Response to referee Jennifer Jacobs

The inclusion of density in the GlobSnow SWE estimates for both snow grain size and final estimate of SWE are a welcome improvement for global estimates of SWE. The key findings, that dynamic density improves SWE estimates, is not surprising. It is interesting that there is little difference among the three methods that are compared. Additional recommendations for the presentation of the comparisons are suggested below. It is also notable that the performance did not improve as much as one might expect overusing a constant density value.

The collection of in situ density values was a massive undertaking. The point made late in the manuscript regarding near real time estimates of SWE not being possible with these same in situ value suggests that a decadal field rather than annual will serve the community better and eliminate the annual collection and QA/QC of density measurements. Since IDWR is recommended and decadal seems to be the most viable and flexible solution, Table 3 should include the performance of decadal IDWR. It is recommended that performance for individual years be assessed using the decadal minus one data set (leave one out) to assess the range of possible performance in any given year. Also, consider making the in-situ density dataset openly available. This resource would extend the value beyond GlobSnow users to the snow community members. For example, there is a rapidly expanding capacity to make snow depth measurements using lidar and structure from motion on airborne and drone platforms that would greatly benefit from insights and data in this current effort.

We thank the reviewer for their time and constructive comments on the manuscript. We take all the comments and suggestions into account. Our replies are written in red and additions to the manuscript are noted in blue.

We will add the performance of the decadal IDWR to tables 2 and 3. We will also include leave one out versions of decadal IDWR densities to table 2.

The snow density data are available from different national agencies. We agree that a combined publicly available dataset would be useful, but this would require agreement to redistribute from each national agency. In the absence of such an agreement, please contact colleen.mortimer@ec.gc.ca. The data for all but Finland are publicly available through the links in Table 1.

L24 Provide a measure of the average or percent improvement

We will add percent improvements.

Overall, the best results were obtained by implementing IDWR interpolated densities into the algorithm, which reduced RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) by about 4 mm (8 % improvement) and 5 mm (16 % improvement), respectively, when compared to the baseline GlobSnow product.

L112 “Around 19 GHz…” is an awkward phrasing. The point being that SSM/I and SSMIS have slightly different frequencies might be stated in a clearer manner.

We will edit the text to read:

The two main data inputs to the algorithm are vertical passive microwave brightness temperatures (Tb) and daily synoptic snow depth (SD) measurements. The satellite Tb data are from the Special Sensor Microwave/Imager (SSM/I) and Special Sensor Microwave Imager/Sounder (SSMIS) instruments on board the Defense Meteorological Satellite Program (DMSP) F-series satellites. Measurements at 37 GHz and 19.40 (SSM/I) or 19.35 GHz (SSMIS) are used for SWE retrieval.

L118 to 119 “removing measurements from stations where the mean March SWE exceeds 150 cm in at least 50% of the years that the station has had at least 20 measurements” This criteria is difficult to follow.

We will rephrase this to read:

The main SD filtering steps include removing grid cells with a height standard deviation according to ETOPO5 greater than 200 m, removing the deepest 1.5 % of SD measurements, removing measurements from stations where the mean SD exceeds 150 cm in March during at least 50% of the years for locations that have more than 20 annual measurements, and removing SD values above 200 cm.

L120 to 121 How was snow wetness determined?

The Hall et al. (2002) dry snow detection algorithm is used to detect dry snow. For dry snow, the following conditions need to be met:

$$SD_i = 15.9 \cdot \left(T_{B,obs}^{19H} - T_{B,obs}^{37H}\right) > 80(mm)$$

$$T_{B,obs}^{37H} < 240K$$

$$T_{B,obs}^{37V} < 250K$$

where SD_i is the snow depth for the pixel under consideration and must be above 80 (mm) and observed brightness temperatures of 37H and 37V below 240K and 250K, respectively, for the pixel to be considered dry snow. Areas identified as wet snow (not dry snow) for the given day are assigned a SWE value based on the kriging-interpolated SD map. We’ll add the reference Hall et al. (2002) to the manuscript.

L180 and others “significant differences” implies a statistical test was performed. Please rephrase.

We’ll rephrase these.

L260 and others Results indicate differences in western versus eastern NA, but are not presented. Perhaps present in supplementary material. Similar for data in Russia later in the manuscript

We have added results of snow density validation for eastern and western North America into Appendix A.

Table 2 and others Add columns for average values of in situ and modeled

Average values will be added to table 2 and 3.

L284 Paragraph break needed starting at “Figure 4”

Paragraph break will be added.

L294 How was the decision made to use a single semivariogram for such large regions, yet a different semivariogram was determined for each day?

The Kriging interpolation is also used interpolate the background SD field in the GlobSnow SWE retrieval. For this SD interpolation, the variogram is fitted separately for North America, Europe, and Asia for each day. Given the successful implementation of this interpolation, a similar approach was used to interpolate snow densities. Variogram fitting was initially tested separately for Europe and Asia but as there is a limited amount of snow density data in (eastern) Asia, especially in early and late winter, fitting the variogram becomes very difficult or even impossible.

L347 “grain”

Noted, thanks.

Figure 6 Excellent figure, shows that performance varies by month. Additional monthly results would be valuable.

Thank you, we’ll replace figure 8 with a new figure that shows monthly and yearly performance.

Section 5.2.2 While it is fine to present a single year, please provide information about why that year was selected and whether it is representative for most of the study regions.

Year 2005 was selected because the performance of the GlobSnow SWE retrieval is average that year. Additionally, the behavior and amount of snow mass was similar to the ten-year average in 2005. We’ll add mention of this to the manuscript.

L373 concentration

Noted, thanks.

Figure 7. Reduce the number of significant digits to 3.

Figure 7 will be updated with 3 significant digits.

Figure 8 A density scatter plots would be more useful. Scatter plots should use the same scale in the x and y -axes (x is much longer). This figure would be valuable to be presented on a monthly basis?

Figure 8 will be replaced with the figure shown below. This shows monthly and yearly correlation, bias and RMSE. This could be replaced with scatter plots, but we feel that figure below is more descriptive.

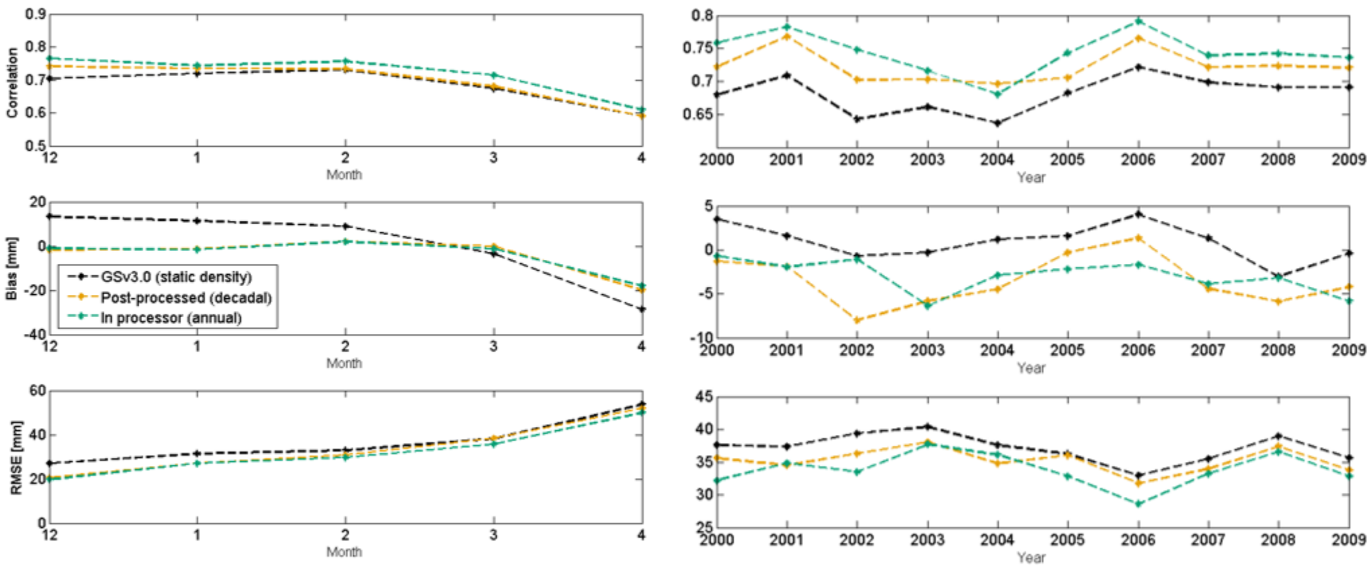

Figure 8: Correlation, bias and RMSE by month (left) and year (right) against validation data.

Figure 10 caption should describe the middle row as well.

Figure 10 caption will be updated.

Figure 10: Top row shows the average monthly difference in SWE between the GSv3.0 product and post-processed product (decadal, Kriging). The middle row shows the average monthly difference between the GSv3.0 product and product with IDWR densities inside the processor. The bottom row shows the average monthly difference between the post-processed product and product with IDWR densities in processor. Note the differing scales on monthly (left) and annual (right) plots. Monthly averages are calculated for years 2000-2009.

The discussion needs to be expanded. This first paragraph is unnecessary because it largely repeats the introduction and the methods rather than putting the work in context. There are a number of topics that would be valuable to consider in the discussion. For example,

1. It appears that performance is not the same globally. One suggestion is to discuss why North America performance is so poor compared to Eurasia. Another is to address the challenges in Russia in greater detail. Also, does performance differ by year – most applications are interested in changes over time or specific years rather than average conditions.
2. There are a number of researchers who have used earlier versions of GlobSnow for applications. The impact and value of these modest improvements in previous research and to the applications in the first paragraph in the introduction could be discussed.
3. The differences between the global snow density product produced here versus other products (global or otherwise) and how the approaches researched for this paper might provide value.

Please consider these to be potential topics that this paper is uniquely qualified to comment on and a request to consider at least one broader topic in the discussion as opposed to a request to discuss all of the examples listed above.

Discussion about the weaker performance of the retrieval in North America will be added to the Discussion section and figure with monthly and annual performance (Figure 8) added to the Results:

It is well documented that the GlobSnow SWE retrieval algorithm performs better in Eurasia than in North America (Mortimer et al. 2020, 2022). The weaker retrieval skill over North America is partially due to higher average SWE in North America. As seen in table 2, the average measured SWE is 132.9 mm in North America compared to 81.8 mm in Eurasia. Locations with a high RMSE tend to have a large negative bias and generally correspond to locations with higher SWE. As seen in figure 8, RMSE increases, and correlation decreases as the bias becomes more negative. Snow densities are larger in North America (274.0 kg m^{-3} in North America and 216.7 kg m^{-3} in Eurasia) and are farther from the static value (240.0 kg m^{-3}). Therefore, we might have expected larger improvements in North America (compared to Eurasia) when moving from a constant to variable density. However, although accuracies improved in both domains, the magnitude of improvement was larger in Eurasia (12.6%/14.2%) compared to North America (7.2%/4.5%) (<500/200mm).

In North America, large errors occur in densely forested high SWE areas in the northeast and in the mountainous west (Mortimer et al. 2022 Figure 7). Dense forest and high SWE are challenging for standalone passive microwave SWE retrievals. Assimilation of in situ SD information from a sufficiently dense observation network can improve SWE estimates in forested deep-snow regions such as Finland (Pulliainen 2006, Takala et al. 2011). However, if the in situ SD network is sparse and the SD variance high, as is the case in northern Quebec, Canada, the SWE estimate is more heavily weighted towards the passive microwave information, which has limited sensitivity to higher SWE (Larue et al. 2017, Brown et al. 2018). Complex terrain is masked out in GlobSnow, but high mountain plateaus, which often have high SWE, are included and can result in large errors in parts of western North America.

L535 Is there a final recommendation on which approach will be used? Will there be a revised GlobSnow dataset in the future or will the algorithm change moving forward?

The development of the SWE retrieval algorithm continues in the ESA SnowCCI+ project and the next version of the dataset is expected to be released later this year (2023). The next version of the product will utilize dynamic snow densities in the retrieval. Similar to Snow CCI+ SWE CRDP2, the upcoming product will be produced in EASE-Grid 2.0 12.5km and regridded to 0.1° lat/lon for distribution. We will add this information to conclusion.

The development of the SWE retrieval algorithm continues in the ESA SnowCCI+ project and, as implementing annual dynamic snow densities into the retrieval improves the retrieval skill, this modification will be used in the production of the next iteration of SnowCCI+ SWE. However, as decadal snow densities are more accurate in North America, they might be preferred for some applications.

Overall, this manuscript presents a clear next generation approach to providing improved estimates of SWE globally. Well done.