Responses to the First Anonymous Review of:
"Independent evaluation of the SNODAS snow depth product using regional scale LiDAR-derived measurements"

Overall Comments:

(1) Specific deficiencies could be identified in the SNODAS model, in order to diagnose the sources of bias in the SNODAS output.

Response: This avenue was discussed among the authors prior to submission, but it was ultimately decided to be beyond the scope of this paper to delve quantitatively into the inner workings of the model. Future work is planned to identify specific shortcomings of the model for different local environmental variables such as elevation, vegetation cover, and slope/aspect by working with the NOHRSC analysts that work with the data assimilation aspect of SNODAS. However, a thorough investigation of regional environmental variables such as weather patterns and long-term climate fluctuations may require higher temporal resolution of the validation datasets (mentioned in #2). Speculations as to the sources of bias within SNODAS will be added to the Discussion section.

We don't have access to the source code so we aren't able to test any of the specific output biases from SNODAS, but we have added more speculation within the Conclusions section.

(2) The analysis could be bridged to additional years of data, either from SNOTEL or other sources in order to add an assessment of interannual model performance to the analysis (although I realize there are major limitations and challenges to this).

Response: The SNOTEL network is much too sparse to perform an analysis of this scope and extent. We will include text to outline the fact that SNODAS is considered a spatial average of conditions over a 1-km^2 pixel, while the measurements that are used for assimilation are only representing one point within that pixel. The major advantage of this LiDAR dataset was the data continuity of the snow depth changes over an area large enough to effectively assess a 1-km^2 resolution snow model. And with regards to other sources, there is simply no available independent data with which to make a valid comparison to SNODAS.

Specific Comments

(3) Page 3145 line 24: ". . .while an assimilation step give analysts the ability to decide every day whether to augment the model estimates. . ." I don't know the details of the SNODAS assimilation approach, but this statement implies that there is manual intervention by analysts with respect to the use of observations and the assimilation is not standalone. Is this true? If so this raises a lot of potential ambiguities.

Response: Yes, there is manual intervention by analysts. All available nationwide electronic point data is used to adjust the estimates created by the downscaled NWPs and NOHRSC Snow Model (NSM). Given the difficulty of keeping an automated sensor functional, the analysts must perform rigorous quality control on the data to make sure it can be used.

(4) The Introduction is well written and thorough. But it's not until the final paragraph of the section that objectives of the paper are referred to. I suggest moving this forward to near the beginning of the section to engage the reader earlier in the goals of this study. The objectives are not explicitly stated at any point in the introduction.

Response: Agreed. Some of the Introduction was reorganized and some statements were added to more succinctly state the objectives of the paper.

(5) Page 3150 paragraph 1: I like the hourglass approach to in situ sampling over a 500 x 500 metre area. But were only ~50 snow depth measurements made at each hourglass? This seems like a very small number. Calculating conservatively, a 500 x 500 metre hourglass is composed of over 2000 metres of linear sampling distance. 50 measurements equates to a snow depth measurement only every 40 metres. Does this capture the local scale variability? You have to walk the whole hourglass anyway, why not make more measurements?

Response: The manual measurement team was stretched fairly thin with the amount of ground they desired to cover and the number of people involved to make the measurements. At the resulting resolution, the surveys definitely did not capture small-scale variability, but it must be remembered that the in situ sampling was not intended to support LiDAR validation. Rather, the surveys were conducted to complement the POLSCAT airborne radar mission, which bore different goals for ground validation.

(6) Page 3150 paragraph 2: Perhaps I'm missing something, but why is it 'paramount' that no snow melt occurred between the two dates of LiDAR acquisition? You have produced a snow depth difference field from all three datasets, and the assessment is of delta depth. Given that the region was not snow free at the first date, it's okay to have a negative change.

Response: You are absolutely correct. The way the text was written implied that no melt could have occurred in order to conduct this study using snapshots of snow distribution. This is simply not the case. What we intended to show by including the melt discussion and figure was that by using the routines already present within SNODAS we could rule out snow depth loss due to melt and sublimation as contributing factors to model bias with regards to this study. This works to narrow down discrepancies between modeled and measured snow depth change to merely LiDAR error and the densification/compaction routines in SNODAS in all the regions where no melt occurred between the survey dates. This portion of the manuscript will be rewritten considerably.

(7) Page 3150 line 22: "…estimates of snow melt due to incoming solar radiation and sublimation…". This statement requires clarification. Snow melt and sublimation are two different phase changes that will reduce the snow mass, and

they are driven by different processes.

Response: Yes, the text here was part of the melt discussion that will be rewritten. The routines within SNODAS that estimate melt incorporate the physics of sublimation due to wind and solar radiation. Therefore, where model estimates of melt were negligible, we can narrow our focus to other processes (i.e. new snow density, densification, and compaction) that may have caused discrepancies between SNODAS and LiDAR.

(8) Figure 5 is effective at showing the tendency for SNODAS to increasingly underestimate snow depth relative to the in situ measurements as snow depth increases. Figure 6 essentially shows a similar pattern for the LiDAR (although lower in magnitude). Why not combine these figures, using different symbols for the in situ vs. SNODAS and in situ vs. LiDAR results? This would provide a direct comparison relative to ground measurements, including the systematic bias over deeper snow.

Response: This is a terrific suggestion. A combination of Figures 5 and 6 condenses the information that both figures are conveying.

(9) Page 3153 line 18: a +/- value of 13 cm is provided for the LiDAR data relative to the in situ measurements. But this is somewhat misleading as there is systematic bias in this comparison: LiDAR snow depth is always shallower than the hourglass measurements. Would it not be possible to bias correct the LiDAR estimates of snow depth based on these results?

Response: For the twelve in situ sites, the upscaled 1-km^2 LiDAR snow depth change is consistently 5-10 centimeters less than the upscaled manual measurements. This does result in a negative bias, which can physically be explained by the combination of a few important factors.

1. It can be assumed that the vast majority of the December LiDAR pulses completely penetrated the low-lying brush canopy layer all the way to the dirt/snow surface. However, this also means a portion of the LiDAR pulses were not able to make it through the brush for a ground reflection, which would result in lower measurements of snow depth when differenced with the February surface.
2. Snow survey probing teams are instructed to record measurements only when rock or wood causes a hard report, or when the probe tip is removed from the snow dirty, which could cause slight overestimates of snow depth.

Consequently, the 5-10 cm underestimation in the LiDAR measurements is well within the vendor-recommended uncertainty and does not ultimately affect the final comparison to SNODAS in a strong manner.

(10) Page 3154 line 15: Is it worth summarizing the regression results in a table? There is not much detail provided here.

Response: The results were not included due to the fact that the LiDAR snow depth change was overwhelmingly the greatest cause in discrepancy between LiDAR and SNODAS. The other predictor variables that were hypothesized to have an impact on SNODAS – LiDAR differences turned out to not be nearly as important as the actual snow depth change.

(11) Figure 8: I suggest using different symbols for the points corresponding to each of the outlined areas in Figure 9. This would explicitly show which points come from which area.

Response: Agreed. Revealing each region specifically within the scatter plot better quantifies the results and strengthens the hypothesis of region-specific uncertainties.

Editorial Comments

(12) Drop 'Independent' from the title

Response: The word 'independent' was included to highlight the fact that 100% non-biased validation datasets for SNODAS are *extremely* rare, especially at the regional scale.

(13) Page 3143 line 2: "...an important ecological component of Earth's water cycle." Not clear what is meant by 'ecological component'.

Response: Removed 'ecological' from the sentence.

(14) Page 3143 line 5: 70% of the water supply to what geopolitical region?

Response: 70% of the water supplied to populated regions of the Western U.S. originates in mountain snowpacks.

(15) Page 3143 line 9: consider changing 'hydrologic snow models' to 'distributed snow models'

Response: Added "Distributed..." to the start of the sentence

(16) Page 3151 line 4: change 'snow height' to 'snow depth'

Response: Changed "...model estimates of snow height change." to "...ΔSNODAS"

(17) Page 3152 line 10: the use of 'substantiate' seems odd here

Response: Re-worded the sentence.

(18) Page 3153 line 20: The relationship between SNODAS and LiDAR snow depth has a R^2 (coefficient of determination) of 0.72, but note that this is described as 'correlation' in the text which should be expresses as r not R^2.

Response: You're correct. We removed all instances where r^2 is referred to as correlation and replaced with explanations of the independent and dependent variables.

(19) Page 3155 line 9: provide a reference to the Figure 8 along with the mention of the 'pink vertical stripe'. I would change 'stripe' to 'shading'.

Response: This was changed to shading.

Responses to the Second Anonymous Review of:
"Independent evaluation of the SNODAS snow depth product using regional scale LiDAR-derived measurements"

Overall Comments:

(1) I do not understand the comment "it is paramount that no snow melt occurred between the survey dates in order to properly assess the snow depth component of SNODAS using LiDAR estimates alone." (p. 3150, l. 20-22). I can understand that snow accumulation is the primary phenomenon being examined, but why is it a problem if snow melt occurred during the 81-day span separating the measurements? In section 4 the integrative nature of the measurements is discussed (p. 3153, l. 21-24), and it does not seem to be an obvious problem if snow melt is another contributing process.

Response: What we were aiming to accomplish by analyzing the melt estimates of SNODAS was to rule out the melt routines within the energy balance model as a contributing factor of uncertainty in the context of this study. If we are confident that we know where melt occurred between the survey dates, then we can narrow down which processes control SNODAS uncertainty in different locations. The wording was incorrect for the draft, and we are rewriting the discussion of melt in the SNODAS description section to convey this message.

(2) SNODAS uses quantitative precipitation as its forcing, and assumes a constant bulk density for newly fallen snow. Therefore, SWE is the primary state variable produced by that system, while snow depth is derived as a function of SWE and snowpack density. This relationship is acknowledged (but with the positions of SWE and depth reversed) in the introduction (p. 3143, l. 14), but the following comment that "snow depth varies considerably more than bulk density over space" (p. 3143, l. 16-17) serves to diminish the importance of modeled snow density. However, this came back to me as I read the discussion of the in situ vs. SNODAS comparisons shown in Figure 5b (p. 3152, l. 14-19). SNODAS appears to underestimate snow depth when the observed snow is deep: is it possible that SNODAS overestimates compaction or initial density in these areas of high accumulation?

Response: While depths vary more than density over space, depths are also much more easily quantified over large extents and high spatial resolution, pointing to the power of the LiDAR data. That comment is just saying that high-resolution information about depth is much more powerful than an equivalent resolution density knowledge. You are correct that the method SNODAS uses to determine depth is based on getting the density right. By comparing the modeled and measured depths we can quantify SNODAS uncertainty in different regions. Yet we still do not have a good assessment of SNODAS ability to model SWE, other than the age-old streamflow measurements that do not explain how or why the model got it right or wrong. However, there is such a substantial amount of uncertainty in modeling snowpack compaction using densification models, that I agree those parts of the NSM could be a major source of uncertainty in the high accumulation regions.

(3) I’m having a little trouble with the comparisons shown in Figure 8. Basically, the RMS difference between the in situ and LiDAR data is treated as a random error, but the data shown in Figure 6 depict a systematic error causing the datasets to differ, as discussed in section 4 (p. 3153, l. 3-18). I suppose the use of the RMS difference between the data sets means that neither is considered authoritative, and is a hedge against preferring one over the other. However, given that the LiDAR depths are systematically lower than the in situ depths, it seems more appropriate to treat the LiDAR data as negatively biased, and for example to depict the uncertainty associated with them as a more narrow range centered around a negative value. Perhaps this is not the ideal approach given that the difference between the two sets of observations seems somewhat proportional in nature (i.e., the slope of the line in Figure 6 is significantly less than 1.0), but further discussion of the choice to treat the RMS difference as a random error estimate seems warranted.

Response: Though Figure 6 does show a systematic underestimation by the LiDAR, we believe that the magnitude of the bias (5cm @ shallow sites – 15cm @ deep sites) is well within the noise of the LiDAR data in the first place. Additionally, the in situ measurements tend to undersample the variability of snow depths more and more as the mean snow depth increases, as shown in this modified Figure 6 below. If we really want to quantify LiDAR biases using the twelve HG sites, we would have to only consider the LiDAR measurements exactly at the probed locations due to the inherent depth variability at short length scales. However, the uncertainty of the reported manual measurement locations is ~7-10 meters due to the mapping-grade GPS receivers that were used at the time.

The scale difference between SNODAS pixels and probe measurements is very apparent in Figure 5a, as they display a very low correlation. For a reason similar to the LiDAR/HG argument made above, the low correlation between SNODAS and HG stems from the undersampling of the HG surveys not only at the hillslope scale but at the regional scale as well. SNODAS does a good job of representing coarse regional scale patterns over very large areas, but naturally has difficulty with the hillslope and micro scale variability that can have substantial effects on water storage. The main point of Figure 5a (which will be combined with Figure 6) was to show how difficult a task it is to evaluate such a large-scale snow model using sparse manual measurements, which up to this point has been the only technique available.

The slope of the best-fit line in Figure 6 can be explained by the undersampling of the manual measurements as the mean site snow depth increased. The point to point comparison was performed only after averaging all ΔLiDAR pixels in a 10m radius surrounding the reported in situ measurement location, resulting in the large support discrepancy that was alluded to in the text.

(4) The discussion for “region #1” (p. 3156, l. 5-19) suggests that SNODAS has failed to account for persistent snowpack sublimation in that region, which is certainly possible, even likely. However, the discussion implies that wind speeds that drive SNODAS and its simulation of snow sublimation are inaccurate because there are no nearby SNOTEL sites. I don’t find this convincing. I don’t see that the proximity of a SNOTEL would make any difference with respect to whether or not wind speeds are well represented in SNODAS forcing data in the area. Isn't it possible that observed wind speeds are assimilated into the forcings used by the SNODAS model, since it uses NWP analyses as its primary forcing data, and that many of these come from observation stations other than SNOTEL sites? Perhaps the winds are well represented, but the SNODAS model nevertheless fails to simulate the extent of snow sublimation occurring in the region.

Response: We will change this discussion of distance from SNOTEL sites, specifically, to distance from any station whatsoever. The North Park region is a very sparsely instrumented area and there isn't any other available forcing data to be found. However, I believe you are getting at the root of the main error source for SNODAS. Density modeling is the most likely explanation for model uncertainty given that estimating new snow density and snowpack compaction over time are the two most difficult processes to quantify by physically based models.

(5) In the discussion for "region #3", why does sub-kilometer scale heterogeneity of snow distribution cause SNODAS to underestimate, and not overestimate, snow accumulation (p. 3156, l. 10-12)?

Response:

Editorial Comments:

(6) P. 3148, l. 1: organizeded -> organized

(7) P. 3155, l. 22-23: the geographic location of the pixels are in a region -> the pixels are in a region

Responses to James McCreight's Comments on:
"Independent evaluation of the SNODAS snow depth product using regional scale LiDAR-derived measurements"

Overall Comments:

(1) A comparison of SNOTEL and SNODAS is lacking which will greatly improve the context of the results. Because SNOTEL roughly governs SNODAS (via data assimilation), this context will provide substance for extended discussion.

Response: This was an avenue that we considered early on, but abandoned it for two reasons. 1) SNOTEL is only a point measurement, and as such the assimilation of SNOTEL data into the model are only approximations for a larger area. SNODAS estimates are considered to be approximations of the average conditions within a 1 km grid cell. 2) We are simply comparing two snapshots of the snow conditions over a large area. Since the LiDAR over flights only covered two SNOTEL sites (Columbine and Rabbit Ears), we would gain context about two wind-sheltered points within two pixels of the survey swath.

However, that is not to say that an analysis of SNOTEL and SNODAS would be inconsequential. Further work examining a time series of SNOTEL and SNODAS could shed some light on how well the model is distributing the assimilated data over space. We did look at SNOTEL depths and swe between the LiDAR flights and compared them to the SNODAS output over that time period, and we will consider adding another figure depicting the time series of snow depths for the two overlapped SNOTEL sites and their corresponding SNODAS pixels.

(2) A discussion should more generally frame the results in terms of potential water yield. More importantly, what are the next steps for improving SNODAS using future LiDAR data sets similar to those in this paper? What aspect of the LiDAR acquisition will be key to get right next time? Are the LiDAR errors actually small enough for a comprehensive validation? This is important to consider because LiDAR snow depth is perhaps the best opportunity to understand and improve the SNODAS products or other, similar model estimates at large spatial scales. Could the LiDAR be assimilated? How much ground truth would be necessary to properly bias correct the LiDAR? Etc.

Response: A main goal of this paper is to foster more discussion about how to effectively use LiDAR snow depth campaigns for model calibration and validation. Even so, we need to be more vigilant about data quality when collecting the data in the first place, and that quality can change drastically depending on who/what is collecting the data. The question "Are the LiDAR errors actually small enough for a comprehensive validation?" is a very important one that we have struggled with over the entire course of this work. Though not collected to specifically validate the LiDAR, the HG in situ surveys were a boon to the study and really added another dimension to the LiDAR, and it was our judgment that the RMSD between the LiDAR and HG snow depth change was a reasonable uncertainty estimate. Additionally, if we had noticed no trends whatsoever in the comparison of LiDAR + SNODAS (Figure 8), we likely could not have continued. However, by quantifying the error for this particular survey and defining regions where SNODAS disagreed with the LiDAR, we felt that this study could be a first step in really nailing down some of the issues facing large-scale energy balance models. All of this speaks to the importance of ground truthing remote sensing data.

Specific Comments

(3) Given the analysis in the paper, my lingering question is "why is SNODAS wrong?" Or, why is SNODAS right?

Response: After performing this comparison, we were extremely surprised at just how well SNODAS was able to predict snow depth change given that the primary goal of the model framework is to predict SWE. The estimation of depth heavily relies on physically based assumptions about the density of new falling snow, and how that snow evolves and compacts over time. With that being said, the three regions containing the largest model/observation discrepancies are definitely the traditional 'problem areas' for snow models. SNODAS is such a complex model framework that it is likely a combination of many small assumptions in the physics of the modeled processes that are causing particular physiographic locations to overestimate depth and other to underestimate. Ideally, higher temporally resolved LiDAR flights at this scale and extent would allow for more of a concise evaluation of the shortcomings of SNODAS.

You are likely correct that SNODAS performs more poorly the further from an assimilation data point, and should be an entire publication in itself. In fact, Region 2 shows a stark contrast between ΔLiDAR and ΔSNODAS as the elevation decreases to the east of the Columbine SNOTEL. Nevertheless, as of yet we can merely show the locations of the discrepancy and do some cursory hand waiving as to the causes. We will include more discussion on this matter in the final manuscript.

It will be impossible to answer this comprehensively because we don't know the exact assumptions in SNODAS. To address comment 1 of reviewer 1: yes, there are "MODS" in SNODAS (at least this is generally believed). This is (still) fairly standard practice for operational products (e.g. Seo et al, 2009). New validation products, such as presented in this paper, will hopefully lead to comparisons of MODS assimilations and automated assimilation procedures and advance the science. The upshot of the MODS is ambiguity in how to improve the results. This will make for challenging speculation in the discussion. However, efforts along these lines could be a significant benefit to the community and help push the science forward.

My thoughts on the initial question:
a) SNODAS is going to be correct owing to SNOTEL observations
b) It's going to be wrong moving away from SNOTEL in space as (MODS) assumptions about variability break down.

To me this explains why SNODAS is not simply biased, but the line of best fit intersects the 1:1 line. I'd guess that the intersection is roughly near the magnitude SNOTEL observations. That's not going to be exactly true, but makes a reasonable story. I think this general idea is sketched on P3154 L3, but it deserves clarification and expansion along with the relationship of the SNOTEL observations to the results. There should be speculation about why the assumptions moving away from SNOTEL are likely wrong and how we might fix that. Wind is mentioned in passing. How about vegetation? Other differences in physiography with the assimilated SNOTEL observations? The entire regression analysis centered on page 3154 strongly suggests to me that the SNODAS assimilation/MODS (which are the errors away from SNOTEL) are not based on any of these explanatory variables which we commonly expect to govern snow depth. This may or may not be true, but it appears that such predictors are not the basis of the MODS. I currently know of now snow depth/SWE assimilation technique that actually uses such variables. So, it's not really surprising.

I know there are already multiple analyses of different products in this paper, but a comparison is needed of the SNOTEL and SNODAS used in the study. This will help to illuminate the above points. For the two SNOTEL locations in the LiDAR footprint, I would suggest also including the LiDAR spread and mean information for the SNODAS pixel and also for a, say, 10-15m radius centered on the SNOTEL.

Also hindering interpretation of the results is that SNOTEL information is somewhat hard to see where it does exist in the paper. While it's nice to see the SNOTEL positions overlaid, it makes it difficult to interpret the colored values in figs 2, 4, and 9 at the location of the SNOTEL, which is of acute interest. This is challenging to fix and make obvious, and is part of the reason I suggest treating these comparisons in a separate figure. Improving the readability of this figure is needed to help with interpretation in the spatial context. I don't have a great suggestion for how best to do this. Smaller symbols would help fix this, but be more difficult to see. Perhaps empty squares or diamonds centered on SNOTEL? Similarly the HG symbols can block the information that they overlay and make interpretation of the underlying values difficult.

Figure 1 is very nice. My concern, again, is about promoting interpretability. Showing the ΔLidar in the figure detracts from our ability to use topography in that region as context for interpretation of results. I suggest removing it here and combining it with Figs 2 and 4, in this order 1, 4, 2 as panels of a single figure. I think being able to compare ΔLidar, ΔSNODAS, and ΔSNODAS-ΔLiDAR in the same figure is important to the interpretation. Flipping pages detracts from the interpretation. I'd also include SNODAS values in another panel in the same figure. This will help promote a coherent discussion where these things are easily compared.

Also, would a vegetation figure (e.g. NLCD or MODIS) contribute to interpretation?

As mentioned above, the results (the difference between LiDAR and SNODAS changes) should be put into the context of difference in potential water yield or potential energy balance effects. I'm talking about back of the envelope calculations with simple assumptions. Other suggestions for discussion questions were also offered above.

The last thing I'm interested in, which maybe less relevant and more of my own pet interest, is the sub-grid distribution of ΔLiDAR for each SNODAS pixel, where the SNODAS value falls in the distribution, and trying to explain the variance using predictor variables. (OK, This could easily be a separate paper).

(4) Horizontal error bars on figures 5 and 6 could place SNODAS more broadly within the measurement context.

Response: Agreed. After combining figures 5 and 6, we will add error bars for the insitu measurement surveys.

(5) (Same as reviewer 1, comment 7; P3153 L18) It nags at my conscience that you're using ±13cm from RMSD as the error range for the LiDAR. I think a simple discussion (1 sentence?) justifying why this is appropriate would be helpful. My concern is that the errors are biased so that they are not symmetric about zero. If you plot the distribution of these errors, the mean is not zero. The assumption of ±13cm is similar to assuming 1 standard deviation of a mean-zero distribution? Also why is 1 standard deviation, or whatever exactly RMSD represents, appropriate? It's not the same as 1 standard deviation if there's a bias. The conclusions are somewhat dependent on this assumption, so it should be clearly argued.

Response: We don't believe that the sample size of just 12 in situ points where the LiDAR was analyzed provides enough spatial representation of the 750-km^2 study area to apply a bias correction. With a bias on the order of the LiDAR uncertainty, we did not feel that the bias correction would largely affect the ΔSNODAS vs. ΔLiDAR comparison results.

(6) (Same as reviewer 1, comment 3; P3150 L20) The argument about why melt being insignificant is lost on me. Why is this important? Clarification needed. Related to this, a time series "spaghetti-plot" of all the SNODAS pixels along with their mean would illuminate SNODAS behavior during the Δtime.

Response: This analysis was intended to simply narrow down the processes that could be causing uncertainties in SNODAS. By eliminating melt as a contributing factor, the focus turns to the densification, compaction, and new snow density routines within the NSM portion of SNODAS.

(7) (Same as reviewer 1, comment 5; P3150 L22) Do you mean just change in depth (what I'm calling Δ) instead of melt? Sublimation doesn't really cause melt; it probably has the opposite effect like sweat cools the body.

Response: Yes these are two completely separate processes, and we removed any mention of sublimation from the text, since we did not perform an analysis on the sublimation products. SNODAS treats processes affecting mass loss individually within the energy balance model, and melt and sublimation are modeled as individual products: sublimation due to wind, sublimation within the snowpack, and melt due to solar radiation.

(8) (P3151 L19) The word "model" is a bit ambiguous. You could change to "its" or "SNODAS".

Response: We changed this to "...SNODAS uncertainty."

(9) (P3152 L26) "...mean HG" wasn't defined as "mean HG difference" previously. I assume that's what you mean. Generally I'd suggest revising the notation to use deltas, it would be clearer: ΔLiDAR, ΔSNODAS, ΔHG.

Response: This suggestion is great. We will go back and change all the surface difference variables to include deltas.

(10) (P3153 L21) Seems like bias and RMSE should be mentioned in this paragraph. It's on the figure and important.

Response:

(11) (P3153 L21) "...potential explanatory physiographic variables" might be a better expression than "...potential physiographic parameters".

Response: This change was applied. Better refers to the regression method that we used.

(12) (P3155 L6) This paragraph would benefit by starting with its final sentence.

Response: This entire paragraph was rewritten.

(13) (P3155 L17) SNOTEL used for assimilation are also in the trees, which affects solar radiation as well. This is a point that seems to be worth exploring or mentioning.

Response: This point was added to the paragraph at the end.

(14) (P3155 L22) "changed" is vague. Is "accumulated more snow" better?

Response: We replaced "changed more than" with "accumulated more snow than".

(15) (P3156 L3) "over-distributing" is just vague. You could use more words if you think the point is important, but I'd just remove "over-".

Response: Yes, it is vague so we removed "over-".

Reference:

Seo, D. J., Cajina, L., Corby, R., & Howieson, T. (2009). Automatic state updating for operational streamflow forecasting via variational data assimilation. Journal of Hydrology, 367(3), 255-275.

Final Responses to Anonymous Reviewers (12/1/14)

Major Comments:

(1) Identification of deficiencies in SNODAS: I agree that detailed analysis of this issue is beyond the scope of the paper, and is obviously hampered by no access to the source code. But you need to be careful, therefore, with the wording in the paper. If you can't diagnose model deficiencies then you need to modify the justification provided on line 397: "…a much more spatially extensive dataset, such as the CLPX-2 LiDAR, is required for determining the underlying causes of model error."

Absolutely true. The intent of that statement is to emphasize the advantage of spatially continuous data over randomly sampled in situ validation methods. It was not meant to imply that LiDAR is the only method for determining causes of error. We modified the sentence to read:

> "Also, without access to the NOHRSC Snow Model source code, it is not possible to discern the physiographic factors that could be influencing such discrepancies using the relatively small sample size of the in situ measurements alone. A much more spatially continuous dataset, such as the CLPX-2 LiDAR, is necessary to begin quantifying the underlying causes of SNODAS uncertainty."

(2) Extend analysis to other uses with SnoTel: OK. I agree with the response that use of SnoTel data is extremely limited due to the pointwise nature of the measurements. I was simply wondering if a long time series of SnoTel versus SNODAS would have any value.

This is something that I have thought about since the first stages of writing this manuscript. I believe that the scope of a complete time series analysis of SNOTEL vs SNODAS would be too large to include simply as a validation detail in this paper. A comparison using a large collection of stations over a long time period could show some very interesting trends due to the data assimilation in SNODAS. Such work could be a powerful complement to this spatially complete study.

(3) Analyst intervention in the SNODAS assimilation: it's still not clear whether these are objective data QC routines (i.e. to remove erroneous observations) or more subjective analyst-driven decisions. Can this be clarified?

In the introduction we added some description of the analyst's role in the data assimilation to explicitly state that they usually only adjust for mesoscale discrepancies.

(4) Clearly state objectives in Introduction: The introduction reads well, but I still think this section would be strengthened by a more detailed/itemized statement of the objectives that occurs before the final sentence of the section.

We altered some of the text in the introduction to give more of a sense why we think the LiDAR is important for examining a model like SNODAS when source code is not available.

(5) Hourglass sampling: OK. But next time give them magnaprobes and tell the samplers to make more measurements!

Duly noted

(6) Snowmelt: OK

(7) Snowmelt/sublimation: OK

(8) Combine Figures 5/6: OK

(9) Potential bias in LiDAR measurements: "…the 5-10 cm underestimation in the LiDAR measurements is well within the vendor-recommended uncertainty and does not ultimately affect the final comparison to SNODAS in a strong manner." I agree that the impact of this bias is minor. But were this random instrument / retrieval, one would expect a randomly distributed error around zero, not a consistent bias. The bias is certainly consistent with a systematic over-probe in the in situ dataset…

On line 422, I suggest the negative bias value be provided in the text.

The bias value of 12 cm was added to the end of the sentence.

(10) Regression results: I don't feel super strongly about this, but in the absence of a summary table with actual statistical results, I'm left with the subjective statement that "…upscaled snow depth changes were overwhelmingly found to best predict the discrepancy." Even if only one variable provided significant results, I still think a summary should be provided.

We added a summary table of the predictor variables to explicitly show the relative importance of each variable to the regression model.

(11) Figure 8/9: I don't see different symbols for each region in the scatterplot…

The scatter plot (Figure 6) was altered to show the pixels within each of the regions that displayed the largest SNODAS/LiDAR differences. Also, Figure 7 was changed to show the same colors for each of the three regions in Figure 6.

Editorial Comments:

(1) Line 60: "…this study purports to examine…" 'Purports' is not the right word choice here.

Yes. This was changed to "aims".

(2) Line 66: "over a vast extent." I'm not sure I would describe this study are as 'vast'.

Changed to large geographic area.

(3) Line 175: change to "…data are used…"

Changed.

New and Revised Figures

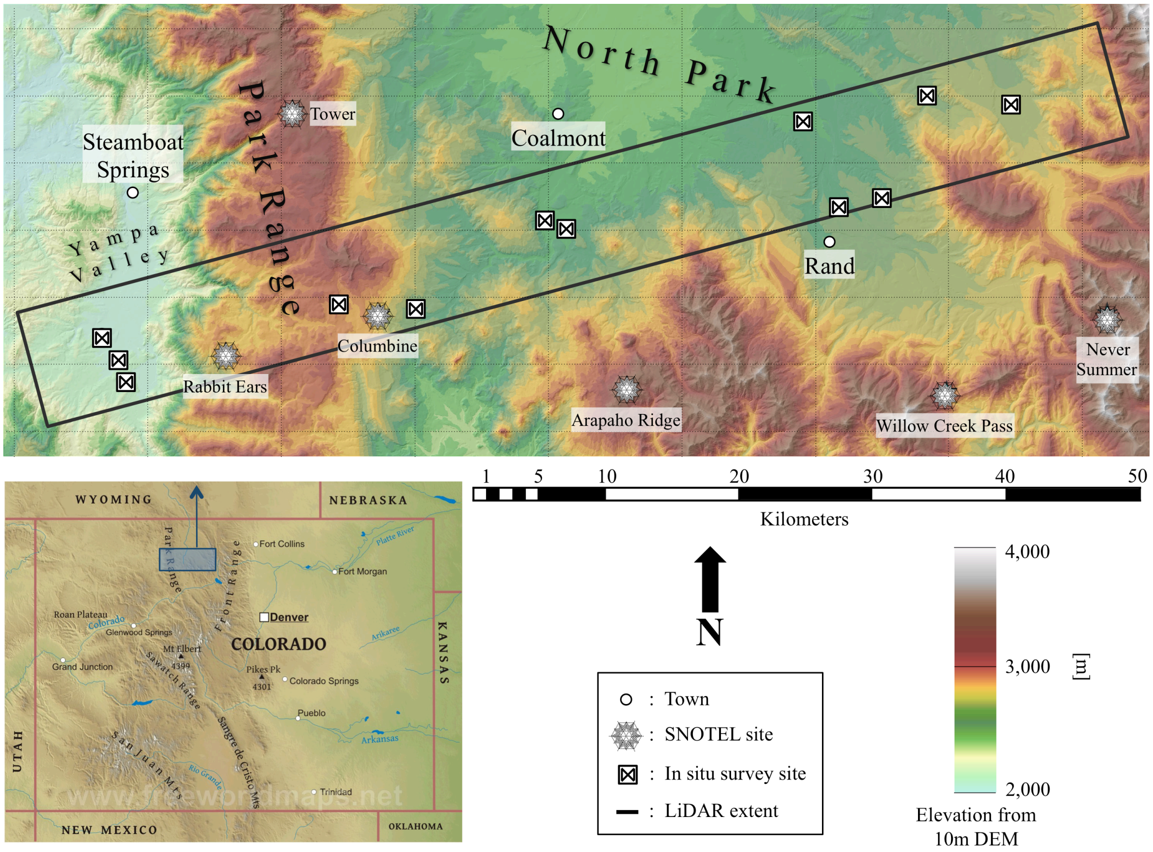


Fig. 1. Location of the CLPX-2 LiDAR footprint in Colorado, USA with nearby towns, SNOTEL sites, and IOP in situ hourglass (HG) measurement transect locations indicated.

Figure 1 was modified to only show elevations on the location and feature map.

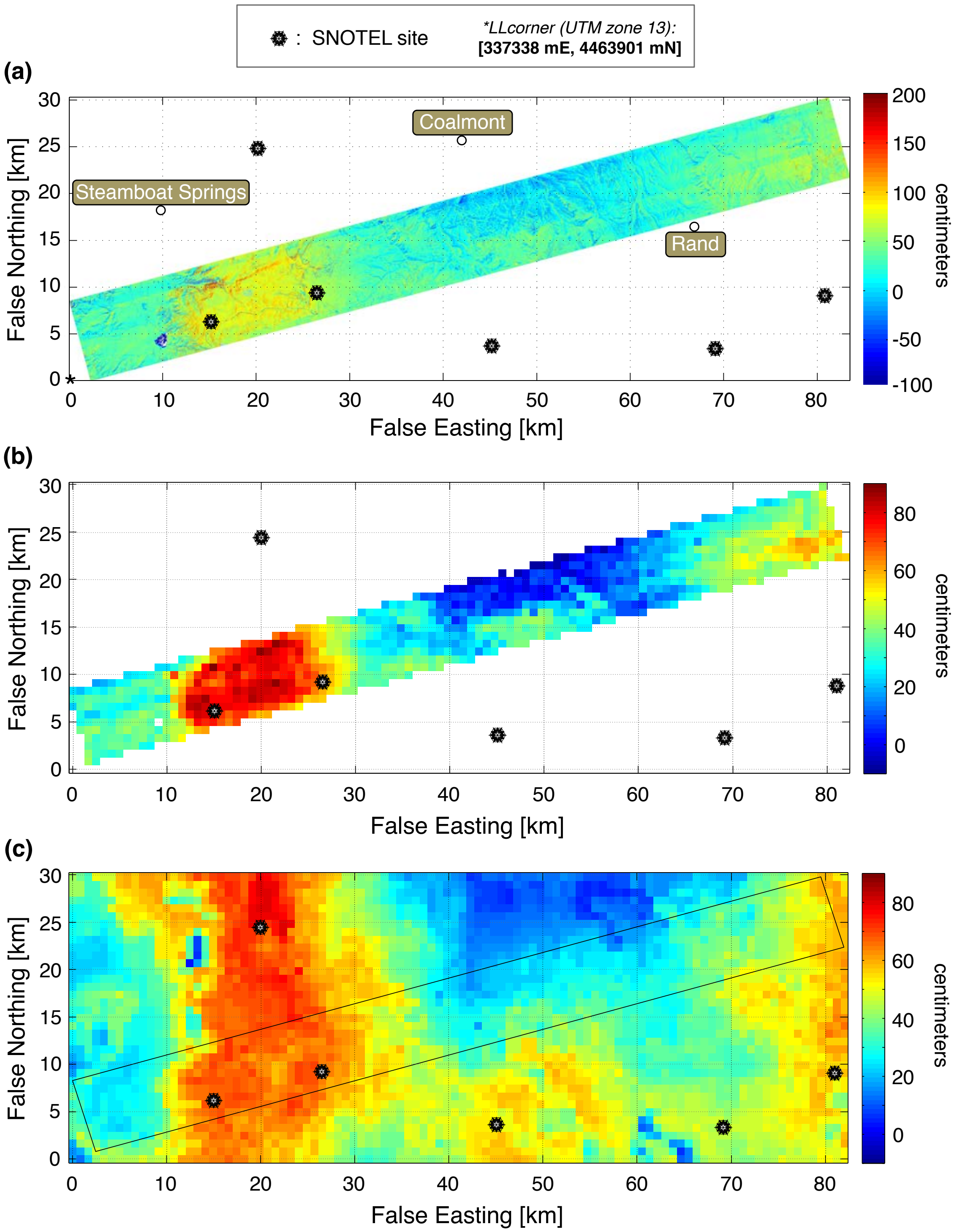
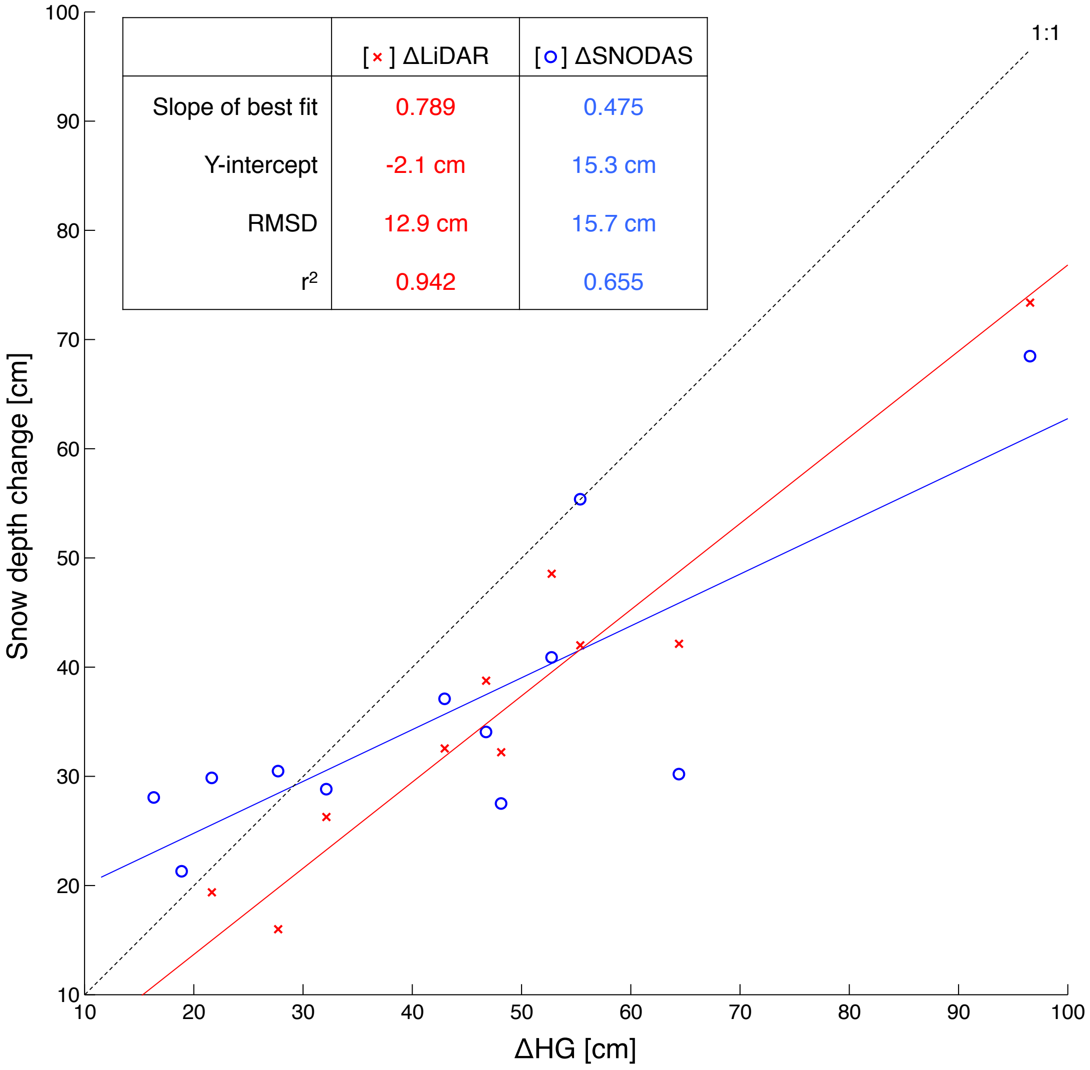


Fig. 2. Estimates of snow depth change between December 3rd, 2006 and February 22nd, 2007 along with the six nearby SNOTEL sites used by SNODAS for data assimilation. (a) represents the 5-meter resolution LiDAR-derived snow depth change, ΔLiDAR, (b) shows the upscaled LiDAR estimates of snow depth change at the 1-km SNODAS resolution, and (c) is the difference in SNODAS estimates of snow depth, ΔSNODAS, on the dates of the LiDAR acquisitions, with the LiDAR footprint outlined for reference.

Figs 2 & 4 and a portion of Fig 1 were combined as a panel figure for easier comparison.



	[×] ΔLiDAR	[○] ΔSNODAS
Slope of best fit	0.789	0.475
Y-intercept	-2.1 cm	15.3 cm
RMSD	12.9 cm	15.7 cm
r^2	0.942	0.655

Fig. 3. ΔSNODAS (blue circles) and ΔLiDAR (red crosses) snow depths evaluated over the centers of the twelve ΔHG measurement transects. The ΔLiDAR points were determined by averaging each reported 5 meter resolution ΔLiDAR snow depth within a 10 meter radius of each reported HG measurement, then averaging again over each HG transect site. The ΔSNODAS estimates were the areal-weighted averages of the four nearest SNODAS pixels to the center of each HG transect site.

Figures 5 and 6 were combined to create a more succinct figure.

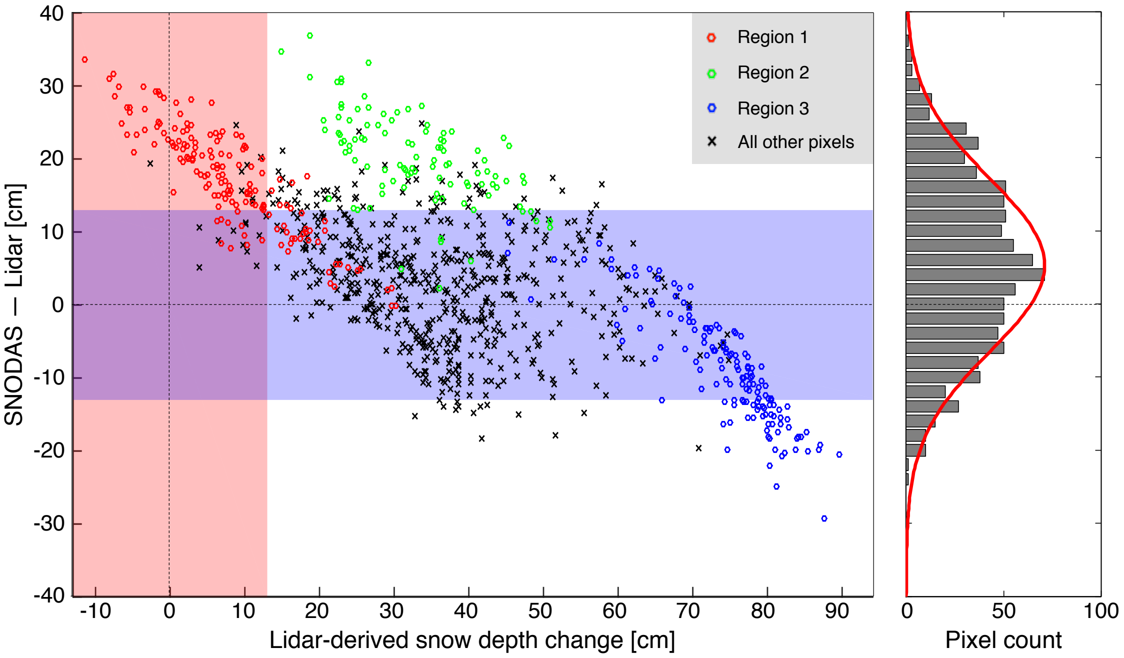


Fig. 4. Pixel by pixel ΔSNODAS–ΔLiDAR differences of snow depth change plotted against the mean ΔLiDAR within each SNODAS pixel. The pink and blue shaded areas represent the ±13cm error threshold for the upscaled LiDAR estimates determined from the CLPX-2 in situ ΔHG measurements. Three grouped regions are depicted that were found to contain model pixels of significant ΔSNODAS–ΔLiDAR discrepancies (|±13|cm). Also plotted is a histogram of differences showing a bias toward higher SNODAS estimates across the CLPX-2 study area.

Figure 6 was changed to display the 3 regions of the study area more succinctly within the scatter plot. The previous version simply used circles to approximate the pixels within the three regions.
Table 1 was added to present the results of the regression analysis.

Predictor Variable	β-coefficient	|t-statistic| (|β| – Std. Error)	Relative Importance
Mean ΔLiDAR (cm)	-17.075	52.936	62.8%
Mean Elevation (m)	3.729	13.208	15.7%
ΔLiDAR IQR (cm)	2.282	6.625	7.9%
Vegetation Cover (%)	-2.048	4.050	4.8%
Mean Vegetation Height (m)	1.989	3.742	4.4%
Vegetation Height IQR (m)	1.097	2.970	3.5%
Elevation IQR (m)	-0.291	0.735	0.8%

Table 1. Summary table of the stepwise multiple linear regression for predicting ΔSNODAS–ΔLiDAR differences using seven explanatory variables from the CLPX-2 LiDAR surveys. Relative importance of the predictors was determined from the ratio of the individual t-statistics to the sum of all t-statistics.