Shown are modeled snow depth and density with the PF for two ensemble creation methods: (a, c) through perturbed meteorological forcing and (b, d) different compaction parameters. OL = open loop; PF = particle filter; RMSE = root-mean-square error.

86% of all points indicated case 1. In this framework, it makes sense that an ensemble of model runs with different precipitation inputs improved both model states for that year, while an ensemble constructed solely of compaction parameters did not (Figure 4). In 2013, 2014, and 2016, most points fell in case 1 (60% in 2014, 78% in 2016).

Year 2015 was the only year when the compaction-only ensemble did not increase RMSE for density and SWE_PF relative to the OL (Figure 4). We can explain this result based on OL errors, as 62% of the points in 2015 fall in the case 2 quadrants (Figure 9c). This is the only water year where most points indicate case 2. Even though varying compaction parameters improved density and SWE, the best results in 2015 occurred when meteorological forcing and compaction were varied together (Figure 4).

6.2. Operational Feasibility

Our analysis with Figures 9b and 9c is possible because we have both depth and density data for validation. Collocated depth and density are observations that do not exist throughout drainage basins. In practice, we suggest the following.

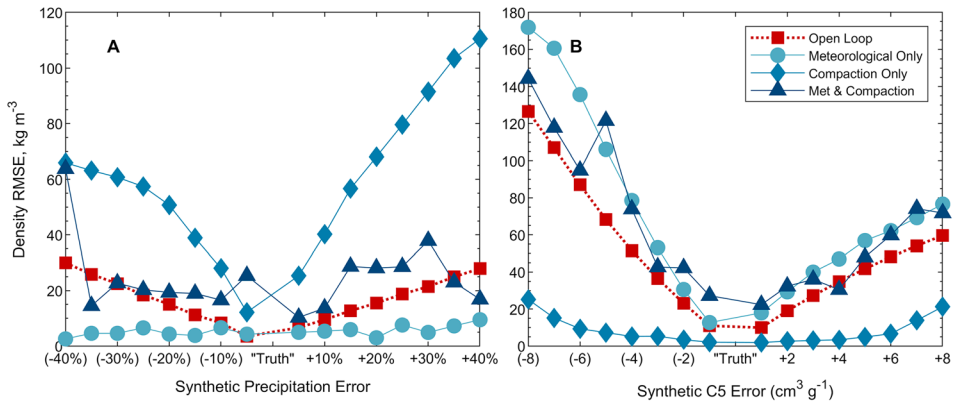

Figure 7. Synthetic experiments: comparison of density RMSE in 2013 for different ensemble creation methods (lines) for (a) scenario 1 (errors in precipitation) and (b) scenario 2 (errors in C5 compaction parameter). In scenario 2, the OL compaction parameter is varied around the "default" value of 21 cm^3/g. OL = open loop; RMSE = root-mean-square error.

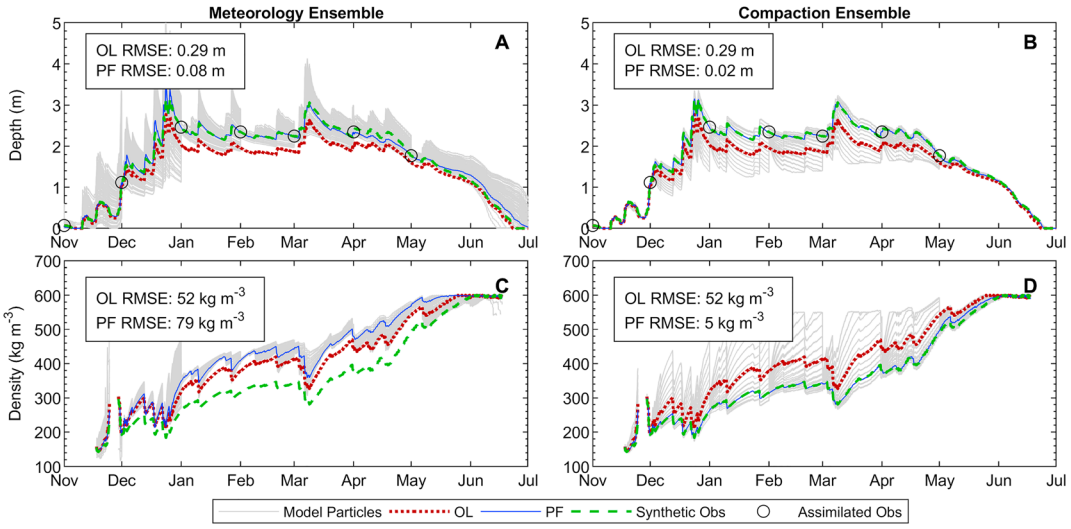

Figure 8. Synthetic experiment, scenario 2, with the synthetic OL C5 parameter value set to 17 cm^3/g in 2013. Shown are modeled snow depth and density with the PF for two ensemble creation methods: (a, c) through perturbed meteorological forcing and (b, d) different compaction parameters. OL = open loop; PF = particle filter; RMSE = root-mean-square error.

1) It may be possible to determine the direction of model depth and density errors and therefore the optimal PF ensemble. Where snow depth observations are available (regardless of the technique), it may be necessary to obtain density measurements from manual snow pits, SNOTEL sites, etc. The depth and density validation data could be used to calibrate a model, but calibration may be difficult if meteorological forcing data are highly uncertain.

2) Varying meteorological forcing and compaction together yields considerable improvement in density and SWE estimates, similar to varying meteorology alone (Table 1). If results at CUES are typical, this suggests that perturbing both meteorology and compaction together is the best approach, especially when the direction of density errors is unknown.

We demonstrate that the PF can overcome errors from coarse-resolution meteorological data (such as NLDAS) without downscaling, and surprisingly, yield error statistics similar to using local observations (Table 1). While downscaling may improve OL performance and may be necessary for other applications

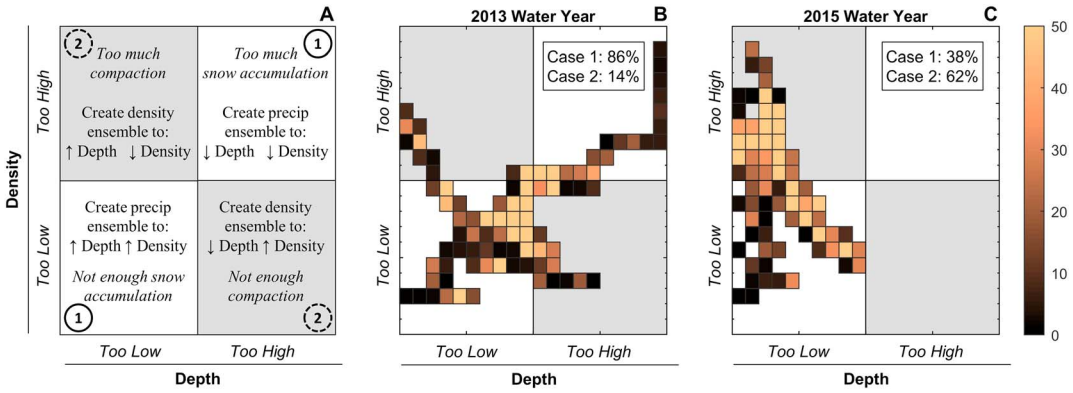

Figure 9. (a) Conceptual diagram of potential model depth and density errors. Points that are "too low" indicate modeled depth or density 0–1 times observed values, and "too high" refers to modeled values 1–2 times observed. In case 1 (white boxes) errors are in the same direction. In case 2 (gray boxes) errors are in opposite directions. Also shown are 2-D histograms of open loop depth (*x* axis) and density (*y* axis) model errors (open loop modeled/observed values) for time steps in (b) 2013 and (c) 2015 water years. Lighter copper colors indicate more points fall within a histogram bin. In 2013 most points indicate case 1, while in 2015 most points indicate case 2.

(e.g., modeling hydrologic processes in the snow-free season), the PF yields the greatest improvements when OL precipitation errors are greatest (Figure 7a), so downscaling operations are likely unnecessary for SWE estimation.

Snobal is an efficient point model which, in our experience, can simulate 1 year on one processor in approximately 3 s. With parallel processing on modern computers, it should be possible to extend our PF framework to a basin in real time, even with snow depth data at a 50-m lidar spatial resolution that produces approximately one million pixels (Figure 1b), if the model simulations are updated each day throughout the snow season. The use of iSnobal (a spatial version of Snobal, more efficient at simulating multiple points) or even faster snow models (e.g., Factorial Snow Model, Essery, 2015) would further enhance practicality.

Snobal was used for illustrative purposes, and we expect similar results from other snow models that have similar density process representations and similar model improvements with the PF when adjusting for precipitation uncertainty (Figure 7a). We did not calibrate the model, focusing instead on OL versus PF performance, and we acknowledge that errors for both are high relative to other papers (e.g., Magnusson et al., 2017) that do use calibrated models. Still, the results are encouraging, indicating that SWE RMSE can be reduced by at least 28% (when using high quality forcing) and as much as 51% (with coarse-resolution forcing).

More work is required to determine the spatial representativeness of our approach. Snow depth is highly variable through space, affected by factors such as canopy interception of snowfall, wind redistribution, aspect, and elevation. While lidar (for example) is a promising platform to resolve this variability, lidar snow depth retrievals have their own sources of uncertainty, especially in areas with dense canopy and/or patchy snowpack. In addition, simulating spatial variations in snow density with models is challenging, particularly in light of spatial variations in forest characteristics. While an analysis at several SNOTEL sites at different elevations and climates may provide useful information, such sites are typically located in flat, open clearings. Future work should implement the PF with spatially distributed, remotely sensed snow depth data —quantifying gains from the approach using spatially extensive snow density observations. The PF is well suited for such spatial applications, as it can account for uncertainty in boundary conditions (e.g., precipitation and radiation), which exhibit high variability across mountainous and forested watersheds.

6.3. Timing of Observations

In this paper, we assume that snow depth observations are collected once per month from November to May, which could realistically represent the timing of remote sensing retrievals or manual snow surveys. As a test, we ran our analysis with different observation timings (e.g., weekly, spring-only, and single observation at peak SWE) and did not find large differences in resulting model estimates. Preliminary results indicate that, while more frequent observations (e.g., weekly) enhance PF estimates of snow depth, more infrequent observations (e.g., monthly) generate greater corrections to modeled density at each assimilation step. Future work should investigate the effects of observation timing (relative to snow storms, for example).

7. Conclusions

Assimilation of infrequent snow depth observations with the PF improves model estimates of snow density and therefore SWE. The improvement is bidirectional: the PF tends to lower estimated snow density when OL density is too high and increases density when the OL density is too low. In both cases, the PF reduces model RMSE most when OL errors are greatest. Overall, the PF improves estimated SWE more than estimated density—because SWE is calculated by multiplying modeled density by observed snow depth, and the PF tends to improve modeled density the most when observed depth is highest.

We generated ensembles by perturbing precipitation, radiation, a model compaction parameter, and combinations of the three. At CUES, the effect of precipitation perturbations dominated the effect of varying radiation—but perturbing both meteorological inputs produced the best estimates of snow depth, density, and SWE. In contrast, introducing compaction variations alone degraded model estimates of density and SWE in three of four years. Varying meteorology together with compaction yielded improvements similar to perturbing meteorology alone.

These results could be explained by the direction of OL depth and density errors in each water year. In 2013, 2014, and 2016, the errors were mostly in the same direction (estimated depth and density both low relative

to observations). In those years, when the PF improved estimated snow depth by weighing particles with higher precipitation inputs, modeled density also increased. In 2015, OL depth and density errors were in opposite directions. This was the only year the compaction-only ensemble improved density and SWE relative to the OL—by favoring particles with lower compaction rates.

Assimilation of snow depth led to greater improvements when the model was forced with coarse-resolution NLDAS data. Similar to Magnusson et al. (2017), this demonstrates the potential to use lidar (or other remotely sensed snow) data with the PF forced by nonlocal meteorological inputs and produces comparable results to assimilation with a model driven by high-quality field station measurements.

The PF is a promising approach for improving density and SWE estimation with remotely sensed snow depth, which may become available from ASO, National Aeronautics and Space Administration SnowEx, or other future missions. The optimal construction of the PF ensemble depends on the cause and sign of model depth and density errors. Future research should examine the representativeness of the results at CUES (e.g., in different climates), potentially including other snow density models.

Acknowledgments
This work was supported by the National Aeronautics and Space Administration (NASA) Terrestrial Hydrology Program under grant NNX17AL41G. We thank Tom Painter, Kat Bormann, and the ASO Team for providing spatial snow depth data (Figure 1). Also, we thank Ned Bair, Robert Davis, and Jeff Dozier for providing the CUES data. We thank Charlie Luce, Ryan Currier, and two anonymous reviewers who provided critical feedback during the peer review process. Thanks to Danny Marks and colleagues for making Snobal code available (https://gitlab.com/ars-snow/ipw). NLDAS-2 forcing data were acquired through NASA GES DISC (https://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing.php).

References

Anderson, E. A. (1976). A point energy and mass balance model of a snow cover. NOAA Technical Report NWS 19, 150.

Andreadis, K. M., & Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, *29*(6), 872–886. https://doi.org/10.1016/j.advwatres.2005.08.004

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2007). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, *50*(2), 723–737. https://doi.org/10.1109/9780470544198.ch73

Bair, E. H., Davis, R. E., & Dozier, J. (2018). Hourly mass and snow energy balance measurements from Mammoth Mountain, CA USA, 2011-2017. *Earth System Science Data*, *10*(1), 549–563. https://doi.org/10.5194/essd-10-549-2018

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., & Dozier, J. (2006). Mountain hydrology of the western United States. *Water Resources Research*, *42*, W08432. https://doi.org/10.1029/2005WR004387

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 U.S. Proceedings of the 69th Annual Meeting of the Western Snow Conference, 1–14.

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., & Picard, G. (2016). On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model. *The Cryosphere*, *10*(3), 1021–1038. https://doi.org/10.5194/tc-10-1021-2016

Conger, S. M., & McClung, D. M. (2009). Comparison of density cutters for snow profile observations. *Journal of Glaciology*, *55*(189), 163–169. https://doi.org/10.3189/002214309788609038

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 Resources Research*, *48*, W01522. https://doi.org/10.1029/2011WR010588

Dong, J., Steele-Dunne, S. C., Judge, J., & van de Giesen, N. (2015). A particle batch smoother for soil moisture estimation using soil temperature observations. *Advances in Water Resources*, *83*, 111–122. https://doi.org/10.1016/j.advwatres.2015.05.017

Dozier, J. (2011). Mountain hydrology, snow color, and the fourth paradigm. *Eos, Transactions American Geophysical Union*, *92*(43), 373–374. https://doi.org/10.1029/2011EO430001

Durand, M., & Margulis, S. A. (2006). Feasibility test of multifrequency radiometric data assimilation to estimate snow water equivalent. *Journal of Hydrometeorology*, *7*(3), 443–457. https://doi.org/10.1175/JHM502.1

Durand, M., & Margulis, S. A. (2007). Correcting first-order errors in snow water equivalent estimates using a multifrequency, multiscale radiometric data assimilation scheme. *Journal of Geophysical Research*, *112*, D13121. https://doi.org/10.1029/2006JD008067

Elder, K., Goodbody, A., Cline, D., Houser, P., Liston, G. E., Mahrt, L., & Rutter, N. (2009). NASA cold land processes experiment (CLPX 2002/03): Ground-based and near-surface meteorological observations. *Journal of Hydrometeorology*, *10*(1), 330–337. https://doi.org/10.1175/2008JHM878.1

Essery (2015). A factorial snowpack model (FSM 1.0). *Geoscientific Model Development*, *5*(1), 219–227. https://doi.org/10.5194/gi-5-219-2016

Feng, X., Sahoo, A., Arsenault, K., Houser, P., Luo, Y., & Troy, T. J. (2008). The impact of snow model complexity at three CLPX sites. *Journal of Hydrometeorology*, *9*(6), 1464–1481. https://doi.org/10.1175/2008JHM860.1

Girotto, M., Cortes, G., Margulis, S. A., & Durand, M. (2014). Examining spatial and temporal variability in snow water equivalent using a 27 year reanalysis: Kern River watershed, Sierra Nevada. *Water Resources Research*, *50*, 6713–6734. https://doi.org/10.1002/2014WR015346

Girotto, M., Margulis, S. A., & Durand, M. (2014). Probabilistic SWE reanalysis as a generalization of deterministic SWE reconstruction techniques. *Hydrological Processes*, *28*(12), 3875–3895. https://doi.org/10.1002/hyp.9887

Gleason, K. E., Nolin, A. W., & Roth, T. R. (2017). Developing a representative snow-monitoring network in a forested mountain watershed. *Hydrology and Earth System Sciences*, *21*(2), 1137–1147. https://doi.org/10.5194/hess-21-1137-2017

Grünewald, T., Schirmer, M., Mott, R., & Lehning, M. (2010). Spatial and temporal variability of snow depth and SWE in a small mountain catchment. *The Cryosphere Discussions*, *4*(1), 1–30. https://doi.org/10.5194/tcd-4-1-2010

Hedrick, A. R., Marks, D., Havens, S., Robertson, M., Johnson, M., Sandusky, M., et al. (2018). Direct insertion of NASA airborne snow observatory-derived snow depth time-series into the iSnobal energy balance snow model. *Water Resources Research*, *54*, 8045–8063. https://doi.org/10.1029/2018WR023190

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, *5*(1), 1–25. https://doi.org/10.1080/10618600.1996.10474692

Kojima, K. (1967). Densification of seasonal snow cover. In *Physics of ice and snow, proceedings of an international conference on low temperature science* (pp. 929–952). Sapporo, Japan: Institute of Low Temperature Science, Hokkaido University.

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. *Journal of Hydrometeorology*, *16*(4), 1736–1741. https://doi.org/10.1175/JHM-D-15-0021.1

Kumar, S. V., Peters-Lidard, C. D., Mocko, D., Reichle, R., Liu, Y., Arsenault, K. R., et al. (2014). Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *Journal of Hydrometeorology*, *15*(6), 2446–2469. https://doi.org/10.1175/JHM-D-13-0132.1

Kumar, S. V., Reichle, R. H., Peters-Lidard, C. D., Koster, R. D., Zhan, X., Crow, W. T., et al. (2008). A land surface data assimilation framework using the land information system: Description and applications. *Advances in Water Resources*, *31*(11), 1419–1432. https://doi.org/10.1016/j.advwatres.2008.01.013

van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Monthly Weather Review*, *137*(12), 4089–4114. https://doi.org/10.1175/2009MWR2835.1

Liston, G. E., Pielke, R. A. Sr., & Greene, E. M. (1999). Improving first-order snow-related deficiencies in a regional climate model. *Journal of Geophysical Research*, *104*, 559–567. https://doi.org/10.1029/1999JD900055

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. *Advances in Water Resources*, *54*, 208–227. https://doi.org/10.1016/j.advwatres.2013.02.005

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

Margulis, S. A., Girotto, M., Cortés, G., & Durand, M. (2015). A particle batch smoother approach to snow water equivalent estimation. *Journal of Hydrometeorology*, *16*(4), 1752–1772. https://doi.org/10.1175/JHM-D-14-0177.1

Marks, D., & Dozier, J. (1992). Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada: 2. Snow cover energy balance. *Water Resources Research*, *28*(11), 3043–3054. https://doi.org/10.1029/92WR01483

Marks, D., Dozier, J., & Davis, R. E. (1992). Climate and energy exchange at the snow surface in the Alpine Region of the Sierra Nevada: 1. Meteorological measurements and monitoring. *Water Resources Research*, *28*(11), 3029–3042. https://doi.org/10.1029/92WR01482

Marti, R., Gascoin, S., Berthier, E., De Pinel, M., Houet, T., & Laffly, D. (2016). Mapping snow depth in open alpine terrain from stereo satellite imagery. *The Cryosphere*, *10*(4), 1361–1380. https://doi.org/10.5194/tc-10-1361-2016

Moller, D., Andreadis, K. M., Bormann, K. J., Hensley, S., & Painter, T. H. (2017). Mapping snow depth from Ka-band interferometry: Proof of concept and comparison with scanning Lidar retrievals. *IEEE Geoscience and Remote Sensing Letters*, *14*(6), 886–890. https://doi.org/10.1109/LGRS.2017.2686398

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

Painter, T. H., Barrett, A. P., Landry, C. C., Neff, J. C., Cassidy, M. P., Lawrence, C. R., et al. (2007). Impact of disturbed desert soils on duration of mountain snow cover. *Geophysical Research Letters*, *34*, L12502. https://doi.org/10.1029/2007GL030284

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., et al. (2016). The airborne snow observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, *184*, 139–152. https://doi.org/10.1016/j.rse.2016.06.018

Pomeroy, J. W., Fang, X., & Marks, D. G. (2016). The cold rain-on-snow event of June 2013 in the Canadian Rockies—Characteristics and diagnosis. *Hydrological Processes*, *30*(17), 2899–2914. https://doi.org/10.1002/hyp.10905

Proksch, M., Löwe, H., & Schneebeli, M. (2016). Density, specific surface area, and correlation length of snow measured by high-resolution penetrometry. *Journal of Geophysical Research: Earth Surface*, *120*, 346–363. https://doi.org/10.1002/2014JF003266

Raleigh, M. S., Lundquist, J. D., & Clark, M. P. (2015). Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrology and Earth System Sciences*, *19*(7), 3153–3179. https://doi.org/10.5194/hess-19-3153-2015

Raleigh, M. S., & Small, E. E. (2017). Snowpack density modeling is the primary source of uncertainty when mapping basin-wide SWE with lidar. *Geophysical Research Letters*, *44*, 3700–3709. https://doi.org/10.1002/2016GL071999

Reba, M. L., Marks, D., Link, T. E., Pomeroy, J., & Winstral, A. (2014). Sensitivity of model parameterizations for simulated latent heat flux at the snow surface for complex mountain sites. *Hydrological Processes*, *28*(3), 868–881. https://doi.org/10.1002/hyp.9619

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, *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., et al. (2009). Evaluation of forest snow processes models (SnowMIP2). *Journal of Geophysical Research*, *114*, D06111. https://doi.org/10.1029/2008JD011063

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 Resources Research*, *35*(7), 2145–2160. https://doi.org/10.1029/1999WR900090

Slater, A. G., & Clark, M. P. (2006). Snow data assimilation via an ensemble Kalman filter. *Journal of Hydrometeorology*, *7*(3), 478–493. https://doi.org/10.1175/JHM505.1

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

Sturm, M., & Holmgren, J. (1998). Differences in compaction behavior of three climate classes of snow. *Annals of Glaciology*, *26*, 125–130. https://doi.org/10.1017/S0260305500014683

Sun, C., Walker, J. P., & Houser, P. R. (2004). A methodology for snow data assimilation in a land surface model. *Journal of Geophysical Research*, *109*, D08108. https://doi.org/10.1029/2003JD003765

Vander Jagt, B., Lucieer, A., Wallace, L., Turner, D., & Durand, M. (2015). Snow depth retrieval with UAS using photogrammetric techniques. *Geosciences*, *5*(3), 264–285. https://doi.org/10.3390/geosciences5030264

Wetlaufer, K., Hendrikx, J., & Marshall, L. (2016). Spatial heterogeneity of snow density and its influence on snow water equivalence estimates in a large mountainous basin. *Hydrology*, *3*(1), 3. https://doi.org/10.3390/hydrology3010003

Willmott, C. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, *63*(11), 1309–1313. https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2

Winstral, A., & Marks, D. (2014). Long-term snow distribution observations in a mountain catchment: Assessing variability, time stability, and the representativeness of an index site. *Water Resources Research*, *50*, 293–305. https://doi.org/10.1002/2012WR013038

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (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. *Journal of Geophysical Research*, *117*, D03109. https://doi.org/10.1029/2011JD016048