Response to referee 2

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 with normal font.

SUMMARY OF THE PAPER

This study concerns the treatment of snow density in the GlobSnow SWE product (post-processing only). It develops and tests new dynamic snow density estimates based on data collected at snow courses and telemetered snow sites in the Northern Hemisphere. These estimates are compared to the existing approach, which assumes constant snow density in space and time. The new dynamic approach is first developed at a subset of measurement location. Snow density is averaged each day of winter over different periods (multi-decadal, decadal, and annual), then interpolated in time and spatially mapped using ordinary kriging interpolation. All three approaches show improved mapping of snow density against observations relative to the constant density assumption, however lower density values (300 kg/m3) are underestimated. The snow density estimates based on annual values are slightly improved over the multi-decadal and decadal versions. Post-processing of the GlobSnow SWE data is then conducted for each of the three dynamic density datasets, with SWE errors lowest for the multi-decadal and decadal values of snow density, and improved over the case of constant density. The decadal version was then applied to the northern Hemisphere GlobSnow and SWE was post-processed. This yielded slightly better agreement in SWE compared to ground truth data. SWE errors were consistently reduced at all reference SWE values in Eurasia but only at SWE greater than 100 mm in North America.

COMMENTS

1. The Sturm density method was briefly mentioned as a candidate approach, but found to not improve retrieval skill notably (Lines 63-64). Was that analysis published? If not, it could be worth including in the current study, as it would be an interesting point of comparison that would be of interest to the community.

Citation to relevant report will be added (below). Also, sort comparison of post-processing with the Strum density method will be added to article (see answer to comment #4).

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

2. For the SNOTEL data, did the authors check to ensure that the snow depth sensor measures snow depth on top of the snow pillow? This is not always the case at SNOTEL stations, and when snow depth is recorded at a location next to the snow pillow, an estimate of density from the SWE and depth data may be biased. Unfortunately, documentation on the layout of sites is inconsistent. Careful inspection of site photos can reveal which SNOTEL stations may have collocated SWE and depth measurements.

We did not check that snow depth is measured on top of the snow pillow and this could potentially affect the results. However, the uncertainties of the SNOTEL data are, in any case, much larger compared to the Eurasian data and this is mentioned in the article. However, in order to come up with a dynamic snow density estimates for the North American sector, this SNOTEL dataset is the best one to our knowledge (in regard to temporal coverage, spatial coverage and its applicability for the task) and it is therefore applied here. Additionally, in cases where snow depth is not measured in top of the snow pillow, it is measured close by the pillow and the bias this causes should not be very large. We will add discussion about SNOTEL to the article.

3. The results in Figures 5 and 6 appear to be contradictory. Figure 5 suggests that that snow density estimation is improved (but only slightly) relative to decadal and multi-decadal across a range of ground truth density values. In contrast, Figure 6 suggests much larger SWE errors associated with the annually-derived snow density when compared to multi-decadal and decadal versions. Given the post-processing (equation 5), the differences in SWE must be due solely to differences in snow density errors and not errors in baseline SWE. However, Figure 5 suggests the snow density errors are lower for the annual estimate. How do you rectify these conflicting results? Are you sure these have been analyzed correctly?

There was a mistake in figure 6, which explains the conflicting results. Fixed figure below.

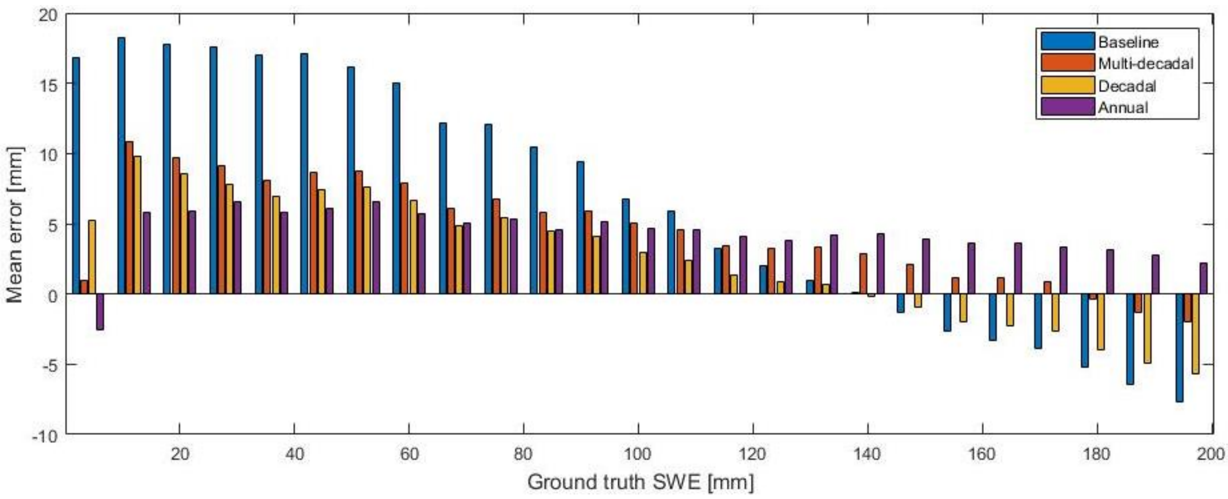

Figure 6. The mean error for the baseline, multi-decadal, decadal, and annual SWE estimates, 2000-2009 Eurasia.

4. Overall, the discussion section is brief and does little to connect the current study to previous work (no other studies are actually cited in the discussion). For example, there have been other studies that have compared the constant density assumption versus dynamic snow density for passive microwave SWE retrievals and found improvements with dynamic snow density (e.g., Tedesco and Narvekar, 2010), and other studies have called for improved density representation for passive microwave (Kelly et al., 2003). How do the results of the current study build on established knowledge?

Discussion section will be expanded. For example, following section about different dynamic snow densities will be added:

Different approaches for varying snow density for satellite based SWE retrievals have been used. The AMSR-E v1.0 product (Kelly, 2009) uses spatially varying but temporally static snow density maps based on snow classes suggested in Liston et al. (1998). However, evaluation of this product by Tedesco and Narvekar (2010) pointed out the need to have also temporal variability in the snow density. The AMSR-E

SWE v2.0 product uses snow spatially and temporally varying density maps based on Strum et al. (2010) for converting snow depth to SWE (Tedesco et al., 2016). However, the densities based Strum et al. (2010) cause large overestimation of small SWE value when used for post-processing the GlobSnow product as seen in Fig. 11, which shows the mean retrieval error for GSv3.0 SWE values post-processed with densities based on Strum et al. (2010) method for Eurasia for 2000-2009.

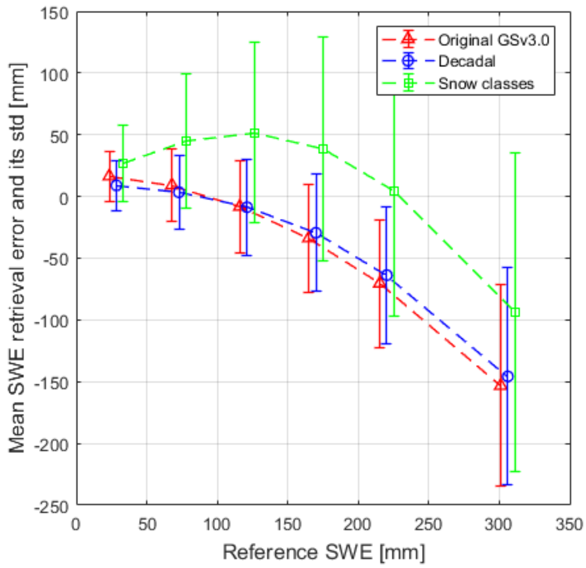

Figure 11. Comparisons of the mean error and standard deviations between baseline and datasets post-processed with decadal snow densities and densities based on snow classed.

5. Ignoring the potential issues with SNOTEL-based values snow density (see comment #2 above), there are issues with the location of SNOTEL stations (mountains where snow is deeper and possibly denser, and GlobSnow does not produce SWE), and differences in temporal sampling. To the latter concern, it would be possible to check whether the temporal sampling of snow density influences the density curve development and validation, compared to the case of less frequent snow course measurements (e.g., Canada and Eurasia). The differences in GlobSnow SWE for constant versus dynamic snow density were smaller in North America (where SNOTEL are located) versus Eurasia (snow courses only), as seen in Figure 7. To what degree do the differences in daily snow density (SNOTEL) versus 10-15 day snow density (elsewhere) influence the result? This could be checked by sampling the SNOTEL density to a similar interval as snow courses and recomputing the statistics.

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 the new version.

Validation statistics were recomputed using resampled SNOTEL data, but no significant changes were detected. However, the worse performance of the GlobSnow SWE retrieval over Canada (and North America) is expected and partly due to larger higher average SWE compared to Eurasia (Mortimer et al. 2020; Luojus et al. 2021).

6. The results seem rather marginal and potentially overstated. An example is Lines 292-293, which state 'the accuracy is considerably better with RMSE for the baseline being 31.87 mm and for the post-processed dataset, the error is 30.87 mm' I question the significance of a 1 mm reduction in SWE error and whether this is 'considerably better'.

We feel that improvements are significant because the global dataset is large and big changes in the RSME are unlikely. Most validation points are in areas where there are no large mistakes in the baseline product but when looking at figures 8-10 or 11 clear improvements are still visible. Fixed figure 6 also shows that significant improvements (5 -10 mm smaller mean error) are obtained with post-processing for SWE < 100 mm and SWE > 170 mm.

To show some of these improvements more clearly, we have added analyses of specific months to table 3. Monthly analyses show, for example, that the improvement in SWE was 5 mm for SWE values up to 500 mm and 7 mm for values up to 150 mm for December. Monthly analysis also reveals significant improvements in bias. The bias is reduced by over 10 mm for Eurasia and about 8 mm for North America for December.

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**

GENERAL COMMENTS

Please consider the number of significant digits for the error statistics reported in the text, figures, and tables. What is the measurement precision of the snow density observations? Are we really confident in density to the tenth or hundredth of a kg/m3? Are we really confident in SWE to the hundredth mm?

Number of significant digits will be reduced.

A histogram of snow density values (development and validation data) would be useful context, especially for interpreting Figure 5.

Figure below can be added to the article.

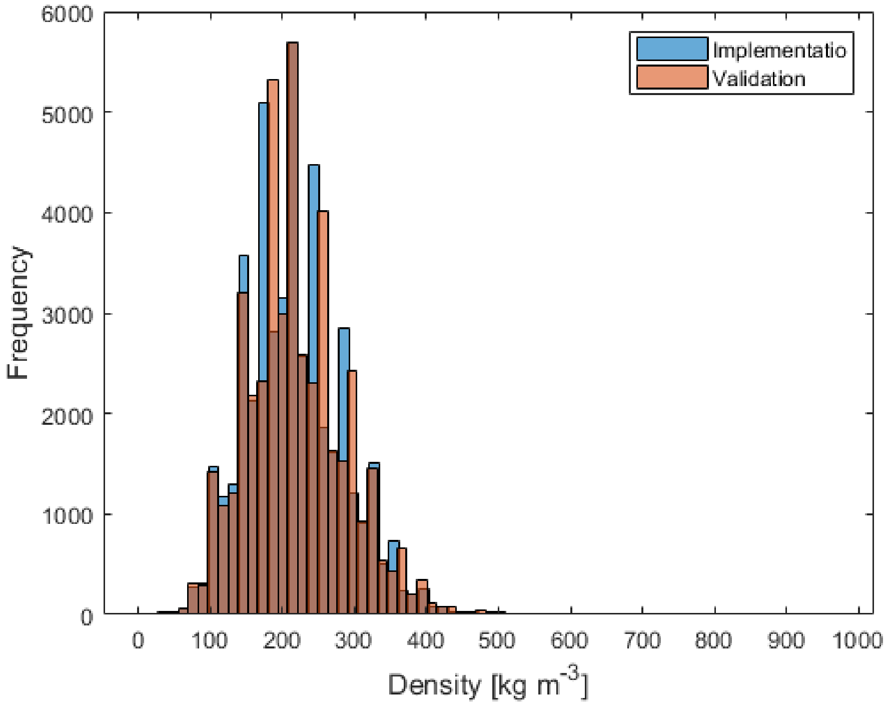

Figure 3. Histogram of implementation and validation densities for Eurasia for 2000-2009.

Figures 8-10 are all referenced in the text before Figure 7 (see Lines 289-304). Consider reorganizing the text or renumbering/reordering the figures.

Figures will be reordered.

The Discussion section and the Conclusion section are mislabeled and should be Sections 4 and 5, respectively.

Thanks, this will be fixed.

There is some discussion-like comments mixed into the results section (e.g., Lines 293-294, 302-304). Please consider reorganizing.

These comments will be moved to discussion section.

TECHNICAL CORRECTIONS

Line 34: Should be 'data are also available' since 'data' are plural.

Line 40: Should be 'in deep snow'.

Line 44: Should be 'observations'.

Edits mentioned will be included, thanks.

Line 59: The Maurice and Harold (1981) citation is not included in the References section. Please add.

Citation will be added.

Line 62: Please rearrange parentheses such that it reads 'Sturm et al. (2010)'.

Line 91: Is a subsection required here? There is a 2.1.1.1 but no 2.1.1.2, so this subsection title may not be warranted.

Line 101: Should be 'were used'.

Line 120: This should be 'Figure 2'.

Line 130: Add 'The' before North American.

Line 132: Should be 'observations'.

Lines 132-133: Can you state why only SNOTEL stations in Alaska and the northwestern USA were selected, and not SNOTEL stations farther south in the USA?

Line 147: Please rearrange the parentheses here such that they are for the year only (two cases).

Line 186: Add 'the' before 'variogram'.

Line 208: Add 'the' before 'baseline'.

All the edits mentioned above will be included, thanks.

Line 224 and Table 1: The period of validation reported here does not correspond to what is described earlier in Lines 100-105. The annual data do not appear to span 2000-2009, and the decadal was previously described as 1999-2009.

These inconsistencies will be fixed, the correct period is 2000-2009.

Line 229: Should be ’decadal’ rather than ’decal’.

Line 243-244: Add ’separately’ after ’performed’ to indicate that the validation was done for two different cases of SWE values.

Line 254: Should be ‘decadal’ rather than ‘decal’.

Edits mentioned above will be included, thanks.

Line 282: Is this supposed to be ‘SWE values up to 500 mm and 150 mm’ rather than ’density values’?

Yes, it is supposed to be SWE.

Line 291: Replace ’improves’ with ‘reduces’.

Line 329: Should be ‘SWE retrievals’ at the end of this line.

Line 330: Replace ’effects’ with ’errors’.

Edits mentioned above will be included, thanks.

TABLE AND FIGURE COMMENTS

Figures 5, 8, 9, 10: The vertical axis should read ‘Estimated’ rather than ‘Estimates’.

Figure 5 caption: Should be ‘decadal’ rather than ‘decal’.

Edits mentioned above will be included, thanks.

Figure 7 would be more effective if a third panel was included that showed the different in SWE estimates from GSv3.0 vs. post-processed. In the current form, I can only tell subtle differences between the two maps.

Figure 7: Note the title above each panel is overlapping text below (90 deg W). Please correct.

Updated figure 7 below

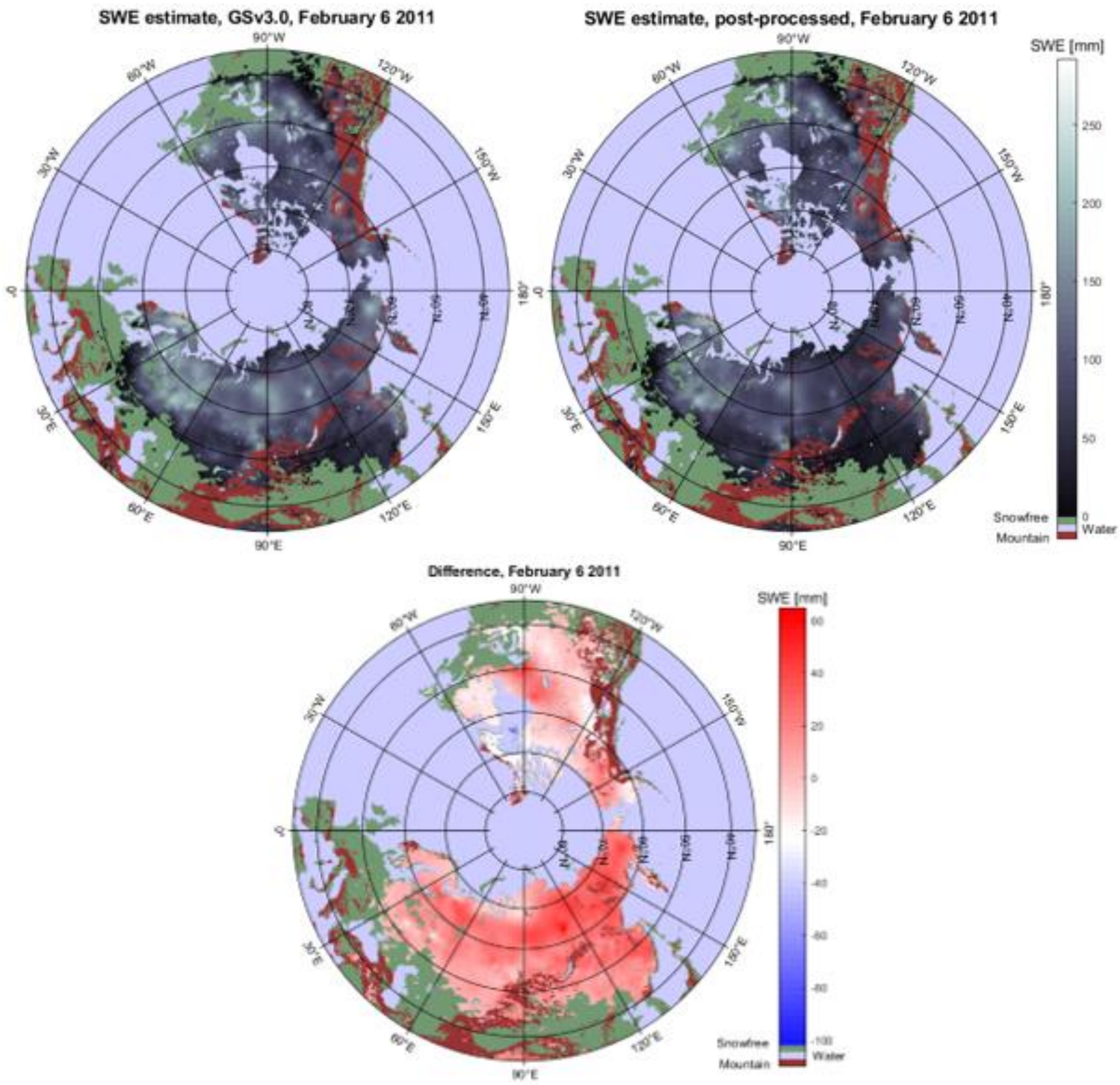

Figure 7. SWE maps for baseline GSv3.0 retrieval (top left), post-processed retrieval (top right) and difference between GSv3.0 and post-processed retrieval (post-processed subtracted from the baseline) for February 6, 2011. The post-processed SWE values are lower as overestimation of small SWE values is reduced.

REFERENCES

Kelly, R. E., Chang, A. T., Tsang, L. and Foster, J. L.: A prototype AMSR-E global snow area and snow depth algorithm, IEEE Trans. Geosci. Remote Sens., 41(2), 230â242, doi:10.1109/TGRS.2003.809118, 2003.

Tedesco, M. and Narvekar, P. S.: Assessment of the NASA AMSR-E SWE Product, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 3(1), 141â159, doi:10.1109/JSTARS.2010.2040462, 2010.