The authors would like to thank the referee for the prompt and precise comments to our reply. We have thoroughly revised the manuscript to include the presentation of the results, as well as the analysis of the intercomparison results. As follows, we have made replies and the accompanying revisions.

The referee's comments 1:

This paper is reporting inter-comparison of various snow products from all different sources. The topic is timely and important – an overall good attempt. However, the manuscript suffers from a bad description of the results, especially Section 3. I recommend revising the whole section to point each argument to relevant evidence (figure or table) to support, also the supplementary materials should be there to support the main results, so suggest avoiding unnecessary explanation of the supplementary materials (that can be in the caption). In the beginning, I was quite excited to read the manuscript but quickly losing the interests due to the way the results being described. Section 3 can be much improved (I think it is Results and Discussions) if the authors restructure and rewrite them carefully. Too much unnecessary description in my view, and too little discussions. I really value the topic, but unfortunately, the manuscript itself is not up to the standard.

Specific comments: Section 3 – Each sentence should be backed with Figures or Tables for the evidence. Many sentences are without pointing to specific figures. See some examples below. Line 284: In particular, "NEOSIM...section" Should point the readers to figures or tables for your argument.

Reply:

A great thanks for the referee's suggestion! Accordingly, we have reformulated the contents of Sect. 3 through Sect.5 to improve the structure, and rewritten the majority of the contents. In the revised manuscript, Sect.3 includes the major results of intercomparison: Sect.3.1 covers basic climatology, Sect.3.2 the seasonal cycle, and Sect.3.3 the long-term trend and inter-annual variability. Sect.4 now includes the validation results: Sect.4.1 covers study with OIB, Sect.4.2 that with buoys, and Sect.4.3 the discussion of representation issues.

Regarding the specific comment, the authors have changed the Line 284 into 'In particular, Figure 1 indicates that the thickest snow in late winter/early spring for NESOSIM manifests in the East Greenland Sea, while in DESS, the deepest snow is concentrated in the Canadian Arctic.'.

The referee's comments 2:

Line 293: "Deeper. . .-LG." "DESS shows. . ." Are you referring to Figure S1 or others? I recommend revising all sentences for this.

Reply:

The authors have revised the whole paragraph to read as follows 'During autumn, for the region north of Greenland and Svalbard, SnowModel-LG runs forced with MERRA-2 show a similar spatial pattern as other reanalysis based modeling systems (i.e. NESOSIM and CPOM), but shallower snow than NESOSIM and slightly deeper snow (Figure S2). DuST also shows the thickest snow pack in this region (16.0 cm mean snow depth), though the spatial coverage is more limited. Spring snow depth, ranging from 25.0cm to 30.0cm in the Arctic domain (Figure S1), exhibits large spatial variability among all products. Relatively thicker snow packs in the North Atlantic sector are evident in all reanalysis-based products except for CPOM. Deeper snow packs are expected in this region as it receives precipitation from the North

Atlantic storm tracks. For comparison, snow is also the deepest (over 35.0cm) to the north of Svalbard in both the W99 and SS18 climatologies. NESOSIM further suggests thick snow over Davis Strait, with spring averaged snow depths greater than 25.0cm. This is in stark contrast to the other data sets over the FYI in that region, and is likely unrealistic given this is a region of first-year ice that does not usually freeze until December/January (e.g. Stroeve et al. 2014), limiting the time over which snow can accumulate on the ice.'

The referee's comments 3:

Line 301: "Modal and distribution. . ." do you need this? It is obvious from the figure.

Reply:

The authors have deleted the sentence in Line 301 and the sentence is rewritten as 'Out of all the reanalysis-based data products, snow depth distributions in NESOSIM are shifted towards slightly deeper snow packs (8.0cm) than those from SnowModel-LG (7.0cm) and CPOM (6.0cm) during autumn, although the shapes of the distributions are similar.'

The referee's comments 4:

Line 304: "Since the. . ." then what are the spatial coverage for other products in comparison?

Reply:

These histogram comparisons are limited to regions below 81.5°N, which is emphasized in Line 298. That means that the histograms of all products are also constrained to the same region (Figure S1.a).

The referee's comments 5:

Line 305: Are you referring Figure 3b or else?

Line 306: "PMW..." where do the readers to look at? "Mean snow depth...2.0 cm higher. . ." Is it spring or autumn? Please be specific and add figures.

Reply:

The whole paragraph describes the histogram comparison in Figure 2. The part is rewritten to highlight the key points and remove the unnecessary information.

The referee's comments 6:

Line 313: "We additionally. . .set used." The whole paragraph is describing a supplementary figure. What is the key point for this? I see a general tendency that the description of the results is heavily on supplementary materials. The supplementary materials are to support the main figures and tables but seems overpowering. For this paragraph, please think whether you need all detailed description of the figure, rather than using them to support the main results. I found this is the problem throughout the manuscript.

Reply:

The authors have rearranged the content of the results part and changed the whole paragraph as: 'Additionally, we examine snow over the three different sectors in spring 2015 (Figure S4). The thickest snow from reanalysis-based snow products is mainly over the North Atlantic while satellite-based products indicate more snow accumulating over CA. Although this is only one year of comparison, it shows that

regional differences in snow accumulation can be quite pronounced depending on data set used.'

The referee's comments 7:

Line 324: "DESS exhibits. . .snow depths." Which figure or table for this?

Reply:

The authors add Sec. 3.3 to discuss the interannual variability and trends in these products. Figure 5 mainly provides detailed knowledge on interannual variability. Specifically, the interannual variability consistency analysis about DESS and other reanalysis-based products is for March from 2011 to 2018.

The referee's comments 8:

Line 325: Are we still in Figure 4?

Reply:

The snow accumulation is discussed in Section 3.2 and newly added Figure 6 further helps the analysis of seasonal cycle of different products. For example, the paper finds that the snow in W99 accumulates more during early winter and thus the seasonal curves are flattened near the end of winter. However, SnowModel-LG, NESOSIM and CPOM share a similar seasonal accumulation curve but different from W99 during late winter.

The referee's comments 9:

Line 325: winter time snow accumulation is largest in SnowModel-LG. . .Are you refer ring to Figure 4? I don't see it clearly.

Reply:

The newly added Figure 6 clearly shows the distinct accumulations in SnowModel-LG, NESOSIM and CPOM. And the paper highlights that the overall difference by the end of winter between the SnowModel-LG and NESOSIM are less pronounced, with respect to their autumn conditions. This may be partially due to the fact that the initial condition for wintertime accumulation in NESOSIM is adapted from W99 climatology, while that of SnowModel-LG is snow-free at the end of July 1979 and accumulates snow after that date. Meanwhile, compared with the CPOM product, the seasonal growth in SnowModel-LG is also larger, resulting in even higher April snow volume.

The referee's comments 10:

Line 331: "the inter-annual variability of monthly averaged..is small among" I don't get this. Small among all snow products? Small compared to what? What do you mean? Also it is difficult to see which points are November or April in Figure 4.

Reply:

Figure 6 is newly added and it serves the analysis of the seasonal cycle and interannual variability. The interannual variability here means for each snow product, the snow depth variability in November (or April) is calculated for the period of 2000-2018. The variabilities among all products are within 2.0cm in November (3.0cm in

April), which are smaller than the estimation of W99 climatology. Therefore, the authors states that these interannual variabilities are small.

The referee's comments 11:

Line 335: "DuST show a significant positive. . ." where is the evidence?

Reply:

Both Figure 3 and Figure 4 show consistent increasing snow depths from 2003-2008 (mean snow depth 16.0cm) to 2013-2018 (mean snow depth 20.0cm) period in DuST for October (autumn) and for March (late winter) (compared dashed and solid red lines). This is in contrast the reanalysis-based snow products.

The referee's comments 12:

Line 337: "This features is also not. . ." where is the evidence?

Reply:

According to Figure 5, there is no obviously trend in PMW Bremen nor in DESS during their respective period.

The referee's comments 13:

Line 339: "We do not find. . ." Where can I see that? Figure 5 shows the trend from 1991 to 2015.

Reply:

Consistent with the above trend analysis, there is no trend of Arctic mean snow depth in all products except DuST after year 2000 (Figure 5). Figure 7 reveals how snow change regionally from a longer time span.

The referee's comments 14:

Line 353: "...directly fitted against OIB". What do you mean? You mean OIB data assimilated into those products?? ". . .show high correlations. . ." where is the number?

Reply:

Based on data description in Section 2.3, the models for constructing snow depth estimations using DuST, PMW Bremen and DMI are directly trained/fitted from OIB products, which are not based on assimilation, however, directly rely on the specific OIB data used. The correlation values of OIB comparison are shown in Table 2.

The referee's comments 15:

Line 357: "Figure 6. . ." Shouldn't this be first mentioned? Or Do you need this sentence? "The corresponding. . ." If you directly reference in the text, why do you need this?

Reply:

The authors have rewritten the paragraph as: 'We assess the snow products against four different OIB snow depth products. We first compare OIB and snow products after gridding both to a common 100 × 100km grid and by evaluating the monthly averages in 2014 and 2015. Results are shown in Figure 10 and Table 2. Taking the quicklook product as an example, there are on average 1,300 OIB 40-m mean measurement samples per grid cell. It should be noted that snow depths from DuST, PMW Bremen and DMI are directly fitted against OIB snow depths, and as a result,

these data show high correlations (over 0.36 for PMW DMI) with OIB as shown in Table 2, which should not be taken as a real validation (or comparison) for these products. Except for NESOSIM and UW, other reanalysis-based products are also to some extent indirectly tuned by OIB snow depths in some years. In sum, all products show consistent high correlation with OIB, whereas the lowest correlation is seen for UW, which only correlates with some versions of OIB data (except quicklook).'

The referee's comments 16:

Line 362: "Not surprisingly..." Why?

Reply:

As mentioned before, although they use different versions of the OIB products to train the model, PMW Bremen and PMW DMI are directly fitted from OIB. Thus, their correlations are the highest when compared with OIB.

The referee's comments 17:

Line 380: What is the key message here? Higher R^2 for coarse resolution but no significant difference for temperature resolution?

Reply:

In order to avoid the potential problem of temporal and spatial averaging and interpolation during the validation, here the paper carries out the comparison for each of the data products on their native grids and native temporal resolutions. The results between monthly and daily scales hints that temporal resolution exerts only minor influences in the OIB comparison, since there are only small changes in R^2 and RMSE, without significant differences. For comparison, spatial resolution affects statistical fittings more, which is related to the lack of representation of OIB to these products at the coarse scale of 100km.

The referee's comments 18:

Line 387: "Given. . ." it seems the results are sensitive to the choice of OIB data, but to me it does show some consistency, e.g., SnowModel-LG R^2 range from 0.27 to 0.47 yet PMW Bremen from 0.56 to 0.70. So, I don't know whether I should agree that it is impossible to conclude which ones perform better. Perhaps it is more related to how the particular products dependent to OIB data in their production?

Reply:

It is indeed a bit confusing to have so many different OIB data products, each with different mean snow depth and spatial patterns. The authors acknowledge that for most products, the correlations with the various OIB products are over 0.2, except for UW. However, the comparison cannot be regarded as a validation for products that use OIB data to constrain the retrieved snow depths (i.e. PMW Bremen, PMW DMI and DuST). Furthermore, those which are indirectly fitted with OIB do not always have high correlations with the original version of the OIB dataset as used in their model development. For example, SnowModel-LG, DuST and PMW Bremen are based on OIB quicklook, while the best correlations are with the SRLD, quicklook and JPL OIB products respectively (although the differences are small). Although it is impossible to conclude which product is the best in this comparison, the paper does

find outliers among the products. Therefore, this paragraph is rewritten as: ‘Given the potential data dependency problem and the sensitivity to the specific OIB data set, it is impossible to conclude which snow product performs best. Clearly snow products that have been produced through tuning with OIB data show higher R^2 and smaller RMSEs. The outlier is the UW product. In summary, there is a need for consensus as to which OIB data products are the most accurate, and also for further independent observations to compare against the various pan-Arctic snow products currently available to the science community.’

The referee’s comments 19:

Line 397: “PMW Bremen and. . .” no variability where I can see that? Figure 7? You mean no correlation? You have Figure 7 as one of the main figure but very little description for that. What is the point showing this?

Reply:

Note this is now Figure 11. No variability here means the spread of snow depth within the basin in PMW Bremen and PMW DMI is quite small compared to buoy data and other products. There is additionally no correlation against buoy measurements. The authors thank the referee suggestion and this paragraph is rewritten as: ‘We further explore how well the snow products represent the temporal evolution of snow depth by comparing against CRREL IMBs and AWI snow buoys. As discussed in Section 2.1.1, 86 buoy tracks (58 tracks are from CRREL and 28 tracks are from AWI) are processed during the period of 2000 and 2017 (Table S1). It is worth noting that due to representation issues, we do not expect an exact match of any of the coarsely (i.e. 100km) resampled products to the local-scale of the buoy data. Therefore, the scatterplots in Figure 11 between monthly mean (March and April) buoy snow depths and those from the various products are based on their native spatial resolution. DuST is excluded due to lack of buoy samples in its more limited spatial coverage. Despite some statistically significant correlations, the correlations are all very low, with slopes close to 0. The highest correlation among the products is 0.16 for DESS. The PMW Bremen and PMW DMI products show essentially no variability/spread compared to the buoy data.’

The referee’s comments 20:

Line 400: I think it is far-fetched to compare each buoy with such products. I wonder why we see no correlation.

Reply:

This part of the analysis focuses on the snow accumulation process which are clearly an advantage of buoy measurements. The results do highlight the shortcomings of current products, and many factors may contribute to the zero correlations, including improper sea ice drift as used in reanalysis-based products. One thing to note is that, daily snow products have the potential in application of snow process investigation in the Arctic. Based on Stroeve et al. (2020), good correlations are witnessed between SnowModel-LG and buoy data following the buoy tracks in each integration step, while no/low correlation here may be the results of large discrepancies in trajectories determined from ice drift product and from buoy especially after a long-term (over three months) integration, which is also highlighted in the paper. And since the paper

cannot re-run each model to simulate snow changes along buoy true trajectories in each time step integration, thus purpose of validation with buoy here does not tend to pick the best product, but provide another potential validation view and more crucially, to state the importance of representation issue. **Therefore, the authors want to ask for the referee's suggestions about whether the buoy validation is suitable here.**

The referee's comments 21:

Section 3.4: Personally, I like the results from this section, and found interesting among other results. The difficulty is that it is like comparing apple with an orange, but it does give us a general conclusion which is useful.

Reply:

The authors have spitted the previous Section 3.4 and further reunited in Sec. 3.1-3.3 to differentiate current products and climatology in (1) mean state and distribution, (2) wintertime snow accumulation trend and (3) interannual variability and trend. The paper finds that (1) snow depth reduces from the 1980s to the 2000s which is consistent among the majority of products but with different amplitudes of reduction; (2) the flattened snow accumulation in W99 near the end of winter are quite different from in the current product; (3) the interannual variability in snow products is about half of previously reported in W99 climatology. All the above indicates that current snow conditions are quite different from the climatology. While a long-term increase of Arctic precipitation is expected in the future decades, it will more likely start to transition to liquid precipitation (Bintanja et al., 2014). Thus, how snow over sea ice changes and how it affects the ice and the climate as a whole are key scientific questions. The above comparisons with climatology are necessary and needed further discussion.

The referee's comments 22:

5. Discussion: I found this is confusing about the scope of this research. Need to make it clear whether you want to show the inter-comparison results or developing a new sampling strategy. The whole discussions are about sampling strategy not about inter-comparison results. This actually tells me whether this is a research article or just for a discussion.

Reply:

We thank the referee's comment on the scope of our study. The authors have moved the representation issues discussion into Section 4.3, following the OIB and buoy validation in Section 4.1 and 4.2. The purpose of this part is NOT to develop a sampling strategy, but to study how representation effects the validation works, especially regarding the low correlations with buoy measurements. The sampling strategies are designed to simulate various observational coverage (including buoy) and the representation's effects on validation, by using OIB dataset. Since this part serve as an important piece of evidence of validations with very limited snow depth measurements such as buoys, we consider it to be necessary. A full, systematic discussion of representation issue is beyond the scope of this paper, and planned as future work.

The referee's comments 23:

6. Conclusions: It is too lengthy. Should be cut down to the key results and messages.

Reply:

The authors have rewritten this part as Summary and Outlook. Some results have reduced in length in order to be more concise. The major conclusions are included. The choice of products that perform consistently during intercomparion and validation are stated explicitly, which answers the referee's comment on conveying key messages. Besides, extensions of the work in the future are also included.

Reference

Bintanja, Richard, and F. M. Selten. Future increases in Arctic precipitation linked to local evaporation and sea-ice retreat. Nature 509.7501 (2014): 479-482.

Kwok, R., Kacimi, S., Webster, M. A., Kurtz, N. T., & Petty, A. A. Arctic Snow Depth and Sea Ice Thickness From ICESat-2 and CryoSat-2 Freeboards: A First Examination. Journal of Geophysical Research: Oceans, 125(3), e2019JC016008, 2020.

Stroeve, J.C., Markus, T., Boisvert, L., Miller, J. and Barrett, A.: Changes in Arctic melt season and implications for sea ice loss. Geophysical Research Letters, 41(4), 1216-1225, 2014.

Stroeve, J., Liston, G. E., Buzzard, S., Zhou, L., Mallett, R., Barrett, A., Tschudi, M., M. Tsamados, P. I., and Stewart, J. S.: A Lagrangian snow-evolution system for sea-ice applications (SnowModel-LG): Part II - Analyses., Journal of Geophysical Research: Oceans, sumitted, 2020.