Response to referee 1

We thank the referee for his/her valuable comments on the manuscript. Author response to the comments and the changes we plan to make in the revised manuscript are detailed below. Author remarks are written in red, additions to manuscript are written in green. Reviewer remarks are written in black.

General Comments:

This manuscript is fairly straightforward and logical. It is mostly well written, but some sections could use clearer language or additional explanation. The study seems worthwhile, because there is a large amount of uncertainty in global snow estimates and assuming a constant snow density over time and space (as in GlobSnow) is not realistic, so any method that could introduce empirical density estimates would be beneficial. However, the improvements in SWE estimation did not seem all that significant/impressive, so I wonder if post-processing is not the best approach. I have some additional specific comments below about the methodology, interpretations, and presentation.

Specific Comments:

The title is a bit misleading, because it makes it sound like the dynamic snow density is incorporated directly into the retrieval, when in fact it is applied in post-processing.

We appreciate the reviewer's comment on this. However, we feel that changing the tile of the article is not necessary. For the end-user, there is no difference where the new dynamic density information is injected in the overall processing chain. The key point is that the dynamic snow density has been applied for the GlobSnow product in the retrieval framework and this provides a significantly improved product for the end users. Approach to include it during the assimilation phase is being studied but for the time being the presented approach has produced the best results and will be the baseline for the next GlobSnow product version.

Line 64: Do the authors have a citation that they can provide to show that the Sturm et al. approach did not improve retrieval skill?

Citation to relevant report will be added:

Luojus, K., Pulliainen, J., Takala, M., Lemmetuinen, J., Kangwa, M., Smolander, T., Cohen, J., Derksen, C.: Preliminary SWE validation report, European Space Agency Study Contract report, 2013.

Lines 141-142: Why was there so much more validation data than implementation data in the Eurasian portion of the analysis, but the opposite in North America?

Large amount of validation data was wanted for Eurasia because different versions of the snow density fields were tested only in Eurasia. We also wanted to use independent SYKE dataset from Finland for validation to see if there were clear biases in the RIHMI-WDC data, which was used both for validation and implementation.

The North American implementation dataset contained 869 sites after all the filtering steps and not all these sites were producing data already in 1979. Thus, the North American implementation set contains only 681 sites for 1979-1989 clustered mostly in few parts of the continent and if there were much less sites available implementing dynamic densities for the whole continent would become challenging.

Line 163: The method does not produce SWE estimates for mountainous areas, but most of the SNOTEL data come from mountainous regions. How do the authors justify using mountain snow to determine the density for snow that is not in the mountains?

SNOTEL sites located in areas for which the SWE is not estimated were removed from the North American implementation dataset. This filtering step was not mentioned in the article by mistake but will be added to updated version.

Figure 5: When the authors compare the in-situ densities used to reprocess GlobSnow to their nearest neighbors (used for ground truth), is a similar pattern seen, or does it better fit the 1:1 line? Is the interpolation smearing out the range/variability of the density values? If so, the authors might search for a different interpolation method. Additionally, by interpolating across large distances (which may have different land cover, elevation, etc.), the authors are effectively taking local density estimates assuming that they are representative of large areas. Do the authors think this is a fair thing to do? Might it be better to use an interpolation method akin to PRISM (Daly et al.) to take land surface properties into account?

We are using kriging interpolation to form a continuous map of the snow densities. Utilization of a more complex interpolation scheme (than the kriging interpolation) could be a potential avenue for future investigations, but the current agreement between the interpolated snow density maps and independent densities (the densities from the validation dataset) is a rather good. Using nearest neighbor interpolation would probably be quite problematic, and therefore we have not applied that. Further, we know that the variability of densities may be quite a bit higher in mountainous regions, so those are masked out, i.e. pixels where standard deviation of elevation exceeds a 200 m threshold are masked out from the consideration.

When comparing in-situ implementation snow densities and their nearest validation densities similar patter to the pattern of the figure 5 (comparison of interpolated densities to validation in-situ data) can be seen (see Fig. 1 below). This indicates that kriging interpolation is not smearing out the variability of the density values significantly. Most snow density values range between 150 kg m^{-3} and 350 kg m^{-3}, which can explain the departure from 1:1 fit for small and large density values.

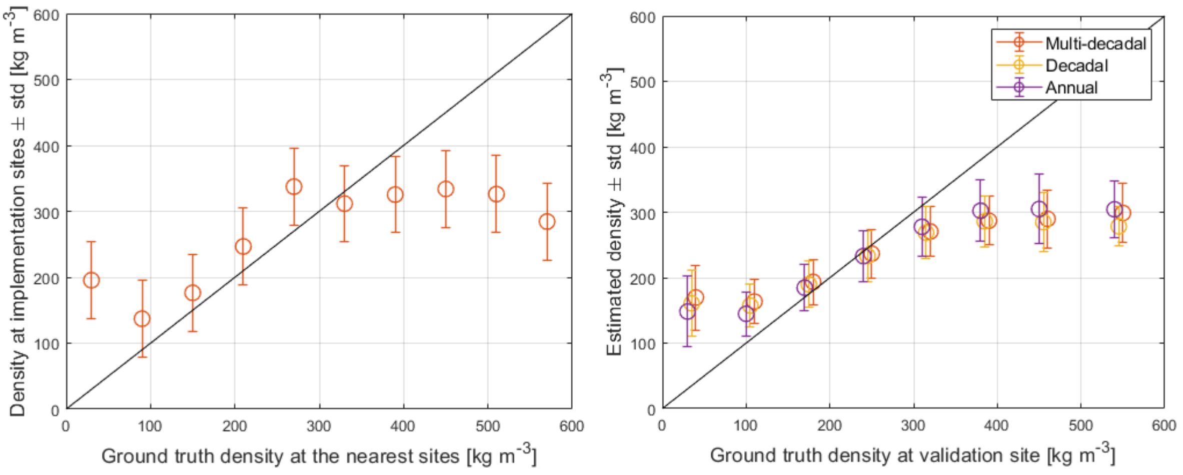

Figure 1. Left: Comparison of in-situ implementation densities and their nearest neighbors. Right: Comparison of interpolated densities and ground truth densities at validation sites (figure 5 from article).

Line 244: "…SWE values up to 500 mm and SWE values up to 150 mm…". This is a confusing sentence, and I was only able to understand what the authors meant by looking at Table 2.

This will be fixed with new sentences:

Two validation were performed, the first validation took into account SWE values up to 500 mm and the second validation considered SWE values only up to 150 mm, as the bulk of the observations are below this value

Line 269: Is GlobSnow calibrated to SWE in any way? If so, then post-processing might not be a good idea? The authors note in lines 338-349 that implementing densities directly into the retrieval may be more beneficial, as a wrong density may lead to reduced retrieval skill. So, then why bother with post-processing?

The GlobSnow 3.0 product used here is not calibrated to SWE (e.g. SWE measured by the snow courses) in any way. The retrieval process does, however, assimilate in-situ measured snow depths from weather stations in the retrieval. In the original scheme these snow depths are converted to SWE using a constant value for density (240 kg/m3), and replacing this value in post processing is actually a fairly straightforward step, and a necessary step, as shown by the results in this manuscript. We thus feel that the improvements obtained with post-processing are significant, in particular for low values of SWE.

A small error is introduced in using only post-processing since simulation of the snow absorption coefficient in inversion of the HUT forward model is still done with the original constant value for density (see e.g. Pulliainen et al., 1999, Pulliainen 2006). However, post-processing is justified as it can be used to study different implementations of the snow densities with relative ease and the results obtained with post-processing are similar to results obtained with implementing dynamic densities in retrieval. Running the full retrieval algorithm is very time consuming, and as such not well suited for testing small changes in methodology. Post-processing can be used to study which densities and methodologies produce the best results and these densities can then be implemented into a final retrieval product.

Also, as some areas have more snow density information available than others, the introduced post-processing methods provide feasible tools to improve the accuracy of the global GlobSnow data based SWE estimates for regional hydrological applications in such areas where snow density information is available.

Line 272: Calculating decadal density maps and applying them to the entire decade seems a bit arbitrary. For instance, the density applied to data from 1981 would be from (approximately) the next 10 years into the future, while for 1989 the density would be from (approximately) the past 10 years. Why not calculate a running decadal average, centered on the year of interest?

A centred decadal average was used because it produces very similar results to a running decadal average but is considerably simpler to produce. The validation parameters for the running decadal average (centred decadal average) snow densities are RMSE = 48.9 (48.9), bias = 0.49 (0.88) and correlation coefficient= 0.71 (0.71) for Eurasia for years 2000-2009. Discussion about using centred decadal averages will be added to the article.

Line 277-278: I don't know what the authors mean in this sentence.

This sentence was aiming to explain that the calculated validation parameters for the decadal densities over the longer period of 1979-2018 are similar to the validation parameters over the shorter period of 2000-2009. However, the sentence is probably unnecessary as the validation parameter are listed in the next sentence.

Table 3: The improvements in post-processed SWE seem marginal, so I wonder if post-processing is worth doing at all. Do the authors have any way to evaluate the statistical significance of the improvements in SWE?

Large improvements in the whole dataset are not expected as most of the validation data are from areas where there are no large mistakes in the baseline product and snow density is close to the constant snow density of 240 km m^{-3} in mid-winter. To highlight the improvements obtained, we have added analyses of specific months to table 3 (shown below). Monthly analyses show, for example, that improvements of 5 mm for SWE values up to 500 mm and 7 mm for values up to 150 mm were obtained for December. Significant improvements are also observed in bias. Figure 2 below (fixed from figure 6 present in the article) also shows that significant improvements (5 -10 mm smaller mean error) are obtained with post-processing for SWE < 100 mm and SWE > 170 mm.

We can also study the statistical differences between the baseline and the post-processed datasets with T-test. The p-value is 3.48E-43 when the T-test is performed for the post-processed and the baseline differences between measured and estimated SWE values. The small p-value indicates that the post-processed and baseline sets differ significantly from each other.

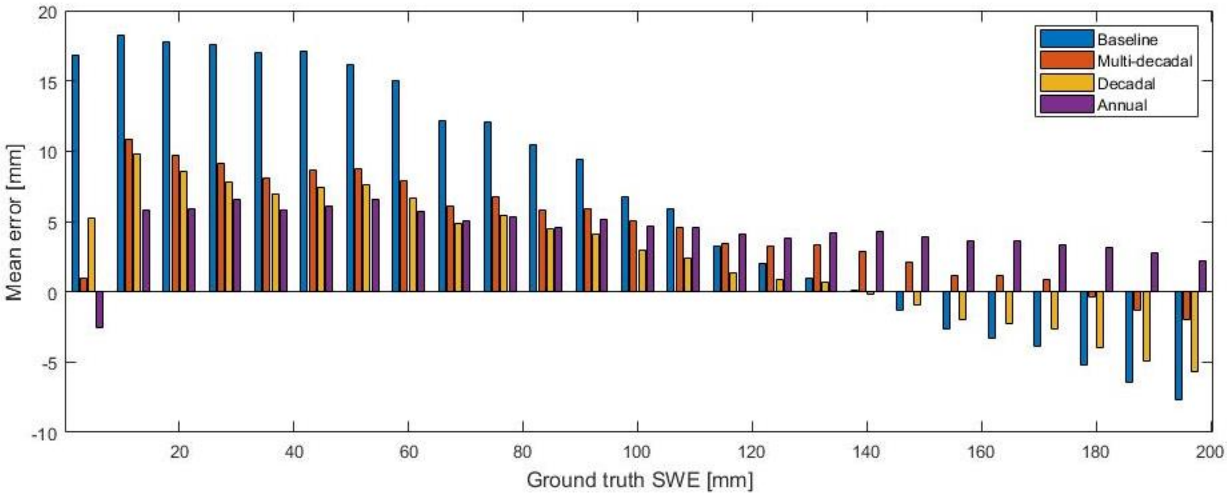

Figure 2. The mean error for the baseline, multi-decadal, decadal, and annual SWE estimates, 2000-2009 Eurasia.

Table 3. Results of validation for whole northern hemisphere, North America, and Eurasia for whole winter, February, April, and December for 1979-2018. Left values are for SWE < 500 mm and bold values are for SWE < 150 mm.

Area	Period	Product	Bias [mm]	RMSE [mm]	MAE [mm]	Correlation coefficient
Northern hemisphere	Winter	GSv3.0	1.5/**9.8**	43.0/**31.3**	29.0/**24.3**	0.71/**0.71**
		post-processed	-1.2/**5.4**	42.0/**30.5**	26.7/**21.9**	0.74/**0.74**
	December	GSv3.0	14.6/**16.1**	29.5/**25.8**	21.8/**20.8**	0.68/**0.75**
		Post-processed	0.1/**1.7**	24.3/**18.5**	14.7/**13.4**	0.69/**0.75**
	February	GSv3.0	11.1/**16.8**	36.9/**30.6**	26.6/**24.3**	0.75/**0.77**
		Post-processed	5.2/**11.2**	36.4/**28.6**	24.6/**21.4**	0.74/**0.76**
	April	GSv3.0	-32.7/**-17.4**	63.9/**40.8**	43.8/**31.3**	0.68/**0.61**
		Post-processed	-20.5/**-9.5**	61.1/**42.4**	41.8/**32.3**	0.68/**0.61**
North America	Winter	GSv3.0	-21.1/**3.1**	72.4/**42.3**	48.6/**32.7**	0.50/**0.49**
		post-processed	-17.3/**4.7**	71.3/**45.3**	47.8/**33.7**	0.53/**0.49**
	December	GSv3.0	14.8/**18.3**	41.2/**36.6**	30.2/**27.6**	0.51/**0.48**
		Post-processed	6.0/**10.3**	38.4/**31.0**	26.2/**23.1**	0.48/**0.51**
	February	GSv3.0	-0.7/**14.7**	51.4/**37.3**	36.7/**29.7**	0.67/**0.63**
		Post-processed	-2.4/**12.5**	52.9/**39.0**	37.2/**29.7**	0.65/**0.60**
	April	GSv3.0	-75.2/**-36.9**	110.8/**60.7**	83.0/**48.9**	0.40/**0.32**
		Post-processed	-60.2/**-25.2**	105.0/**62.3**	77.8/**49.7**	0.40/**0.32**
Eurasia	Winter	GSv3.0	2.5/**10.0**	41.2/**30.8**	28.2/**24.0**	0.73/**0.72**
		post-processed	-0.5/**5.5**	40.2/**29.8**	25.8/**21.4**	0.75/**0.75**
	December	GSv3.0	11.0/**12.4**	29.5/**26.1**	22.0/**21.1**	0.67/**0.72**
		Post-processed	0.0/**1.5**	23.9/**18.1**	14.5/**13.1**	0.70/**0.76**
	February	GSv3.0	10.5/**15.5**	36.5/**30.9**	26.5/**24.5**	0.75/**0.76**
		Post-processed	5.6/**11.1**	35.5/**28.1**	24.0/**21.1**	0.74/**0.76**
	April	GSv3.0	-30.2/**-16.8**	59.6/**39.6**	42.3/**30.4**	0.72/**0.64**
		Post-processed	-18.3/**-8.8**	57.6/**41.4**	39.8/**31.5**	0.71/**0.63**

Figure 8: What does the colorbar in this figure show? It is not labelled.

Colorbar shows number of points within a certain area. Label will be added to the figure.

Figures 8-10: Why are these 3 separate figures? Why not one figure with 6 different panels?

These 3 figures will be combined into one figure with 6 panels.

Lines 339-347: Do the authors have a citation(s) for this?

Relevant citations will be added (Pulliainen et al., 1999 and Takala et al. 2011).

Lines 350-351: To what methods are the authors referring? This paragraph needs to be further explained.

We are referring to alternative interpolation methods. Additional information will be added to the chapter:

In this research, linear interpolation was used for temporal interpolation to obtain snow density values for days without any measurements. However, snow density measurements may contain errors, and these errors can influence the results of the linear interpolation. Thus, alternative interpolation methods for determining behaviour of the snow density throughout the snow season, such as higher degree polynomial interpolation of averaged values or yearly values, could be evaluated in future investigations.

Technical Corrections

Line 17: “was found to produce the best results”?

Will be corrected.

Lines 58-59: Snow also undergoes metamorphism that can cause density to increase.

Mention of metamorphism will be included to the section.

Line 130: “The North American dataset”?

Line 133: “The SNOTEL dataset”?

Caption of Figure 4: “for snow transects”?

Line 231: “differs from the other two versions more” – could improve grammar

Figure 8, y-axis label: “Estimated SWE”?

Edits mentioned above will be included, thanks.

Figures 8 & 9: The font sizes of the titles of these two figures are different.

These figures will be combined, and font sizes will be fixed.