AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2016GL071999

Key Points:
- Models portrayed dissimilar distributions of snow density across the Tuolumne basin
- Snow density uncertainty was the primary source of SWE uncertainty where snow depth exceeded 60 cm, covering >70% of the Tuolumne basin
- Snow density uncertainty dominated SWE uncertainty across a range of snowpack conditions at peak accumulation in the western USA

Supporting Information:
- Supporting Information S1
- Data Set S1

Correspondence to:
M. S. Raleigh,
mark.raleigh@colorado.edu

Citation:
Raleigh, M. S., and E. E. Small (2017), Snowpack density modeling is the primary source of uncertainty when mapping basin-wide SWE with lidar, *Geophys. Res. Lett.*, *44*, 3700–3709, doi:10.1002/2016GL071999.

Received 17 NOV 2016
Accepted 27 MAR 2017
Accepted article online 30 MAR 2017
Published online 22 APR 2017

©2017. American Geophysical Union. All Rights Reserved.

Snowpack density modeling is the primary source of uncertainty when mapping basin-wide SWE with lidar

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

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

Abstract Lidar-measured snow depth and model-estimated snow density can be combined to map snow water equivalent (SWE). This approach has the potential to transform research and operations in snow-dominated regions, but sources of uncertainty need quantification. We compared relative uncertainty contributions from lidar depth measurement and density modeling to SWE estimation, utilizing lidar data from the Tuolumne Basin (California). We found a density uncertainty of 0.048 g cm^{-3} by comparing output from four models. For typical lidar depth uncertainty (8 cm), density estimation was the dominant source of SWE uncertainty when snow exceeded 60 cm depth, representing >70% of snow cover and 90% of SWE volume throughout the basin in both 2014 and 2016. Density uncertainty accounts for 75% of the SWE uncertainty for a broader range of snowpack characteristics, as measured at SNOTEL stations throughout the western U.S. Reducing density uncertainty is essential for improved SWE mapping with lidar.

1. Introduction

In many catchments worldwide, seasonal snowpack is an important determinant of the timing and magnitude of water availability for human use and natural ecosystems. A key variable is the spatial distribution of snow water equivalent (SWE, the amount of water in the snowpack), but there is a historical lack of SWE data [*Bales et al*., 2006]. The high spatial variability of snowpack makes extrapolation of available SWE data problematic [*Clark et al*., 2011; *Rice et al*., 2011]. Satellite remote sensing does not provide direct measurements of SWE in all settings [*Nolin*, 2010; *Dozier*, 2011]. Hence, advances in snowpack monitoring are needed to refine understanding of SWE distributions for research and watershed operations [*Bales et al*., 2006; *Viviroli et al*., 2011].

A path for quantifying SWE variations across large basins is through measuring snow depth with airborne lidar and estimating bulk snowpack density with models (e.g., NASA Jet Propulsion Laboratory Airborne Snow Observatory (ASO) [*Painter et al*., 2016]). Airborne lidar can measure submeter variations in depth, with typical vertical uncertainties reported in the 2 to 30 cm range [*DeBeer and Pomeroy,* 2010; *Grünewald et al*., 2010; *Deems et al*., 2013; *Harpold et al*., 2014; *Grünewald and Lehning*, 2015; *Painter et al*., 2016]. Lidar snow depth uncertainty varies with factors such as vegetation, topography, flight characteristics, sensor specifications, and horizontal resolution of the gridded snow depth [*Deems et al*., 2013]. Generally, snow depth uncertainty decreases as lidar data are aggregated to coarser scales. Other depth measurement approaches [*Kinar and Pomeroy*, 2015; *Sturm*, 2015] are available at equivalent or lower uncertainty and are appropriate at different spatiotemporal scales (see Text S1 and Figure S1 in the supporting information) [*Larson et al*., 2009; *Varhola et al*., 2010; *Kerkez et al*., 2012; *Parajka et al*., 2012; *Pohl et al*., 2014; *Vander Jagt et al*., 2015; *Buhler et al*., 2016]. These generally are unable to match the capacity of lidar to map snow depth in space.

In contrast to this technological revolution in snow depth measurement, there has been no concurrent advance in the measurement of snowpack bulk density across space. Snow pit profiles remain the most reliable measurement of bulk density [*Kinar and Pomeroy*, 2015], but measurements can differ by 10% [*Conger and McClung*, 2009; *Proksch et al*., 2016]. For a 100 cm snowpack with 0.30 g cm^{-3} density, this translates to uncertainties of 0.03 g cm^{-3} in density and 10 cm in SWE. Although snow depth varies more spatially than density [*Balk and Elder*, 2000; *López-Moreno et al.,* 2013; *Wetlaufer et al*., 2016], density measurement in snow pits is disproportionately limited in space relative to depth measurement. Intensive field campaigns can sample density at less than 10^2 snow pits per day [*Elder et al*., 2009], whereas airborne lidar systems can sample

snow depth at 10^5 points per second [*Deems et al.*, 2013]. Ground-penetrating radars (GPR) [e.g., *Marshall and Koh*, 2008] can increase spatial sampling of density, but are not reliable in all conditions [*Lundberg et al.*, 2016]. Furthermore, GPR remains a specialized research tool and yields data that are time and labor intensive to postprocess.

A solution is to model snow density across the domain of the snow depth data set [e.g., *Painter et al.*, 2016]. Statistical [*Elder et al.*, 1998; *Wetlaufer et al.*, 2016], empirical [*Jonas et al.*, 2009; *Sturm et al.*, 2010; *Bormann et al.*, 2014; *McCreight and Small*, 2014], and physically based models [*Jordan*, 1991; *Feng et al.*, 2008; *Shi et al.*, 2009; *Painter et al.*, 2016] have been developed and evaluated against observations (see Text S2 for review). While uncertainty depends on the model, location, evaluation period, and metric, density model uncertainty is generally in the 0.04 to 0.10 g cm^{-3} range for root-mean-square differences (RMSD) and the 0.02 to 0.08 g cm^{-3} range for mean absolute differences (MAD). For context, continental snowpack density typically starts near 0.15 g cm^{-3} early in the season and increases to values exceeding 0.35 g cm^{-3} during snowmelt. The ~10% uncertainty in manual density measurements implies that minimum uncertainties in modeled density must be 0.025 g cm^{-3} or greater. Hence, uncertainty in density measurements imposes a fundamental challenge on improving snow density models.

Quantifying and minimizing the uncertainty of SWE estimated from snow depth and density requires identifying the dominant sources of uncertainty. To date, there has been little attention to the relative uncertainty contributions of depth and density. Given that density measurements are limited in number and tend to be biased toward easily accessible locations (e.g., flat forest clearings), the full range of conditions are not usually sampled. Hence, density models are difficult to test everywhere in a basin. A more spatially comprehensive approach to characterize uncertainty in density is to examine variations across multiple models, a common approach in climate and hydrology studies [e.g., *Rodell et al.*, 2011]. This approach is appropriate for SWE estimation from lidar snow depth, as any application of this approach requires selection of a density model (and associated parameters) with limited evaluation data to guide those decisions.

Here we compare uncertainty in lidar snow depth measurement to that from modeled snow density for SWE estimation across basins. We analyze airborne lidar snow depth data from ASO near-peak conditions during two years (2014 and 2016) over the Tuolumne River Basin (California). We consider a range of uncertainties in lidar measurement based on the literature. For tractability, we focus on density model selection rather than uncertainty due to meteorological data [*Raleigh et al.*, 2015, 2016] or model parameters [*Reba et al.*, 2014]. The analysis is relevant to current efforts to map SWE with lidar, including the ASO campaign and the NASA SnowEx experiment.

2. Methodology

2.1. Snowpack Data Sets

We used airborne lidar snow depth and elevation data from ASO [*Painter et al.*, 2016] over the Tuolumne River Basin (California), gridded to a horizontal resolution of 50 m (Figure S2). The stated uncertainty of the 50 m snow depth data is RMSD < 2 cm, which is at the bottom of the range in reported lidar depth uncertainty. The drainage basin has an area of 1180 km^2 and varies in elevation from 1080 m to 3940 m and has been described previously [e.g., *Rice et al.*, 2011; *Lundquist et al.*, 2016].

We examined two ASO data sets for the Tuolumne Basin: 7 April 2014 and 16 April 2016. We selected early April for our analyses because this is near the typical peak in snow accumulation. After filtering out thin snow cover (<5 cm depth), the mean snow depth across the basin was 94 cm on 7 April 2014 and 117 cm on 16 April 2016. In both cases, approximately 80% of the basin had snow deeper than 10 cm. The snow volume during the 2013–2014 winter ranked second lowest in the 30 year period from 1985 to 2015 [*Margulis et al.*, 2016], while the 2015–2016 snowpack was near average (i.e., ~10% below average).

As a measure of snowpack conditions near the 2014 ASO acquisition date, we examined snow density and SWE measured with a federal sampler at seven snow courses across the basin by the California Department of Water Resources (CDWR). The snow courses ranged in elevation from 2042 to 2987 m and were generally flat meadows. Snow course data were taken 1.5 to 2 weeks before the 7 April ASO flight. The CDWR snow course data were not appropriate for density model evaluation—approximately 50 to

150 mm (water equivalent) of new snow fell in the final week of March, followed by compaction and melt. However, these data offered a glimpse of density variations in the basin.

To place the 2 year ASO analysis in the context of a broader range of snowpack conditions, we also examined snow depth and density near-peak accumulation from NRCS SNOTEL data [*Serreze et al*., 1999]. We screened SNOTEL sites with valid SWE and depth data, resulting in 811 sites and $n = 9013$ station years. For each sample with valid data, we first found peak SWE and then divided that value by depth to estimate bulk snowpack density. We excluded cases with depth less than 20 cm at peak SWE.

2.2. Snow Density Models and Uncertainty

Estimating SWE from lidar requires selecting a density model. Every model will yield different density estimates, and thus, model selection introduces uncertainty to the final SWE estimates. To gauge snow density uncertainty due to model selection, we applied two empirical models and two physically based models to simulate spatial variations in density across the basin on the analysis dates. We randomly selected 1000 analysis points in the basin where ASO snow depth was greater than 5 cm and ran all four models at each of these points. These analysis points reasonably represented snow depth across the basin (Figure S2).

We selected the empirical models of *Sturm et al.* [2010] and *Jonas et al.* [2009]. Both models require snow depth (taken from ASO lidar data) and day of year. Additionally, the Sturm model requires seasonal snow climate classification, which was taken from the *Sturm et al.* [1995] map at 0.5° resolution. We classified 83% of the 1000 analysis points in the maritime class and 17% in the alpine class. The Jonas model applies different parameters depending on elevation zone. Because over 99% of the 1000 points were in the Jonas high-elevation zone, we simplified the model by classifying all analysis points as high elevation.

For physically based models, we selected Snobal [*Marks and Dozier*, 1992; *Marks et al*., 1992, 1999] and SHAW [*Flerchinger and Saxton*, 1989a, 1989b]. We selected Snobal for consistency with ASO model selection [*Painter et al*., 2016]. We selected SHAW to include a more detailed multilayer snow model of compaction and densification, in contrast to the two-layer Snobal model. Whereas Snobal uses empirical density-time curves to represent snow densification with compaction and snowmelt [*Sandells et al*., 2012], SHAW utilizes physically based parameterizations similar to *Anderson* [1976]. The models also represent new snowfall density differently: Snobal indexes a look-up table based on dewpoint temperature, while SHAW uses an empirical relationship with wetbulb temperature. We used standard IPW routines for computing albedo and net shortwave radiation as input into Snobal, while SHAW simulated albedo within the model. Although SHAW is more complex, model complexity does not guarantee improved representation of bulk snowpack properties.

Both Snobal and SHAW were run at an hourly time step with the same meteorological forcing data at each of the 1000 study points. To ensure identical forcing between models at each point, we disabled the canopy in SHAW and hence vegetation had no influence on model forcing or outputs. Forcing data originated from the 1/8° gridded NLDAS-2 data set [*Xia et al*., 2012], which we downscaled to the 50 m ASO grid using the forcing preprocessor of the Alpine3D system [*Lehning et al*., 2006; *Bavay and Egger*, 2014] to account for fine-scale topographic influences and gradients (see Text S4 and Figures S3 and S4). By supplying Snobal and SHAW with the same forcing, the differences in density were due to differences in snow model physics and parameters. An auxiliary analysis found that a simplified downscaling of meteorological forcing (i.e., based on lapse rates from PRISM [*Daly et al*., 1994]) had a negligible effect (<0.01 g cm^{-3}) on density differences between models (see Text S4 and Figure S5).

2.3. SWE Uncertainty: Measures and Maps

At each study point, we calculated the "best estimate" of SWE and associated uncertainty (ΔSWE) as

$$SWE \pm \Delta SWE \tag{1}$$

where SWE is the product of the lidar-measured snow depth (H_s) and a multimodel mean estimate of density (ρ_{mean})

$$SWE = H_s \times \rho_{mean} \tag{2}$$

We use delta (Δ) to denote uncertainty of a variable. ΔSWE was calculated in quadrature (i.e., assuming independent errors) based on fractional uncertainties in lidar snow depth and modeled density:

$$\Delta SWE = SWE \times \sqrt{\left(\frac{\Delta H_s}{H_s}\right)^2 + \left(\frac{\Delta \rho}{\rho_{mean}}\right)^2} \quad (3)$$

From (3), we calculated the relative (i.e., percent) contribution of snow depth (f_H) and density (f_ρ) uncertainties to SWE uncertainty at each point. For the physically based models, we utilized lidar-measured snow depth to estimate SWE (equation (2)) and ΔSWE (equation (3)), not the modeled snow depth.

Different metrics of uncertainty (Δ) are used in the literature for both snow depth and density. We used RMSD because it is a commonly reported metric for both parameters. Quantification of lidar snow depth uncertainty (ΔH_s) is still an active area of research and thus we considered uncertainty ranging from 2 to 30 cm. Within this range, we highlighted RMSD values reported in five specific studies, including (i) 2 cm [*Painter et al.*, 2016], (ii) 5 cm [*Grünewald et al.*, 2010], (iii) 8 cm [*Painter et al.*, 2016], (iv) 17 cm [*DeBeer and Pomeroy*, 2010], and (v) 23 cm [*Harpold et al.*, 2014]: the smaller uncertainties are at grid resolutions of 10^1 to 10^2 m and larger uncertainties are at the meter scale. The analysis focused on the median depth uncertainty (8 cm) in this range. We applied snow depth uncertainty uniformly to all points, regardless of snow depth or geophysical characteristics (e.g., elevation and vegetation). For snow density uncertainty ($\Delta\rho$), we examined RMSD across all four models, an approach similar to *Rodell et al.* [2011]. In each case, we treated the mean across all four models as the best estimate of density (ρ_{mean}), similar to studies that examined density measurement uncertainty [*Conger and McClung*, 2009; *Proksch et al.*, 2016].

3. Results

3.1. Spatial Estimates of Snow Density and SWE

We focused on the dry, low-snowpack of 2014, as this represented a lower bound on the effects of density uncertainty on SWE estimation (described below). The four density models yielded contrasting distributions of snowpack density across the basin (Figure 1). The two physically based models produced lower mean density (Snobal = 0.267 g cm^{-3}, SHAW = 0.298 g cm^{-3}) but higher standard deviations in space (0.041 g cm^{-3} for Snobal and 0.033 g cm^{-3} for SHAW). SHAW density was greater than Snobal in part because SHAW had more snowmelt by early April 2014. In contrast, the empirical models had higher mean density (Jonas model = 0.339 g cm^{-3}, Sturm model = 0.381 g cm^{-3}) and lower standard deviation in space (0.005 g cm^{-3} for Jonas and 0.012 g cm^{-3} for Sturm). Because the empirical models were developed for regional applications, they did not explicitly represent processes that influence density at local scales (e.g., radiation variations with slope). Across all models and analysis points, the mean density was 0.321 g cm^{-3} and the mean uncertainty across density models ($\Delta\rho$) was 0.048 g cm^{-3} in terms of RMSD (0.041 g cm^{-3} for MAD). The same level of density uncertainty (0.048 g cm^{-3}) was found in the near-average snowpack of 2016 (Text S5 and Figure S6).

The CDWR snow courses from late March 2014 showed density varying from 0.292 g cm^{-3} to 0.417 g cm^{-3} (Figure 1c), with a mean of 0.351 g cm^{-3} and a standard deviation of 0.043 g cm^{-3}. The empirical models captured the mean CDWR density better while the physically based models captured the CDWR spatial variability better. Given limitations in the CDWR data (e.g., small sample size and differences in acquisition date), these comparisons were only qualitative.

We applied the mean multimodel density and lidar snow depth at each point to estimate SWE magnitude (Figure 1d and equation (2)) and uncertainty (Figure 1e and equation (3)). In this analysis, we assumed a lidar depth uncertainty of 8 cm. Other magnitudes of depth uncertainty are discussed below and in the supporting information (Text S6 and Figures S7 and S8). Mean basin-wide SWE was 30.3 ± 5.6 cm (18.5% relative SWE uncertainty) in 2014. Not surprisingly, the spatial and statistical distributions of SWE were similar to that of snow depth (compare Figure S2 and Figure 1). SWE uncertainty (Figure 1e) exhibited a similar spatial pattern as the magnitude of SWE (Figure 1d), and most locations (86% in 2014, 74% in 2016) had relative SWE uncertainty between 10 and 30%.

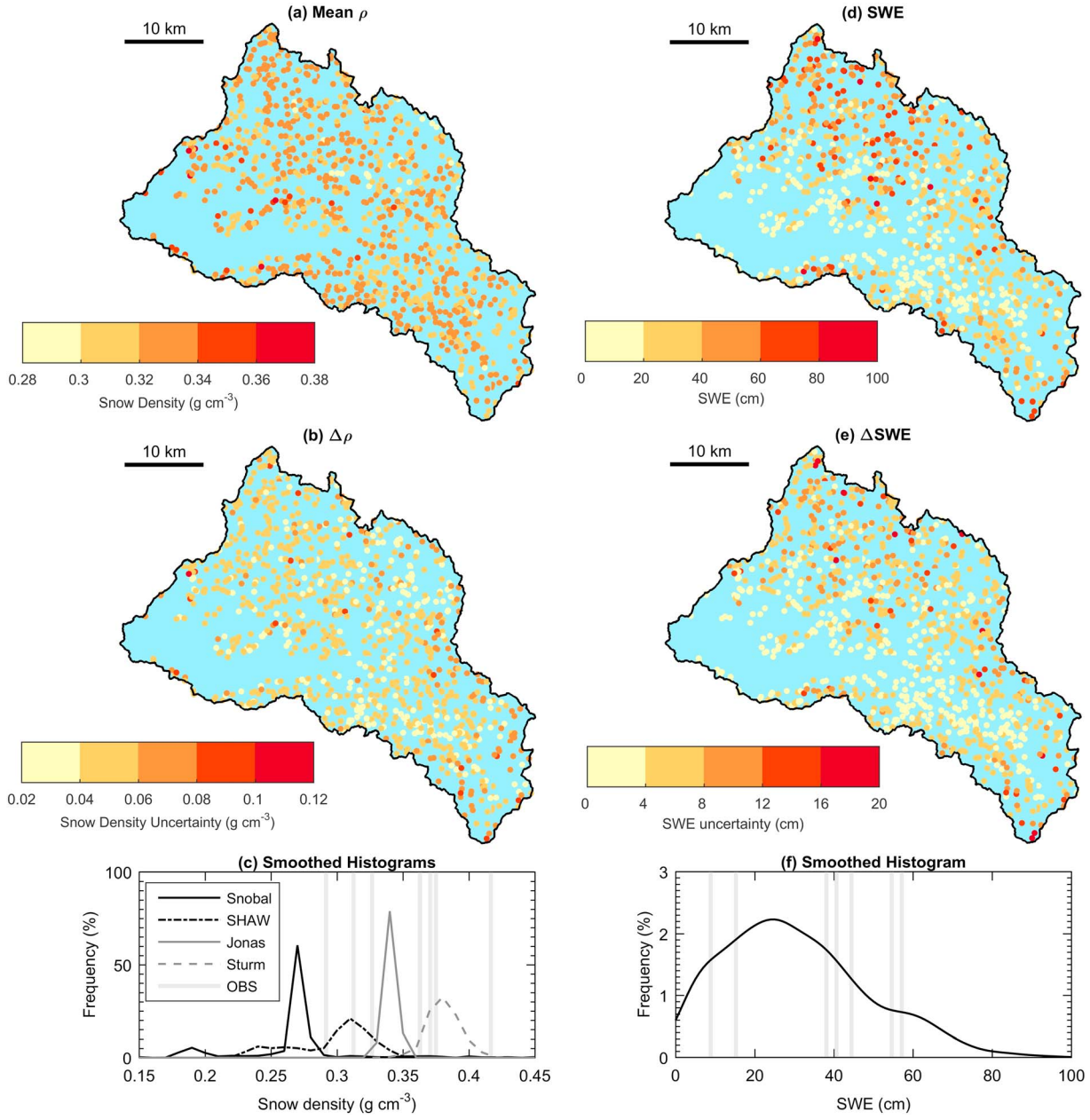

Figure 1. (a) Multimodel mean snowpack density, (b) density model uncertainty, and (c) smoothed histograms for density models on 7 April 2014 at all analysis points (n = 1000) over the Tuolumne River basin. Combining modeled density with lidar snow depth results in maps of (d) mean SWE and (e) SWE uncertainty. (f) A SWE histogram is shown. Both Figures 1c and 1f display CDWR snow course observations (vertical lines) taken at seven locations in the basin over 25–28 March 2014. The histograms bin sizes are 0.01 g cm^{-3} for density (Figure 1c) and 1 cm for SWE (Figure 1f).

3.2. Relative Contributions to SWE Uncertainty

Absolute SWE uncertainty was greater at locations with deeper snow (gray points in Figure 2a). The fraction of ΔSWE due to $\Delta\rho$ was also greater at locations with deeper snow (red points in Figure 2a), and thus the contribution of ΔH_s to ΔSWE diminished with increasing depth (blue points). Figure 2a shows a specific case with ΔH_s = 8 cm at all points and $\Delta\rho$ characterized by the spread of the four models at each point (Figure 1b). We found a crossover point at a snow depth of 60 cm above which $\Delta\rho$ (0.048 g cm^{-3} on average) dominated ΔSWE and below which ΔH_s (8 cm) dominated. Seventy percent of the analysis points in the 2014 analysis and 80% in the 2016 had snow depths exceeding 60 cm and were hence

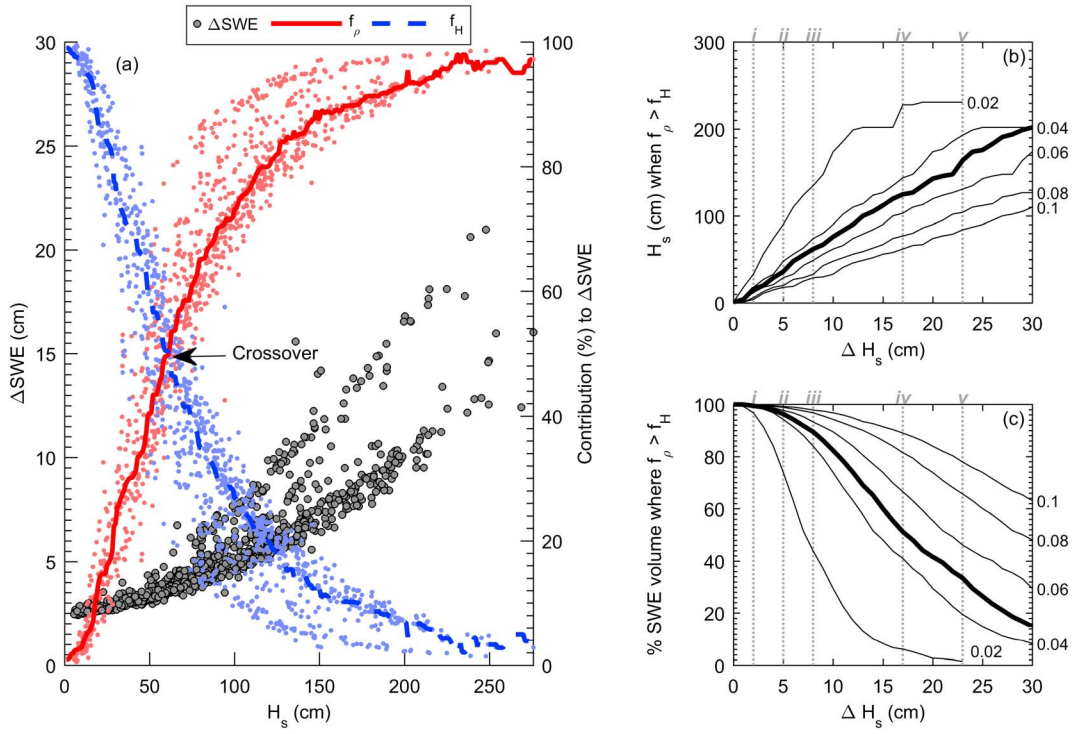

Figure 2. (a) ASO snow depth (H_s) versus uncertainty in estimated SWE (ΔSWE) on 7 April 2014 in the Tuolumne Basin, with relative contributions of uncertainties in modeled density (f_ρ, red line and markers) and lidar snow depth (f_H, blue line and markers) to ΔSWE. ΔSWE and contributions to ΔSWE are calculated in Figure 2a assuming ΔH_s = 8 cm and $\Delta\rho$ as the uncertainty among four models (0.048 g cm^{-3} on average). (b) Crossover snow depth above which f_ρ contributes more than f_H to SWE uncertainty versus different uncertainty levels in snow depth (*x* axis) and density (lines, units = g cm^{-3}). (c) Percentage of SWE volume contained by areas where $f_\rho > f_H$ for different uncertainty levels in snow depth (*x* axis) and density (lines). In Figures 2b and 2c, the bold line shows the density uncertainty in this study (0.048 g cm^{-3}). Five levels of snow depth uncertainty from the literature are traced (see section 2.3).

in the zone where $\Delta\rho$ was dominant. Because the role of $\Delta\rho$ was greater in deeper snowpacks that also store more water (Figure 2a), $\Delta\rho$ was the most important determinant of uncertainty in SWE volume for the basin. For these density and depth uncertainties, density was the dominant source of uncertainty for 90% of the SWE volume basin-wide (Figure 2c). The results were replicated in the 2016 analysis (no figures shown).

Across a wider range of uncertainties in snow depth and density, the crossover point where $f_\rho > f_H$ shifted to lower snow depths with decreasing ΔH_s or increasing $\Delta\rho$ (Figure 2b). This corresponded to changes in the percent of basin SWE volume where $\Delta\rho$ dominated (Figure 2c). For example, ΔH_s = 17 cm had a crossover point of 125 cm snow depth (Figure 2b). Approximately 25% of the basin in 2014 (38% in 2016) had snow depth exceeding 125 cm, and this zone comprised 50% of the SWE volume (both years) in the basin (Figure 2c). If ΔH_s was reduced to 5 cm (e.g., through spatial aggregation), the crossover point reduced to 35 cm snow depth (Figure 2b). About 72% of the basin snowpack was deeper than 35 cm in both years, and this comprised >95% of the basin SWE volume in both years (Figure 2c). As accuracy in snow depth improved, density uncertainty increased in importance.

3.3. Generalization Over a Range of Snowpack Conditions

We compared the relative contributions of ΔH_s and $\Delta\rho$ to ΔSWE across a broader range of snowpack conditions and uncertainty levels (Figures 3, S7, and S8). In Figure 3, the red region denotes conditions where $f_\rho > f_H$, the blue region denotes conditions where $f_\rho < f_H$, and the dashed line is the crossover point (see Figure 2). The gray contours show absolute ΔSWE (cm). Generally, there was a greater range of possible conditions where $\Delta\rho$ dominated over ΔH_s (compare red versus blue areas). In terms of absolute uncertainty,

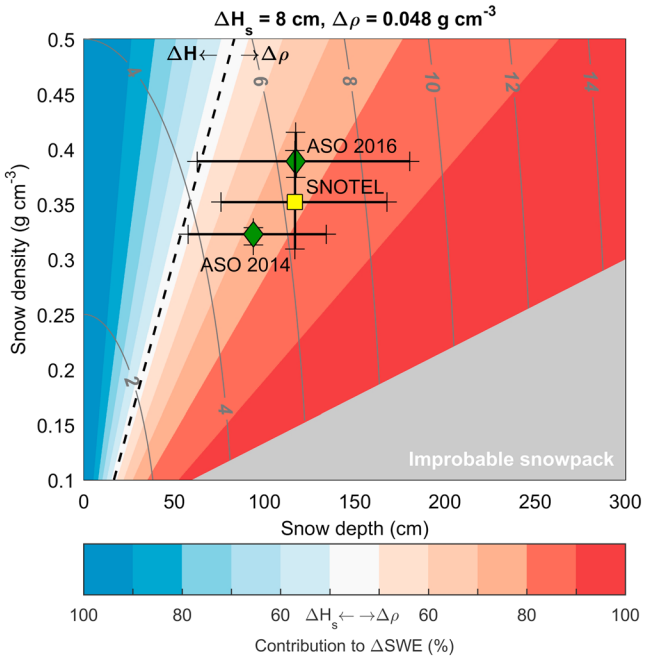

Figure 3. Percent contribution to SWE uncertainty (ΔSWE) from uncertainties in modeled snow density ($\Delta\rho$) versus lidar-measured snow depth (ΔH_s) across a theoretical range of snow depths and densities. Here ΔH_s = 8 cm and $\Delta\rho$ = 0.048 g cm^{-3}. The dashed line is the crossover point (see Figure 2), which divides conditions where snow depth (blue region) or density (red region) uncertainties dominate SWE uncertainty. Median snow depth and density for the ASO analyses in 2014 and 2016 (green diamonds) and the SNOTEL network (yellow square, n = 9013 station years) and interquartile ranges (lines) are shown for reference. Gray contours show absolute ΔSWE (cm) from $\Delta\rho$ and ΔH_s combined.

ΔSWE was greatest for snowpack with higher depth and density (Figure 3). ΔSWE and the relative zones of dominant uncertainty changed with different levels of ΔH_s and $\Delta\rho$ (see Figures S7 and S8).

The competing contributions to ΔSWE were compared for the typical depth and density conditions found at peak SWE in the SNOTEL network (Figure 3). The vast majority of observations from times of peak SWE fell in the zone where $\Delta\rho$ dominated ΔSWE. Approximately 70–90% of SWE uncertainty was due to density uncertainty. The SWE uncertainty was typically about 5 to 7 cm for snowpack conditions sampled by SNOTEL. The results from the 2014 and 2016 ASO surveys were within the range of the SNOTEL data. The contribution from $\Delta\rho$ was slightly greater in the 2016 ASO survey than for 2014. Snow depth was greater in 2016, and thus the depth errors were smaller on a percentage basis. This supported the general result that the contribution of density uncertainty to SWE uncertainty increased at greater snow depths, either in portions of a basin where snow was deep or in years when snow accumulation was higher.

4. Discussion and Conclusions

Lidar makes it possible to map snow depth with uncertainty of ~10 cm across a catchment. Our results show that snow density estimation becomes the dominant source of uncertainty in lidar-based SWE mapping when snow depth exceeds ~50 cm. The importance of snow density uncertainty increases with snow depth, greatly exceeding snow depth uncertainty in areas (or years) with deeper snowpack and hence zones with greater SWE volume (Figures 2 and 3). Snow density uncertainty exceeds depth uncertainty even in a historically dry and low snowpack year (2014). Because snow depth uncertainty is typically greater for lidar than many other techniques, density uncertainty will also dominate SWE uncertainty for estimating SWE with other approaches.

The typical snow density uncertainty across the four models here (~0.05 g cm^{-3}) falls within the range of model uncertainty reported in the literature (see Text S2 and Table S1). This level of uncertainty was found in both an extremely dry year (2014) and a near-average year (2016). We therefore consider our results a reasonable approximation of the uncertainty expected from density estimation. This result is based on the assumption that differences between models can be used as a proxy for uncertainty and that the specific models we selected portray typical intermodel differences. In actual applications, the uncertainty in modeled density may be higher. We only considered one source of uncertainty (model selection), and ignored other factors, including canopy influences on snow density, and errors in forcing data. The physically based models typically differed by 0.06 g cm^{-3} (based on MAD), which is lower than differences documented in other studies (on the order of 0.10 g cm^{-3}) [*Feng et al.*, 2008]. This clearly illustrates the importance of model

uncertainty, suggesting that density model selection can introduce large uncertainty into SWE estimates in the absence of site-specific tuning. Prior model intercomparison studies [*Etchevers et al*., 2004; *Rutter et al*., 2009] evaluated representation of SWE, snow depth, and energy balance variables, but there has been less attention to modeled bulk density. A more systematic analysis [e.g., *Essery et al*., 2013] is needed to isolate the structural and parametric reasons for these large differences in modeled density.

The uncertainty in modeled density (prior to elevation correction) documented by *Painter et al*. [2016] is about half of the model uncertainty reported here and is the lowest found in our literature review (see Text S2). Uncertainty in modeled density can be reduced when in situ measurements of snow density are available to tune model parameters or develop a model correction [*Painter et al*., 2016]. Our analysis assumed no knowledge of density conditions on the ground and hence reflected a general scenario of SWE mapping. While in situ data can constrain snow density models, there are shortcomings to this practice. First, a correction is only straightforward when the model residuals exhibit a coherent relationship with a geophysical parameter (e.g., elevation). Second, corrections are not likely transferable to other catchments. Density uncertainty is also likely to be underestimated, given the tendency for sampling locations to be biased to easily accessible flat clearings. Corrections based on these data may not be applicable to other areas in a basin. Finally, corrections are bound to be model-specific and do little to identify specific model deficiencies.

Our multi-model approach can spatially map model agreement (Figure 1b) to guide selection of field evaluation sites for more targeted testing of models. There have been few efforts to evaluate models by sampling snow density systematically across a range of model uncertainty levels and physiographic settings [*Bormann et al*., 2014]. With the development of technologies like lidar that measure snow depth through space, it is important to assess how models represent the mean and spatial variation of snow density for mapping SWE.

We recognize that uncertainties in lidar-derived snow depth are not constant in space, as assumed above. Lidar uncertainty varies with landscape and measurement characteristics [*Deems et al*., 2013]. Uncertainty in lidar snow depth may increase by 10 cm or more in forests [*Deems et al*., 2013; *Harpold et al*., 2014]. Ongoing research is quantifying variability in lidar uncertainty across diverse landscapes. Uncertainty in snow density estimation may also be enhanced in these same areas (e.g., slopes and forests), as most prior studies evaluated density models in flat clearings, and thus more targeted evaluations are needed.

Historically, in situ measurements of snow depth have outnumbered SWE and density measurements by a factor of 30 [*Sturm et al*., 2010]. As the snow depth measurement revolution continues, the disparity in availability of snow depth versus density data will widen by many more orders of magnitude. Likewise, the accuracy of snow depth measurements will improve with technological advances. Considering the dominance of snow density uncertainty over depth uncertainty and the ongoing proliferation of depth measurements, advances are needed in the measurement and modeling of snow density to resolve specific landscape influences on density. More plentiful and more accurate density measurements in space are essential for process understanding and for reducing uncertainty in modeled density.

Acknowledgments
This work was supported by NSF-EAR 1521474. The data used in the ASO analysis are described and included in the supporting information. We thank Tom Painter and the ASO team for sharing the lidar snow depth and elevation data. Thanks also to Danny Marks for providing a thoughtful review that improved the manuscript and for making the Snobal code available. Thanks to Gerald Flerchinger for making the SHAW model code available. We thank Ryan Webb, John Knowles, and Dave Barnard for providing preliminary feedback on the manuscript.

References

Anderson, E. (1976), *A Point Energy and Mass Balance Model of a Snow Cover*, *NOAA Tech. Rep. NWS 19*, pp. 1–150, Silver Spring, Md.

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

Balk, B., and K. Elder (2000), Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed, *Water Resour. Res.*, *36*(1), 13–26, doi:10.1029/1999WR900251.

Bavay, M., and T. Egger (2014), MeteoIO 2.4.2: a preprocessing library for meteorological data, *Geosci. Model Dev.*, *7*(6), 3135–3151, doi:10.5194/gmd-7-3135-2014.

Bormann, K. J., J. P. Evans, and M. F. McCabe (2014), Constraining snowmelt in a temperature-index model using simulated snow densities, *J. Hydrol.*, *517*, 652–667, doi:10.1016/j.jhydrol.2014.05.073.

Buhler, Y., M. S. Adams, R. Bosch, and A. Stoffel (2016), Mapping snow depth in alpine terrain with unmanned aerial systems (UASs): Potential and limitations, *Cryosphere*, *10*(3), 1075–1088, doi:10.5194/tc-10-1075-2016.

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

Conger, S. M., and D. M. McClung (2009), Comparison of density cutters for snow profile observations, *J. Glaciol.*, *55*(189), 163–169, doi:10.3189/002214309788609038.

Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical–topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*, 140–158.

DeBeer, C. M., and J. W. Pomeroy (2010), Simulation of the snowmelt runoff contributing area in a small alpine basin, *Hydrol. Earth Syst. Sci.*, *14*(7), 1205–1219, doi:10.5194/hess-14-1205-2010.

Deems, J. S., T. H. Painter, and D. C. Finnegan (2013), Lidar measurement of snow depth: A review, *J. Glaciol.*, *59*(215), 467–479, doi:10.3189/2013JoG12J154.
Dozier, J. (2011), Mountain hydrology, snow color, and the fourth paradigm, *Eos. Trans. AGU*, *92*(43), 373, doi:10.1029/2011EO430001.
Elder, K., W. Rosenthal, and R. E. Davis (1998), Estimating the spatial distribution of snow water equivalence in a montane watershed, *Hydrol. Process.*, *12*(1011), 1793–1808, doi:10.1002/(SICI)1099-1085(199808/09)12:10/11<1793::AID-HYP695>3.3.CO;2-B.
Elder, K., D. Cline, G. E. Liston, and R. Armstrong (2009), NASA Cold land processes experiment (CLPX 2002/03): Field measurements of snowpack properties and soil moisture, *J. Hydrometeorol.*, *10*(1), 320–329, doi:10.1175/2008JHM877.1.
Essery, R., S. Morin, Y. Lejeune, and C. B Ménard (2013), A comparison of 1701 snow models using observations from an alpine site, *Adv. Water Resour.*, *55*, 131–148, doi:10.1016/j.advwatres.2012.07.013.
Etchevers, P., et al. (2004), Validation of the energy budget of an alpine snowpack simulated by several snow models (SnowMIP project), *Ann. Glaciol.*, *38*(1), 150–158, doi:10.3189/172756404781814825.
Feng, X., A. Sahoo, K. Arsenault, P. Houser, Y. Luo, and T. J. Troy (2008), The impact of snow model complexity at three CLPX sites, *J. Hydrometeorol.*, *9*(6), 1464–1481, doi:10.1175/2008JHM860.1.
Flerchinger, G., and K. Saxton (1989a), Simultaneous heat and water model of a freezing snow-residue-soil system I. Theory and development, *Trans. Am. Soc. Agric. Eng.*, *32*(2), 565–571.
Flerchinger, G., and K. Saxton (1989b), Simultaneous heat and water model of a freezing snow-residue-soil system II. Field verification, *Trans. Am. Soc. Agric. Eng.*, *32*(2), 573–578.
Grünewald, T., and M. Lehning (2015), Are flat-field snow depth measurements representative? A comparison of selected index sites with areal snow depth measurements at the small catchment scale, *Hydrol. Process.*, *29*(7), 1717–1728, doi:10.1002/hyp.10295.
Grünewald, T., M. Schirmer, R. Mott, and M. Lehning (2010), Spatial and temporal variability of snow depth and ablation rates in a small mountain catchment, *Cryosphere*, *4*(2), 215–225, doi:10.5194/tc-4-215-2010.
Harpold, A. A., et al. (2014), LiDAR-derived snowpack data sets from mixed conifer forests across the western United States, *Water Resour. Res.*, *50,* 2749–2755, doi:10.1002/2013WR013935.
Jonas, T., C. Marty, and J. Magnusson (2009), Estimating the snow water equivalent from snow depth measurements in the Swiss alps, *J. Hydrol.*, *378*(1–2), 161–167, doi:10.1016/j.jhydrol.2009.09.021.
Jordan R. (1991), A one-dimensional temperature model for a snow cover: Technical documentation for SNTHERM.89, p. 58, Special Report 91–16, US Army CRREL, Hanover, NH, USA.
Kerkez, B., S. D. Glaser, R. C. Bales, and M. W. Meadows (2012), Design and performance of a wireless sensor network for catchment-scale snow and soil moisture measurements, *Water Resour. Res.*, *48*, W09515, doi:10.1029/2011WR011214.
Kinar, N. J., and J. W. Pomeroy (2015), Measurement of the physical properties of the snowpack, *Rev. Geophys.*, *53*, 481–544, doi:10.1002/2015RG000481.
Larson, K. M., E. D. Gutmann, V. U. Zavorotny, J. J. Braun, M. W. Williams, and F. G. Nievinski (2009), Can we measure snow depth with GPS receivers?, *Geophys. Res. Lett.*, *36*, L17502, doi:10.1029/2009GL039430.
Lehning, M., I. Völksch, D. Gustafsson, T. A. Nguyen, M. Stähli, and M. Zappa (2006), ALPINE3D: A detailed model of mountain surface processes and its application to snow hydrology, *Hydrol. Process.*, *20,* 2111–2128, doi:10.1002/hyp.6204.
López-Moreno, J. I., S. R. Fassnacht, J. T. Heath, K. N. Musselman, J. Revuelto, J. Latron, E. Morán-Tejeda, and T. Jonas (2013), Small scale spatial variability of snow density and depth over complex alpine terrain: Implications for estimating snow water equivalent, *Adv. Water Resour.*, *55*, 40–52, doi:10.1016/j.advwatres.2012.08.010.
Lundberg, A., D. Gustafsson, C. Stumpp, B. Kløve, and J. Feiccabrino (2016), Spatiotemporal variations in snow and soil frost—A review of measurement techniques, *Hydrology*, *3*(3), 28, doi:10.3390/hydrology3030028.
Lundquist, J. D., et al. (2016), Yosemite hydroclimate network: Distributed stream and atmospheric data for the Tuolumne River watershed and surroundings, *Water Resour. Res.*, *52*, 1–12, doi:10.1002/2016WR019261.
Margulis, S. A., G. Cortés, M. Girotto, and M. Durand (2016), A Landsat-era Sierra Nevada snow reanalysis (1985–2015), *J. Hydrometeorol.*, *17*(4), 1203–1221, doi:10.1175/JHM-D-15-0177.1.
Marks, D., and J. Dozier (1992), Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada: 2. Snow cover energy balance, *Water Resour. Res.*, *28*(11), 3043–3054, doi:10.1029/92WR01483.
Marks, D., J. Dozier, and R. E. Davis (1992), Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada: 1. Meteorological measurements and monitoring, *Water Resour. Res.*, *28*(11), 3029–3042, doi:10.1029/92WR01482.
Marks, D., J. Domingo, D. Susong, T. Link, and D. Garen (1999), A spatially distributed energy balance snowmelt model for application in mountain basins, *Hydrol. Process.*, *13*(12–13), 1935–1959, doi:10.1002/(SICI)1099-1085(199909)13:12/13<1935::AID-HYP868>3.0.CO;2-C.
Marshall, H. P., and G. Koh (2008), FMCW radars for snow research, *Cold Reg. Sci. Technol.*, *52*(2), 118–131, doi:10.1016/j.coldregions.2007.04.008.
McCreight, J. L., and E. E. Small (2014), Modeling bulk density and snow water equivalent using daily snow depth observations, *Cryosphere*, *8*(2), 521–536, doi:10.5194/tc-8-521-2014.
Nolin, A. W. (2010), Recent advances in remote sensing of seasonal snow, *J. Glaciol.*, *56*(200), 1141–1150.
Painter, T. H., 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 Sens. Environ.*, *184*, 139–152, doi:10.1016/j.rse.2016.06.018.
Parajka, J., P. Haas, R. Kirnbauer, J. Jansa, and G. Blöschl (2012), Potential of time-lapse photography of snow for hydrological purposes at the small catchment scale, *Hydrol. Process.*, *26*(22), 3327–3337, doi:10.1002/hyp.8389.
Pohl, S., J. Garvelmann, J. Wawerla, and M. Weiler (2014), Potential of a low-cost sensor network to understand the spatial and temporal dynamics of a mountain snow cover, *Water Resour. Res.*, *50*, 2533–2550, doi:10.1002/2013WR014594.
Proksch, M., N. Rutter, C. Fierz, and M. Schneebeli (2016), Intercomparison of snow density measurements: Bias, precision, and vertical resolution, *Cryosphere*, *10*(1), 371–384, doi:10.5194/tc-10-371-2016.
Raleigh, M. S., J. D. Lundquist, and M. P. Clark (2015), Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, *Hydrol. Earth Syst. Sci.*, *19*(7), 3153–3179, doi:10.5194/hess-19-3153-2015.
Raleigh, M. S., B. Livneh, K. Lapo, and J. D. Lundquist (2016), How does availability of meteorological forcing data impact physically-based snowpack simulations?, *J. Hydrometeorol.*, *17*(1), 99–120, doi:10.1175/JHM-D-14-0235.1.
Reba, M. L., D. Marks, T. E. Link, J. Pomeroy, and A. Winstral (2014), Sensitivity of model parameterizations for simulated latent heat flux at the snow surface for complex mountain sites, *Hydrol. Process.*, *28*(3), 868–881, doi:10.1002/hyp.9619.
Rice, R., R. C. Bales, T. H. Painter, and J. Dozier (2011), Snow water equivalent along elevation gradients in the Merced and Tuolumne River basins of the Sierra Nevada, *Water Resour. Res.*, *47*, W08515, doi:10.1029/2010WR009278.

Rodell, M., E. B. McWilliams, J. S. Famiglietti, H. K. Beaudoing, and J. Nigro (2011), Estimating evapotranspiration using an observation based terrestrial water budget, *Hydrol. Process.*, *25*(26), 4082–4092, doi:10.1002/hyp.8369.

Rutter, N., et al. (2009), Evaluation of forest snow processes models (SnowMIP2), *J. Geophys. Res.*, *114*, D06111, doi:10.1029/2008JD011063.

Sandells, M. J., G. N. Flerchinger, R. J. Gurney, and D. Marks (2012), Simulation of snow and soil water content as a basis for satellite retrievals, *Hydrol. Res.*, *43*(5), 720, doi:10.2166/nh.2012.028.

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

Shi, X., M. Sturm, G. E. Liston, R. E. Jordan, and D. P. Lettenmaier (2009), SnowSTAR2002 transect reconstruction using a multilayered energy and mass balance snow model, *J. Hydrometeorol.*, *10*(5), 1151–1167, doi:10.1175/2009JHM1098.1.

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

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

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

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

Varhola, A., J. Wawerla, M. Weiler, N. C. Coops, D. Bewley, and Y. Alila (2010), A new low-cost, stand-alone sensor system for snow monitoring, *J. Atmos. Oceanic Tech.*, *27*(12), 1973–1978, doi:10.1175/2010JTECHA1508.1.

Viviroli, D., et al. (2011), Climate change and mountain water resources: Overview and recommendations for research, management and policy, *Hydrol. Earth Syst. Sci.*, *15*(2), 471–504, doi:10.5194/hess-15-471-2011.

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

Xia, Y., 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, *J. Geophys. Res.*, *117*, D03109, doi:10.1029/2011JD016048.