

Response to Reviewers

**Comments are indented and responses are left justified. Line numbers of modified text refer to the line numbers in the revised manuscript.

Anonymous Referee #1

Received and published: 6 November 2017

General comments

The authors present a comprehensive study on how Snow Water Equivalent distribution in mountain areas can be derived from terrain topographic characteristic, vegetation height and fractional snow cover data from satellites. The methodology is tested with a high quality database from the Airborne Snow Observatory. The spatio-temporal extension of this database over Toulumne Basin provides a unique opportunity to test the methodology. Results obtained from the statistical approach they describe, show the high potential of the methodology. As far as I know, this work presents the first method intended to obtain SWE distribution combining satellite information and terrain features over extended areas, showing quite promising results. The article is well written, with methods and results in general adequately articulated. The database and the methodology described are suitable for the field of research. Results are sufficiently discussed with suitable references included. The article also suites the scope of the journal, introducing a novel approach which may be applied in other mountain areas with feasibly good results. However, I consider some issues must be clarified and also some complementary hypothesis could be tested, in order to present a more compelling methodology. Thus I recommend its publication after moderate changes described here in after.

We thank the reviewer for their time and positive feedback. We have taken care to address their concerns in the manuscript and below each specific comment.

Major points to be included/discussed:

- 1. Necessity to include more information for some methods/databases presented in the manuscript. SWE distribution maps are generated from snow depth information directly obtained from ASO flights (elevation difference between snow and free snow acquisitions) and an energy balance model as it is stated from lines 55-56. Painter et al. 2016, describe the methodology to generate SWE distribution maps providing detailed information. Nevertheless since this database is the main observation to adjust the linear regression models and also to test the methodology, I encourage manuscript authors to provide more information about how density is simulated, which in-situ data are integrated (are snow pillows observations included?) and how the final SWE product is generated.*

We have included the following additional text about generating SWE estimates on L53-58:
ASO measures snow depth by differencing the lidar-derived elevations from snow-on and snow-off flights and infers albedo and snow extent based on spectroradiometric measurements. Density is computed using the iSNOBAL energy-balance model (Marks et al., 1999) based on inferred albedo, local meteorological measurements, and constrained by in situ snow pillow observations and manual snow course observations (Painter et al., 2016). Estimated snowpack density is multiplied by the lidar-derived snow depth to estimate SWE.

2. *Since SWE information is available on a spatial resolution of 50 m, the vegetation height in a 3 m resolution and terrain topographic variables could be derived from a high spatial resolution DEM (I guess in this study area a 3 DEM may be available), exploiting fSCA information; the statistical models could be intended to generate same spatial resolution of ASO final SWE product. Has been tested this hypothesis?, does manuscript authors planned to do so? Argue in discussion section future work regarding the latest advances on deriving snow absence/presence from fSCA and terrain topographic characteristics (Cistera et al., 2017) to increase the spatial resolution of simulated SWE distribution.*

We do not currently have plans to run a similar analysis at 50 m resolution despite the availability of SWE from ASO at 50m. We specifically limited our analysis to 500 m because it is the only resolution for which operationally available fsca products are available. Downscaling fsca based on terrain characteristics is a relatively recent development and has not been tested with satellite data- Cistera et al. (2017) show that their method is most sensitive to errors in input fsca and thus we do not think it is appropriate for this study. Nonetheless, we have included a statement in the discussion about the potential for higher resolution SWE estimates based on downscaled fsca.

L492-494:

...and recent advances in downscaling fSCA may further improve SWE estimates at the end of the melt season as only small patches of snow persist (Walters et al. 2014, Li et al. 2015, Cristea et al. 2017).

3. *In section 4.2, it is stated that models obtained in a particular year are not used to obtain SWE distribution in other dates of the same year. I understand that authors want to test models without including any information about SWE distribution in a particular year. However, from my point of view, having some information about SWE distribution on a particular date could be quite interesting for many applications and may reduce the uncertainty on determining SWE. Indeed, somehow, in the discussion this idea is supported (lines 486-488). This way if SWE observations obtained on early snow season are used to generate a model for the same year, it could be more accurate on describing SWE distribution within this snow season. I encourage manuscript authors to explore this hypothesis in methodology and result sections and also to argue about it on discussion section.*

We agree with the reviewer that SWE estimates based on observations earlier the same year could potentially be useful for some applications. We investigated this hypothesis and now include the following text on L515-523:

One might also consider the value of a single flight each year for estimating SWE distribution on other dates in the same year, and in this case we found the mean r^2 to be 0.82, %MAE to be 39%, and %Bias to be 7% when considering the best model for each date. These values are similar to the yearly values from the best transferred model from other years (Tables 3, 4), which supports the prior assertion that the good or equivalent performance of models from other years suggests a strong degree of persistence in snow patterns. We note the positive %Bias and suggest that this perhaps should not come as a surprise since we only used prior flights for the same year, which measured deeper snowpacks and greater fSCA than those modeled; ASO flights typically start around peak SWE. Future work should identify the importance of the fSCA state for transferring a model, and whether an ASO flight in the accumulation season might yield better results.

4. Some topographic variables are not sufficiently explained. For instance the “Vector ruggedness measure” and the “Topographic Position Index” both of them have a high dependence on the searching distance. Could you please specify if you have performed preliminary analysis to determine which is the best searching distance within your study area or if you have used values from bibliography? Please, include more details on how these variables have been obtained; this will help potential readers to apply the methods described in different study areas.

The following details have been added to the Methods to explain the topographic variables including the window size for the Vector Ruggedness Measure (VRM), the Topographic Position Index (TPI), and the standard deviation of slope L161-163:

We used the Global Multi-resolution Terrain Elevation Data 2010 digital elevation model (data available from the U.S. Geological Survey at <https://lta.cr.usgs.gov/GMTED2010>) reprojected to the 500 m SWE grid to compute the physiographic predictor variables (Table 1).

L180-191:

The explanatory variables we consider are ASO-observed fSCA and topographic variables previously used in the literature (Table 1). The topographic variables were directly computed from the 500 m digital elevation model with the exception of vegetation height which was derived from the 3 m ASO vegetation height by mean aggregating to 501 m and bilinearly resampling to 500 m. Of note, we use three different windowed variables (topographic position index (TPI), vector ruggedness index (VRM), and standard deviation of slope); to the best of our knowledge these have not previously been used for resolutions larger than 25 m. Comparison with ASO SWE at 500 m shows mean correlations of 0.3, -0.26, and -0.31, respectively, and we consider these variables indicative of preferential deposition from orographic updraft (Dadic et al., 2010; Garvert et al., 2007; Lehning et al., 2008). The window sizes were chosen based on the maximum correlation between ASO SWE and the variable, where the marginal increase in correlation from the next smaller window size was greater than 10%. We tested each odd window size from 3x3 to 15x15 pixels which equate to 1.5 km to 7.5 km scale, respectively.

Minor comments

Line 17: Move performance values, in brackets; to lines 21-22 to show the final performance after showing the mid decrease between the best model and the selected one.

We have moved the performance values as requested.

Lines 33- 34: Add one or two references that exemplify the poor data availability of SWE observations over large areas (i.e. SNOTEL program)

We have included citations for Meromy et al. (2012) and Rice et al. (2010) to reinforce the statement about sparse data coverage. However, we would like to note that, although the same issues apply, we do not use the SNOTEL network because the snow pillows in California are run by the California Dept. of Water Resources as opposed to the National Resources Conservation Service (NRCS).

Line 35-36: Cite: López-Moreno, J. I., Fassnacht, S. R., Heath, J. T., Musselman, K. N., Revuelto, J., Latron, J., Morán-Tejeda, E., and Jonas, T. 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, 2012

This has been added.

Line 42: Cite: Prokop, A.: Assessing the applicability of terrestrial laser scanning for spatial snow depth measurements, Cold Reg. Sci. Technol., 54, 155–163, 2008

This has been added.

Line 55: Please cite which energy-balance model.

ASO density is computed with iSNOBAL (Marks and Dozier, 1992). This has been added.

Line 64: Here and throughout the whole manuscript; I encourage changing “physiography” by “topography” and “vegetation height” since it is often used on snow literature.

This has been changed.

Line 74: Cite: Revuelto J, López-Moreno JI, Azorin-Molina C, Vicente-Serrano SM. 20014b. Topographic control of snowpack distribution in a small catchment in the central Spanish Pyrenees: intra- and inter-annual persistence. The Cryosphere 8(5): 1989–2006. DOI:<http://dx.doi.org/10.5194/tc-8-1989-2014>.

This has been added.

Line 76-78: Remove or merge sentence with lines 43-45.

The statement from 43-45 was removed and merged to L79-82:

The lidar measurements provide snow depth, and coincident density measurements or estimates are required to estimate SWE. However, snow depth measurements still capture the majority of the variability in SWE because snow depth varies an order of magnitude greater than density (Mizukami et al 2008, Jonas et al. 2009, Lopez-Moreno 2013)

Line 85-86: Please include here some information about the spatial extent of ASO; time duration and total number of observations available.

We have included the basin size here (now L89-91)

The spatio-temporal dataset of SWE from ASO, which covers the full range of physiographic variables present in a 1,175 km² watershed, provides an unprecedented opportunity to develop relationships between SWE and topography and test their persistence across several years.

and introduce the time duration of the mission earlier when ASO is first introduced

L50-53:

Since 2013, the National Aeronautics and Space Administration (NASA), Jet Propulsion Laboratory, Airborne Snow Observatory (ASO) has acquired weekly observations of snow properties from approximately the time of annual peak SWE to the end of the snowmelt season in the Tuolumne Basin, California totally 28 flights from 2013-2016 Painter et al. 2016.

Line 90: Include other more recent citations such as: Molotch N.P. and Margulis S.A. 2008: Estimating the distribution of snow water equivalent using remote sensed data and a spatially distributed snowmelt model: a multi-resolution, multi-sensor comparison. Advances in Water Resources Research, 13-1503-1514

We have added Cline et al. (1998) and Molotch and Margulis (2008).

Line 98: Include these two references. Sturm M, Wagner AM. 2010. Using repeated patterns in snow distribution modeling: an Arctic example. Water Resources Research 46, W12549. DOI: <http://dx.doi.org/10.1029/2010WR009434>, and Revuelto, J., Jonas, T., & López-Moreno, J. I. (2016). Backward snow depth reconstruction at high spatial resolution based on time-lapse photography. Hydrological Processes, 30(17), 2976- 2990.

We have added these two citations.

Line 117: You say “. . .unsampled dates of interest”. However dates considered here have been sampled by the ASO. Please rephrase.

We have added “hypothetically” this sentence, now L122:

*Lastly, we present a methodology for identifying which models of SWE distribution, from the ensemble of historical ASO acquisitions, best represents the SWE distribution for **hypothetically** unsampled dates of interest.*

Line 122-Figure 1: Add in this figure an extra map with a DEM showing the heterogeneity of the study area. This information will help to interpret SWE distribution shown in Figure 2. Is there any ice body in the study area? If it is, in the land cover map it is really difficult to see, please highlight it or remove this land cover class from the legend.

We have updated Figure 1 to include a DEM and point to the two small glaciers in the basin.

Line 124 - 127: Please, include scientific names of the trees occurring in the different forests types.

These have been included, now L130:

*....red fir (*Abies magnifica*) and lodgepole pine (*Pinus contorta*).*

Study area section: Climatic information of the study area is highly desirable to understand the importance of snowpack evolution in this site. Please, include a paragraph describing the main climatic characteristics and also the average snowpack values observed during the study years. This information will show the characteristics of the analyzed years (below average snowpack conditions).

We have inserted the following climatic description L136-142:

Analysis of SWE measurements from snow pillows show that historical mean peak SWE ranges from 0.3 m to 1.5 m with a mean of 0.8 m. Each of the study years was characterized by below average snowpack with 2014 and 2015 characterized as a severe dry snow drought (Harpold et al., 2017), experiencing only 36% and 35% of the climatological mean peak SWE, respectively. The years 2013 and 2016 also experienced below average snowpack conditions, but less severely, with the data from the snow pillows reporting 62% and 71% of climatological peak SWE, respectively. Typically, minimum temperatures range from -12°C in winter to 3°C in summer and maximum temperatures range from 4°C in winter to 22°C in summer (Cristea et al., 2017).

Line 139-140: How snow depth and SWE maps are generated? Why these do not have same spatial resolution? I encourage manuscript authors to include here more information about the model used to generate SWE products from snow depth.

We agree with the reviewer it is important to understand how these products were developed, but since we are not involved in producing these products we do not think it is appropriate to discuss in detail. We refer an interested reader to Painter et al. (2016). For your information, snow depth

and SWE do not have the same resolution because SWE is limited by the resolution of the modeled density whereas snow depth comes is the direct result of differencing two lidar scans.

Table 1: Source column for VRM, Lopez et al. (2014) should be Lopez-Moreno et al. (2014).

We apologize for this error. This has been fixed.

Line 180: “discrete points of time” could be confusing. Please reword to make clear it is only considering specific dates.

This has been changed to “...for a given day...”

Line 188: I guess you mean you split the SWE distribution data from each ASO flight. Please clarify in the text.

We apologize for this error. This has been fixed on L208-209:

We randomly split the data for each date into a training (80%) and test (20%) dataset to evaluate overall model performance on this date.

Line 187-189: In subsequent sections you compare “split-sample models” with results obtained from “best transferred model”. Please present in these lines “split-sample models” similarly as you do later with “best model” or “transferred model”.

The terminology has been revised consistently throughout the paper. We now use “same-day models” instead of “split-sample models”.

Line 217: When you talk about nearby snow pillows, you consider only these of Figure 1? all of them?, do you take into account their distance to the study area? their elevation? Please clarify all these questions.

We are using all snow pillows within 20km of the basin, as shown in Figure 1. Their contribution to MAE is not weighted based on distance or elevation. We have clarified this point in the text L236-238:

Here we quantify the SWE similarity between dates using the mean absolute error (MAE) of SWE recorded at snow pillows within 20 km of the basin as shown in Figure 1; station data was not weighted by distance to the basin or elevation.

Line 222: “best model” on italics Please, maintain this criteria along the text. This is, when you talk about “best model”, “selected model”, “split mode”. . . , do it always in italics since it is a convention adopted in methods section.

This has been adjusted throughout the text.

Line 252: Change “alpha” by the Greek symbol as presented before.

This has been adjusted.

Figure 2. It is highly desirable to include elevation contour lines on SWE maps and also in the difference maps to better interpret results.

Contour lines have been added to the SWE maps.

Line 284-287 A reduced map with Land Cover information included on Figure 2 could really help to interpret results.

We thank the reviewer for their suggestions, but we are unable to insert a land cover map inset while maintaining legibility. Further, we would refer the reader to Figure 1 for land cover information.

Line 288: You can also say that in these pixels you observed the higher SWE values (at least on 2014-03-23). Indeed, since Google Earth information is not a database exploited in this article, I suggest removing this sentence and state that these pixels correspond to areas in which you have observed high SWE values from ASO data and also long snow presence from MODIS data.

This text has been removed and is no longer a point of discussion because it is not relevant with the revised results from that include TPI and VRM with different windows.

Figure 3, (4, 5): Since you present box plots of r^2 values for 18 to 24 potential models for each date (depending on the year), to me, it is not needed to include diamonds showing the best model, because it is the 95th (or the 5th, depending on the figure) percentile in all cases. Thereby I would remove it to show graphs easier to interpret. If finally you decide to maintain diamonds, include it on the legend of the figures.

We have included the diamonds in the legend in Figures 3, 4, 5.

Line 343: Remember here that "Selected models" are those selected from the similarity on snow pillows observations.

We have included the following reminder in the text on L367:
(i.e. those chosen based on snow pillow similarity)

Line 358: I guess you mean Figures 3, 4 and 5 and not Figs.s,4,5. Line 359: I guess you meant open circles in Figure 7.

We apologize for this error. This has been corrected.

Line 363: If I am right, figure 6a shows two dates with same %MAE, one in 2015 and one in 2016 (both for the last ASO observation). Please verify this and include this appropriately in the text.

We apologize for this error. There are 2 dates with the same error for %MAE and 1 date for %Bias. This has been confirmed and adjusted in the text.

Figure 7: Please, change open circles by a different symbol because in some cases it is difficult to see their position. Please include this symbol in the legend.

We have changed the symbol to be easier to see and added it to the legend.

Line 404: I suggest starting this section talking about snowpack and citing other recent works that also exploit fSCA such as: Cristea et al., 2017 and Walters et al., 2014. Afterwards, you can talk about SWE distribution as you do. This will show that many researches are working on combining the influence of topography on snowpack distribution with satellite observations with different approaches.

We have added citations for Walters et al. 2014 and Cristea et al. 2017 and included the following sentence to explicitly show that we have extended previous work on the influence of topography on fSCA to estimating SWE distribution.

L426-429:

The basis for this approach is relatively well established given that fSCA is sensitive to topographic complexity and the spatial distribution of SWE (Donald et al., 1995; Fassnacht et al., 2016; Niu and Yang, 2007, Walters et al, 2014, Cristea et al, 2017). We extend the concept of linking fSCA with topography from these previous works to estimating SWE distribution.

Line 416-416: I consider it is a bit presumptuous to say that “these papers, which cover only a few square kilometers, represent a far more simplistic problem”. Some of the works you cite in the previous sentence cover large areas and others, despite the smaller extension, have a different spatial scale, data base etc,. ... Please remove this sentence, you have already shown that your methods performs well and are pushing forward snow science.

We have removed the offending sentence.

Line 437: I guess SNOTEL observations correspond to SWE data from snow pillows, please include before this information in the text (Figure 1, 4.2 section when you talk about “selected model”...)

We have replaced the word “SNOTEL” with “snow pillow” to more generally refer to SWE observations from snow pillows. Although most snow pillows in the western United States are SNOTEL stations run by the Natural Resources Conservation Service, those in California are operated by the California Dept. of Water Resources (CADWR). The snow pillow observations

used in this study are from the CADWR network, as such it is not appropriate to include the suggested information in the text.

Line 434: I guess you meant you are going to present the comparison between SNODAS and PHV-FSCA results in this section. Please rephrase, it could be confusing this sentence.

We have improved the transition to our SNODAS comparison on L455-458:

*The selected model in this study is a simple linear regression that can be applied in real-time, thus we consider our results valuable for applications where real-time estimates of SWE distribution are needed. **In this regard**, we compare our selected model results to SWE estimates from the U.S. National Weather Service's operational Snow Data Assimilation System (SNODAS).*

Line 442: When you state "the yearly mean", it is from results obtained in this study or from SNODAS-ASO comparison? Please clarify.

The SNODAS-ASO comparison presented was our own analysis. We have clarified this on L464-465:

***In our own analysis**, the yearly mean r^2 between ASO and SNODAS ranges from 0.04 in 2016 to 0.36 in 2015 with a mean of 0.17.*

Line 499-500: When did you exactly perform the similarity in remote sensed fSCA? Please introduce it before in methods section, when you talk about similar SWE distribution from snow pillows.

Respectfully, we do not think it is appropriate to include the similarity in remotely sensed fSCA in the methods because it had no role in the results of the paper. We think that including this additional similarity metric in the methods will distract and confuse the reader regarding which method was actually used in all of the results. We included it in the discussion, however, because we think it is a natural next step of the work and could be of interest to the reader.

Conclusion section: Since at the end of the introduction section you state three main questions, I suggest to directly including the question and their answer in the conclusions. The answer is somehow stated, but main conclusions will be more directly linked to questions previously stated.

We have changed the text to more directly answer the first two questions at the beginning of the conclusion L556-558:

We estimated the relationships between SWE, physiography, and fSCA from ASO data. Our results show that fSCA information strongly improves statistical SWE models, and these models can be reliably transferred directly from one year to another due to strong persistence in snow accumulation and melt patterns.

References.

Cristea, N. C., Breckheimer, I., Raleigh, M. S., HilleRisLambers, J., & Lundquist, J. D. (2017). An evaluation of terrain-based downscaling of fractional snow covered area datasets based on Lidar derived snow data and orthoimagery. *Water Resources Research*.

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M., Seidel, F Winstral, A.: 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,

Walters, R. D., Watson, K. A., Marshall, H. P., McNamara, J. P., & Flores, A. N. (2014). A physiographic approach to downscaling fractional snow cover data in mountainous regions. *Remote sensing of environment*, 152, 413-425

Anonymous Referee #2

General Comments

This paper uses observational snow data collected as part of the Airborne Snow Observatory (ASO) in the Tuolumne Basin, California over multiple years and at different times of the melt period to examine temporal persistence in the patterns of snow water equivalent (SWE) and their relationship with various terrain parameters. Multiple regression analysis is used to model the relationship between SWE and the other parameters as independent variables—not from the original high resolution ASO products (3 m horizontal resolution), but from coarser spatially aggregated (to 500 m resolution) representations of these data. The main reported contribution is in using fractional snow cover area (fSCA) information from the ASO dataset to improve the transferability of the regression models to other dates when the detailed ASO snow observations are lacking, with the premise being that other available fSCA products, such as from MODIS, could be used instead at these times.

While the purpose and rationale seem clear enough and there is value in using the exemplary dataset from ASO to develop better approaches for predicting SWE and snowcover patterns when detailed observations are not available, the study suffers from major conceptual and methodological flaws that greatly limit its practical usefulness. From a conceptual perspective, it makes little sense and it is not particularly useful to develop a set of different regression models for each of the ASO acquisition dates, and then to later determine which one to use based on some similarity criteria from snow pillow observations. It would be more useful, for example, to develop a single model or approach to characterize spatial patterns of SWE where observations such as SWE at the snow pillows and remotely sensed snow cover information are used in combination to inform this. Or more simply, why not just use the original ASO datasets themselves based on similarity with the snow pillow data? Of course the limits with this are clearly apparent. Given the empirical nature and lack of any incorporation of snow accumulation, redistribution, and ablation processes in the approach, the results are not likely to transfer well outside of this basin, and quite possibly not even to other years and seasons under altered climate conditions.

Regression-based techniques for interpolating SWE from point measurements for each individual date is a common technique in the literature (e.g. {Elder:1998hu, Fassnacht:2003kl, Rice:2010wc, Schneider:2016cr}) and the method presented here extends this to interpolate SWE from each grid cell as measured by ASO. At 500 m, this represents over 7000 more measurement points from which we can establish robust relationships to estimate SWE. Our results show that leveraging multiple years of ASO data with widely available SCA products can be used to estimate SWE distribution at accuracy levels that are significantly greater than what they would be if ASO data have not been leveraged. Consider previous studies with real-time SWE estimates that reported mean RMSE of 0.23 m {Fassnacht:2003kl} and 0.12 m {Schneider:2016cr} compared with 0.08 m in this study. Importantly, for a water manager

interested in employing ASO, if they fly ASO for 5 years they can then use this approach to obtain meaningful snowpack information for years beyond the ASO flights.

Given our results, we respectfully disagree that an ensemble of regression models “makes little sense”. A universal and singular model is a good idea but is likely not doable given the exact points raised by the reviewer that spatial patterns can vary seasonally. An ensemble of models provides a greater chance to use an appropriate model with similar relationships between topographic variables and SWE. We do not think using the original ASO dataset as an estimate on another date makes sense because it does not consider the physically-based relationships, in particular, the relationship between SWE and time-variant fSCA.

Even more concerning is the methodological approach. The ASO data provide an opportunity to explore patterns of SWE variability and their changes over time at a high level of detail, and yet the approach here has been to aggregate these data and lose that valuable information. The local scale patterns of SWE accumulation are no longer captured, which is the whole purpose of relating SWE to terrain parameters (i.e. the drifts form in sheltered locations while exposed sites are more wind scoured, and much of this variability occurs at scales from meters to 10s of meters in complex mountain terrain). Moreover, the terrain parameters that are derived from this aggregated DEM become physically meaningless. How useful is average vegetation height over a 500 m grid? What do vector ruggedness or topographic position index really mean at this scale, for example? Certainly nothing from a physical sense that relates to patterns of SWE accumulation due to drifting, wind scouring, trapping of blowing snow by exposed vegetation, etc., and especially when relating to the averaged SWE over a 500 m grid.

While we agree with the reviewer that snow processes at high resolution are important, from a water resources perspective, SWE information at moderate resolution (500-m) is extremely useful and there are dozens of papers in the literature that attempt to estimate SWE at these scales or greater {e.g. Carroll:1996cr, Carroll:1997tr, Garen:1997jh, Brasnett:1999cq, Fassnacht:2003kl, Rice:2011fl, Schneider:2016cr, Broxton:2016es, Magnusson:2014gl, JorgHess:2014jo, Molotch:2008ja, Sturm:2010df}. SWE variability exists at scales from meters to 10s of meters to 1000s of meters. One needs only to go to the mountains in the winter to realize that the terrain parameters we derived at 500 m are not meaningless. In fact, the single variable correlations between the terrain variables and ASO SWE in our dataset range from -0.4 to 0.6 therefore we disagree with the reviewer that these variables are “meaningless”.

As we noted in Table 1, vegetation height ($r = -0.4$ with ASO SWE) is used as a proxy for forest canopy density which is correlated to SWE on the ground {Molotch:2007kf, Fassnacht:2003kl}. Other studies have noted the importance of vegetation presence for SWE variability (e.g. {Rice:2011fl}). Moreover, comparison of Figures 1 and 2 shows that our regression estimates are strongly driven by vegetation presence.

The windowed terrain variables mentioned by the reviewer (VRM and TPI) are actually quite important, exhibiting correlations of 0.3 and -0.26, respectively, with ASO-derived SWE at 500 m. Although not previously used in the literature at 500 m resolution with a scale of 4.5km, we suggest they indicate orographic effects and preferential deposition of snowfall. The scale of 4.5 km compares well with the findings from a numerical weather model in which the maximum

horizontal resolution for adequately simulating updraft and the resultant orographic effect for snowfall was 4 km {Garvert:2007jr}. The scales for VRM and TPI were chosen based on the maximum correlations with ASO SWE, where the marginal increase in correlation from the next smaller window size was greater than 10%. We tested each odd window size from 3x3 to 15x15 pixels which equate to 1.5 km to 7.5 km scale, respectively.

In addition to these concerns, the approach itself seems quite overcomplicated for what is essentially just an examination of the temporal persistence in SWE patterns over time, and ultimately an attempt to determine which regression model best fits the conditions of a given time based on a set of in situ SWE observations. The methods are not well described and hard to follow in some parts, leaving some major doubts. I will explain more in my specific comments below.

Respectfully, we also disagree with the reviewer that the study is “just an examination of the temporal persistence in SWE patterns over time” (but even that would be a valuable exercise!). The goal of estimating SWE distribution at unsampled time points is a complicated problem, but we provide an approach that is actually quite simple computationally and in terms of data requirements. Once the ensemble of models is established, a researcher or water manager could apply this approach relatively easily to estimate SWE in the basin. We have actually exploited the temporal persistence in SWE patterns, as observed with an unprecedented data set, toward a meaningful means to estimate SWE during unsampled time periods. This is by no means trivial given the needs for SWE estimates for water resource management and the very high cost of obtaining ASO data. Lastly, we have made several changes to the methods description to improve clarity. These are detailed below each specific comment below.

In the end, the study makes only a very incremental contribution to this important topic. While the ASO data provide unique opportunities to explore relationships between SWE, terrain, and snow cover patterns, and to advance understanding of the process and terrain controls in a way that could lead to improved predictability of SWE patterns, this study takes this in a different direction and instead focuses on coarser resolution regression analyses that largely miss this. Although the authors argue that this approach is novel and no one has previously used fSCA in such a regression, there is no major theoretical advancement, no practical utility as a result of the flawed methodology, and the results are severely limited in their broader applicability elsewhere. I would therefore recommend this paper to be rejected.

We agree with the reviewer that ASO data provides a unique opportunity to explore the relationships between SWE, terrain, and snow cover patterns, and improving predictability. Our work is conducted at 500-m resolution by necessity as higher resolution operational snow-covered area products do not exist. While there are means to obtain snow-covered area information at higher resolution (e.g. Landsat, WorldView, others), these are not operationally produced. We disagree that our approach is “incremental”. Several previous papers have modeled SWE at these scales and our results show a major improvement in accuracy. Text in the paper on L423-436, L437-451, L457-472 speak explicitly to the points made here.

We respectfully disagree with the reviewer's point that "there is no major theoretical advancement". The reviewer does not dispute our point that no previous work has used fSCA in this type of a regression. The fact that fSCA data improves predictability of SWE distribution is deeply rooted in theoretical concepts associated with interactions between topography, wind redistribution, the snowpack mass balance. In this regard, the reviewer seems to be missing a main point of our paper which is that fSCA data, when related to ASO observations, provides a means to estimate SWE during unsampled time periods. Given the high cost of ASO data and wide availability of 500-m MODIS fSCA data, this point / contribution is far from trivial and thus we disagree with the reviewers point here. Moreover, the reviewer's point that there is "no practical utility" in the approach is baseless. Quite the contrary, what we show here is that relating ASO-observed SWE to readily available fSCA and terrain information has significant utility with regard to estimating SWE during unsampled time periods. We interface with dozens of water managers in California who routinely request the exact type of information that our methodology generates. Text in the paper on L423-436, L473-487, L488-494 speak explicitly to the points made here.

Specific Comments

Page 5, 6 - Data Sources: The ASO data includes a 3 m snow-free DEM, 3m snow depth, 3 m vegetation height, and 50 m SWE information. From this it would be straight- forward to estimate 3 m SWE, and this is the appropriate scale to work at in deriving relationships between snow cover, SWE, and terrain. It makes little or no sense to aggregate to 500 m. This greatly affects the subsequently derived terrain parameters and alters the SWE distribution through averaging. The rationale in this study is to move to a scale where fSCA can be used as another variable in the analysis, but this is done at the cost of the detailed information on SWE patterns, which is key for such an analysis.

We respectfully disagree with the reviewer that it would be "straight forward" to estimate SWE at 3 m. Estimating the spatial variability of density is non-trivial and uncertainty in ASO density estimates is currently larger than that for snow depth {Painter:2016f}. There are several studies that use regression-based techniques at 500 m or greater to interpolate SWE based on topography, so this methodology is not unprecedented {Fassnacht:2003kl, Broxton:2016es, Schneider:2016cr, Rice:2011fl}. Given the goal of the estimating SWE distribution for dates without an ASO flight, it is essential that we use a resolution for which remotely sensed fSCA is operationally available.

Page 7, lines 173-174: The "PHV" model is a multiple regression that uses only physio- graphic variables as independent variables. This is fine, but to use such an approach in a predictive sense would require other information to adjust the results for a particular time (i.e. SWE observations at snow pillows, other in situ data, etc.). What is instead done here is to develop a whole suite of regression models for the different ASO dates and cross-compare against the observations on other dates to pick which fits best.

This is correct that we develop an ensemble of regression models for each ASO date. This allows use to capture the differences in the relationships between terrain and SWE as they change throughout the season. This technique is the base-line for SWE interpolations in the literature {Fassnacht:2003kl, Harshburger:2010he, Schneider:2016cr}. In order to adjust the results for a particular time, as suggested by the reviewer, we introduce fSCA as a new time-variant variable in the PHV-FSCA model. We agree there are other ways to perform the time adjustment but use this documented technique for comparison with our time-variant method.

Page 7, lines 178-179: If you are masking the regression estimates to the ASO observed snow cover areas, are you not greatly influencing the results and the reported performance of the model? Shouldn't areas without snow cover be included? This is not clear. And what about when the model is to be used in a predictive sense when there are no ASO observations?

The point the reviewer makes about masking the regression estimates with ASO is EXACTLY the reason we performed this analysis at 500 m so that we could use reliable remotely sensed fSCA from MODIS for masking the regression results when we do not have a corresponding ASO observation. The use of remotely sensed fSCA for masking regression results is well documented {Fassnacht:2003kl, Rice:2011fl, Schneider:2016cr, Bavera:2009cv}.

Page 8, lines 281-222: To pick which model dates exhibits the greatest similarity with the transfer date, the mean error in SWE at snow pillows is used. As currently described in the paper, this is unclear. Is this error between the model predicted SWE (for 500 m grid squares) and the snow pillows? How would this be later used in a predictive sense? As I understand, to find the selected model, you need to compare the results of all of them against the snow pillow observations.

This is incorrect and we apologize for the confusion. The mean absolute error in SWE at snow pillows is between the snow pillow data on the model date and the snow pillow data on the simulation date, since, as the reviewer noted, we would not have a SWE estimate for the grid cells. This has been clarified in multiple places:

L367

(i.e. those chosen based on snow pillow similarity)

L236

Here we quantify the SWE similarity between dates using the mean absolute error (MAE) of SWE recorded at snow pillows on both dates

Page 15, lines 358-359: the text refers to selected models being shown as open circles in Fig. 6. There are no open circles in Fig. 6.

We apologize for this error. This has been corrected. The selected models are shown in Fig. 7

Page 18, lines 432-434: The selected model is a simple linear regression to be applied in real time. How? As noted above, this is not clear. Do you need to simply look at all of the models and see which fits best to the SWE at the snow

pillows, and then choose that one to predict SWE patterns? How exactly is this helpful at all and how does it advance the science? For instance, if this is indeed the case, I would argue that it would be far better to look at the original ASO data, with its fine scale detail and high accuracy, and simply compare that against the snow pillow observations for a given date.

The real-time model is chosen from the ensemble of models based on the mean error in SWE at snow pillows between the snow pillow data on the model date and the snow pillow data on the simulation date. We do not choose the model date based on validating the resulting estimate with the snow pillow data, which we recognize would be subject to scale issues. Since our goal is the estimate SWE when we do not have an ASO observation, it is not justified to look at the original ASO data to select a model.

If all that is gained is to use fSCA as another parameter in the model for prediction, what about the point the authors make on page 20, lines 499-501 that there are anomalous SWE and fSCA distributions that limit the usefulness of using fSCA as an indicator for model selection?

It is true that anomalous fSCA distributions, and for that matter uncertainty in the fSCA data, may cause problems for this method. We recognize and discuss this point L534-539 and provide a potential alternative for selecting a model from the ensemble. Nonetheless, we believe that the results speak for themselves – they show that in the four years we analyzed our selected model results are significantly better than previous real-time estimates of SWE from snow pillow data. Certainly, future work should seek to mitigate the impact of issues with remotely sensed fSCA.

How are predictions to be made with any confidence in circumstances where the observed SWE (and perhaps fSCA) differs greatly from the previous conditions observed during ASO campaigns, and thus how can these results extend the ASO record as suggested? This is a fundamental weakness of the study.

We acknowledge that unprecedented conditions may be difficult to capture with this method, but this is a weakness with all statistical methods, including a universal statistical model suggested by the reviewer. All of our study years had below average snowpacks so we do not know how well the models from these four years would translate to years with larger, above average snowpacks; we note, however, that our results show good transferability from the driest year (2015) to the wettest year (2016). Nonetheless, it is wise to heed caution when predicting outside the scope of previous conditions and we have included a warning to this point.

L485-487:

Also, it is unclear as to the impact of climate non-stationarity with respect to the ability to transfer models to future years. Even so, for each year in this dataset there exists a corresponding year from which accurate predictions can be made.