This content about the authors' response to 'Inter-comparison of snow depth over sea ice from multiple methods' includes:

The authors would like to thank the referee for the prompt and precise comments to our reply. We want to re-emphasize that the purpose of the study is to investigate the current status-quo of the community in the reconstruction and retrieval of snow on Arctic sea ice. The variety of the methods used in the products, as well as the potential lack of independent data pose unique challenges, but we consider this a timely update of the community's efforts, and the collateral trade-offs during the analysis necessary at this stage of research progress. Moreover, the general consistency among many of the products reflect the exciting progress that has been made, and provide confidence for further improvements of snow depth retrieval. Also, messages are conveyed relating to both observation and representation issues. In specific, with respect to the comments from the referee, we have made replies and accompanying revisions as follows.

The referee's comments 1:

Review of Inter-comparison of snow depth over sea ice from multiple methods by Zhou, Stroeve and others.

In this paper the authors explore how well various methods of determining the snow cover (mostly depth) across the sea ice of the Arctic Basin match each other, and where possible, match in-situ data. The methods consist of a) in-situ depth values from two ice buoy systems, b) IceBridge airborne measurements from radar, c) the Warren et al. (1999) and newer climatologies based on Russian measurements, d) satellite retrievals and e) physical models driven from reanalysis products. The authors find similarities between products in some geographic areas and at some periods of the winter, but also large discrepancies in both absolute values and patterns of depth. This comes as no surprise given the diversity of the ways in which the snow fields are derived, and the limitations within the various methods.

This is a useful paper, and quite an impressive job in merely handling the massive data sets involved, but it does not go far enough in examining the "whys" of the discrepancies, and for that reason I think needs another round of revision. To simply state that radical differences in footprint sizes and spatial scales of the various snow depth products are at the root of the observed differences, while perhaps correct, seems of little practical use.

Part of the problem with the paper seems to start with this statement, given as a motivation for the study: (line 48) [to] provide an inter-comparison of these products so that recommendations can be made to the science community as to which data product best suits their needs. I think I understand what the authors mean here, but it is not what they actually wrote. One would hope that the science community is not only looking for products that suit their needs, but in fact, are as accurate as possible. So while we all can accept that with footprint size and scaling issues, snow depth "truth" may be elusive, conceptionally it exists and it really is what users, and snow product developers, need to strive for. This murkiness in purpose reappears throughout the paper in that none of the models or methods are ever labeled as "wrong," even when the results seem to be utterly improbable. I understand no one wants to "slag" a model in print, but clearly one conclusion from the paper is that some of the models, in some situations, should be avoided for now (until improved) and the authors could be more explicit in saying so when that conclusion is clear (e.g. The PMW-DMI model has 13 cm as the end-of winter average across the entire Basin: Fig. 3 left).

Reply:

The authors understand the urge for the 'accurate' snow product for the community. The motivation of the paper is to investigate the status and (to try) to attribute the discrepancies among current snow products constructed from multiple methods, and further validate against available observations. However, as discussed in Section 4.3, due to the resolution differences of the various snow depth measurements, there exist distinctive representation issues when local snow depth measurements are used. Therefore, it is hard to judge which is the "true" snow depth based on limited validation datasets.

Although it is impossible to pick one "perfect" snow product, we do find outliers. For example, the positive trend in DuST snow depths during ICESat and CryoSat-2 periods is caused by the limitation in processing the calibration with OIB in the product, which is inconsistent with other products. In spite of lacking thorough analysis of independency and representative issues, we pointed out "unreasonable" performances in some products, including low correlation with OIB in UW product and no correlation and small snow variability in the PMW Bremen and PMW DMI product and buoy comparison although existing representation issues. The above four products are therefore not recommended.

According to the comments, the revised paper further emphasizes the choices of the candidate products with consistent performance according to our analysis. Regarding the manuscript, relevant revisions are highlighted in yellow and also denoted in the following:

Line 313: *The shallowest spring snow in PMW DMI, with a mean snow depth < 15cm, and the thickest autumn snow in DuST, are the two extremes among current snow products.*

Line 379: *It should be also noticed that DuST shows a significant positive snow depth bias from the Envisat period to the CryoSat-2 period, likely a result of using OIB to calibrate the snow depth estimates. This positive snow trend is missing in the passive-microwave based snow products (e.g., PMW Bremen), and is not observed in DESS (time-series in PMW DMI is too short to be assessed).*

Line 461: *Snow products that have been produced through tuning with OIB data show higher R2 and smaller RMSEs. The PMW DMI product performs best, despite not being tuned with OIB data for the time period to which it is compared. The outlier is the UW product.*

The referee's comment 2:

Before getting to the heart of my review, I wanted to raise a point that I am not an expert in but I believe is important. In any inter-comparison of models, the comparison will be skewed when the models are being forced by different reanalysis products, with the precipitation forcing, notoriously difficult to get right in the Arctic, varying a lot in both time and space. How did the authors sort out input reanalysis precipitation differences from model biases. It seems to me that the models need to be run on the same input.

Reply:

The authors are working with data products as provided by various contributors. For the four reanalysis-based products, contributors use different reanalyses as input. Furthermore, the methodologies differ a lot as well. Regarding reanalyses, a comprehensive assessment of available products was made recently by Barrett et al. (2020). They found inter annual variability in reanalysis precipitation was consistent among the different products, but that MERRA-2 was wetter than ERA-I or ERA-5. Spatial patterns are however consistent. Further, Stroeve et al. (2020) also found that running SnowModel-LG with ERA-5 and MERRA-2 gave snow depths within a few cm of each other. We anticipate that the differences between the various reanalysis-based estimates has more to do with how snow processes related to sea ice are considered, such as the snow accumulation or loss and physical process, rather than the reanalysis product used. The distinctive snow initial conditions between SnowModel-LG and NESOSIM is another example of hurdles in aligning these products. More aligned comparison of the reanalysis-based products, which is suggested by the referee, is beyond the scope of this study, including attributing to uncertainties (or differences) to driving datasets and methods. This would potentially require another round of intercomparison with much higher coordinated activities across the contributing institutes.

The referee's comment 3:

I found the results section of the paper quite good, and the graphics, of which perhaps there are a few too many, both useful and explanatory. If some shortening were to occur here, I would suggest the spatial trends in Autumn (Fig. 1), and the temporal trends over the period of record (Fig. 5) could be deleted. Autumn is a tough time to model or map snow, since open water (no platform) determines snow depth to a large extent, depending how it is handled in the model or with the satellite products. Figure 5, while interesting, is a dangerous figure to put out there, as it is likely to be cited as depicting real changes, when in fact the point of the paper is that the models and methods need significant improvement, as attested to in Figures 6, 7, 8, 9, 10 and 11. Personally, I would remove Figure 5. But if it is to be left in (see the next paragraph for this same point), then it would be more useful to explore why all of the models/methods converge in getting increasing snow depth in the Greenland-Canadian sector, while the Russian-Siberia sector is losing snow then to just put it out there as "trends." Can this convergence be traced back to some basic climate variables that enters all of the various models? What drives the convergence?

Reply:

Thank you for referee's suggestion. Figure 1 is removed from the manuscript, and moved into the supplementary material as Figure S1. Further explanation is added about newly Figure 7 (denoted as Figure 5 in old version). As mentioned in the paper, snow depth decreases over Eurasia as a result of delayed freeze-up while the increasing over Greenland-Canadian sector from the three reanalysis-based products (SnowModel-LG, NESOSIM and CPOM) use precipitation (or snowfall) from reanalysis. Existing studies of reanalysis-based precipitation in the Arctic indicate more snowfall and ensuing accumulation, hence thicker snow. Serreze et al. (2012) found that in summer and early autumn, the precipitable water from MERRA, CFSR and ERA-Interim both show positive trend in the Canadian Arctic Archipelago, and Stroeve et al. (2020) suggests that widely open water may increase winter precipitation and further affect snow accumulation. Further, the snow accumulation process largely depends on ice motion fields and the above all three products (SnowModel-LG, NESOSIM and UW) use NSIDC ice drift, which may potentially cause similar pattern in snow depth trend.

The referee's comment 4:

And that type of analysis is what I think the paper currently lacks. For example, starting with Figure 3, depth histograms offer a wealth of interpretive information if examined closely. SnowModel and NESOSIM show a distinct drift shoulder (deeper snow) in Autumn that indicates they allow the snow to drift immediately, while CPOM and DuST show little or none. So this feature is about drifting: is it real? Should it be there? Strangely, the drift shoulder disappears by Spring for SnowModel and NESOSIM, yet appears in UW and DuST. Therein lies some behavioral differences that could provide insights for model/method improvement.

Reply:

In general, the authors want to explain that the observable modes in the PDF is caused by snow accumulation on different ice types. The shoulder shape in Figure 2 mentioned by the referee is the second mode in PDF. This bimodal PDF is more obvious in Figure S3 both for SnowModel-LG, NESOSIM and CPOM as a result of different snow accumulation over FYI and

MYI. MYI may accumulates initial snow cover early in autumn, while FYI won't be able to catch any snow until it forms. Another minor issue in the comparison is that, the shoulder shape in Figure 2.a is due to the smaller common coverage area compared with Figure S3. The common coverage without DuST (shown in Figure S2) covers more over Canadian coastal regions, where MYI manifests. All spring snow PDFs in Figure S3.b show the bimodal or long-tail features, which is also found in other observations (Kwok et al., 2011).

The referee's comment 5:

Figure 4 offers similar analysis possibilities. A lot is known at least at local scales about the patterns of snow build up on sea ice in various locations. The slope of these seasonal trajectories, and whether they curve off late in the season or not, is a diagnostic that could readily provide insight into what is or isn't working. The W99 curve is suggestive; only the CPOM curves seem to carry that shape, yet these curves fail to reach reasonable depth values by Spring. That should tell us something. Finally, considering Figure 10, which is fascinating, the paragraph (lines 426-431) discussing it barely scratches what could be gleaned from the data. Not only is W99 significantly deeper than the model/method products, in some cases by more than 2X, but also the histogram shapes are so different as to look like they are from totally different fields. SnowModel, DuST, and DESS all have a zero snow fraction, while the others do not. PMW-DMI is as peaked as the W99 data, but is about 1/3rd as deep. Surely, contained in this plot, which took considerable effort to develop, are many useful suggestions for model and method improvement (most of the same comments apply to Fig. 11 and SS18), but to get to them requires thinking about why the histograms have the shape they do. In the old days, a lot of emphasis was place on interpreting skewness and kurtosis, and that literature might be of use here.

Reply:

The authors have included more analysis of the seasonal cycle in the manuscript (now a dedicated section, Sec. 3.2). Regarding to the comment on the specific details of intercomparison, the authors agree that the seasonal curve shape in CPOM is similar with that in W99, however, SnowModel-LG, NESOSIM also has the similar curve shapes in some years, especially from 2011 to 2012 and from 2015 to 2016. However, the interannual variability of this seasonal shape varies widely in all reanalysis-based products. As sea ice freeze-up has happened later and later over the last 40 years, it is expected that the snow accumulation during the early stage of winter is no longer similar to W99.

As in Figure 3 and S5 (which are Figure 8 and 9 in the previous version of the manuscript), the authors apologize that the previous result of the comparison between SnowModel-LG and W99 was not confined into Arctic basin, which is not the same region as the other products. These two plots are updated into the paper and the discussion part is revised. Figure 3 basically shows the results of Figure 2 but for different regions. In Figure 2, common regions only include Baffin Bay, north of Barents and Kara Sea but in Figure 3 and S5, only the Arctic basin is included (as shown by W99 in Figure 1). Within the Arctic basin, thin snow (less than 10cm) is observed in SnowModel-LG, DuST and DESS, while NESOSIM, CPOM, UW, PMW Bremen and PMW DMI show deeper snow packs. Other studies have indicated that snow depth is decreasing over the last 30 years (Webster et al., 2014), and that the snow depth since 2000s is thinner than in W99 and SS18. The majority of the snow products are consistent with this decreasing snow depth, but their mean values and slopes differ.

As measurements in W99 are mainly over the western of Arctic, where more of the ice was MYI, SS18 is adopted for providing more details over eastern Arctic, where FYI dominates. The shapes of W99 and SS18 PDF are different as a result of spatial coverage and the inclusion of FYI. Since the ice is getting younger, the histogram shapes of the various snow products are expected to differ from W99 and SS18 especially after 2000. Figure 3-4 also provide snow depth estimations during 1980s or 1990s from reanalysis-based products, suggesting how snow depth has changed in each product over the longer time series. The mode in SnowModel-LG is decreasing while that in UW is still unchanged over the 40 years. The authors agree that the comparison with climatology is helpful for model improvements and we have revised relevant parts, highlighted in red and denoted in the following:

Line 320: *In autumn (Figure S5), bimodal snow distributions are noticeable both in SnowModel-LG and NESOSIM using the data post 2000, with a large proportion of thin snow not seen in the W99 climatology. By the end of winter, the differences between all products and W99 are larger (Figure 3) than in autumn, with the minimum decreasing is 10.0cm for UW and the maximum is over 15.0cm for PMW DMI in spring. SnowModel-LG, DuST and DESS all have snow depths below 10cm in March/April. For reanalysis-based products, snowpack is still significant shallower against climatologies in the 1980s (1990s in the case of CPOM) when W99 is partly collected. In addition to mean snow depth, the skewness of the snow distribution, especially in spring, among all products are mainly positive as a result of the larger presence of thin snow cover over FYI dominated in the current era, while that of W99 is in the opposite.*

As for the skewness and kurtosis of snow depth distribution, Kwok et al, (2011) found that the snow depth is left skewed from OIB and snow depth from ICESat-2 and CryoSat-2 (Kwok et al., 2020) also shows the left skewness especially during early winter. PDF shapes from the snow products and OIB observations is quite different from W99 and SS18.

Summary comments:

In summary, I recommend this manuscript be returned to the authors for revisions without acceptance, that they be commended for undertaking a very useful and difficult set of analyses, and that they be urged now to reap the rewards of that analyses by gleaning more pertinent and useful information from their work. That that could prove very useful in improving existing and future models and methods of extrapolating snow over the Arctic Basin.

Reply:

The authors sincerely thank the referee for the comments. Revisions to the manuscript have been made to: (1) restructure Section 3 to 5 for better clarity, with each subsection covering an aspect of intercomparison, and (2) include more analyses of the intercomparison results, especially for the consistency of the products and PDF of snow depth and comparison with climatology, and (3) explicitly picking out the consistent products and outliners. We sincerely hope that through these revisions, we can convey more clear and informative results.

Reference:

Barrett, A. P., Stroeve, J., and Serreze, M. C.: Arctic Ocean Precipitation from Atmospheric Reanalyses and Comparisons with North Pole Drifting Station Records, Journal of Geophysical Research: Oceans, 2020.

Kwok, R., et al. Airborne surveys of snow depth over Arctic sea ice. Journal of Geophysical Research: Oceans 116.C11, 2011.

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.

Serreze, M. C., Barrett, A. P., and Stroeve, J.: Recent changes in tropospheric water vapor over the Arctic as assessed from radiosondes and atmospheric reanalyses, Journal of Geophysical Research: Atmospheres, 117, 2012.

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, in revision, 2019.

Webster, M. A., Rigor, I. G., Nghiem, S. V., Kurtz, N. T., Farrell, S. L., Perovich, D. K., and Sturm, M.: Interdecadal changes in snow depth on Arctic sea ice, Journal of Geophysical Research: Oceans, 119, 5395–5406, 2014.

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 (Figure S2). DuST also shows deep snow packs 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. Further, relatively thick 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 winter 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 depth over the three different sectors in spring 2015 (Figure S4). Deepest snow packs from reanalysis-based snow products mainly occur over the North Atlantic, while satellite-based products indicate more snow accumulating over the 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 × 100 km 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. Therefore, we do not consider this a suitable validation (or comparison) for these products. For the PMW DMI product, however, OIB data from the period from March 2014-2015 was not used during model development, and hence we have more confidence in the high correlation observed. Except for NESOSIM and UW, the other reanalysis-based products are also to some extent indirectly tuned by OIB snow depths in some years. Overall, all products show reasonably high correlations with the different OIB snow estimates except for UW, which only shows a slight correlation with some versions of the OIB data.'

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 in some years. 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 not possible to conclude which snow product performs best. Snow products that have been produced through tuning with OIB data show higher R^2 and smaller RMSEs. The PMW DMI product performs best, despite not being tuned with OIB data for the time period to which it is compared. The outlier is the UW product. In summary, there is a need for a 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 mean state of small scale snow depth on Arctic sea ice 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) were processed from 2000 to 2017 (Table S1). 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.

Inter-comparison of snow depth over sea ice from multiple methods

Lu Zhou[1], Julienne Stroeve[2,3,4], Shiming Xu[1,5], Alek Petty[6,7], Rachel Tilling[6,7], Mai Winstrup[8,9], Philip Rostosky[10], Isobel R. Lawrence[11], Glen E. Liston[12], Andy Ridout[2], Michel Tsamados[2], and Vishnu Nandan[3]

[1]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China
[2]Centre for Polar Observation and Modelling, Earth Sciences, University College London, London, UK
[3]Centre for Earth Observation Science, University of Manitoba, Winnipeg, Canada
[4]National Snow and Ice Data Center, University of Colorado, Boulder, CO, USA
[5]University Corporation for Polar Research, Beijing, China
[6]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[7]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
[8]DTU Space, Technical University of Denmark, Lyngby, Denmark
[9]Danish Meteorological Institute (DMI), Copenhagen, Denmark
[10]University of Bremen, Institute of Environmental Physics, Bremen, Germany
[11]Centre for Polar Observation and Modelling, University of Leeds, UK
[12]Colorado State University, Cooperative Institute for Research in the Atmosphere (CIRA), Fort Collins, CO, USA

Correspondence: Julienne Stroeve (stroeve@nsidc.org)

Abstract. In this study, we compare eight recently developed snow depth products that use satellite observations, modeling or a combination of satellite and modeling approaches. These products are further compared against various ground-truth observations, including those from ice mass balance buoys (IMBs), snow buoys, snow depth derived from NASA's Operation IceBridge (OIB) flights, as well as snow depth climatologies from historical observations.

Although we observe consistency in general structure among different snow depth data sets, mean snow depth discrepancies are observed over the Atlantic and Canadian Arctic sectors. There is no significant trend in the mean snow depth among all snow products since the 2000s, despite overall shallower snow packs in recent years compared to snow depth climatologies. The delaying in Arctic freeze-up could bring about the differences in early snow accumulation from products against the climatology. Among the products evaluated, the University of Washington (UW) snow depth product produces the deepest spring (March-April) snow packs, while the snow product from the Danish Meteorological Institute (DMI) provides the shallowest spring snow depths. Four products, SnowModel-LG, NASA Eulerian Snow on Sea Ice Model (NESOSIM), Centre for Polar Observation and Modelling (CPOM) and Department of Earth System Science in Tsinghua University (DESS) are found to be consistent between each other, spatially and temporally. SnowModel-LG and NESOSIM, also provide estimates of snow density: snow density in SnowModel-LG is generally higher than climatology, whereas NESOSIM density is generally lower. Arctic-wide, these two density products show no significant trend since the 2000s.

Inconsistencies in reconstructed snow parameters among the products, as well as differences between in-situ and airborne observations can in part be attributed to differences in effective footprint and spatial/temporal coverage, as well as insufficient observations for validation/bias adjustments. Our results highlight the need for more targeted Arctic surveys over different

spatial and temporal scales to allow for a more systematic comparison and fusion of airborne, in-situ and remote sensing observations.

1 Introduction

Snow on sea ice plays an important role in the Arctic climate system. Snow provides freshwater for melt pond development and when the melt ponds drain, freshwater to the upper ocean (Eicken et al., 2004). In winter, snow insulates the underlying sea ice cover, reducing heat flux from the ice-ocean interface to the atmosphere and slowing winter sea ice growth (Sturm and Massom, 2017). Snow also strongly reflects incoming solar radiation, impacting the surface energy balance and under-ice algae and phytoplankton growth (Mundy et al., 2009). Furthermore, sea ice thickness cannot be retrieved from either laser or radar satellite altimetry without good knowledge of both the snow depth and snow density (Giles et al., 2008; Kwok, 2010; Zygmuntowska et al., 2014).

Despite its recognized importance, snow depth and density over the Arctic Ocean remain poorly known. Most of our understanding comes from measurements collected from Soviet North Pole drifting stations and limited field campaigns. The Warren et al. (1999) climatology and the recently updated Shalina and Sandven (2018) climatology, hereafter $W99$ and $SS18$, respectively, have provided a basic understanding of the seasonally- and spatially-varying snow depth distribution over Arctic sea ice. However, these data were collected from stations between 1950 and 1991 and were largely confined to multi-year ice (MYI) in the central Arctic. Hence, they are not representative of the Arctic-wide snow cover characteristics of recent years (Webster et al., 2014). Today, the Arctic Ocean has transitioned from a regime dominated by thicker, older MYI to one dominated by thinner and younger first-year ice (FYI) (Maslanik et al., 2007, 2011). In addition, the length of the ice-free season has increased (delayed freeze-up and early melt-onset/break up), reducing the amount of time over which snow can accumulate on the sea ice (Stroeve and Notz, 2018). IIn response to this data gap, several groups are working to produce updated assessments of snow on sea ice using a variety of techniques.

From satellites, several studies have used passive microwave brightness temperatures to retrieve snow depth, on FYI (Markus et al., 2011) and MYI (Rostosky et al., 2018; Kilic et al., 2019; Braakmann-Folgmann and Donlon, 2019; Winstrup et al., 2019). Other studies have modelled snow accumulation over sea ice using various atmospheric reanalysis products as input (Blanchard-Wrigglesworth et al., 2018; Petty et al., 2018; Liston et al., 2020; Tilling et al., In preparation). Some promise has also been shown in mapping snow depth by combining satellite-derived radar freeboards from two different radar altimeters (Lawrence et al., 2018; Guerreiro et al., 2016) and from active - passive microwave satellite synergies (Xu et al., In preparation). With the launch of ICESat-2, additional possibilities exist to combine radar and lidar altimeters to directly retrieve snow depth (Kwok and Markus, 2018). Such approaches are paving the way for proposed future satellite missions (i.e., ESA's CRISTAL (Kern et al., 2019)).

In recognition of the numerous new snow data products available, it is timely to provide an inter-comparison of these products so that recommendations can be made to the science community as to which data product best suits their needs. Towards this end, we provide a comprehensive intercomparison between eight new snow depth products and evaluate them against various in situ observations and different NASA's Operation IceBridge (OIB) snow depth products. Since these datasets do not have common spatio-temporal resolutions, we limit our comparisons to monthly averages between October-November (from now referred to as the autumn period) and March-April (spring period) from 2000 to 2018 and also limit our region to the Arctic basin (i.e., we exclude regions such as the Sea of Okhotsk, Bering Sea, Baffin Bay/Davis Strait, and the East Greenland Sea).

The paper is organized as follows. The next section describes the details of each snow product, snow depth observations and climatology data used for comparison. Comparisons among snow products and between climatologies are shown in Section 3. Section 4 discusses validation against OIB/buoy measurements and representation issues. The conclusion and further implications for snow are found in Section 5.

2 Data and methods

In this section, we introduce: (1) observational datasets of snow depth used for validation/comparison, (2) climatological snow depth products, (3) passive and active microwave snow depth products and (4) reconstructed snow depth estimates from models. The inter-comparison among all snow products and comparison between these products and measurements are based on their native spatial and temporal resolution. However, the spatial and temporal resolution varies considerably between products. Details on resolution are provided in Table 1.

An additional comparison was made between OIB and the different snow depth products at the coarsest spatial (100×100 km) and temporal (monthly) resolution. Below we briefly describe each data set. We refer the reader to the references for each data product for more detailed information on the individual algorithms.

2.1 Measurements used for comparison

2.1.1 In situ Observations

Ice Mass Balance Buoys (IMBs) are designed to provide snow depth and ice thickness information, and have generally been deployed within undeformed MYI (Richter-Menge et al., 2006). In this study, we use snow depth measurements from IMBs deployed by US Army Cold Regions Research and Engineering Laboratory (CRREL). Each buoy is equipped with acoustic sounders above and below the ice with an accuracy 5 mm for depth measurements (Richter-Menge et al., 2006). Following Perovich et al. (2009), snow depth measurements greater than 2 m or less than 0 m are removed. In total, 58 CRREL buoy tracks are used between 2010 and 2015 (available at: http://imb-crrel-dartmouth.org). To align the IMB data with the daily and monthly snow depth products, as well eliminate effects from missing measurements, we average the IMB data into monthly averages after first creating daily averages from the four-hourly observations. See, Table S1 for a listing of buoys used, along with their dates/time periods of operation.

Snow buoys have been deployed by the Alfred Wegener Institute (AWI) since 2010 (Nicolaus et al., 2017), and provide snow depth estimates from 4 separate snow depth pingers. These are averaged together to provide one snow depth value at each buoy location. Here, we use snow depths from 28 AWI snow buoys between 2013-2017 (accessible at: http://data.meereisportal.de), which are also listed in Table S1. Similar to the CREEL buoys, the snow depths are first averaged into daily averages before monthly averages are derived.

2.1.2 OIB Airborne observations

Since 2009, NASA Operation IceBridge (OIB) has conducted airborne profiling of Arctic sea ice every spring, generally across the western Arctic. Snow depth observations are derived with an ultra-wideband quicklook snow radar (Paden et al., 2014), capable of retrieving snow depth from the radar echoes from both air-snow and snow-ice interfaces (Kurtz and Farrell, 2011). Snow depth can then be retrieved through retracking and compensation from the radar echogram. The OIB campaigns provide unique large-scale and high spatial resolution observations, however the swath is limited due to the airborne nature of the measurements. Furthermore, most flight tracks cover a limited area from the north of Greenland towards Alaska, and data covers only a limited time period, namely March and April.

Several algorithms have been developed to derive snow depth from this radar system (Kwok et al., 2017). While different algorithms show general agreement in regional snow depth distributions, larger interannual variability is observed among these algorithms than found in $W99$ and mean snow depth differences result from different ways of detecting the snow-air and snow-ice interfaces from the radar returns (Kwok et al., 2017). Taking these differences into consideration, we use four OIB snow depth products from: (1) Sea Ice Freeboard, Snow Depth, and Thickness data Quicklook product (quicklook) available from NSIDC website and examined in King et al. (2015) and Kwok et al. (2017); (2) NASA Goddard Space Flight Center (GSFC) (Kurtz et al., 2013, 2015); (3) Jet Propulsion Laboratory (JPL) (Kwok and Maksym, 2014) and; (4) snow radar layer detection (SRLD) (Koenig et al., 2016). These four OIB products overlap only in 2014 and 2015, and thus we limit our comparisons with OIB-derived snow depths to those two years. Each OIB product is first regridded to the 100×100 km polar stereographic grid, and then all daily flight tracks are averaged together to produce monthly mean OIB snow depths (March or April) for each year. We also compare snow reconstructions on their native spatio-temporal resolution with OIB measurements in the Supplementary Material.

2.2 Snow depth climatologies

In addition to the airborne and in-situ snow measurements, we use two types of conventional large-scale snow products that are often used in the derivation of sea ice thickness from radar or laser altimetry (Ricker et al., 2014; Kwok and Cunningham, 2008), the Warren et al. (1999) ($W99$) and the Shalina and Sandven (2018) ($SS18$) climatologies. $W99$ provides distributions of snow depth and density for each calendar month by assembling "North Pole" (NP) drifting station observations from the 1950s to 1990s. A two-dimensional quadratic function is adopted to fit the measurements to the Arctic basin. $W99$ also provides a climatology of snow water equivalent (SWE) from January to December. This is derived using snow depths and densities measured along snow lines, and if unavailable, Arctic-mean density for that month is used. Like snow depth, a two-dimensional quadratic fit was applied to the SWE data.

$SS18$ further combines NP data (as in $W99$) with additional snow data from the Soviet airborne expeditions (Sever) obtained primarily during the 1960s to 1980s to produce spring (March-April-May) snow depth fields. Since the aircraft would land on level FYI, $SS18$ is not limited to MYI in the central Arctic (as $W99$), but includes FYI in the Eurasian seas as well (Shalina and Sandven, 2018). The spatial resolution of the $SS18$ climatology is 100×100 km within Arctic basin.

2.3 Satellite- and model-based snow depth products

Eight snow depth data sets are included in this inter-comparison study (Table 1). They mainly fall into 2 categories: (1) snow reconstruction using atmospheric reanalysis data as input to a snow accumulation model together with snow redistribution by sea ice drift; (2) snow depth retrieved from satellite data, including passive microwave-based snow retrieval, blended satellite-derived radar sea ice freeboards at two different frequencies, and active-passive satellite (combining CryoSat-2 and SMOS) data synergy. Here, snow depth is defined as the average thickness of snow over the entire grid-cell area for eight products, not just over the sea ice-covered fraction.

The first category includes four different new products: the distributed snow evolution model (SnowModel-LG) (Liston et al., 2020; Stroeve et al., 2020); NASA Eulerian Snow on Sea Ice Model (NESOSIM) (Petty et al., 2018); the Centre for Polar Observation and Modelling (CPOM) model (Tilling et al., In preparation); and the Lagrangian Ice Tracking System for snowfall over sea ice from University of Washington (UW) (Blanchard-Wrigglesworth et al., 2018). The second category includes the following snow depth products: the products from the University of Bremen (PMW Bremen) (Rostosky et al., 2018) and the Danish Meteorological Institute (PMW DMI) (Winstrup et al., 2019) rely on satellite passive microwaves at multiple frequencies/polarizations for their snow depth retrieval algorithms. The dual-altimeter snow thickness (DuST) product (Lawrence et al., 2018) is derived from combining data from CryoSat-2 (Ku-band) and AltiKa (Ka-band) satellite radar altimeters. The DuST product also combines Envisat (radar altimeter: Ku-band) and ICESat (laser altimeter) data during their period of overlap. Finally, Department of Earth System Science, Tsinghua University (DESS) combines Ku-band altimeter (CryoSat-2) and L-band passive microwave radiometer (SMOS) to retrieve snow depth and sea ice thickness based on two physical models (Xu et al., In preparation). Each approach uses vastly different methodologies and provide snow information at different spatial and temporal resolutions.

2.3.1 Reanalysis-based snow depth reconstruction

SnowModel-LG

SnowModel-LG is a prognostic snow model originally developed for terrestrial snow applications, now adapted for snow depth reconstruction over sea ice using Lagrangian ice parcel tracking (Liston et al., 2020). Physical snow processes are included such as blowing snow redistribution and sublimation, density evolution and snow pack metamorphosis. SnowModel-LG is used in a Lagrangian framework to redistribute snow around the Arctic basin as the sea ice moves. Tracking begins on August 1st 1980 assuming snow-free initial conditions, and accumulates snow until July 31st of the next year. On July 31st, any remaining snow that is isothermal and saturated with meltwater becomes superimposed ice and is no longer identified as ‘snow’. SnowModel-

LG is the only data product that includes snow depth during the melt season. The essential inputs to this data product are atmospheric reanalysis estimates of precipitation, 2m air temperature, wind speed and direction, and weekly ice motion vectors from NISDC (Stroeve et al., 2020). Weekly ice motion vectors are linearly interpolated to daily resolution. Output relevant to this study are the snow depth and bulk snow density (Liston et al., 2020), and are provided on a 25×25 km EASE grid. A recent study (Stroeve et al., 2020) evaluated SnowModel-LG using NASA MERRA2 (Gelaro et al., 2017) and ERA5 atmospheric reanalysis fields (Hersbach and Dee, 2016) against several data sets including OIB, IMBs, MagnaProbe and passive microwave estimates. They found that the model captured observed spatial and seasonal variability in snow depth accumulation, while also showing statistically significant declines in snow depth since 1980 during the cold season. The temporal resolution of SnowModel-LG is daily between 1st August 1980 and 31st July 2018.

NESOSIM

The NASA Eulerian Snow On Sea Ice Model (NESOSIM) is a three-dimensional, two-layer (vertical), Eulerian snow budget model (Petty et al., 2018). NESOSIM includes two snow layers: old compacted snow and new fresh snow. Wind-packing and snow loss to the leads are included, and were used to calibrate NESOSIM with historical snow depth and density observations from the NP drifting station data. NESOSIM was run using several atmospheric reanalyses, including ERA-I (Dee et al., 2011), JRA55 (Ebita et al., 2011) and MERRA (Rienecker et al., 2011) and a median of these daily reanalysis snowfall estimates is used in this study. The model is also forced with near-surface wind fields from ERA-I, NSIDC Polar Pathfinder sea ice drift vectors (Tschudi et al., 2019) and Bootstrap passive microwave ice concentrations (Comiso, 2017). Snow accumulation is initialized at the end of summer (default of August 15th) and run until the following spring (May 1st). To initialize snow depth in mid-August, NESOSIM linearly scales the August snow depth in the $W99$ climatology based on the ratio between duration of the summer melt season and the climatological summer duration. The duration of the summer melt season is defined based on ERA-I air temperatures, and climatological summer melt duration is from Radionov et al. (1997). Snow depth is further equally divided into the 'old' and 'new' snow layers, with snow transferred from the 'new' to the 'old' snow layer based on the wind conditions. Snow is then accumulated and evolves dynamically with sea ice motion through a divergence-convergence and an advection term. Daily snow depth (mean depth over the sea ice fraction) and snow density and snow volume (per unit grid-cell) are available from August 15th to April 30th for each year, at a spatial resolution of 100×100 km in a polar stereographic grid. NESOSIM provides the effective snow depth of the ice covered fraction, and the mean grid-cell snow depth. In order to be consistence with snow depth estimates from the other methods, we will use the latter in this study.

CPOM

The CPOM snow depth product (Tilling et al., In preparation) is initialized on a Lagrangian grid with a spacing of 10 km running from $40°$N to the pole. Snow accumulation begins on 15th August each year using the $W99$ August snow depth and a fixed density of 350 kg/m^3 on all ice covered (sea ice concentration > 15%) grid points. This initial snow layer is kept separated from any accumulated snow after the model has started running. The model then steps daily through the winter, accumulating snow in SWE. Snow parcels are moved using NSIDC Polar Pathfinder ice motion data, and any parcels moving outside the ice-covered region are removed. New parcels covered by expansion of the ice-covered region become active with no initial snow. Where the 2 m ERA-I air temperature is above freezing point, the daily ERA-I SWE of snowfall is added to the already accumulated column. A fraction of the accumulated snow is removed when the wind speed exceeds 5 ms^{-1} using a function proportional to wind speed and lead fraction (see (Schröder et al., 2019), table 1). Finally, the total column of accumulated SWE at each Lagrangian point is converted to snow depth using a daily snow density function constructed in a similar way to Kwok and Cunningham (2008), which is added to the initial snow layer (if present). The irregularly spaced snow data from the Lagrangian grid are re-gridded onto a regular $10km^2$ polar stereographic projection using an averaging radius of 50 km to give a snow depth map for each day of winter.

UW

The last reanalysis-based snow reconstruction is from University of Washington (UW). The algorithm accumulates cold-season snowfall along sea-ice drift trajectories using 12-hourly snowfall from ERA-I, weekly NSIDC sea ice vectors and weekly-averaged NOAA-NSIDC sea ice concentration (Meier et al., 2013). A Lagrangian Ice Tracking System (LITS) (DeRepentigny et al., 2016) is used to backward track each grid point from the first week of April to the last week of the previous August (Blanchard-Wrigglesworth et al., 2018). This tracking system has a claimed accuracy of $50km$ after 6 months of tracking (DeRepentigny et al., 2016). Once the 6-month trajectories are established for each ice parcel, the algorithm accumulates weekly averaged snowfall along each parcel trajectory. A sea ice concentration correction is further imposed every week (one time-step). Specifically, if the ice concentration drops below 15%, the accumulation stops and the trajectory is ended at the previous time-step. Only monthly snow depths in April are available for the period from 1980 to 2015. Spatial resolution of the data set is 75$\times$75 km on a polar stereographic grid.

2.3.2 Satellite-based snow depth retrieval

DuST

Lawrence et al. (2018) derives snow depth by utilizing the difference in freeboards retrieved from radar altimeter satellites operating at different frequencies. Specifically, satellite data from ESA's CryoSat-2 (CS-2, Ku-band radar satellite altimeter operational since 2010) and CNES/ISRO's Altika (Ka-band radar satellite altimeter, 2013-present) are used. The deviation of each satellite's return from its "expected" dominant scattering horizon (the snow surface for Ka-band and the ice/snow interface for Ku-band) is quantified using independent snow and ice freeboards from OIB. Using a spatially variable correction function, AltiKa and CS-2 freeboards are calibrated to the snow surface and snow/ice interfaces respectively, allowing snow depth to be estimated as the difference between the two. A caveat to the approach is that since OIB data are only available during March/April and cover limited regions of the Arctic, the calibration of the CS-2 and AltiKa freeboards with OIB may not be valid during other months and/or regions. The same methodology has also been applied to ICESat and Envisat satellites, whose active periods overlap between 2003 and 2009, and these data are used to extend the snow depth product further back in time.

This Dual-altimeter Snow Thickness (DuST) product produces: (1) monthly snow depth maps from October to April during CS-2/Altika period (2013-present), and (2) bi-monthly annual snow depth maps (March-April or October-November) during the ICESat-Envisat period (2003-2008). For both time-periods, the snow depth is gridded on a $1.5°$ longitude $\times 0.5°$ latitude grid. The data is limited to below $81.5°$N due to the upper latitude of AltiKa/Envisat.

PMW Bremen

The University of Bremen's PMW snow depth product described in Rostosky et al. (2018) on FYI is available during the AMSR-E/2 period. The algorithm is adapted from Markus and Cavalieri (1998), which was derived from a series of passive microwave sensors, such as the Scanning Multichannel Microwave Radiometer (SMMR) (from 1979), and continuing on through the Special Sensor Microwave Imager (SSM/I) and SSM/I Sounders (SSMIS). The latter algorithm computes the spectral gradient ratio between 18.7 and 37 GHz vertical polarization brightness temperatures (Tbs) to generate 5-day averaged snow depth estimations over FYI (Markus et al., 2011). This algorithm has limitations for wet snow spring and multiyear ice (Comiso et al., 2003) and large sensitivity to surface roughness (Stroeve et al., 2006). Snow depths over smooth FYI were found to be accurate in comparisons with OIB during 2009 and 2011 ($RMSE$ < 0.06 m over a shallow snow cover), while significant biases were found over rougher FYI or MYI (Brucker and Markus, 2013).

Rostosky et al. (2018) further extends the approach of Markus and Cavalieri (1998) to also include snow depth over MYI using the lower frequencies of 6.9 GHz from the NASA Advanced Microwave Scanning Radiometer AMSR-E and the JAXA Global Change Observation Mission-Water (GCOM-W) AMSR2 instrument. We refer to the resulting product as the PMW Bremen snow depth product. The gradient ratio between vertical polarized brightness temperatures (Tbs) at 18.7 GHz and 6.9 GHz helps to mitigate the retrieval sensitivity problems over MYI during March and April. Specifically, robust linear regressions were derived based on fitting 5 years' (2009, 2010, 2011, 2014, and 2015) NOAA Wavelet ("WAV") Airborne Snow Radar-Snow Depth on Arctic Sea Ice Data Set (Newman et al., 2014) and the polarized Tb gradient ratio. Snow depths over different ice types were fitted separately, using OSI-SAF sea ice type map to distinguish between FYI and MYI. No evaluation/validation were performed in regions outside of OIB measurements. The spatial resolution is similar to typical passive microwave measurements, at 25×25 km in the polar stereographic grid. Daily snow depth maps for winter months from November to April since 2002 are available. It should be noted that snow depth over MYI is only available during March and April when the OIB data were available to constrain the model. Snow depth uncertainty is estimated to be between 0.1 and 6.0 cm over FYI and between 3.4 and 9.4 cm over MYI.

PMW DMI

In Winstrup et al. (2019), snow depth is derived by a random forest regression model based on passive microwave Tbs from AMSR-E and AMSR2. The model was trained using a Round Robin Data Package (Pedersen et al., 2018), with OIB campaigns snow thicknesses provided by NSIDC including IDCSI4 and quicklook products and collocated brightness temperatures. Training was performed on 2/3 of the data set, using OIB data from the period April 2009-March 2014, leaving the remaining OIB data (period: March 2014-April 2015) for validation purposes. Specifically, multi-channel Tbs ranging from C-band (6.9 GHz) to 89 GHz are included as predictors, at both vertical and horizontal polarization. The random forest consists of 500 regression trees, each derived from bootstrapping the input data, and using a maximum limit of five predictors for each leaf in the regression trees. The most important predictors, as found by the algorithm, were the following channels: 18.7 GHz (27%), 23.8 GHz (17%) and 10.7 GHz (11%), all vertical polarizations. Derived snow depths were found to be in good agreement with the OIB data retained for validation, with an average error of 0.05 m, and an accuracy of 78.7%. The passive microwave snow depth product from the Danish Meteorological Institute (PMW DMI) constructs spring-time (March and April) daily snow depth between 2013 and 2018 at 25×25 km spatial resolution on the EASE-Grid 2.0 (Brodzik et al., 2012).

DESS

The Department of Earth System Science in Tsinghua University (DESS) algorithm retrieves both sea ice thickness and snow depth simultaneously by using sea ice freeboard from CS2 and L-band (1.4 GHz) Tbs from the Soil Moisture and Ocean Salinity (SMOS) satellite (Xu et al., In preparation). The active period of these two satellites are both from 2010 to present. Specifically, this algorithm combines a hydrostatic equilibrium model and improved L-band radiation model (Xu et al., 2017). Sea ice freeboard is converted to sea ice thickness based on hydrostatic equilibrium (Laxon et al., 2003), using assumptions on snow, ice and water densities. Tbs from SMOS can be used to retrieve thin sea ice thickness (Kaleschke et al., 2012) and snow depth over thick ice (Maaß et al., 2013). The L-band radiation model is further improved by adding vertical structure of temperature and salinity in sea ice and snow (Zhou et al., 2017). In order to obtain the missing measurements resulting from limited upper latitude in the SMOS satellite, L-band Tbs spanning inclination angles from $0°$ to $40°$ and from $85°$N to $87.5°$N is approximated using Tbs of all frequencies in AMSR-E and AMSR2 through a back-propagation machine learning process. By combining the two observational datasets, the uncertainties in both sea ice thickness and snow depth are largely reduced. Unlike optimal interpolation-based sea ice thickness synergy, as applied in Ricker et al. (2017), the uncertainty in ice thickness is reduced through an explicitly retrieved snow depth.

Both sea ice thickness and snow depth are available in the DESS product. Here we use the monthly snow depth maps available for March of each year since 2011, which are provided at a spatial resolution of 12.5×12.5 km on a polar stereographic grid.

3 Snow products inter-comparison

In this section, we carry out a systematic comparison of the eight snow products, using the two climatological snow datasets ($W99$ and $SS18$) as a reference for all comparisons and analyses. Section 3.1 presents the results for mean snow depth and its spatial distribution, Section 3.2 introduces the seasonal cycle, whereas the long-term trend and inter-annual variability is presented in Section 3.3. For the two products that model an evolving snow density, we also intercompare these against climatology (Section 3.4). While we focus on basin-wide estimates for the whole Arctic, we also explore consistency and mean state of all products over three sub-regions: the Canadian Arctic sector (CA) including Canadian Archipelago, the Atlantic (Atlantic) and the Pacific & Central Arctic (Pacific) sectors (regions outlined in Figure S1). Over these regions, sea ice conditions vary considerably, with mostly thick MYI within the CA, and thinner FYI elsewhere. The North Atlantic has generally more precipitation as a result of proximity to the North Atlantic storm tracks.

Given the relatively long time period and basin-scale coverage provided by SnowModel-LG, this snow depth product is used as the reference product when carrying out the regional consistency checks. To align the temporal and spatial resolution between all snow products, comparisons are mainly carried out after year 2000 and focused on the early and late winter months of October/November and March/April. The products containing data for longer time periods are further explored against climatologies and their long-term trends are assessed.

3.1 Mean state and distribution of snow depth

Mean snow depth, which is a direct indicator of total snow volume over sea ice, is of fundamental importance for the characterization of the Arctic hydrological cycle. Spatial maps of monthly mean snow depth across all data products, as well as the $W99$ and $SS18$ climatologies, are shown in Figure 1 (and S2) in spring (autumn) months for the season of 2014-2015. Products are shown with their native resolution and grids. The spatial patterns in all products are in broad agreement that thicker snow occurs north of Greenland and the CA sector and thinner snow in the seasonal ice zones (i.e., Baffin Bay and marginal seas of the Eurasia continent). Some products also show thicker snow in the East Greenland Sea (NESOSIM in particular) and the Atlantic sector of the Arctic. Despite some general agreements, mean spring snow pack and regional discrepancies are evident in Figure 1. In particular, the thickest snow in late winter/early spring for NESOSIM occurs in the East Greenland Sea, while in DESS, the deepest snow is concentrated in the Canadian Arctic.

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 (Figure S2). DuST also shows deep snow packs in this region (16.0 cm mean snow depth), though the spatial coverage is more limited. Spring snow depth, ranging from 25.0 cm to 30.0 cm in the Arctic domain (Figure S1), exhibits large spatial variability among all products. Further, relatively thick 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 winter precipitation from the North Atlantic storm tracks. For comparison, snow is also the deepest (over 35.0 cm) 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 25.0 cm. 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 histogram of time-averaged snow depth in the different products is shown in Figure 2 for the period 2000 to 2018 during the spring and autumn periods, respectively. Out of all the reanalysis-based data products, snow depth distributions in NESOSIM are shifted towards slightly deeper snow packs (8.0 cm) than those from SnowModel-LG (7.0 cm) and CPOM (6.0 cm) during autumn, although the shapes of the distributions are similar. The deepest snow packs during October/November are found in DuST, with a mode of the distribution at about 17.0 cm. This is much larger than other products and seems unlikely to be realistic early in the snow accumulation season. During spring (Figure 2.b), PMW DMI exhibits the overall smallest snow depth in late winter $(14.5cm)$, while UW shows the largest snow depth $(25.2cm)$ in the common region of the snow products. Conclusions are similar if we omit DuST, and extend the analysis up to $87.5°$N. Doing so, but the bimodal snow depth distribution in autumn is more evident, suggesting the separation of snow cover over old and newly formed sea ice. As expected, snow depths shift to overall deeper snow packs in spring, especially when including the higher latitudes (Figure S3). The shallowest spring snow in PMW DMI, with a mean snow depth < 15.0 cm, and the thickest autumn snow in DuST, are the two extremes among current snow products.

Additionally, we examine snow depth over the three different sectors in spring 2015 (Figure S4). Deepest snow packs from reanalysis-based snow products mainly occur over the North Atlantic, while satellite-based products indicate more snow accumulating over the 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.

Discrepancies between snow products and the climatologies both in spring (Figure 3) and in autumn (Figure S5) are evident. In autumn (Figure S5), bimodal snow distributions are noticeable both in SnowModel-LG and NESOSIM using the data post 2000, with a large proportion of thin snow not seen in the $W99$ climatology. By the end of winter, the differences between all products and $W99$ are larger (Figure 3) than in autumn, with the minimum decreasing is 10.0 cm for UW and the maximum is over 15.0 cm for PMW DMI in spring. SnowModel-LG, DuST and DESS all have snow depths below 10.0 cm in March/April. For reanalysis-based products, snowpack is still significant shallower against climatologies in the 1980s (1990s in the case of CPOM) when $W99$ is partly collected. In addition to mean snow depth, the skewness of the snow distribution, especially in spring, among all products are mainly positive as a result of the larger presence of thin snow cover over FYI dominated in the current era, while that of $W99$ is in the opposite.

Apart from $W99$, the Shalina and Sandven (2018) climatology provides additional snow depth information in the marginal seas, especially over the Eurasian seas. Overall $SS18$ has lower snow depths in the central Arctic compared with $W99$. Figure 4 reveals that snow distribution in $SS18$ includes two modes: 18.0 cm and 32.0 cm, which correspond to snow over FYI and MYI respectively. Generally, differences between the $SS18$ climatology and the various snow products are similar to the comparison results with $W99$, but $SS18$ tends to exhibit similar bimodal distribution as seen in some of the snow products.

To summarize, there is general agreement among the products (except for UW) that there is a distinct difference in snow depth on perennial and seasonal ice. It is worth noting that this agreement is across the two distinctive methods for snow depth reconstruction, i.e., the reanalysis based numerical integration, and satellite based retrievals. This is also reflected in both the basin-scale snow depth averages and the spatial patterns. However, the areas of where the thickest/thinnest snow depths are found tend to differ between the two types of methods. The heavy snow in reanalysis-based products falls primarily over the East Greenland Sea and Atlantic sector as a result of frequent storm tracks over this area, whereas from active or passive microwave satellite data the thickest snow is detected over the Canadian Arctic sector due to the different features of old *vs.* newly formed ice. Finally, the intercomparison of the different snow products, and further against two climatological data sets, reach similar conclusions: (1) all snow products have similar structure such that snow over perennial ice is thicker, but quite large regional discrepancies in mean value, (2) snow depths in the products evaluated show less snow in recent years relative to the climatologies.

3.2 Seasonal cycle of snow depth

Time-series (2000-2018) of monthly mean snow depths during winter (September to April) averaged over regions up to $81.5°$N are displayed in Figure 5, whereas Figure 6 shows the seasonal mean and spread. The $W99$ is included as reference (see Figure S6 for the results when the region is extended up to $87.5°$N).

In NESOSIM, April snow depths are higher than in SnowModel-LG (Figure 5), especially after 2012. We further see from Figure 6 that the initial snow depth before autumn is thinner in SnowModel-LG, but there is more wintertime snow accumulation in SnowModel-LG than in NESOSIM. Thicker autumn snow in NESOSIM is a result of using $W99$ for the initial conditions, while SnowModel-LG tracks the snow through the summer melt season, removing any remaining snow at the end of summer when the snow pack is saturated and isothermal (this then becomes decomposed ice). The end result is that by the end of winter, overall differences between the two products are less pronounced despite large differences in total snow accumulated over winter.

Seasonal snow accumulation is also larger in SnowModel-LG compared with CPOM. In contrast, the seasonal accumulation in DuST and snow changes from March to April in PMW Bremen are unexpectedly small, even negative, while the PMW DMI shows similar seasonal changes as the reanalysis-based products. Overall, the largest seasonal snow accumulation occurs in $W99$ and the deepest snowpacks are observed in March, the month when snow reaches its maximum depth, whereas many of the other products reach their maximum snow depths in April. However, based on Section 3.1, compared with $W99$, all snow cover products in early winter have thinner snow packs than at the end of the winter, which implies either that the intensity of snow accumulation is weakening or that the snow accumulation period has shortened. We further note that $W99$ snow accumulates more snow during the early part of winter and thus the seasonal curves are flattened near spring. SnowModel-LG, NESOSIM and CPOM on the other hand share similar seasonal accumulation curves, with accumulation continuing to increase through winter. This seasonal pattern of winter snow accumulation finds support in a recent study by Kwok et al. (2020) that found accumulation later in winter in 2018-2019. This implies that there may be limited efficacy of the climatological seasonal pattern of snow accumulation in the current Arctic climate. As sea ice freeze-up continues to delay further over the last 40 years, it is expected that the early snow accumulation will continue to differ from that reflected in $W99$.

3.3 Trend and inter-annual variability of snow depth

All reanalysis-based snow reconstructions, namely SnowModel-LG, CPOM, NESOSIM and UW have consistent inter-annual variability in spring snow volume from 2000 to present (see details in Table S2), with statistically significant (confidence level 99%) correlations of the interannual snow depth/volume between the data sets. DESS also exhibits generally similar inter-annual variability as the reanalysis-based products (R^2= 0.42 to SnowModel-LG; 0.68 to NESOSIM and; 0.32 to CPOM), whereas PMW Bremen and DuST do not.

While there is interannual variability in snow accumulation, the variability is overall quite small, ranging from 2.0 cm in November to about 2.0 cm to 3.0 cm in April. This also holds for results as averaged up to $87.5°$N (Figure S6). The interannual variability seen in the data products is about half of that previously reported in $W99$, where interannual variability was estimated to be about 4.3 cm in November and 6.1 cm in April. It is important to note, however, the climatological estimation of inter-annual variability in $W99$ includes snow depth uncertainties, and should be treated as an upper bound for the inherent physical interannual variability. It should be also noticed that DuST shows a significant positive snow depth bias from the Envisat period to the CryoSat-2 period, likely a result of using OIB to calibrate the snow depth estimates. This positive snow bias is missing in the passive-microwave based snow products (e.g., PMW Bremen), and is not observed in DESS (time-series in PMW DMI is too short to be assessed).

As averaged over the common regions of all data sets, no significant trend is observed in most snow products except DuST since 2000. However, regionally the reanalysis-based snow products exhibit regions of statistically significant positive and negative snow depth trends over a longer time-period. For example, positive snow depth trends are found north of Greenland and the Canadian Archipelago from 1991 to 2015 (common period) (Figure 7). These positive trends may be a result of more autumn precipitation (Serreze et al., 2012)or changes in the proportion of MYI *vs.* FYI in the region. On the other hand, snow depth trends in spring are mostly negative within the rest of the Arctic basin, statistically significant for SnowModel-LG in FYI regions, and for CPOM in Barents Sea. In autumn, the regions with statistically significant negative snow depth trends in SnowModel-LG are larger, which is likely a result of delays in freeze-up (Markus et al., 2009; Stroeve and Notz, 2018). CPOM also shows negative trends in these regions, but not as large as those from SnowModel-LG. April and November mean snow depth trends from SnowModel-LG as computed over the entire Arctic basin are -0.5 $cm/decade$ and -0.9 $cm/decade$, respectively, although some regions show larger trends. In CPOM, basin-mean significantly negative trend (-0.47 $cm/decade$) is only found in November.

Finally, we synthesize long-term changes in snow depth in relationship to the climatology products. As mentioned in Section 3.1, the minimum and maximum differences of mean snow depth in March/April between products in the current years and climatology are 10.0 cm and 15.0 cm respectively over the last 40 years, thus the inter-decadal snow depth changes would be in the range of -0.25 $cm/year$ and -0.375 $cm/year$. These estimates span the value of -0.29 $cm/year$ in Webster et al. (2014). However, one should keep in mind that there are large uncertainties in the snow climatology data sets and the interannual variability is larger than in the snow products.

3.4 Snow density comparison

Apart from snow depth, we further investigate snow density in the two products that provide snow density estimates: SnowModel-LG and NESOSIM. Since the $W99$ climatology contains both snow depth and SWE, we can compare against the $W99$ snow densities. Snow density in $W99$ is limited to the Arctic basin. SnowModel-LG suggests that snow is denser than in NESOSIM in both November and April, as shown in 8. Considering that the snow depth in SnowModel-LG is thinner than NESOSIM, we find that the two models provide a broadly equivalent SWE (not shown). Snow over the Atlantic sector, especially within the

East Greenland Sea, is the densest in SnowModel-LG, with mean density values above 370 kg/m^3 in November and April. In contrast, NESOSIM has mostly smaller snow densities and considerably less spatial variability. For $W99$, the snow is denser over the Atlantic sector in November, while in April, the denser snow is over the Pacific sector.

Time-series of wintertime mean snow density within the Arctic basin in SnowModel-LG, NESOSIM and $W99$ are summarized in Figure 9. Both SnowModel-LG and NESOSIM show an increase in snow density into the winter, which is consistent with $W99$. However, snow density in SnowModel-LG is consistently higher than that of NESOSIM, with more pronounced differences at the end of winter; $W99$ falls between the two estimates at 320 kg/m^3. Seasonally, SnowModel-LG densities increase more from October to April than in NESOISM and $W99$. Neither the NESOSIM nor SnowModel-LG densities suggest any long-term changes in snow density, yet SnowModel-LG shows considerable interannual variability of spring and winter snow density, which is not present in NESOSIM or $W99$.

4 Comparison of snow depth products against observations

In this section, we compare the snow depth products against the observational datasets. Since snow depth data from OIB play an important role in the development of some of the snow products, the comparison and validation potentially suffer from the problem of data dependency. This applies to both reanalysis based snow reconstructions and the satellite-retrieved snow depth fields. Buoy-based comparisons are free from the data dependency problem, yet the limited local spatial coverage of buoys hinders direct comparisons with the products in study, which are all on much larger spatial scales (>10 km). Therefore, after the intercomparisons we discuss the complexities in validating different snow products and observations with in-situ and airborne measurements in Section 4.3.

4.1 Comparison against OIB

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 100×100 km 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 significantly correlations (over 0.36 for PMW DMI) with OIB as shown in Table 2. Therefore, we do not consider this a suitable validation (or comparison) for these products. For the PMW DMI product, however, OIB data from the period from March 2014-2015 was not used during model development, and hence we have more confidence in the high correlation observed. Except for NESOSIM and UW, the other reanalysis-based products are also to some extent indirectly tuned by OIB snow depths in some years. Overall, all products show reasonably high correlations with the different OIB snow estimates except for UW, which only shows a slight correlation with some versions of the OIB data.

It should be noted that OIB snow depth is itself a derived product, and thus caution is warranted when interpreting the results of the comparison with snow products. In fact, a strong dependence of our validation results to the specific OIB data product is evident by the different linear fitting slopes shown in Figure 10, as well as the R^2 values, $RMSE$ and normalised $RMSE$ ($NRMSE$: $RMSE/(max-min)$) in Table 2. The fit is best for the PMW Bremen and PMW DMI data, followed by the CPOM and DESS products, though this depends on the choice of OIB data set used for evaluation. For example, CPOM performs best against SRLD (R^2=0.61) and worst against GSFC (R^2=0.43); DESS also performs best against (R^2=0.59) but worst against quicklook (R^2=0.26). SnowModel-LG and NESOSIM have similar R^2 ranging from a low R^2 of 0.27 and 0.29, respectively against the quicklook product to a high of 0.47 (SnowModel-LG *vs.* SRLD) and 0.39 (NESOSIM *vs.* GSFC). The UW snow product, on the other hand, performs poorly against all OIB snow depth estimates (maximum R^2 of 0.31 with SRLD). Among all non-directly OIB-fitted snow products (SnowModel-LG, CPOM, DESS, UW and NESOSIM), RMSE in CPOM is the lowest, while in DESS the RMSE is over 10.0 cm. The distribution/variability in UW is narrow compared with other products related to the lack of spatial variability of snow depth across the basin.

In order to avoid biases introduced by temporal and spatial averaging and interpolation during the validation, we also carry out the comparison for each of the data products on their native grids and native temporal resolution (Figure S7). In contrast to the coarser resolution comparisons shown in Figure 10 and Table 2, more outliers, lower R^2 values and larger $RMSE$s are evident in these native spatial-temporal resolution comparisons. The comparison results still depend on the choice of OIB data product, and in some instances the statistical correlation improves. There is still no significant correlation between UW snow depths and those from OIB, but the overall $RMSE$ is the lowest recorded at less than 4.0cm. PMW DMI has the lowest $NRMSE$, followed closely by all other snow products, with the largest $NRMSE$ in DESS. All snow products except NESOSIM and UW show higher R^2 under spatial coarser resolution compared to Figure 10, Figure S7 and Table S3. The comparisons between monthly and daily scale estimates suggests that temporal resolution has little influence on OIB comparisons, since there are only small changes in R^2 and $RMSE$, without significant differences. However, spatial resolution does impact the statistical fits, which is related to the lack of representation between OIB and these products at the coarse scale of 100 km (discussed further in section 4.3).

Given the potential data dependency problem and the sensitivity to the specific OIB data set, it is not possible to conclude which snow product performs best. Snow products that have been produced through tuning with OIB data show higher R^2 and smaller $RMSE$s. The PMW DMI product performs best, despite not being tuned with OIB data for the time period to which it is compared. The outlier is the UW product. In summary, there is a need for a 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.

4.2 Comparison with buoy data

We further explore how well the snow products represent the mean state of small scale snow depth on Arctic sea ice 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) were processed from 2000 to 2017 (Table S1). 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.

Next, we focus on temporal evolution as we do not expect the local-scale point measurements of the IMBs/snow buoys to match with the coarse spatial scale of the data products. We compare snow depth differences in the products along the buoys' drifting tracks against the accompanied snow accumulation measured by buoys during winter (Figure 12). Specifically, the three daily-resolution products (SnowModel-LG, NESOSIM and CPOM) are interpolated onto daily geolocations of the buoys. Only buoys with valid measurements from October until the following February are considered. Snow accumulation is then calculated by subtracting the mean snow depth within the first seven days and that in the last seven days. None of these products show significant correlation with buoy accumulation because several outliers weaken the overall correlation. Based on Stroeve et al. (2020), good correlations are witnessed between SnowModel-LG and buoy data using the buoy location in each integration step. Therefore, the lack of correlation we observe may be due to large discrepancies in trajectories determined from the ice drift products and from buoys, especially after a long-term integration.

In general, the comparison with snow depth measurements from buoys highlight two important limitations of these sorts of comparisons. First, pronounced representation issues are present when comparing point measurements to coarse-resolution products despite buoys being intentionally deployed to represent the adjacent areas, and despite the fact that they potentially traverse large geospatial ranges. Second, sea ice drift is a major factor in determining the uncertainty of snow depth estimations during the numerical integration of reanalysis-based approaches.

4.3 Study of representation issues

Comparison and validation of snow depth products with observations of vastly different spatial resolutions and coverage is complex. For OIB, although the ground tracks cover large regions, the aggregate footprint size is small. For example, an average of 1300 OIB samples are used to compute the 100×100 km cell-mean snow depth and the aggregate area is about 0.6 km^2, which is 0.006% of the 100 km cell. However, the spatial coverage of OIB mitigates this problem, by introducing large-scale variability in computing the cell-mean values of snow depth. For comparison, buoy measurements are the extreme case of limited representation for products with coarse resolution. Although the absolute error for each buoy measurement is low, the practice of using it for validation purposes should be scrutinized.

n order to study the effect of limited spatial coverage on validation, we utilized an OIB dataset to simulate various spatial resolutions of products that can be used for snow depth validation, including the extreme case of point measurements (e.g. buoys). Specifically, we divide OIB scans into segments of about 37.5 km, which is the typical resolution of the snow products evaluated in the study. The mean snow depths from all OIB measurements within each segment are computed and treated as the "true" now depth (hereinafter referred to as Hs), which is then used as the snow depth measurement to be validated. Re-sampling, that is multiple selections from samples, of the OIB measurement in each segment is carried out, in order to simulate the various observations with limited coverage. By studying the behaviors of fitting snow depth (Hs) values to samples, we observe the change in the slopes and qualities of the statistical fits. Three sampling strategies are designed: (I) random re-sampling, for which 40 OIB samples are randomly chosen within the segment region and used to compute the mean Hs; (II) segment-based re-sampling, for which a local OIB segment track with 40 consecutive samples is chosen and used to compute mean Hs; and (III) single sample, for which a single OIB Hs measurement is chosen to represent the locally-measured snow depth. Strategy III essentially mimics a buoy observation in one day within typical passive microwave satellite spatial resolution (i.e., limited spatial coverage). Specifically, in order to further increase the validity of Strategy III for mimicking buoy measurements, we make sure that the thickness of the random single sample in Strategy III is within 1 standard deviation of the mode of the OIB ice thickness distribution.

Figure 13.a shows the snow depth distribution from strategy I (red lines) and strategy II (blue lines) for an example segment, in order to demonstrate the simulation strategies. A typical case of the fitting in strategy III is shown in Figure 13.b, with each dot representing a local region of 37.5 km. After extracting and averaging the sample in different strategies, snow distribution is much narrower in strategy I and II than in the original measurements, especially in random re-sampling. Using Monte-Carlo simulations with re-sampling, we compute the distribution of fitting slopes for the three strategies, shown in Figure 13.c. In strategy I (random re-sampling), due to the small number of OIB samples chosen (40 samples), the slopes of the linear fitting lines are lower than 1 (around 0.91). This indicates that even if there is overall good coverage of the measurements, the limited footprint still causes the fitting lines to flatten. If we limit the measurement of Hs to a local segment (40 consecutive samples: strategy II), the slopes drop to about 0.73, which indicates that the limited representation of the local segment to regional variations further affects the validations. This is also in direct contrast with strategy I. When we further limit the comparison to a single OIB sample (strategy III), the slopes further decrease to around 0.24. It is important to note that, although the fitted lines are very flat (an example in Figure 13.b), the fittings are all statistically significant.

This result agrees with the buoy validation results in Section 4.2, where the fitted slope between snow products and buoy data (Figure 11) is considerably flatter when compared with the results involving OIB observations, as a result of both large aggregate footprint and wider spatial coverage of OIB. Yet, the study with OIB data suggests that the product validation with buoy data should yield statistically significant correlations. This result also indicates that when tuning the prognostic/statistical models against airborne or in-situ measurements for reconstructing snow, the uncertainty concerning limited representation should be addressed in order to avoid over-fitting. The systematic study of uncertainty quantification for limited representation is beyond the scope of this paper, and planned as future work.

5 Summary and outlook

This paper offers a detailed assessment of current snow products over sea ice. Although general consistency in snow structure is witnessed among the products in terms of deeper snow over perennial ice, regionally large spatial and temporal discrepancies do occur. Further, the magnitude of seasonal snow accumulation differs, as do long-term trends. Among the products evaluated, UW and PMW DMI provide the upper and lower extremes of snow volume/depth at the end of the freezing season. Wintertime snow accumulation is overall less than than in the $W99$ climatology, and is postponed towards springtime in the reanalysis-based products. On the other hand, the satellite-derived DuST product does not exhibit a distinct wintertime snow accumulation. Inter annual variability of percipitation in the majority of reanalysis-based products were found to be consistent (Barrett et al., 2020) and thus we may have expected the reanalysis-based snow depth to share similar inter-annual variability in snow depth. However, different methods for accumulating the precipitation in the reanalysis, as well as how that precipitation is redistributed within the Arctic basin under ice motion leads to pronounced differences among the reanalysis-based products. The March/April interannual variability for all snow products was about half that previously estimated by $W99$ (3.0 cm $vs.$ 6.1 cm). Furthermore, none of the current snow products display a trend in overall snow depth since 2000, although they all suggest significantly thinner snow in spring than found in the climatologies. Over a longer time-period however, several of the reanalysis-based snow products do show statistically significant trends towards thinner snow.

We further compared each product against OIB and buoy observations. All snow reconstructions demonstrate significant correlation with OIB observations, except for UW, which only shows a slight correlation with some versions of the OIB data. Correlations are lower when compared against buoy measurements (no significant correlation in PMW Bremen and PMW DMI), with the highest correlation of 0.4 in DESS. We suggest that poor correlations are at least in part caused by representation issues when trying to validate coarse observations against finer resolution measurements. More effort is required to simulate how snow accumulates locally since currently no product has the ability to model snow depth at local spatial scales, as captured by observational platforms such as buoys. For the reanalysis-based products, improvement in the ice drifting algorithm and snow redistribution parameterisation may result in more accurate representation of snow distribution. Based on the results, we pick SnowModel-LG, NESOSIM, CPOM and DESS as the products with the highest consistency. While we cannot reach conclusive results on which snow product best represents the "true" snow depth, we emphasize the importance of further coordinated intercomparisons, and call for community effort to collect more independent observations for snow over sea ice.

Despite its importance for polar climate studies and sea ice altimetry, the snow cover on sea ice remains a scarcely observed parameter. Airborne surveys such as OIB provide a good trade-off between spatial resolution and coverage, yielding one of the richest snow depth observations for the polar regions. Although it is a derived product, OIB snow depth has become a major reference for model tuning, as well as the benchmark for many snow depth reconstruction products. Therefore, although the comparisons with OIB snow depth differs among the products (Section 4.1), there remain inherent limitations regarding data dependency. Given the status quo of limited snow depth products as reference data sets, more independent observations are required. The problems associated with validation and data dependency would be avoided by using separate training and testing datasets (such as subsets of OIB snow depth data) for the model tuning of snow reconstructions.

Due to the resolution differences of the various snow depth products, spatial and temporal representation issues need to be explored further to facilitate detailed comparison and validation. This is especially true when validating coarse spatial resolution data with local snow depth measurements such as buoy or a single OIB sample (Section 4). In order to address these issues, the multi-scale characteristics of the snow should be studied in a systematic manner, including: (1) its distribution and interaction with sea ice (Xu et al., 2020), and (2) the scaling properties for a wide range of measurement footprints, including in-situ measurements with snow probes and snow stakes (footprint < 0.1 m^2), airborne measurements (100 m^2), and typical satellite passive microwave imagers (100-1000 km^2). Uncertainty quantification can then be carried out and incorporated in the development of new snow remote sensing techniques, as well as product intercomparison and validations.

With climate change and Arctic warming, the snow cover on sea ice has undergone drastic changes. A longer melting season, later freeze-up and shrinking of the perennial sea ice cover have resulted in an overall thinning of the snow cover (Section 3.1). However, the polar hydrological processes have become more pronounced with warming, with stronger and more frequent polar storms inducing more snowfall (Webster et al., 2019). In addition to these competing factors influencing the total amount of snow on Arctic sea ice, the snow stratigraphy and potential changes are a challenge for both remote sensing and modeling of the snow cover. With the combined factors of more snowfall and thinner sea ice, the resulting snow-ice formation causes changes in the thermodynamic and radiometric properties of the snow cover, as witnessed in N-ICE2015 (Gallet et al., 2017; Merkouriadi et al., 2017). In addition, the repeated intrusions of warm air causes snow stratigraphy changes and even thawing-refreezing cycles, bringing challenges to both airborne and satellite remote sensing of the sea ice and snow cover. Furthermore, the long-term increase of Arctic precipitation is projected to continue in the future decades, with potential transition in the phase of precipitation (Bintanja and Selten, 2014). The complex processes involving snow stratigraphy pose further challenges for the modeling of active/passive radiometry of snow and sea ice, as well as that of prognostic modeling of snow over sea ice.

Current and future in-situ and satellite campaigns contribute to an ever-growing and enriching catalogue for snow and sea ice observations. In-situ campaigns, such as MOSAiC, provide annual, process level studies of the atmosphere-ice-ocean coupled system for the Arctic. Novel satellite campaigns such as the multi-frequency radar altimeter CRISTAL (Kern et al., 2020) will provide unique capabilities for snow depth measurement and sea ice altimetry. The abundance of the methodologies and techniques that are investigated in this study show that there is general consistency among the products derived by both numerical and satellite-based approaches, and snow on Arctic sea ice is significantly thinner against observations from the last century. However, some spatial/temporal differences among these products calls for more attention for more independent snow observations as well as modeling improvements.

Data availability. CRREL and AWI buoy data are downloaded from the website (http://www.imb-crrel-dartmouth.org/imb.crrel) and (http://data.meereisp IceBridge data were provided by A. Petty. NESOSIM data are available from https://neptune. gsfc.nasa.gov/csb/index.php?section=516, UW data are from http:// www.atmos.washington.edu/ ed/ data/snow, DuST data are available from http://www.cpom.ucl.ac.uk/DuST and snow data from SnowModel-LG, CPOM, DESS, PMW Bremen and PMW DMI are available from the related authors upon request.

Author contributions. LZ and SMX contributed the DESS data. JS and GEL contributed the SnowModel-LG data. AP contributed the NESOSIM and three different OIB data (JPL, GSFC and SRLD). RT and AR contributed the CPOM data. MW contributed the PMW DMI data. PR contributed the PMW Bremen data. IRL and MT contributed the DuST data. LZ and JS wrote the paper with inputs from co-authors, and all authors contributed to editing of the manuscript.

Competing interests. The authors declare that they have no conflict of interest

Acknowledgements. This work is partially supported by the National Key R & D Program of China under the grant number 2017YFA0603902 and the National Science Foundation of China under the grant number 42030602. This work is also partially supported by Center for High-Performance Computing and System Simulation, Pilot National Laboratory for Marine Science and Technology (Qingdao). The work of P. Rostosky was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project Number 268020496 TRR 172, within the Transregional Collaborative Research Center ArctiC Amplification: Climate Relevant Atmospheric and SurfaCe Processes, and Feedback Mechanisms $(AC)^3$. MT acknowledges support from the European Space Agency Project in part by project “Polarice” under Grant ESA/AO/1-9132/17/NL/MP and in part by the project “CryoSat+ Antarctica” under Grant ESA ESA AO/1-9156/17/I-BG.

References

Barrett, A. P., Stroeve, J., and Serreze, M. C.: Arctic Ocean Precipitation from Atmospheric Reanalyses and Comparisons with North Pole Drifting Station Records, Journal of Geophysical Research: Oceans, 2020.

Bintanja, R. and Selten, F.: Future increases in Arctic precipitation linked to local evaporation and sea-ice retreat, Nature, 509, 479–482, 2014.

Blanchard-Wrigglesworth, E., Webster, M., Farrell, S. L., and Bitz, C. M.: Reconstruction of snow on Arctic sea ice, Journal of Geophysical Research: Oceans, 123, 3588–3602, 2018.

Braakmann-Folgmann, A. and Donlon, C.: Estimating snow depth on Arctic sea ice using satellite microwave radiometry and a neural network, The Cryosphere, 13, 2421–2438, 2019.

Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., and Savoie, M. H.: EASE-Grid 2.0: Incremental but significant improvements for Earth-gridded data sets, ISPRS International Journal of Geo-Information, 1, 32–45, 2012.

Brucker, L. and Markus, T.: Arctic-scale assessment of satellite passive microwave-derived snow depth on sea ice using Operation IceBridge airborne data, Journal of Geophysical Research: Oceans, 118, 2892–2905, 2013.

Comiso, J. C.: Bootstrap Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS, Version 3, boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/7Q8HCCWS4I0R. [Date Accessed: 2020-10-19], 2017.

Comiso, J. C., Cavalieri, D. J., and Markus, T.: Sea ice concentration, ice temperature, and snow depth using AMSR-E data, IEEE Transactions on Geoscience and Remote Sensing, 41, 243–252, 2003.

Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 137, 553–597, 2011.

DeRepentigny, P., Tremblay, L. B., Newton, R., and Pfirman, S.: Patterns of sea ice retreat in the transition to a seasonally ice-free Arctic, Journal of Climate, 29, 6993–7008, 2016.

Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi, K., et al.: The Japanese 55-year reanalysis“JRA-55”: an interim report, Sola, 7, 149–152, 2011.

Eicken, H., Grenfell, T., Perovich, D., Richter-Menge, J., and Frey, K.: Hydraulic controls of summer Arctic pack ice albedo, Journal of Geophysical Research: Oceans, 109, 2004.

Gallet, J.-C., Merkouriadi, I., Liston, G. E., Polashenski, C., Hudson, S., Rösel, A., and Gerland, S.: Spring snow conditions on Arctic sea ice north of Svalbard, during the Norwegian Young Sea ICE (N-ICE2015) expedition, Journal of Geophysical Research: Atmospheres, 122, 10–820, 2017.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, 2017.

Giles, K. A., Laxon, S. W., and Ridout, A. L.: Circumpolar thinning of Arctic sea ice following the 2007 record ice extent minimum, Geophysical Research Letters, 35, 2008.

Guerreiro, K., Fleury, S., Zakharova, E., Rémy, F., and Kouraev, A.: Potential for estimation of snow depth on Arctic sea ice from CryoSat-2 and SARAL/AltiKa missions, Remote Sensing of Environment, 186, 339–349, 2016.

Hersbach, H. and Dee, D.: ERA5 reanalysis is in production, ECMWF newsletter, 147, 5–6, 2016.

Kaleschke, L., Tian-Kunze, X., Maaß, N., Mäkynen, M., and Drusch, M.: Sea ice thickness retrieval from SMOS brightness temperatures during the Arctic freeze-up period, Geophysical Research Letters, 39, 2012.

Kern, M., Ressler, G., Cullen, R., Parrinello, T., Casal, T., and Bouffard, J.: Copernicus polaR Ice and Snow Topography Altimeter (CRISTAL) Mission Requirements Document, version 2, available from the European Space Agency, ESTEC, Noordwijk, The Netherlands, 72 pp., 2019.

Kern, M., Cullen, R., Berruti, B., Bouffard, J., Casal, T., Drinkwater, M. R., Gabriele, A., Lecuyot, A., Ludwig, M., Midthassel, R., et al.: The Copernicus Polar Ice and Snow Topography Altimeter (CRISTAL) high-priority candidate mission, The Cryosphere, 14, 2235–2251, 2020.

Kilic, L., Tonboe, R. T., Prigent, C., and Heygster, G.: Estimating the snow depth, the snow-ice interface temperature, and the effective temperature of Arctic sea ice using Advanced Microwave Scanning Radiometer 2 and ice mass balance buoy data, The Cryosphere, 13, 1283–1296, 2019.

King, J., Howell, S., Derksen, C., Rutter, N., Toose, P., Beckers, J. F., Haas, C., Kurtz, N., and Richter-Menge, J.: Evaluation of Operation IceBridge quick-look snow depth estimates on sea ice, Geophysical Research Letters, 42, 9302–9310, 2015.

Koenig, L. S., Ivanoff, A., Alexander, P. M., MacGregor, J. A., Fettweis, X., Panzer, B., Forster, R. R., Das, I., McConnell, J. R., Tedesco, M., et al.: Annual Greenland accumulation rates (2009-2012) from airborne snow radar, The Cryosphere, 10, 2016.

Kurtz, N., Farrell, S., Studinger, M., Galin, N., Harbeck, J., Lindsay, R., Onana, V., Panzer, B., and Sonntag, J.: Sea ice thickness, freeboard, and snow depth products from Operation IceBridge airborne data, The Cryosphere, 7, 1035–1056, 2013.

Kurtz, N., Studinger, M., Harbeck, J., Onana, V., and Yi, D.: IceBridge L4 Sea Ice Freeboard, Snow Depth, and Thickness, Version 1, boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center, https://doi.org/10.5067/G519SHCKWQV6, 2015.

Kurtz, N. T. and Farrell, S. L.: Large-scale surveys of snow depth on Arctic sea ice from Operation IceBridge, Geophysical Research Letters, 38, 2011.

Kwok, R.: Satellite remote sensing of sea-ice thickness and kinematics: a review, Journal of Glaciology, 56, 1129–1140, 2010.

Kwok, R. and Cunningham, G.: ICESat over Arctic sea ice: Estimation of snow depth and ice thickness., Journal of Geophysical Research: Oceans, 113, 2008.

Kwok, R. and Maksym, T.: Snow depth of the Weddell and Bellingshausen sea ice covers from IceBridge surveys in 2010 and 2011: An examination, Journal of Geophysical Research: Oceans, 119, 4141–4167, 2014.

Kwok, R. and Markus, T.: Potential basin-scale estimates of Arctic snow depth with sea ice freeboards from CryoSat-2 and ICESat-2: An exploratory analysis, Advances in Space Research, 62, 1243–1250, 2018.

Kwok, R., Kurtz, N. T., Brucker, L., Ivanoff, A., Newman, T., Farrell, S. L., King, J., Howell, S., Webster, M. A., Paden, J., et al.: Intercomparison of snow depth retrievals over Arctic sea ice from radar data acquired by Operation IceBridge, The Cryosphere, 11, 2571–2593, 2017.

Kwok, R., Kacimi, S., Webster, M., Kurtz, N., and Petty, A.: Arctic Snow Depth and Sea Ice Thickness From ICESat-2 and CryoSat-2 Freeboards: A First Examination, Journal of Geophysical Research: Oceans, 125, e2019JC016 008, 2020.

Lawrence, I. R., Tsamados, M. C., Stroeve, J. C., Armitage, T. W., and Ridout, A. L.: Estimating snow depth over Arctic sea ice from calibrated dual-frequency radar freeboards, The Cryosphere, 12, 3551–3564, 2018.

Laxon, S., Peacock, N., and Smith, D.: High interannual variability of sea ice thickness in the Arctic region, Nature, 425, 947–950, 2003.

Liston, G. E., Itkin, P., Stroeve, J., Tschudi, M., Stewart, J. S., Pedersen, S. H., Reinking, A. K., and Elder, K.: A Lagrangian Snow-Evolution System for Sea-Ice Applications (SnowModel-LG): Part I–Model Description, Journal of Geophysical Research: Oceans, p. e2019JC015913, 2020.

Maaß, N., Kaleschke, L., Tian-Kunze, X., and Drusch, M.: Snow thickness retrieval over thick Arctic sea ice using SMOS satellite data, The Cryosphere, 7, 1971–1989, 2013.

Markus, T. and Cavalieri, D. J.: Snow depth distribution over sea ice in the Southern Ocean from satellite passive microwave data, Antarctic sea ice: physical processes, interactions and variability, 74, 19–39, 1998.

Markus, T., Stroeve, J. C., and Miller, J.: Recent changes in Arctic sea ice melt onset, freezeup, and melt season length, Journal of Geophysical Research: Oceans, 114, 2009.

Markus, T., Cavalieri, D. J., and Ivanoff, A.: Algorithm theoretical basis document: Sea ice products: Updated December 2011, cryospheric Sciences Laboratory, NASA Goddard Space Flight Center. Retrieved from: https://nsidc.org/sites/nsidc.org/files/files/amsr_atbd_seaice_dec2011.pdf, 2011.

Maslanik, J., Fowler, C., Stroeve, J., Drobot, S., Zwally, J., Yi, D., and Emery, W.: A younger, thinner Arctic ice cover: Increased potential for rapid, extensive sea-ice loss, Geophysical Research Letters, 34, 2007.

Maslanik, J., Stroeve, J., Fowler, C., and Emery, W.: Distribution and trends in Arctic sea ice age through spring 2011, Geophysical Research Letters, 38, 2011.

Meier, W., Fetterer, F., Savoie, M., Mallory, S., Duerr, R., and Stroeve, J.: NOAA/NSIDC climate data record of passive microwave sea ice concentration, version 2, National Snow and Ice Data Center, 10, N55M63M1, 2013.

Merkouriadi, I., Cheng, B., Graham, R. M., Rösel, A., and Granskog, M. A.: Critical role of snow on sea ice growth in the Atlantic sector of the Arctic Ocean, Geophysical Research Letters, 44, 10–479, 2017.

Mundy, C., Gosselin, M., Ehn, J., Gratton, Y., Rossnagel, A., Barber, D. G., Martin, J., Tremblay, J.-É., Palmer, M., Arrigo, K. R., et al.: Contribution of under-ice primary production to an ice-edge upwelling phytoplankton bloom in the Canadian Beaufort Sea, Geophysical Research Letters, 36, 2009.

Newman, T., Farrell, S. L., Richter-Menge, J., Connor, L. N., Kurtz, N. T., Elder, B. C., and McAdoo, D.: Assessment of radar-derived snow depth over Arctic sea ice, Journal of Geophysical Research: Oceans, 119, 8578–8602, 2014.

Nicolaus, M., Hoppmann, M., Arndt, S., Hendricks, S., Kaltein, C., König-Langlo, G., Nicolaus, A., Rossmann, L., Schiller, M., Schwegmann, S., et al.: Snow height and air temperature on sea ice from snow buoy measurements, Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Bremerhaven, PANGAEA, https://doi. org/10.1594/PANGAEA, 875638, 2017.

Paden, J., Li, J., Leuschen, C., Rodriguez-Morales, F., and Hale, R.: IceBridge Snow Radar L1B Geolocated Radar Echo Strength Profiles, Version 2, boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/FAZTWP500V70.[Date Accessed: 2020-02-07], updated 2019, 2014.

Pedersen, L., Saldo, R., Ivanova, N., Kern, S., Heygster, G., Tonboe, R., Huntemann, M., Ozsoy, B., Ardhuin, F., and Kaleschke, L.: Rasmus Reference dataset for sea ice concentration, Fileset, figshare, https://doi.org/10.6084/m9.figshare.6626549.v6, 2018.

Perovich, D., Richter-Menge, J., Elder, B., Arbetter, T., Claffey, K., and Polashenski, C.: Observing and understanding climate change: Monitoring the mass balance, motion, and thickness of Arctic sea ice, URL http://imb. crrel. usace. army. mil, URL http://imb. crrel. usace. army. mil, 2009.

Petty, A. A., Webster, M., Boisvert, L., and Markus, T.: The NASA Eulerian Snow on Sea Ice Model (NESOSIM) v1.0: initial model development and analysis., Geoscientific Model Development, 11, 2018.

Radionov, V. F., Bryazgin, N. N., and Alexandrov, E. I.: The Snow Cover of the Arctic Basin., Tech. rep., WASHINGTON UNIV SEATTLE APPLIED PHYSICS LAB, 1997.

Richter-Menge, J. A., Perovich, D. K., Elder, B. C., Claffey, K., Rigor, I., and Ortmeyer, M.: Ice mass-balance buoys: a tool for measuring and attributing changes in the thickness of the Arctic sea-ice cover, Annals of Glaciology, 44, 205–210, 2006.

Ricker, R., Hendricks, S., Helm, V., Skourup, H., and Davidson, M.: Sensitivity of CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation, The Cryosphere, 8, 1607–1622, 2014.

Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J., and Haas, C.: A weekly Arctic sea-ice thickness data record from merged CryoSat-2 and SMOS satellite data, Cryosphere, 11, 1607–1623, 2017.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., et al.: MERRA: NASA's modern-era retrospective analysis for research and applications, Journal of climate, 24, 3624–3648, 2011.

Rostosky, P., Spreen, G., Farrell, S. L., Frost, T., Heygster, G., and Melsheimer, C.: Snow depth retrieval on Arctic sea ice from passive microwave radiometers-Improvements and extensions to multiyear ice using lower frequencies, Journal of Geophysical Research: Oceans, 123, 7120–7138, 2018.

Schröder, D., Feltham, D. L., Tsamados, M., Ridout, A., and Tilling, R.: New insight from CryoSat-2 sea ice thickness for sea ice modelling, The Cryosphere, 13, 125–139, 2019.

Serreze, M. C., Barrett, A. P., and Stroeve, J.: Recent changes in tropospheric water vapor over the Arctic as assessed from radiosondes and atmospheric reanalyses, Journal of Geophysical Research: Atmospheres, 117, 2012.

Shalina, E. V. and Sandven, S.: Snow depth on Arctic sea ice from historical in situ data, The Cryosphere, 12, 1867–1886, 2018.

Stroeve, J. and Notz, D.: Changing state of Arctic sea ice across all seasons, Environmental Research Letters, 13, 103 001, 2018.

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

Stroeve, J., Liston, G. E., Buzzard, S., Zhou, L., Mallett, R., Barrett, A., Tschudi, M., Tsamados, M., Itkin, P., and Stewart, J.: A Lagrangian Snow-Evolution System for Sea Ice Applications (SnowModel-LG): Part II-Analyses, Journal of Geophysical Research: Oceans, p. e2019JC015900, 2020.

Stroeve, J. C., Markus, T., Maslanik, J. A., Cavalieri, D. J., Gasiewski, A. J., Heinrichs, J. F., Holmgren, J., Perovich, D. K., and Sturm, M.: Impact of surface roughness on AMSR-E sea ice products, IEEE Transactions on Geoscience and Remote Sensing, 44, 3103–3117, 2006.

Sturm, M. and Massom, R. A.: Snow in the sea ice system: Friend or foe, Wiley-Blackwell New York, NY, 2017.

Tilling, R., Ridout, A., Shepherd, A., Tsamados, M., and Buzzard, S.: A model of snow depth an Arctic winter sea ice., In preparation.

Tschudi, M., Meier, W. N., Stewart, J. S., Fowler, C., and Maslanik, J.: Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors, Version 4, boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/INAWUWO7QH7B. [Date Accessed: 2020-10-19], 2019.

Warren, S. G., Rigor, I. G., Untersteiner, N., Radionov, V. F., Bryazgin, N. N., Aleksandrov, Y. I., and Colony, R.: Snow depth on Arctic sea ice, Journal of Climate, 12, 1814–1829, 1999.

Webster, M. A., Rigor, I. G., Nghiem, S. V., Kurtz, N. T., Farrell, S. L., Perovich, D. K., and Sturm, M.: Interdecadal changes in snow depth on Arctic sea ice, Journal of Geophysical Research: Oceans, 119, 5395–5406, 2014.

Webster, M. A., Parker, C., Boisvert, L., and Kwok, R.: The role of cyclone activity in snow accumulation on Arctic sea ice, Nature communications, 10, 1–12, 2019.

Table 1. Summary of investigated snow reconstruction products.

Product	Time span	Temporal resolution	Spatial resolution	Projection type	Method type	Reference
SnowModel-LG	1980-2018	All year (daily)	$25 \times 25km$	EASE grid	Reanalysis-based	Liston et al. (2020) Stroeve et al. (2020)
NESOSIM	2000-2017	Aug to Apr (daily)	$100 \times 100km$	Polar stereographic grid	Reanalysis-based	Petty et al. (2018)
CPOM	1991-2017	All year (daily)	$10 \times 10km$	Polar stereographic grid	Reanalysis-based	Tilling et al. (In preparation)
UW	1980-2015	Apr (monthly)	$75 \times 75km$	Polar stereographic grid	Reanalysis-based	Blanchard-Wrigglesworth et al. (2018)
DuST	2003-2008, 2013-2018	Bi-monthly (2013-2018) & monthly (2003-2008)	1.5° longitude $\times\ 0.5^\circ$ latitude	Up to 81.5° N	Active satellite-based	Lawrence et al. (2018)
DESS	2011-2019	Mar (monthly)	$12.5 \times 12.5km$	Polar stereographic grid (up to 87.5° N)	Active & passive satellite-based	Xu et al. (In preparation)
PMW Bremen	2003-2018	Mar & Apr (Daily)	$25 \times 25km$	Polar stereographic grid	Passive satellite-based	Rostosky et al. (2018)
PMW DMI	2013-2018	Jan to Apr (Daily)	$25 \times 25km$	EASE grid 2.0	Passive satellite-based	Winstrup et al. (2019)

Table 2. R^2(in bold), $RMSE$ (left in bracket, units: cm) and $NRMSE$ (right in bracket) of various average monthly snow depth products in comparison with four OIB snow depth products, using $100 \times 100km$ monthly comparison.

OIB Product	SnowModel-LG	NESOSIM	CPOM	UW	DuST	DESS	PMW Bremen	PMW DMI
quicklook	**0.27**	**0.29**	**0.54**	**0.00**	**0.36**	**0.26**	**0.59**	**0.54**
	(9.5,0.23)	(10.0,0.22)	(6.2,0.15)	(3.1,0.15)	(4.6,0.14)	(14.0,0.41)	(4.4,0.11)	(4.4,0.10)
GSFC	**0.30**	**0.39**	**0.43**	**0.10**	**0.30**	**0.48**	**0.56**	**0.37**
	(9.5,0.23)	(9.0,0.21)	(7.4,0.19)	(3.6,0.13)	(5.2,0.14)	(12.4,0.34)	(4.8,0.11)	(5.4,0.13)
JPL	**0.41**	**0.38**	**0.59**	**0.17**	**0.35**	**0.51**	**0.70**	**0.52**
	(8.7,0.15)	(9.0,0.15)	(6.2,0.13)	(3.5,0.10)	(5.0,0.13)	(12.2,0.35)	(4.0,0.09)	(4.8,0.08)
SRLD	**0.47**	**0.38**	**0.61**	**0.31**	**0.21**	**0.59**	**0.63**	**0.43**
	(8.4,0.10)	(9.1,0.10)	(6.1,0.10)	(3.2,0.06)	(5.6,0.11)	(11.1,0.23)	(4.4,0.08)	(5.2,0.06)

Winstrup, M., Tonboe, R., Lavergne, T., Rasmussen, T., Saldo, R., Tietsche, S., and Pedersen, L. T.: Retrieval of spring-time snow thicknesses on Arctic sea ice from AMSR-2 microwave radiometer data., eSA Living Planet Symposium 2019, May 13-17, Milan, 2019.

Xu, S., Zhou, L., Liu, J., Lu, H., and Wang, B.: Data Synergy between Altimetry and L-Band Passive Microwave Remote Sensing for the Retrieval of Sea Ice Parameters?A Theoretical Study of Methodology, Remote Sensing, 9, 1079, 2017.

Xu, S., Zhou, L., and Wang, B.: Variability scaling and consistency in airborne and satellite altimetry measurements of Arctic sea ice., Cryosphere, 14, 2020.

Xu, S., Zhou, L., Zhu, W., Wang, B., and Liu, J.: Arctic Wintertime Sea Ice Thickness and Snow Depth with physical synergy of CryoSat-2 and SMOS., In preparation.

Zhou, L., Xu, S., Liu, J., Lu, H., and Wang, B.: Improving L-band radiation model and representation of small-scale variability to simulate brightness temperature of sea ice, International Journal of Remote Sensing, 38, 7070–7084, 2017.

Zygmuntowska, M., Rampal, P., Ivanova, N., and Smedsrud, L. H.: Uncertainties in Arctic sea ice thickness and volume: new estimates and implications for trends, The Cryosphere, 8, 705–720, 2014.

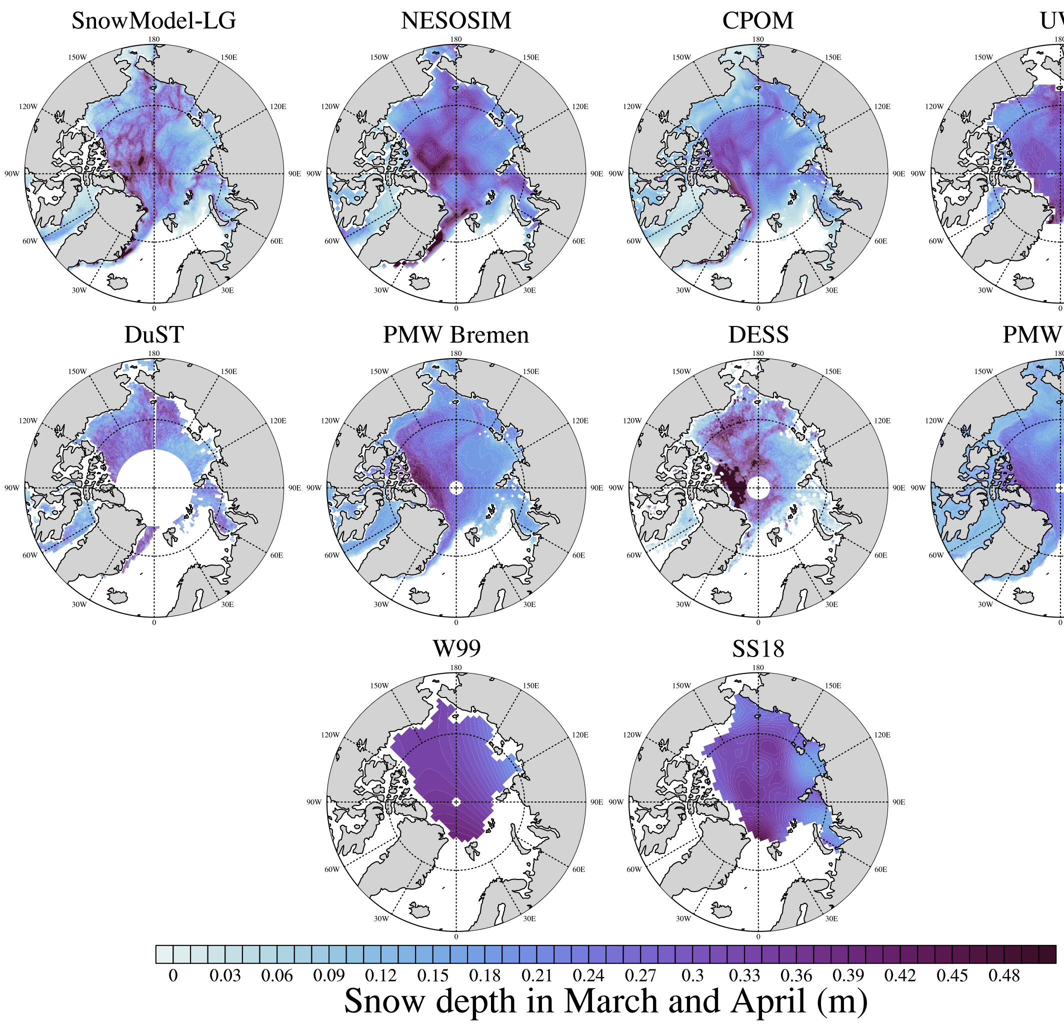


Figure 1. Mean snow depth (units: m) depth in spring (March-April) 2014 for eight snow products, $W99$ and $SS18$.

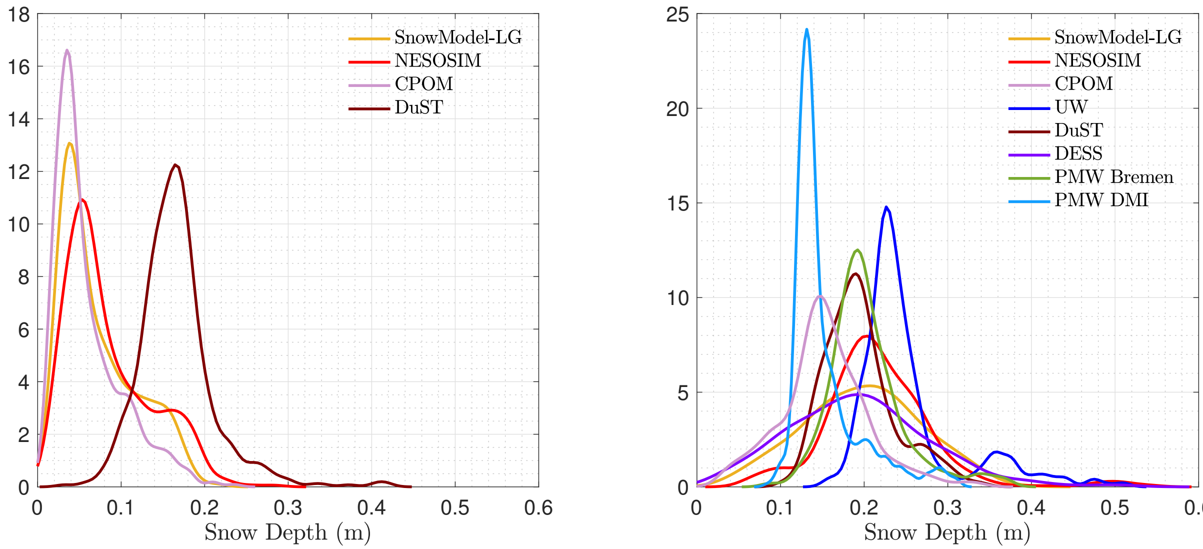


Figure 2. Snow distribution comparison within the common regions (in Figure S2(a)) in all snow products during the period 2000-2018 (different products cover different periods) in October-November (left: a) and March-April (right: b).

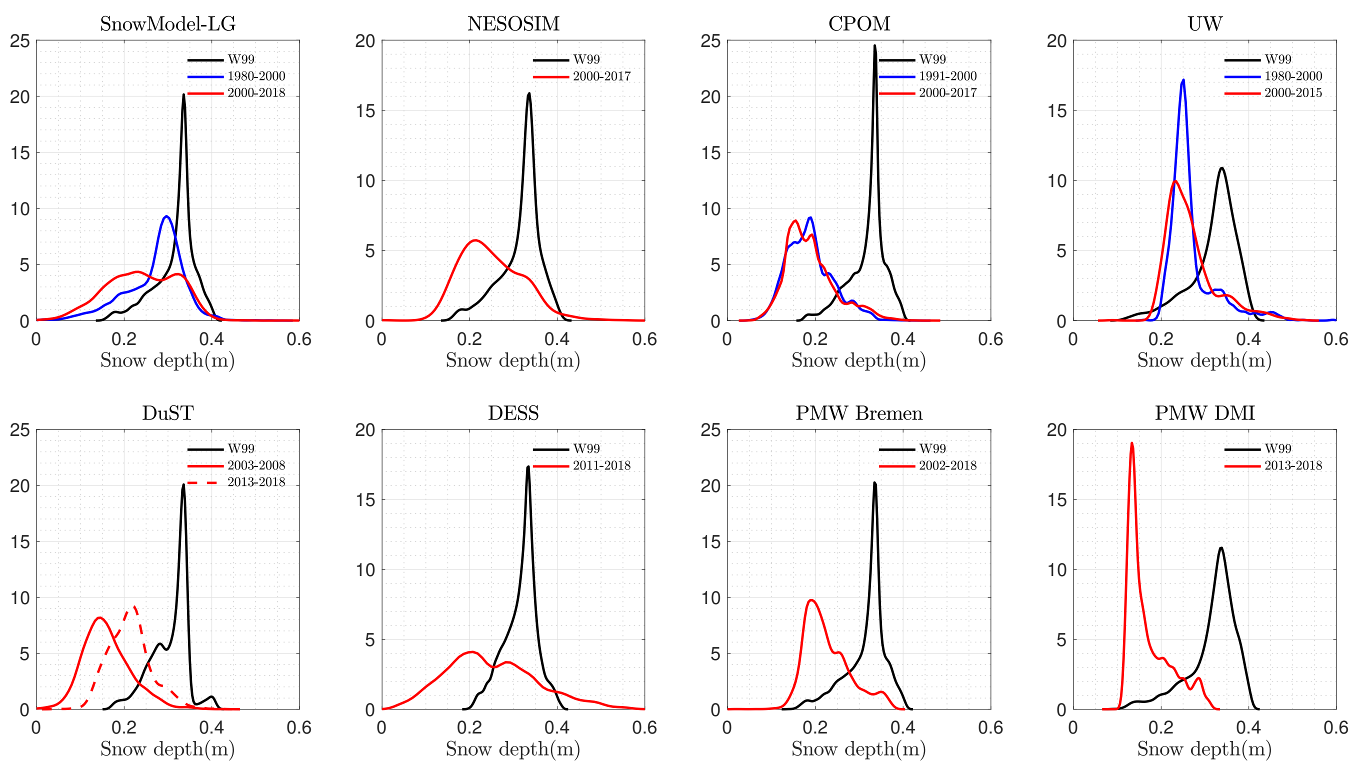


Figure 3. Snow distribution comparison between $W99$ and snow products in spring (March/April).

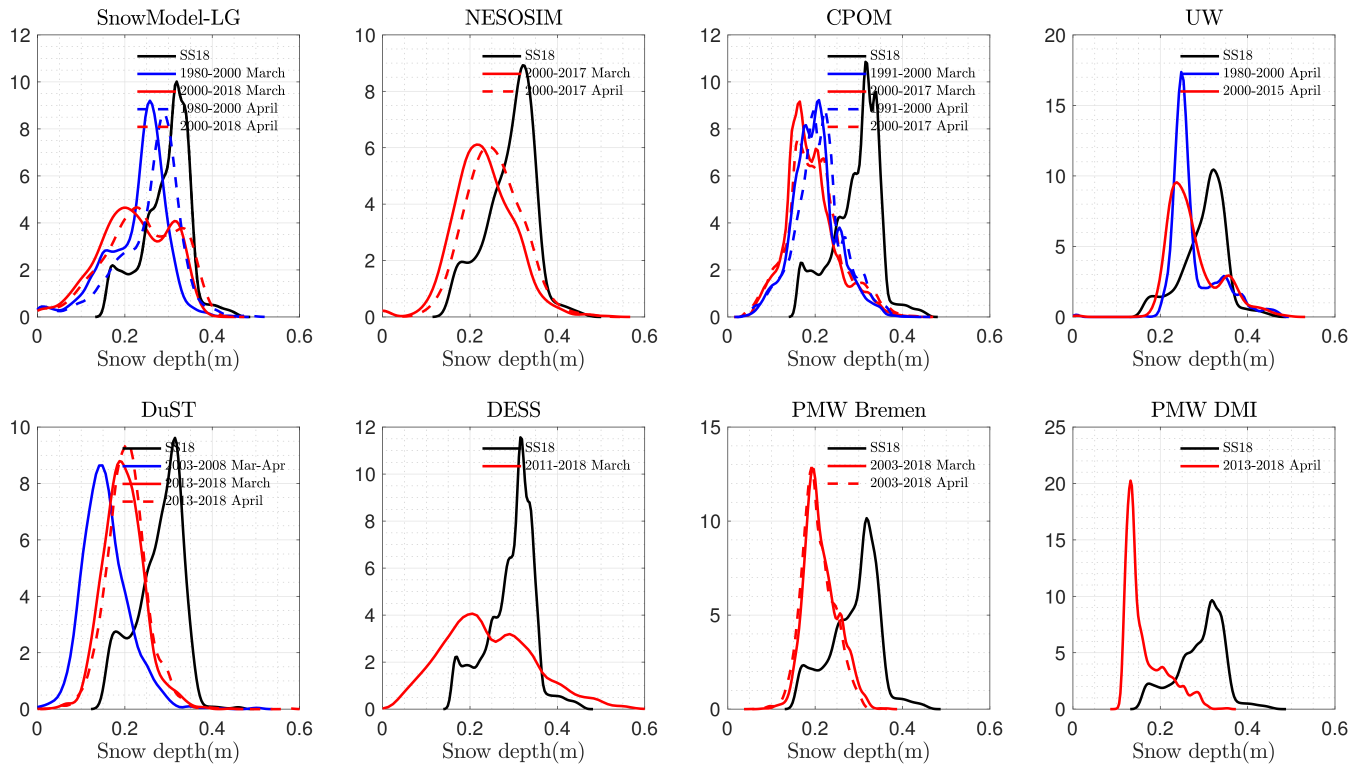


Figure 4. Snow distribution comparison between $SS18$ climatology and snow products.

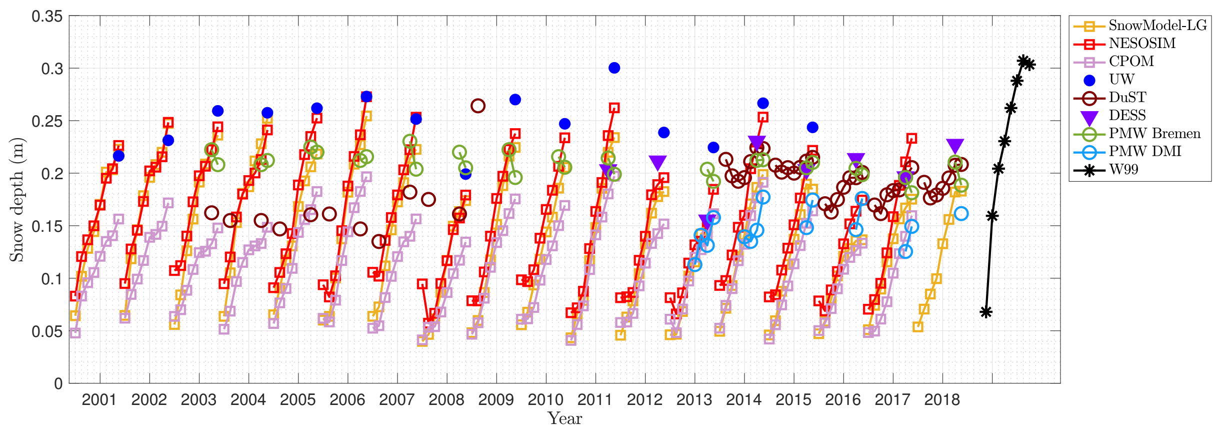


Figure 5. Time series of average monthly snow depth in each snow product within the common regions since 2000. Only winter time (September to next April) is shown.

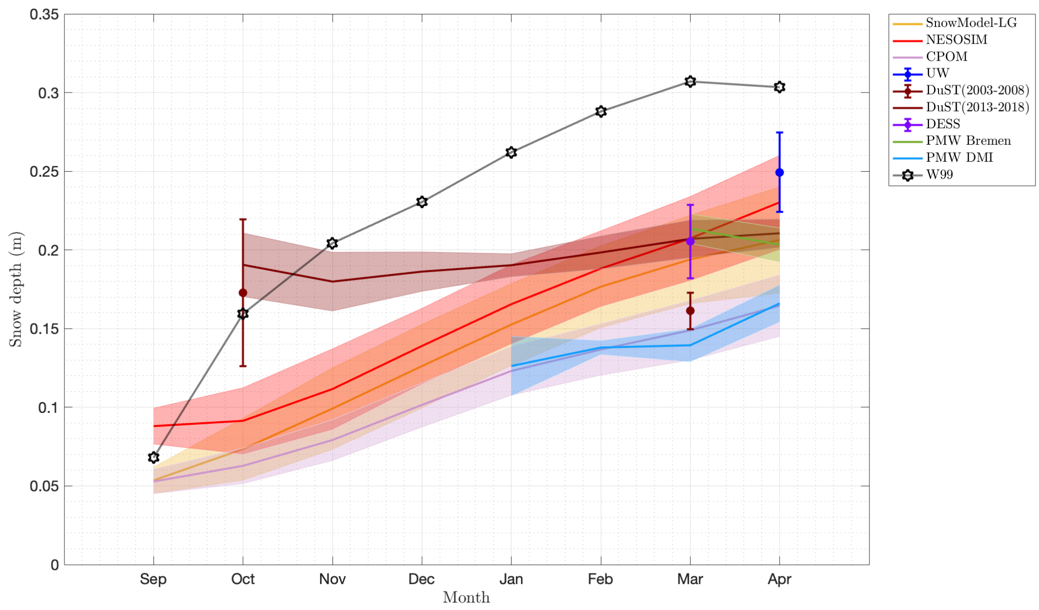


Figure 6. Wintertime snow accumulation among eight products and $W99$ since 2000.

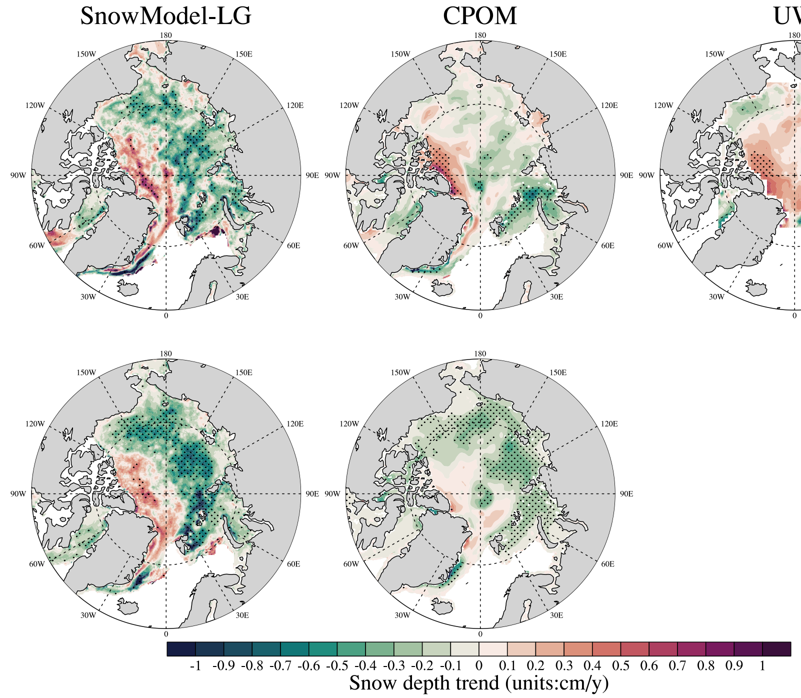


Figure 7. Trend of snow depth (Units: $cm/year$) for SnowModel-LG, CPOM and UW in spring (April) (first row) and autumn (November) (second row) during the period from 1991 to 2015. Areas with significant trends are shown as dotted areas (confidence level 95%).

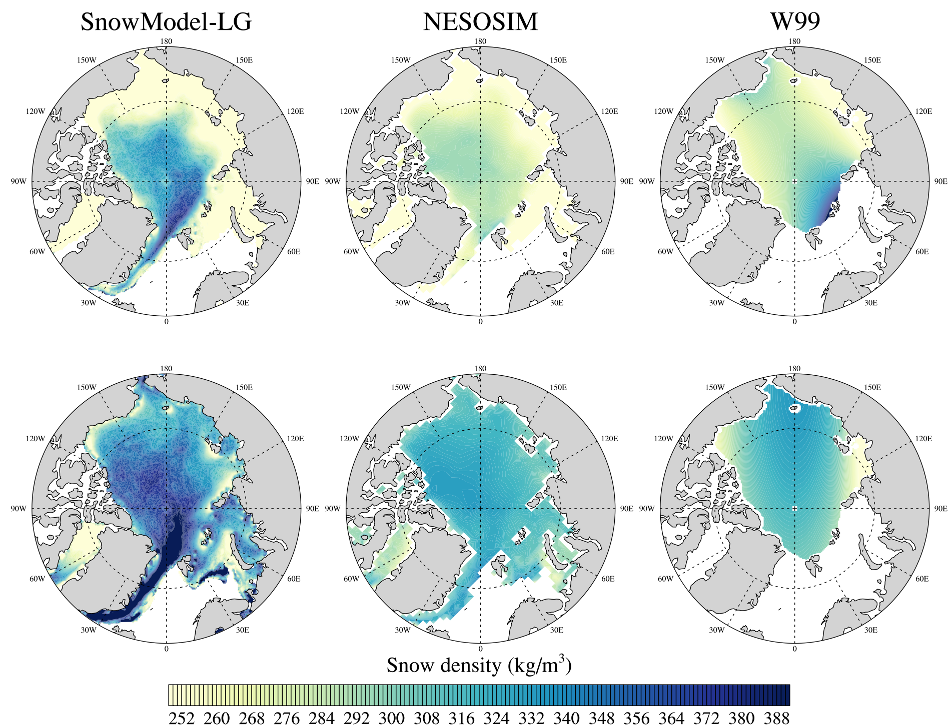


Figure 8. Mean snow density (Units: kg/m^3) according to SnowModel-LG, NESOSIM and $W99$ in November (first row) and the next April (second row) since 2000.

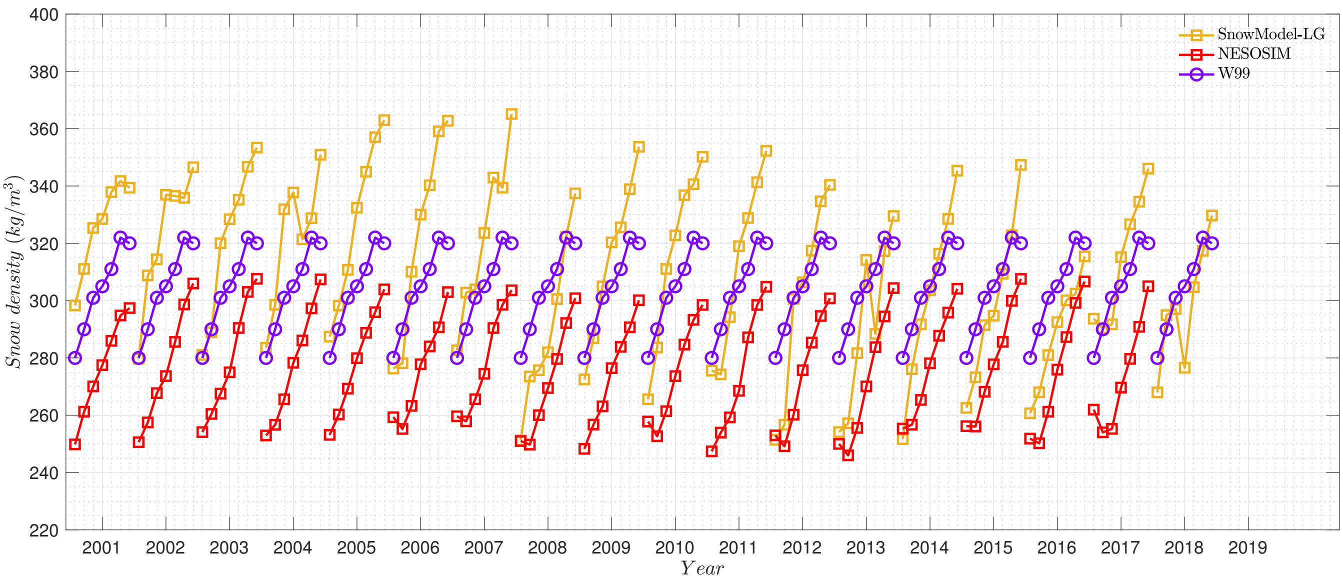


Figure 9. Time series of mean snow density (Units: kg/m^3) within Arctic basin (common region in Figure S1) comparison in SnowModel-LG, NESOSIM and $W99$ since 2000.

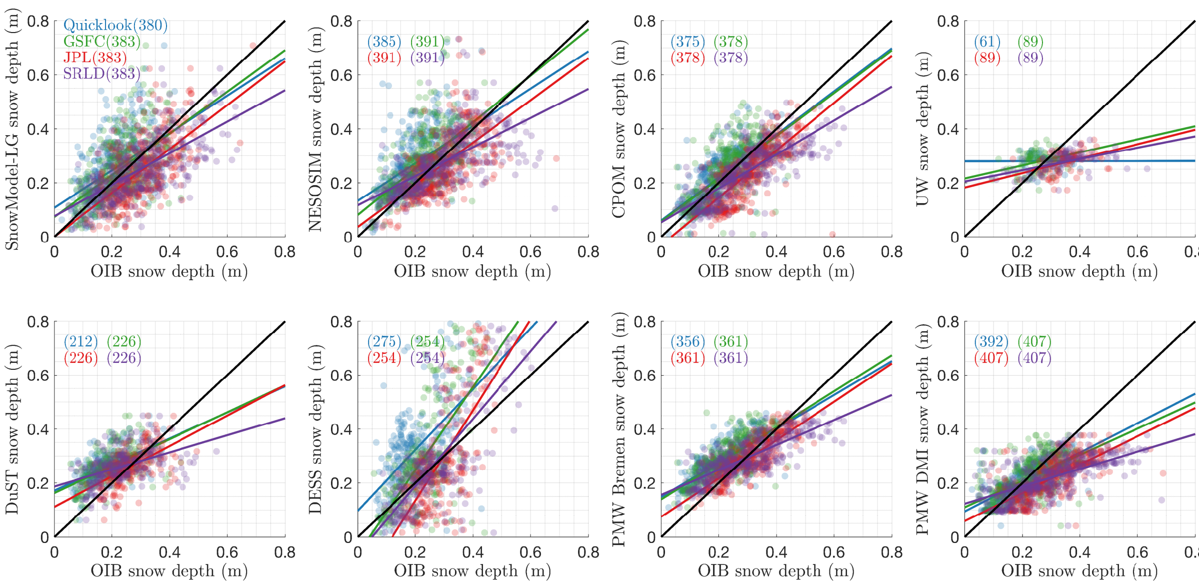


Figure 10. March and April comparison (only April in UW and March in DESS) between four monthly OIB products (quicklook in blue, GSFC in green, JPL in red and SRLD in purple) and all snow products in monthly onto $100km$ grid in the period of 2014 and 2015. Lines are best linear fits. Numbers in left corners indicate the valid count in the linear fitting.

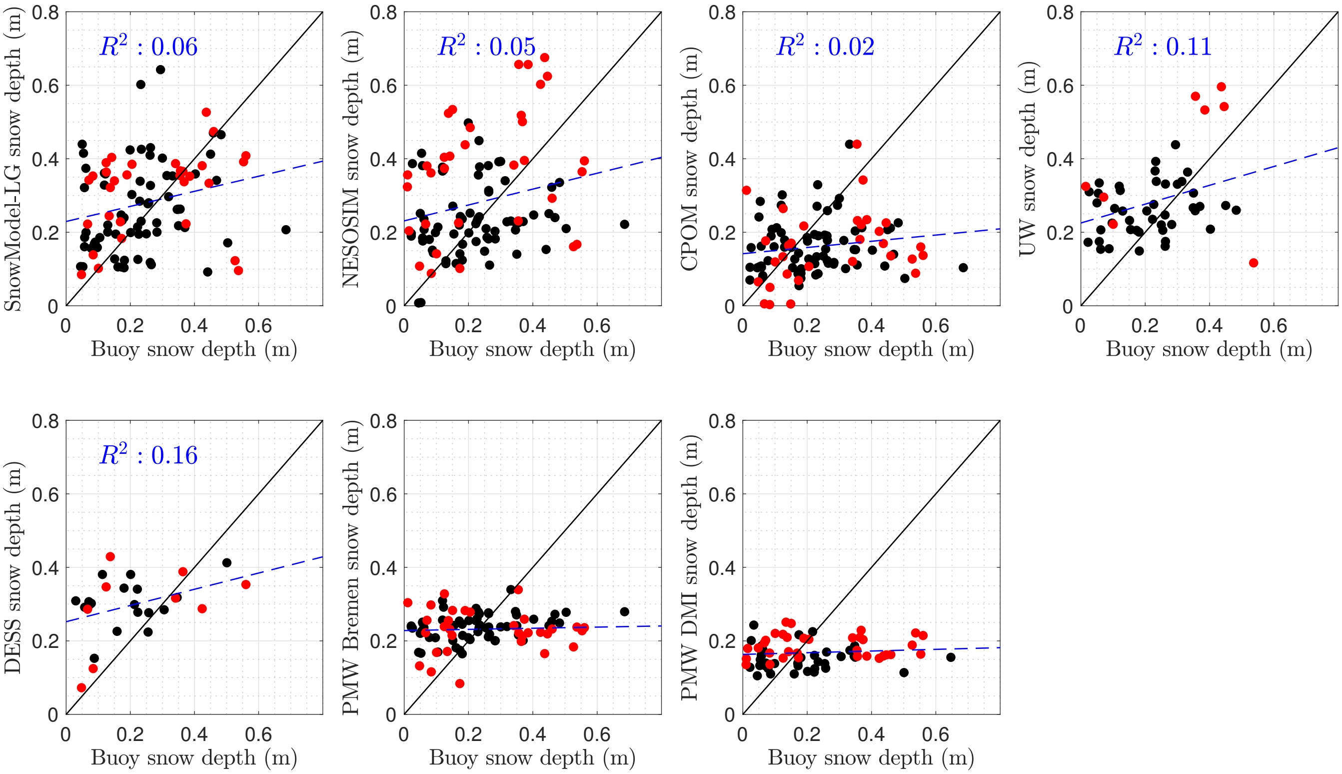


Figure 11. Comparison of average March/April snow depth in snow products in native spatial resolution versus the AWI (red dots)/IMBs (black dots) buoys. All significant correlation (R^2) are given in blue.

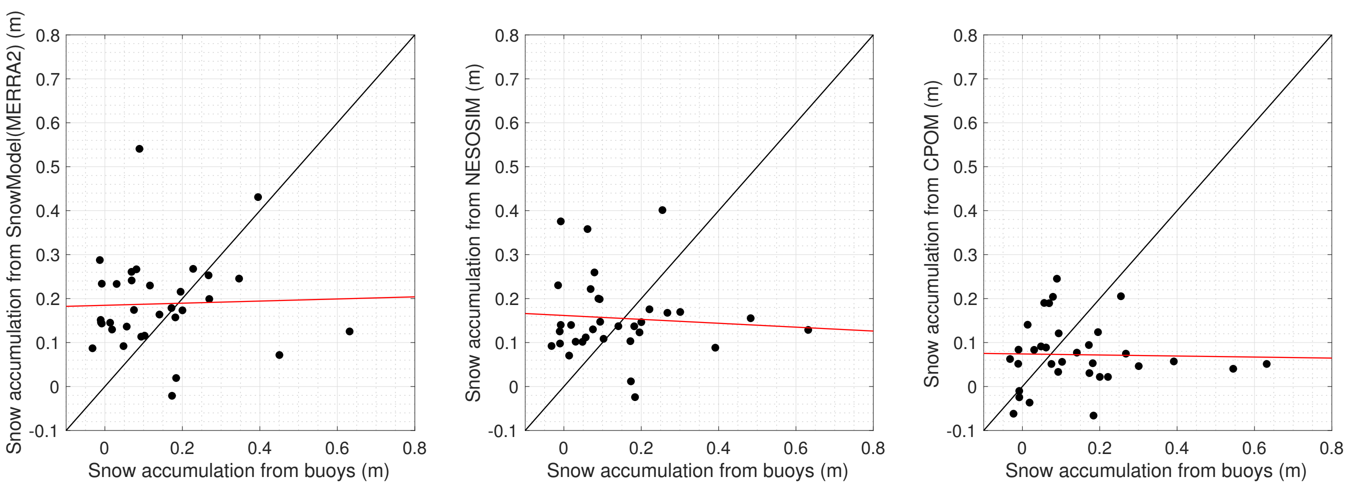
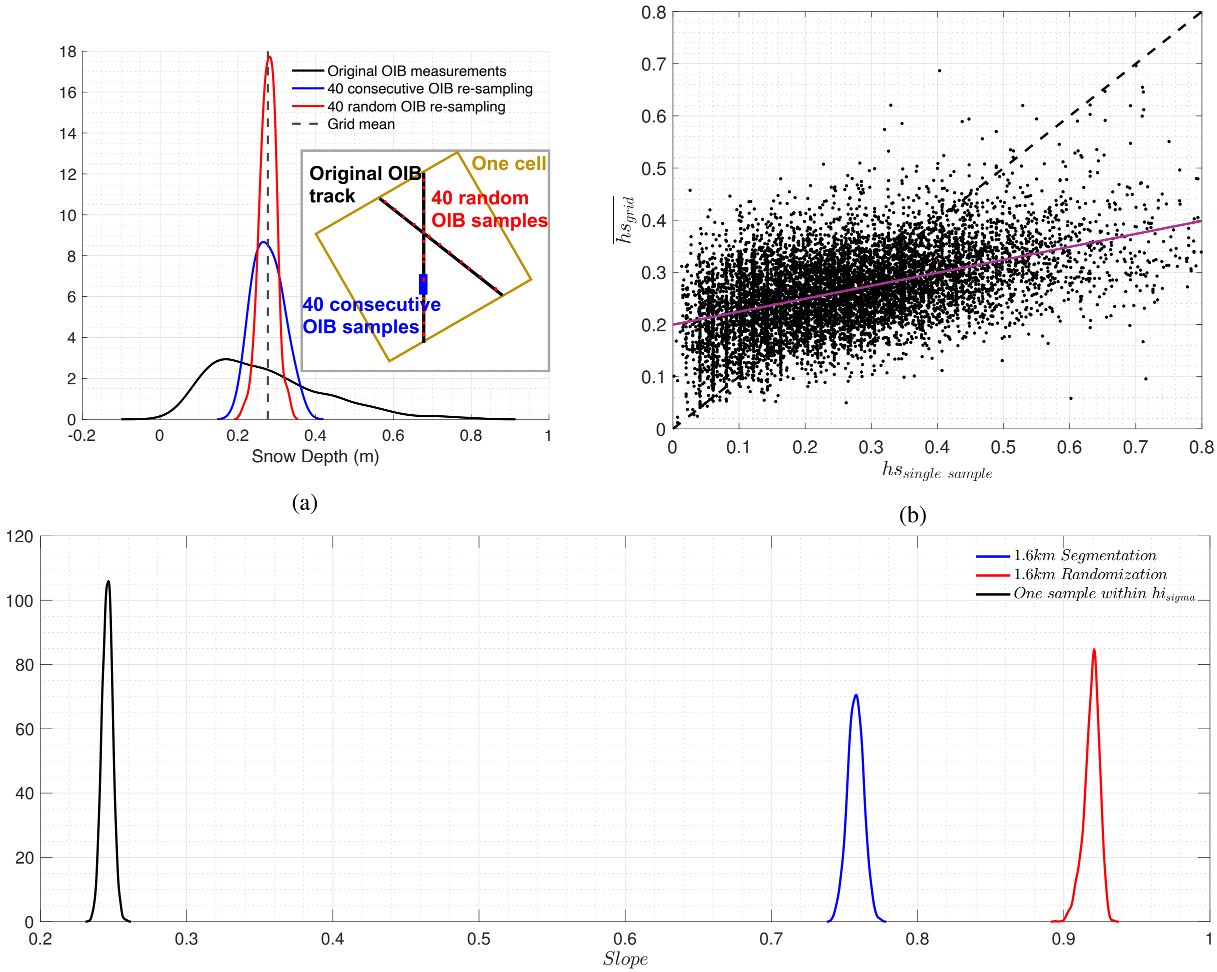


Figure 12. Comparison of total winter snow accumulation (starts from early winter to the next year) between buoys and three snow products based on daily snow products.



(c)

Figure 13. The study with OIB on the effects of sampling strategy for observations. Snow distribution within typical region (a) under different sampling method embedded with schematic on different strategies deployment: original OIB measurements (black), 40 consecutive OIB re-sampling (strategy II: blue) and 40 random OIB re-sampling (strategy I: red). Mean snow depth are in dashed lines. Typical fitting (b) between cell mean and a single sample snow depth for all regions, with fitting line in purple. Overall distribution of slopes (c) in three re-sampling strategies, fitting between OIB cell mean and 40 random OIB samples (strategy I), between OIB cell mean and 40 consecutive samples (strategy II) and between OIB cell mean and single sample within 1 sigma of the mode of the ice thickness (strategy III: black).