We would like to thank Dr. Kern for his thoughtful and constructive review; we believe that his feedback will undoubtedly improve the quality of our manuscript.

We have appended an ammended manuscript to this document which illustrates the changes we have made to our submission in response to the feedback. Below we quote Dr. Kern's comments in blue, and our responses follow in red.

We would first like to first bring to the reviewer's attention a mistake made in the original manuscript. Due to a programming error we inadvertantly used radar freeboard data from a different product (that of Landy et al., 2020) in the winter of 2017/18. Because this product generally exhibits higher radar freeboard values than those used in the rest of the study due to a different retracking algorithm, we misidentified this winter at one point as 'a trend bucking year' for radar freeboards. We have now fixed this error and updated our statistics. This has had the following results:

- Regional declines in radar freeboard and resulting sea ice thickness are generally smoother.
- Negative trends in several regions are slightly increased.
- Negative trends are therefore more frequently statistically significant at the 5% level.
- Trends when calculated with SnowModel-LG in the 2002-2018 period are now in better agreement with those calculated from NESOSIM in the 2002-2015 period.

Despite these changes, the central thesis of our paper remains unchanged: the use of a snow product with regional variability and trends propagates into variability and trends in regional sea ice thickness.

Summary: This very interesting paper illustrates the potential improvement in the credibility of trends in and inter-annual / intra-basin variations of Arctic sea-ice thickness estimates from satellite radar altimetry. This is achieved by a comprehensive intercomparison of the contribution of snow on sea ice on the retrieval of sea-ice thickness from radar freeboard when using the Warren et al. (1999) snow climatology on the one hand and a physical model for snow properties driven by atmospheric reanalyses' precipitation and other relevant meteorological parameters on the other hand. As expected, the inter-annual variability of the snow contributions based on the model data is considerably larger than the one based on the Warren et al. (1999) data. The paper further convincingly demonstrates that the more realistic inter-annual variability and spatio-temporal development of the Arctic Ocean sea-ice thickness.

The paper is generally well written and will have considerable impact on the scientific community. It would benefit from some re-organization (see GC1). It is furthermore quite light when it comes to descriptions of data and methodologies used (GC2). Currently, one would not be able to re-produce the work done. The inclusion of Kara and Barents Sea I find quite a hypothetical move based on the data availability and suggest to consider removing those from the analysis (GC3). Finally, there is a number of open points to discuss when it comes to the illustration and interpretation of the results presented. In the following you will find my list of general comments (GC), specific comments and some suggestions to mitigate typos and editoral issues - all for the main manuscript - followed by a short list of things I found worth to consider in the supplementary material.

Title: While your main conclusion supports the title in general, it is in some way misleading. The main focus of the paper is on the illustration that a snow depth climatology is not well suited to compute credible trends in sea-ice thickness estimates derived from

satellite altimetry with such a snow depth as input for the freeboard to thickness conversion. In your paper, this is illustrated by usage of data from a numerical model which has experienced limited validation. Hence, albeit the improvement using these model data is obvious it is not necessarily the truth either. Hence, instead of formulating the title as a fact I suggest to include points of the above-stated.

In response to this feedback we would append a clarifying clause, so that the title reads:

'Faster decline and higher variability in the sea ice thickness of the marginal Arctic seas *when accounting for dynamic snow cover*'

GC1: I strongly recommend to re-organize the paper. Most of the explanations / motivations given in the subsections 1.1 and 1.2 are tied relatively close to Section 3 and should be combined with that section. In addition, subsections 1.1 and 1.2 refer to data and regions denoted in Section 2. Hence: Remove 1.1 and 1.2 and put it into Section 3. Let Section 2 start right behind the "true" introduction. That way the data sets used in 1.1 and 1.2 would be introduced adequately beforehand which eases reading and which reduces the number of open questions.

We have rearranged these sections accordingly. Section 1.1 (on our method of separating the impacts of snow and radar freeboard data on thickness determination) has been moved to the methods section (Section 3). We agree with the reviewer that our illustration of the limitations of W99 in Sects 1.2.1 & 1.2.2 would have been better placed after the data description (Sect 2). Rather than put this in the Methods section (Sect 3), we have moved it to the beginning of our Results section (Sect 4). We hope this is satisfactory to the reviewer, but if not we will of course reconsider his original suggestion of Sect 3.

GC2: Both, the description of the data used as well as of the methodologies used lack some clarity and/or do not contain all information required. One good example: The ESA-CCI radar freeboard data set used comprises data of two different satellites with some overlap. It is not clear from the description in the data how long the Envisat and how long the Cryosat-2 part of the data used is - plus a motivation of the choice made - plus a discussion about the biases between the radar freeboards of these two satellites, which have a different sign based upon the region. Some of the descriptions also appear to contain errors which ask for re-phrasing.

We agree that not specifying the transition point from Envisat to CryoSat-2 radar freeboard data was an oversight. We would add the following clarifying information into the "Radar Freeboard Data" subsection:

CS2 carries a delay-Doppler altimeter that significantly enhances along-track resolution by creating a synthetic aperture. For this reason as well as its higher latitudinal limit, we used CS2 radar freeboard measurements over Envisat's during the period when the missions overlapped (November 2010 - March 2012).

We have also made changes to our data description and discussion sections with regard to the potential effect of biases between the radar freeboard measurements of the two satellites. To "Data Description" (Sect 2.2) we have added:

To create a radar freeboard product that is consistent between the Envisat and CS2 missions, Envisat returns are retracked using a variable threshold retracking algorithm. This variable threshold is calculated from the strength of the surface backscatter and the width of the leading edge of the return waveform such that the inter-mission bias is minimised (Paul et al., 2018). The results are comprehensively analysed in the Product

Validation & Intercomparison Report (ESA, 2018). One key finding of this report is that while Envisat radar freeboards are calculated so as to match CS2 freeboards during the period of overlap over the whole Arctic basin, there are biases over ice types. In particular, Envisat ice freeboards (not radar freeboards) are biased 2-3 cm low (relative to CS2) in areas dominated by MYI, and 2-3 cm high in areas dominated by FYI. We discuss the implications of these biases in Sect. (5.3).

To our Discussion section, we have added a subsection "Inter-Mission Bias between Envisat and CryoSat-2" (5.3). This reads:

An extensive validation exercise for the merged products indicated that although Envisat radar freeboards match well with CS2 freeboards in the Arctic overall, some biases do exist over specific ice types (ESA, 2018). In particular, analysis of the inter-mission overlap period indicates that Envisat freeboards were biased low (relative to CS2) in areas dominated by MYI, and high in areas dominated by FYI.

We first make the point that this will have a relatively minimal effect on our findings regarding interannual variability, as  $\overline{\text{Snow}}$  is unaffected by this and  $\sigma^2_{RF}$  is likely relatively independent of the absolute magnitude of  $\overline{\text{RF}}$ .

With regard to trends, if Envisat radar freeboards (and thus  $\overline{\text{RF}}$ ) are in fact biased high over FYI between 2002-2010 (relative to CS2), then the total trend in many regions dominated by FYI could potentially be smaller than calculated in this manuscript.

We do however add that our findings regarding the impact of declining Snow is unaffected by any inter-mission bias in RF. Because the trend in SIT is determined by both Snow & RF, the trend in SIT will always be more negative when calculated with downward trending data for Snow.

GC3: The overall credibility of the paper would benefit from a more critical consideration of the application area of the Warren et al. (1999) climatology. Sampling density, number of observations, as well as the distribution of the snow depth observations over time combined with the usage of a polynomial fit limits the usefulness of these observations in the regions Kara Sea and Barents Sea. One good solution would be to omit these regions.

Following the suggestion of the reviewer we analysed the original positional data (found within the meteorological observations) from drifting stations NP3 – NP31. This was all the data available to us, and was supplied previously in a personal communication by the NSIDC.

After plotting the tracks of these 27 drifting stations, we counted the number of stations that visited each region in each month:





We note in the above figure that repeat visits by stations on the same or consecutive years do not add to the tally. For example, NP22 lasted for four years and visited the same regions in the same months on consecutive years. However it then became apparent that some stations were not making snow measurements during some regional 'visits', and this should be included in the consideration of the sampling density as suggested by the reviewer. Instead we identified all distinct dates on which snow stake data was gathered by each NP station. We then cross-referenced this with the positional data found within the met data, to break down the number of distinct stake-measurement-days in each region by month. We believe this is a suitable metric for the spatial sampling of the drifting stations.



We have now included this figure and a description of the analysis within the W99 component of the Data Description section.

With regard to omitting undersampled marginal seas such as the Kara and Barents Seas from our analysis, we note that poor drifting station coverage has not stopped sevaral authors from using (m)W99 to derive sea ice thickness in these marginal seas (e.g. Sallila et al., 2019; Li et al., 2020; Li et al., 2020b; Belter et al., 2020) particularly in the pursuit of estimating sea ice volume (Tilling et al., 2015; Tilling et al., 2018; Laxon et al., 2013). We therefore would like to consider these regions in this manuscript, but with the clear caveat (which is now made explicit to the reader) that mW99 is likely not representative of the snow conditions. The subsection that we have added reads:

## 2.3.1 Drifting Station Coverage Illustration

At this point it is instructive to briefly illustrate the coverage of the drifting stations from which W99 was compiled. We analysed position and snow depth data from the twenty-eight drifting stations that contributed to W99 (Fig. 2a). It is clear that the vast majority of these operated in the Central Arctic or in the East Siberian Sea, with very little sampling done in most other marginal seas. But while these tracks illustrate the movements of the drifting stations, it is important to note that the stations were not always collecting snow data which would contribute to the W99 climatology. To assess the spatial distribution of snow sampling, we cross-referenced the position data with days on which the drifting stations recorded the snow depth at their measuring stakes. We then calculated the number of `measurement-days' in each region-month combination (Fig. 2b). We note that when two

drifting stations were operating at the same day, we count this as two distinct days (as they were rarely so close together so as to collect redundant data).

This reveals that no snow measurements were taken in the Barents and Kara Seas, and none in the Laptev Sea for four of the seven winter months. While `snow-line' transect data also contributed to W99 (and indeed was used in preference to stake data where possible), we find that snow-line data was overwhelmingly collected on days where stake-data was also collected.

Figure 2 illustrates that the quadratic fits of W99 are not appropriate for use in several of the marginals seas. However we note that a number of authors have still used the climatology for sea ice thickness retrievals in these regions, often in the course of sea ice volume calculations (e.g. Sallila et al., 2019; Li et al., 2020; Li et al., 2020b; Belter et al., 2020; Tilling et al., 2015; Tilling et al., 2018; Laxon et al., 2003; Laxon et al., 2013). We therefore consider these regions in this manuscript, but with the understanding that mW99 is likely not representative of the snow conditions.

## Specific comments

Line 18: " ... it determines whether floes ridge or raft ... "  $\rightarrow$  My take on this would be that this happens at rather small ice thicknesses, i.e. around 20-30 cm. I am therefore not so sure whether this is such an important physical role of the sea-ice thickness.,

We have removed this line.

Line 21: "... with thin ice favoring melt pond formation ..." Why is that? Because there is little snow, which melts away more quickly than on thick is with a thicker snow cover? Otherwise I don't see a pressing reason why melt pond formation, which is basically driven by downward short- and longwave radiation should occur more easily on thin than on thick sea ice.

Thin ice favors melt pond formation because it is generally more level. Level ice favors melt pond formation because a given amount of water covers a greater surface area. We have removed this line.

Line 76: "below the waterline" –> Why this part of the total SIT is from below the waterline? I find this addition confusing because it implies that the sea ice is thicker when it is snow covered - which is not necessarily the case. Equation (2) refers to a SIT which is similar in both cases, bare or with snow cover. It could be 1 m, 2 m, whatsoever, with or without snow. The quantities that change are the sea-ice freeboard and the radar freeboard. It seems you want to express that in case of a snow cover the part of the sea ice that is below the waterline is larger than in case of bare ice. Usually this part below the waterline is called draft. It might hence make sense to re-phrase this sentence a bit to avoid confusion.

While the presence of snow on sea ice doesn't causally make the ice thicker, when we try to estimate the thickness of ice *with a known radar/ice freeboard*, we assume that it is thicker if it has thicker snow on it. That's because to support the weight of the snow while maintaining radar/ice freeboard, it must be more buoyant and therefore thicker.

For instance an MYI floe with a known radar freeboard of 5cm and no snow cover is likely around 36cm thick. But if it has 6kg/m<sup>2</sup> of snow on it (~2cm @ 300kg/m<sup>3</sup>), we estimate it to be around 44cm thick (Eq. 2). If it has 18kg/m<sup>2</sup> of snow (~6cm), we estimate a 60cm thickness. It is in this sense that six centimeters of snow can 'add' twenty-four centimeters of sea ice thickness to our estimation.

We can assess exactly how much 'extra' sea ice thickness is inferred by a snow thickness by calculating  $d(SIT)/d(m_s) = (\rho_w/\rho_w-\rho_i) * 1.81 * 10^{-3} = 1.3$ cm per kg/m<sup>2</sup> of additional snow cover (at MYI ice density).

As for the reviewer's point about whether this 'extra' thickness is added 'below the waterline', he is right that this is confusing. This line was originally included with the following schematic in mind:



But there is no reason that the 'extra' thickness inferred should be conceptualised as 'appearing' at the bottom, as illustrated. We therefore clarify by removing the phrase 'below the waterline'.

To further reduce confusion surrounding this issue, we have clarified throughout the paper that we are considering the relative contributions of snow and RF data to the thickness *determination*. That is, their contribution is to the calculation, as opposed to their literal contribution in space.

Line 83: "assumes total radar penetration of overlying snow" -> How about penetration of the Ku-Band signal into the sea ice? Given the different near-surface sea-ice salinity and densities of MYI compared to FYI one might need to also make a comment on this issue?

This is an interesting point. In-situ studies by Willatt et al. (2010, 2011) suggest that a very low amount of Ku-band radar energy returns from below the sea ice surface. Further in-situ measurements from the MOSAiC expedition published by Stroeve et al. (2020) are in agreement with this. The physical basis for this is the strong dielectric contrast between the snow and the sea ice.

Regarding the differences between MYI and FYI, it likely that an FYI-snow interface will have a higher normalised radar cross-section (and would therefore be penetrated less). This is because (as alluded to by the reviewer) the density, salinity, and therefore permitivity of the ice is higher relative to that of snow or air. We do however note that due to the process of upward brine migration into the snow from the FYI surface, an 'impedance matching' effect can be produced (Perovich et al., 1998) which lowers the dielectric contrast at the ice-snow interface. However the strength and prevalence of this effect are not well understood.

Lines 109/110: "as quadratic fits ... without corresponding fits of density." -> I don't agree. Warren et al. (1999), page 8, writes about 2-dimensional density fits. Yes, only a May map is shown but maps are derived for all months. And it is the SWE which is computed, not the other way round. Please rewrite this paragraph accordingly.

The reviewer is right to highlight that a quadratic fit for May snow density is shown in Fig. 10 of Warren et al. (1999), and it was density that was measured in-situ, not SWE (as shown in Fig. 2 of Warren et al., 1999). We would like to state that we have conducted an extensive search for the original density fits (one of which is displayed in Fig. 10 of Warren et al.), but have not been able to find them. This search included contacting the first two authors of Warren et al. 1999 and going through the Fetterer and Radionov (2000) data hosted publicly by the NSIDC (which only contain fits for SWE and depth). We also received additional, raw data on snow line and stake data from the NSIDC in a private communication which did not include the quadratic density fits. But it is clearly

the case that the fit for May is published in Warren et al., so we revise our wording to: "quadratic fits of density … are not publicly available for all months".

Lines 121-124: I suggest to provide information whether and to which extend this modification of W99 has been implemented by follow-on studies, e.g. Tilling et al. ? Kwok and Cunningham 2015 ...

We have elaborated on this in the text:

Mainstream CS2 thickness retrieval products have generally used this approach since (e.g. Tilling et al., 2018; Kurtz et al., 2014; Hendricks et al., 2018). Kwok and Cunningham (2015) also investigated the effect of multiplying by a factor of 0.7 (rather than 0.5) with some success.

Lines 126-127: "are not currently used in sea ice thickness retrievals" —> I am wondering whether these variabilities are input into the uncertainty estimates provided alongwith the ESA-CCI SIT products? It might be worth to check.

This information is contained in the <u>D2.1 Algorithm Theoretical Basis Document for the ESA-CCI</u> <u>freeboards</u> (Sect. 2.9.4). The IAV values do indeed contribute to the uncertainty estimates provided with the ESA-CCI product and we have added this information to our manuscript.

Lines 130-133: "As such, ... values." –> Will this described in more detail in section 2? No it will not. Are positions of real drift stations used? If yes, which? If not: Isn't taking into account ALL ice covered grid cells of a 25 km grid providing a substantially different statistics - compared to the few drift stations used in the W99 climatology (their Fig. 1)? What is the time period considered? The description of this analysis step is lacking key details and should be re-written.

We have comprehensively rewritten this section to clarify. We now stress that the IAV values presented in W99 represent **the mean of many (positive) IAV values from individual, point-like drifting stations**. To compare mW99 and SnowModel-LG values to these in a rigorous way, we therefore must similarly take **the mean of many (positive) IAV values from individual points**. It would be wrong to take the IAV of the mean value of many points (which may each have cancelling positive or negative anomalies). This is a nuanced distinction that was not fully spelled out, but we believe to now be remedied. We also note that the logic above does not hold regarding the trends observed at drifting stations (because individual drifting station trends are not always positive), but because none the trends reported in Warren et al. (1999) are statistically significant we have not reported or visualised them in our manuscript.

Figure 2: Fig. 2 could be connected more easily to Fig. 1 if you'd use snow depth instead of SWE. It is not entirely clear how these maps are derived.

We have now changed this figure to snow depth and added clarifying information in the caption and the main text (see below):



Caption: "Snow depth variability at each EASE grid point over the 2002-2018 period. This is calculated by generating a timeseries of snow depth at each point and then calculating the standard deviation of that timeseries. High variability is displayed in a band where sea ice type typically fluctuates from year to year. IAV is zero in areas that do not exhibit sea ice type variability, introducing unphysically low variability in SIT in these areas."

I note that the ice edge in the Kara / Barents Seas looks a bit weird for months DEC and JAN.

These sharp edges were formed by an unexpected bug involving interaction between the "zero-line" of the W99 quadratic fits and our masking technique in these two months. A small change of a boolean ( $x<0 \rightarrow x<=0$ ) has fixed this.

Line 136: "this results" -> Which? Not clear to what this refers.

We have changed this to:

"We present this analysis of the point-like snow variability to illustrate that mW99..."

Line 138: "where the ice type typically varies from year to year" -> How did you define this?

We have clarified this with the following sentence:

"This band represents areas where the ice-type is not typically either FYI or MYI. Instead it is either switching between the two, or it is an area where FYI has replaced MYI during the period of analysis."

Lines 157-162: "We find ... (S3)." -> One question upfront: What is exactly the region you are considering here? The central Arctic Ocean? Laptev Sea included? Kara Sea?

In the manuscript we stated (L155):

"We instead compare the trends in basin-wide snow depth..."

But we appreciate it may have been unclear what we meant by "basin-wide". We have therefore added a set of clarifying parentheses so that the sentence now reads:

"We instead compare the trends in basin-wide (all shaded regions of Fig. 1) snow depth..."

I am just asking because at a certain point in winter the entire region considered should be ice covered and the FYI fraction hence be only a function of the MYI extent. I buy that there is a the decreasing fraction of FYI relative to the total ice extent in October due to later freeze-up. I don't agree, however, that this is the sole reason for your observation with respect to the trend mW99 snow depth and SWE. I believe an issue to consider is that the MYI coverage retreats more and more to those regions where W99 has maximum snow depth. Hence the relative fraction of MYI grid cells with comparably thick snow is increasing which to my opinion can result in a higher mean W99 snow depth (and SWE) for the MYI part of the sea ice in October.

This is an interesting point that we had not considered. To investigate the effect of the ice retretating into a region of high W99 snow depth, we repeated our analysis with a 'flattened' W99 climatology. To do this we first calculated monthly basin-wide SWE and depth averages from the W99 climatology. We then assigned those values to every point in the basin, rather than using the quadratic fits. This has the effect of removing the areas of high snow depth/SWE. We then halved the snow depth/ SWE over FYI as before to produce an mW99 product without the quadratic fitting.

We then performed our trend analysis again and found that the October trends were lower in the 'flattened' case, indicating that retreat into a higher-depth region of W99 is indeed playing a role in the positive snow trends. We do however note that the October trends in SWE and snow depth are both still significant at the 5% level in the 'flattened' case.





We therefore modify our manuscript to say:

"This increasing trend in snow depth is in part due to the diminishing area of October FYI relative to that of MYI (Fig. S4), and in part due to the retreat of the October ice into the Central Arctic where W99 exhibits higher snow depths and SWE."

Lines 164-166: "Several ... year to year. -> I don't find this formulation particularly clear. I find the "cannot accumulate snow from year to year" not to well chosen in the light of mostly complete snow melt during summer - also on multiyear ice. I'd state that there are two reasons for the observations with SnowModel-LG: 1) The MYI area shrinks. Hence your sentence about "a lower [smaller] ice area is exposed to snowfall in September/October fits well. 2) Freeze-up commences later, hence new seasonal ice either has not yet formed or is too thin to carry / accumulate snow resulting in a substantial amount of the precipitation falling as snow being dumped into open water. In short also here your sentence about "a lower ice areas ..." applies, meaning you can, to my opinion, delete that extension "also the later ... year."

We agree with this observation and have changed the manuscript to the following:

"We identify two processes as responsible for this decreasing trend: the MYI ice area is shrinking, so a smaller MYI sea ice area is present during during the high snowfall months of September and October (Boisvert et al., 2018); also freeze-up commences later, so a lower FYI area is available in these months and more precipitation falls directly into the ocean"

Line 167: "Webster et al. ..." -> I am wondering whether the "in situ sources" mentioned in the context of Webster et al. (2014) are i) also representing FYI and ii) aren't complemented with information from airborne operation ice bridge data [in which case these are not "in situ" anymore]. Please check! If their data indeed represent FYI and MYI then it might be worth to mention that explicitly in your manuscript.

This comment has led to our revision of this sentence. In particular, we highlight that the figure cited from Webster et al. (2014) includes raw data from the NP stations, so cannot be seen as independent of W99. We have also ammended the sentence to highlight that airborne measurements contributed significantly to the figure.

"Webster et al. (2014) observed a -0.29cm/yr trend in Western Arctic spring snow depths using both airborne and *in situ* sources. This airborne contributions to this figure included data over both ice types, and the in-situ contributions included data from individual Soviet drifting stations from the Western Arctic."

Lines 187-189: "Where sea ice ..." -> Could you please comment on whether this second data set is similar to / consistent with the OSI-SAF one? What is the basis? Given the fact that you investigate quite short time series in your paper and put quite some weight on different ice types it is important that thes two data sets are consistent to each other, i.e. provide a seamless continuous spatial FYI/MYI fraction distribution without a jump in total regional FYI and MYI extent from, e.g., Feb 2005 to March 2005.

The OSISAF-403c Global Sea Ice Type data is an operational product that is only available since 2005. Unfortunately the NSIDC product (which spans a longer time period) is not suitable because of its weekly time resolution (see next point), so we wanted to use this product as our primary source of ice-type data.

To extend our ice-type data for three years to reach the beginning of the radar freeboard timeseries we used the Copernicus Sea Ice Type Climate Data Record. Since our original submission this data has been relocated, the documentation has improved and it has been assigned a DOI (10.24381/cds.29c46d83). This documentation (<u>here</u>) includes a dedicated appendix (App. B) summarising the main differences between the OSI SAF operational product and the C3S product. The C3S product's underlying algorithm is adopted from the OSISAF operational processing chain, but then modified to produce a consistent record compatable with reanalysis.

The other key difference is that the OSISAF record uses active sensors such as the ASCAT scatterometer whereas the C3S product is confined to 'passive' satellite radiometers. This allows the OSISAF product to be delivered at a higher spatial resolution. This advantage is generally lost in our analysis as we would downsample to a 25km grid for compatability with our radar freeboard and SnowModel-LG data. But the reviewer is absolutely right to ask (a) whether these differences affect our study and (b) is it appropriate to switch from C3S to OSISAF when the latter becomes available for the winter of 2005?

We compared the two ice-type datasets over the period 2005-2019 and found the differences to be very small (once we'd classified the ambiguous pixels). Nonetheless we now opt to now use the C3S dataset over the entire period of our study. This is because our study is based around trends and variability, so a consistent record is of particular importance. We found after switching that this had a minimal impact on our results, and removes the need for a detailed assessment of (a) the abrupt transition from the CDS to the OSISAF product and (b) the impact on trends from the transition over the period.

We have rewritten the description of our ice type data to reflect this information.

I note that you also use NSIDC ice-age data and one could ask the question: why didn't you use ice-age data throughout the entire study?

We do not use the NSIDC ice-age data in this study. The only NSIDC product cited here is the concetration data that is used to drive the NESOSIM model. Although the NSIDC product has a long time span (1984-2019), it is only available at a weekly temporal resolution. This makes it unsuitable for use in conjunction with the monthly data from the ESA-CCI product. This unsuitability stems from ambiguity about how to split certain weeks of data that span two months.

Section 2.3: Please comment on two issues.

1) The sampling on which W99 is based has large regional variations with substantial differences between marginal seas such as the Kara or Barents Seas compared to the central Arctic. How does the lack of reliability of W99 in these partly undersampled regions influence your results - particularly in the two regions mentioned above?

Our response above to GC2 highlights the exceptionally poor (and sometimes non-existant) sampling of the Kara, Barents and Laptev seas by Russian drifting stations. This uncertainty undoubtedly propagates into W99 and mW99. However, this has not stopped sevaral groups from using W99 to derive sea ice thickness in these marginal seas (e.g. Li et al., 2020a,b; Belter et al., 2020) particularly in the pursuit of estimating sea ice volume (Tilling et al., 2015; Laxon et al., 2013).

We do not argue here that using (m)W99 in these regions is an appropriate approach, but simply point out the difference in trends and variability of SIT when calculated with the two different methods. We also point out that even if the climatology has an absolute error in these regions, it does not significantly affect our findings regarding variability and trends (which are both low or effectively zero in (m)W99 regardless of the snow depth value itself). We have now included this argument in our manuscript (see GC2 response).

2) Snow depths / SWE in the Kara / Barents Sea do - for the same reason - depend a lot on the extrapolation / fit function used. Howdoes this influence your results?Aren't the snow depth values in these regions too hypothetical to be adequately used in your study? Wouldn't it make sense to exclude the Kara and Barents Seas? To my opinion it would make your study considerably more credible. And it would potentially safe some space.

The extrapolation/fit function is indeed highly uncertain in these marginal seas, although this does not affect our central arguments concerning trend and variability. This is because the trends and variability in the implementation of (m)W99 are not sensitive to the extrapolation/fit functions, but rather the lack of trends that stem from the climatological approach.

Section 2.4: Isn't the EASE grid a polar aspect of the Lambert Azimuthal Equal area grid? I'd suspect no re-gridding is required.

The reviewer is correct that EASE is a polar aspect of the Lambert Azimuthal Equal Area projection, and our description of the data was confusing in this respect. Both the ESA-CCI freeboards and the SnowModel-LG data are supplied on a 25 km EASE grid. The ESA-CCI data were previously being subjected to a "regridding" process that did not change the positions of the relevant coordinates at all (but did remove low-latitude 'nan' datapoints to make the grids the same dimension). This regridding process has now been removed from our processing chain (with no impact on the analysis).

What is the grid resolution of the radar freeboard data?

As mentioned above, it supplied on a 25x25 km EASE grid, and this is now mentioned in our data description section.

What is the time period (years, months of the year) for which these data are available and used by you?

The radar freeboard data is available from the CCI website from October in the winter of 2002/03 until then end of winter 2016/17. However the CS2 radar freeboard component of the CCI product is functionally identical to the radar freeboard product made available by the Alfred Wegener Institue (Hendricks and Ricker, 2019). We have checked this manually for the period of overlap. We were therefore able to extend the timeseries of radar freeboards for one year until the end of winter 2017/2018, which is the end of our SnowModel-LG data series. All of these details have now been added to the Data Description of our revised manuscript.

How did you treat the overlap of Envisat and Cryosat-2?

A description of this was originally omitted in our manuscript. We direct the reviewer to our response above to GC2 where we have answered this question.

Key information is lacking here. It might also be worthwhile to take a look into the validation report of the SIT / freeboard data set used (see e.g.:

https://icdc.cen.uni-hamburg.de/fileadmin/user\_upload/ESA\_Sea-IceECV\_Phase2/ SICCI\_P2\_PVIR-SIT\_D4.1\_Issue\_1.1.pdf ). It provides some information about how "consistent" the two "merged" data sets are. Taking this information into account and discussing the potential biases (which still exist) I rate mandatory for a paper which so much relies on the analysis of this 17-year long time-series with a change of sensor right in the middle of the time series.

We note first for readers of this review that this document is now found <u>here</u>. We have now added relevant information to Section 2 on how the Envisat retracker threshold is varied to match CS2 in the period of overlap, as well as a citation to the PVIR document and some information on the relevant biases.

We have also added the following subsection to our Discussion:

## "Inter-Mission Sea Ice Type Bias between Envisat and CryoSat-2

An extensive validation exercise for the merged products indicated that although EnviSat radar freeboards match well with CS2 freeboards in the Arctic overall, some biases do exist over specific ice types (ESA, 2018). In particular, analysis of the inter-mission overlap period indicates that Envisat freeboards were biased low (relative to CS2) in areas dominated by MYI, and high in areas dominated by FYI.

We first make the point that this will have a relatively minimal effect on our findings regarding interannual variability, as  $\overline{\text{Snow}}$  is unaffected by this and  $\sigma^2_{\overline{\text{RF}}}$  is likely relatively independent of the absolute magnitude of  $\overline{\text{RF}}$ .

With regard to trends, if Envisat radar freeboards (and thus  $\overline{RF}$ ) are in fact biased high over FYI between 2002-2010 (relative to CS2), then the total trends in many regions dominated by FYI could potentially be smaller than calculated in this manuscript.

We do however add that our central findings regarding the impact of declining  $\overline{\text{Snow}}$  is unaffected by any inter-mission bias in  $\overline{\text{RF}}$ . Because the trend in  $\overline{\text{SIT}}$  is determined by both  $\overline{\text{Snow}}$  &  $\overline{\text{RF}}$ , the trend in  $\overline{\text{SIT}}$  will always be more negative when calculated with downward trending data for  $\overline{\text{Snow}}$ ."

Line 206: "ice motion vectors" –> Which ice motion vectors? Please provide this information - including the temporal resolution and the version of that ice motion data set used - in your manuscript.

The ice motion vectors used were v4 of the polar pathfinder ice motion vectors (Tschudi et al., 2020). These are supplied on the 25x25 km EASE grid and were analysed at weekly time resolution. This information has been included in our revised manuscript.

In addition: "pan-Arctic snow depth and density distributions" -> Please provide a spatial and a temporal resolution as well as the domain. While the paper focuses a lot on snow depth I am wondering how snow densities obtained with SnowModel-LG compare to W99 ones?

The SnowModel-LG output is supplied on the 25 km EASE grid at daily time resolution. From this we produced monthly gridded fields for combination with the monthly radar freeboard data. This information is now included in our revised manuscript.

Line 210: "snow-ice accumulation" –> Please explain what "snow-ice accumulation" is. Do you refer to snow-ice formation at the basal snow layer?

We were actually referring to superimposed ice and used the wrong term. Snow-ice in its conventional meaning results from flooding of the base of the snow by depression of the ice surface below the waterline. This is not modelled in SnowModel-LG, as the model does not model or assimilate ice freeboard. We have removed this reference to snow-ice from our revised manuscript.

Lines 216/217: "snow depth differences ... than 5 cm" –> This is a quite global statement. Is this an Arctic mean value? Is this the mean difference in SnowModel-LS realizations just for the grid cells co-located with the OIB data? What is the standarddeviation of this difference?

In our submission we stated:

"snow depth differences between the reanalysis products were found to be less than 5 cm"

By this we were referring to the difference in SnowModel-LG snow depths when forced with MERRA-2 and ERA5 reanalysis data. We have clarified this in our revised manuscript to:

"snow depth differences between the *ERA5 and MERRA2* reanalysis products were found to be less than 5 cm"

The 5 cm figure represents the Arctic as a whole and is illustrated by Figure 4 of Stroeve et al. (2020). The standard deviation of this value was not reported by the authors. We investigate this here.

We first visualise the "mean bias", which we calculate as the difference in the snow depths calculated when SnowModel-LG is run with ERA5 and MERRA2, averaged over several years. In this scheme, if the ERA5 run is larger in one year and smaller in another, this cancels out. This indicates the degree of *bias*. We also calculate the average absolute difference. For this, we subtract MERRA2 SnowModel-LG output from that of ERA5, and then take the absolute value. We then plot the time-mean of this distribution. We also summarise these data by region and by month in the lower right hand panels of the below figures:



We place this figure in the supplement and reference it from the main mauscript.

Do the OIB data used to tune SnowModel-LG represent FYI conditions adequately?

Figure 7 of Stroeve et al. (2020) illustrates the locations of the OIB flights used to tune SnowModel-LG. The vast majority are over the Beaufort Sea and the Greenlandic side of the Central Arctic, which is generally consituted by multiyear ice. It is therefore conceivable that the scaling factor would be different if FYI were better sampled by OIB. We have raised this point in our revised manuscript.

We do however believe that the scaling factor is not paticularly relevant to our central observations concerning trends and variability. It is certainly the case that if the scaling factor is off then we are under/overestimating snow and sea ice thickness. However the impact on this on trends and variability is potentially very small.

Section 2.6: For a better understanding it might make sense to explicitly state whether precipitation and/or snow fall are assimilated into NESOSIM as well.

It does assimilate reanalysis precipitation directly and we have now explicitly stated this.

The snow pack initialisation, is this covering both snow depth and density? As W99 data are monthly values, is this initialisation only done monthly, or are monthly values interpolated to daily values with which the model is initialised henceforth?

In this study we use data from a NESOSIM run initialised on the 15<sup>th</sup> August for each year (Sect 2.1 Petty et al., 2018). The initial depth was produced by a "near-surface air-temperature-based scaling of the August W99 snow depth climatology". This is a linear scaling based on the duration of the preceeding summer melt season. Snow density was initialised using the August snow-line observations of Soviet NP drifting stations 25, 26, 30 and 31. Data from the most recent publicly available drifting stations were chosen to maximise their relevance in a changing climate. We have ammended our draft manuscript to include this information.

You explicitly mention depth-hoar and wind-packed layers in the context of NESOSIM. Does this imply that SnowModelLG does not represent such features? If not, then I suggest to be more specific in the description of what SnowModel-LG can do and what not.

SnowModel-LG does include these aspects of snowpack evolution in a multi-layered scheme. We have added a line in the SnowModel-LG descrition to reflect this, and also modified our subsequent description of NESOSIM to draw a direct comparison.

Section 3.2: Please provide more details. How many grid cells with valid SIT measurements are requied to compute a regional mean SIT value?

(related comment) How many valid observations are required for the results broken down into ice types (see Figures S4 and S5)?

Snow, RF & SIT were calculated when any grid cells were present in a region, and we have now added this information to our manuscript. We have also now conducted analysis on the number of measurements that feed into our analysis. Although the number of snow and radar freeboard data points in a region were closely related (as both are tightly coupled to the ice area in the region), we found that there were generally fewer radar freeboard measurements. We have visualised this below:



We add to our manuscript:

"Snow, RF & SIT were calculated where any valid grid points existed on the 25x25 km EASE grid. Because of this, no average values were computed in the Kara Sea in October 2009 or 2012. Furthermore, no October values were generally available in the Barents Sea after 2008 (with the exception of 2011 and 2014). The impact of this on our analysis is clearly visible in Fig. (10).We do not exclude the Barents Sea in October from our analysis because of the low number of valid points, but we do highlight the undersampling issue here. We continue to consider it because we do not find statistically significant declining trends with the data we have, so essentially we are reporting a null result. Our calculations of interannual variability in this month is inherently adjusted for the small sample size, but we nonetheless urge caution in interpretation of the value."

How about regional means of radar freeboard and snow depth / SWE? Did you compute these as well?

These were calculated during our analysis and plotted alongside each other as a figure in the supplement. We have now added an explicit reference to this figure to this section.

Please provide a reference for the "Wald test".

Although we were correct to use this term, its use (and technical definition) are not particularly enlightening for the reader. Instead we have opted now to just describe the test as a hypothesis test (and state the null-hypothesis).

Line 249: I suggest to stress here once more what "Snow\_overbar" is, that it is not the snow depth but the snow-depth contribution to the SIT retrieved from altimetry

We have reiterated the definition here.

Line 259: "individual years, regions and months" –> Not clear what you did. You used detrended time-series of monthly, region-mean values of RF and snow and computed the correlation between these time series separately for every month and every region?

Yes, this is correct. We have added the following text to clarify this:

To do this we calculated a monthly timeseries of  $\overline{\text{RF}}$  and  $\overline{\text{Snow}}$  for each region over the time-periods (2002-2018, with the Central Arctic being 2010-2018). Because we considered eight regions and seven months, this led to to 56 pairs of timeseries for  $\overline{\text{RF}}$  and  $\overline{\text{Snow}}$ . We then detrended each of them. We then calculated the correlation between each of the pairs of detrended timeseries.

Lines 263/264: "The Barents Sea ... correlation." -> I suspect this observation is based on two completely different causes. For the Central Arctic the time series is just 9 years long. For the Barents Sea, neither is mW99 overly reliable nor are RF values overly reliable - especially during the Envisat period.

We have added a brief clause to this line: "- the reasons for this are discussed in Sect. (5.4)".

Another, more general comment: The RF data for region Central Arctic are considerably more robust in terms of the number of valid observations contributing to the RF values used.

Figure 6: Not clear what is shown magnitude-wise on x- and y-axes. The same applies to Figures S6 and S7.

Because correlation statistics are not sensitive to the choice of axes, units or linear scalings of the values, we decided to not display axes ticks or labels and scale the axes to fit the rectangular panels of the figure. However we clearly should have stated this in our submission and we now have added the following text:

"We note here that the correlation between the timeseries is dependent on their relative position to a linear regression. These correlation statistics are thus independent of the absolute magnitude of the values, their units, or any linear scaling of the axes. We therefore choose to present the correlations in Fig (7) without axes and scaled to the rectangular panels, so as to best show the relative positions of the points without extraneous numerical information."

Line 286: "but analysis ... regions." -> Not clear what you mean here.

We have reworded this sentence for clarification:

"but analysis of this grouping conceals more significant variation at the scale of the individual group members" Line 287: "The covariability ... contribution" –> This discussion focuses on radar freeboard. It does not comment on the observation that at the beginning of winter (Oct.) the fraction of SIT IAV that is explained by RF-Snow covariance is larger than snow IAV.

We have added the following text to include this information:

We note that in contrast to  $\sigma^2_{\overline{RF}}$ ,  $\sigma^2_{\overline{Snow}}$  is almost always larger than the covariability component. A noticeable exception to this generalisation is in October for the Marginal Seas grouping, where the covariability contribution to  $\sigma^2_{\overline{SIT}}$  is around twice as large as the contribution from  $\sigma^2_{\overline{Snow}}$ 

Figure 7: Can you please check whether your representation of "Fraction of Total Variance (%)" is correct? I mean, ok, if the dimensionless factor rho is negative then the covariance term in Eq (5) gets a negative sign. Therefore you plot negative bars in panels (b).

We gave the visualisation of the data in Figure 7 considerable thought, but agree that a succinct characterisation of what is actually being plotted on the y axis is challenging. It is possible that "Fraction" is not the best word, as it carries connotations of being less than a whole. Given that the quantity on the y axis is the normalised contribution to  $\sigma^2_{STT}$ , then we propose that this is put as the axis label to minimise confusion.

However, looking at the Central Arctic, November, this results in a fraction of radar freeboard IAV of about 110%, also the one for October is larger than 100%. I get a headache with this because a fraction cannot be negative (have you ever had a negative piece of cake?) and it can also not be larger than 100%, i.e. larger than the total (only if you order a medium size Pizza and get a large one instead). This applies then also to Figure S15 where the deviations from 100% are even larger. Again, I can see from Eq. 5 that it is mathematically correct. However, a positive covariability means that sigma\_RF and sigma\_snow are positively correlated while a negative one means that these quantities are anti-correlated. If we assume a very strong negative covariability of, say -90%, does that mean that the IAVs of RF and snow need to sum up to a fraction of the total of 190%?

We consider the more intuitive case of apples and oranges being delivered in a truck. If the number of apples is determined randomly and truck space is limited, it is the case that when more apples are delivered then less oranges are delivered. This corresponds to a strong negative covariability between the n\_apples and n\_oranges – they are random, dependent variables just like RF and Snow. For the case that 100 extra apples delivered means that you lose the space for 90 oranges, then the variability in n\_apples or n\_oranges individually is many times larger than the variability in the total number of pieces of fruit. If your standard deviation in n\_apples = 500, then your standard deviation in n\_oranges is around 450, but the standard deviation in total pieces of fruit is just 50. This case illustrates that the  $\sigma^2_{RF}$  and  $\sigma^2_{Snow}$  can logically be much larger than the standard deviation of the sum of RF and Snow (so long as their covariability is negative).

Lines 327/328: "Perhaps more significantly, ..." –> This I don't find too convincing - also given the unknown uncertainty of these regional mean SIT values. I suggest to only mention these three new trends but do not hypothesize about the main reason.

We have removed our hypothesis about this being driven by the years 2003 and 2004.

Line 350: "... truncated SIT distribution ... thicker ice." -> This relatively global statement is not supported by Fig. S10 for all months.

We have now qualified this statement by adding "in the months January – April".

Particularly, I would not use the word "truncated". Truncated means that below or above a certain SIT values the area occupied by these SIT bins is abruptly zero

We have now removed all uses of the word "truncated" from our manuscript. In the case of the CS2 observational period, we have used the word "shorter". In this case of the distribution of sea ice thickness distributions, we have used the word "narrower".

Lines 351/352: Please see my comment at Figure S10: You need to provide more details about how you derived this Figure. What is missing are binsizes and borders as well as the time-period for which the Figure is valid (2010-2018 I assume) as well as a statement here that this Figure is now showing a classical pdf but expresses the distribution in form of sea-ice area. In order to avoid confusion with the classical definition of sea-ice area which is sum of the area of ice covered grid cells weighed with the actual sea-ice concentration, you might want to rename your y-variable.

Line 353: "The regional, seasonal growth rate ..." -> What is the period considered?

We've added the following sentence to clarify this:

"These rates were calculated over the period 2002-2018 with the exception of the Central Arctic which was restricted to the period 2010-2018."

Lines 372-374: I suggest to refer to the Boisvert et al. (2018) paper here (about the difference between Merra-2 and ERA-Interim) and in addition take into acount this paper: <u>https://doi.org/10.1029/2019GL086426</u>.

We have now added the following text:

"Boisvert et al. (2018) conducted a similar analysis with drifting ice mass balance buoys, and found the interannual variability of the data sets to also be similar (although the authors found larger discrepancies in magnitude). These differences in magnitude however cannot be physical (as there is only one Arctic), and Cabaj et al. (2020) were able to bring precipitation estimates into better alignment using CloudSat data with a scaling approach. However this scaling approach preserved the interannual variability of the data sets, which Barrett et al. (2020) and Boisvert et al. (2018) found to be in comparatively good agreement.

I am wondering whether Figure S12 is required. After all it confirms that your choice combining SnowModel-LG runs of ERA5 and Merra-2 is a good one.

We agree that Fig. S12 is not strictly required, although we would prefer to keep it in the supplement (as were the reader to not see it, they may doubt our choice).

Line 387: "replicates the higher contribution ..."  $\rightarrow$  Even though for the Central Arctic the data are just based on the period 2010-2015 = 6 years.

The word "replicates" does perhaps imply that the data are more directly comparable than they are. We therefore explicitly note the timeframe difference and change "replicate" to "also exhibits" as such:

"Despite the shorter timeframe, NESOSIM also exhibits an increasingly ... and also a higher..."

Lines 390/391: "underlying trends ... period" –> Ok ... but at the same time you use a shorter time period for the Central Arctic anyways. So this argument is not conclusive.

In response we add:

"(by comparison to regions where all relevant data is available from 2002-2018)"

Lines 394-399: "As such ... 2018)." –> I suggest to shorten this part and perhaps either delete Figure 11 or move it as well to the Appendix. - which however, already contains a lot of additional material. But I feel that the paper would be still understandable with a few sentences highlighting the common findings between NESOSIM and SnowModelLG

We have removed this paragraph and the figure.

Line 427-428: "... these investigations ... warmer temperatures ..." -> I do not quite agree to this statement because, according to my knowledge, among the cases investigated and presented in Nandan et al. (2017) as well as in the later paper by Nandan etal. (2020) is a sufficiently high number of cases with cold snow; hence the observation of a rising scattering horizon is not uniquely tied to warmer temperatures.

In response to this comment we have modified this statement:

"However, these investigations were **often (but not exclusively)** carried out at the end of the winter season or in the Sub-Arctic, when warmer temperatures **may have increased** the snow's brine volume fraction..."

Lines 434-440: "Knowledge ... depth." -> I can follow the physical reasoning here and also how it is connected with Equation (2). However, it might not be as straightforward as it is formulated, given the fact that equation (2) transforms into equation (4) in Kwok and Cunningham (2008, https://doi.org/10.1029/2008JC004753 ) for a lidar - hence introducing a different factor in front of the snow depth. Perhaps formulating these lines more like being your own hypothesis than being fact might be more appropriate?

To improve the rigor of this section we have elected to remove the following sentence:

"This incidentally raises the possibility that radar waves with a certain relative penetration depth may allow the estimation of SIT without requiring any knowledge of the snow depth."

We hope that by doing this we remove the hypothesis-based element of this part of the discussion.

Line 443: In the context of the paragraph ending here you could also comment on the impact of i) the increased likelihood of a flooded basal snow layer in regions of comparably thin sea ice / thick snow / high SWE as, e.g. the Barents Sea and part of the region Central Arctic facing the Atlantic and ii) of a potentially increased likelihood for enhanced snow metamorphism, like is typical for the Antarctic, even in the middle of winter due to intrusions of marine air-masses into the Arctic, e.g. via the Fram Strait.

We have addressed these issues with the following paragraph:

"We finally note the potentially confounding influence of negative freeboard in regions such as the Atlantic sector of the Central Arctic region and the Barents Sea. In the case of high snowfall and low ice thickness, the ice surface can be depressed to the waterline or below. Beyond this point Eq. (5) no longer function. The prevalence of negative freeboards has been studied by Rosel et al. (2018) and Merkouriadi et al. (2020), but has yet to be incorporated into any radar-altimetry based sea ice thickness retrievals. This situation can be driven by storm tracks entering the Arctic from the Atlantic (but also the Bering Strait). These intrusions of warm air can also drive snow grain metamorphism, which may well affect radar penetration through the snowpack."

Line 469: "in the high precipitation months ..." -> One could argue, though, that thanks to later freeze-up and the concomitant change in atmospheric moisture content (and perhaps also circulation) also shifts the maximum of the precipitation to a later time - and with that the seasonal sea ice would still be able to accumulate a fair amount of snow - particularly supported by the fact that a warmer atmosphere can hold more moisture and with that can lead to increased precipitation rates. This is hypothetical of course but among the possible scenarios.

Lines 504/505: "negative ... this is not seen." –> I thought the negative covariances between RF and Snow shown in Figures 7 and S15 are an indication of exactly this observation?! The fact that these only occur in early winter makes a lot of sense as well.

Yes, "this is not seen" was an oversimplification. We have ammended this to read:

"This corresponds to a negative covariability term in Eq. (5) and is represented by purple bars in Fig. (8). Negative values are generally not seen, with the exception of October and November in the Central Arctic, November in the Barents Sea and December in the Chukchi Sea."

Lines 505-507: "... snow is a highly ... weeks." -> I am wondering whether it wouldmake sense to to have a thought experiment to check whether the magnitudes of the the changes involved fit this hypothesis. While Snow overbar can be easily computed based on the snow properties and is independent of the sea-ice thickness and its radar freeboard, RF\_overbar is not. It is a function of both, ice growth and snow load. An experiment one could think of is, e.g. an initially 80 cm thick ice floe with i) 5 cm snow depth and ii) 20 cm snow depth (and similar snow densities) grows at -30 degC over a month. What is the RF for both cases at the beginning and what is the change in RF over the month? Without further snow accumulation Snow overbar remains constant. How would the change in RF be modified if one would add another 5 cm of snow in the middle of the month and again at the end of the month? My hypothesis is that RF overbar and Ssnow overbar are correlated well in case a thin to moderately deep, slowly increasing snow cover allows adequate ice thickening so that Snow\_overbar increases over time but RF increases as well over time. Further, my hypothesis is that RF\_overbar and Snow\_overbar are not well correlated in case an already thick snow cover hampers adequate ice thickening while it further deepens over time. In that case Snow overbar increase over time like in the above-mentioned example, but the RF increase by increasing SIT and hence sea-ice freeboard is counterbalanced by an increasing radar range such that the increase is considerably slower than in the first example or even zero.

We found the reviewers comments of great interest, and they have clear bearing on the seasonal correlations between RF and Snow We point out that while there is an immediate (theoretical) negative impact between snow accumulation and radar freeboard, this relationship is weakened by a negative feedback in which snow insulates theice and reduces subsequent growth. We believe that that this monthly-to-seasonal interplayof snow and radar freeboard is fertile ground for further study, and draw the reader's attention recent

conference presentations by Lawrence et al on this topic, as well as previous modelling work by Stroeve et al. (2018) and Petty et al. (2018).

Lines 513-517: "Freeze-up ... further study" -> I doubt that for the months you consider this is an issue. Melt ponds in the Central Arctic begin to freeze over in mid August, latest in September; hence any snow falling in October falls on solid ice. I suggest to delete this part.

In response we have removed this section.

Typos / Editoral remarks

Lines 5/6 " ... with the conventional method ... " –> I suggest to tie this better the usage of a snow climatology.

We have ammended this line to do this

Line 91: I suggest to use a different letter for the dimensionless factor than rhoto avoid confusion with a density.

We have ammended this line to do this

Also "sigma\_SIT\_overbar" should possibly read "sigma\_Snow\_overbar"

Yes, this was a typo and has now been fixed.

Line 110: "radar speed" -> "radar wave speed" or "speed of the radar waves"

This line has now been restructured in response to another comment and "radar speed" no longer appears.

Line 118: If I am not mistaken, then Giles et al. (2008b) is purely dealing with the Antarctic, hence application of the W99 climatology appears to be unlikely.

Yes, we have removed this reference.

Line 119: I agree that Eq. (2) contains the snow contribution to the sea-ice thickness; however, the term Snow\_overbar is used in Eq. (4).

Yes, we have now changed this.

Line 134: "average half of the value" -> In order to allow a more fluent reading I suggest to write "about 50%" instead of "average half of the value" in Line 134. That way it will be easier to connect this finding with the SnowModel findings.

We have now changed this.

Line 139 and 140: "consistent" -> What is a consistent ice type in this context? Could it be that you wanted say "constant" or "unchanged"?

We have changed "temporally consistent" to "temporally unchanging"

Line 187: "2005" -> "March 2005"

We've added this as suggested.

Lines 205: "assimilates reanalysis weather data" -> "is capable to assimilate meteorological data from different atmospheric reanalyses (see below)"

We've changed this as suggested.

Lines 214: "results of reanalysis" -> "representation of the actual distribution of relevant meteorological parameters by atmospheric reanalyses"

We've changed this as suggested (and added a citation to Boisvert et al., 2018)

Line 228: "Eq. (2)" -> Same issue as for Line 119 (see above)

We've changed this as suggested.

Line 238: "The Central Arctic region exists above the latitudinal limit of the Envisat orbit" -> "The Central Arctic region is not sufficiently well observed by the Envisat radar altimeter (see Fig. 4)"

We've changed this as suggested.

Line 255: "SnowModel's" -> I suggest to always keep the full name.

We've changed this as suggested.

Line 257: "Having calculated ..." –> I suggest to refer to Figure 5 and the standard deviation shown therein.

We've added a reference to Figure 5 (detrended timeseries of  $\sigma^{2}_{\text{Snow}}$ )

Line 260: "between" -> "between detrended"

We've added this as suggested.

Line 278: "in for each" -> "for each"

We've changed this as suggested.

Lines 286/287: "grouping by comparison to" -> "grouping in comparison to"

We've changed this as suggested.

Lines 287-289: I suggest to add "radar" to all mentioning of "freeboard"

We have been through each usage of the word "freeboard" and added "radar" where appropriate.

Figure 7 caption: For better readability of the figure and the text refering to it I suggest to somewhere re-introduce IAV here as the inter-annual variability. The last time IAV was used was in Figure 2.

We have reintroduced the term on L293.

I cannot see (a) and (b) in the Figure.

We've now added these annotations.

I suggest to write "panel" instead of "figure in Line 293.

We've made this change as suggested.

Line 294: "continuous" -> "contiguous"

We've made this change as suggested.

Line 296: "was the" -> "was in the"

We've made this change as suggested.

Line 306 "2002 - 2018": I suggest to add something like: "except for the Central Arctic: 2010-2018" in this caption as well as in all other figure captions where the Central Arctic region is shown along with results of the other regions.

We've made this change as suggested.

Line 315: "Fig. 9" -> I suggest to refer to the panels highlighted in green.

We've made this change as suggested.

Line 320: "declining" -> "negative"

We've made this change as suggested.

Line 324: "mW99" -> Suggest to refer to Fig. 9 and the red panels

We've made this change as suggested.

Line 325: "declining months to four" -> "months with a decline in SIT to four"

We've made this change as suggested.

Line 326: "after 2006" -> I suggest to write "after 2003, except 2006"

We've made this change as suggested.

Lines 333/334: "are only ones" -> "are the only ones"

We've made this change as suggested.

Line 337: "Laptev Sea." -> I suggest to add "when using SnowModel-LG instead of mW99."

We've made this change as suggested.

Line 340: "statistically significant months" -> When SIT is computed using SnowModelLG?

We've made this change as suggested.

Line 344: "snow" -> "show"

We've made this change as suggested.

Figure 9, caption: I suggest to add the notion that y-axes of Central Arctic and East Siberian Sea differs from all other regions.

We've made this change as suggested.

I suggest to make "Where trends are ... superimposed" the second sentence of the caption.

We've made this change as suggested.

I suggest to add a note that unlike all other regions Central Arctic data are based solely on Cryosat-2 data.

We've made this change as suggested.

Line 356: "Siberian there" -> "Siberian Sea there"

We've made this change as suggested.

Line 404: "Chukchi" -> "Chukchi seas"

We've made this change as suggested.

Line 405: "one month" -> "only one month"

We've made this change as suggested.

Line 427: "raise" -> "rise"

We believe that we have used the word 'raise' correctly here (with the difference being that *raise* is a transitive verb while *rise* is not – the object of the sentence being "the scattering horizon").

Line 442: "diminishing show cover" -> I suggest to add: ", i.e. actual values of snow depth and SWE that are smaller than the climatological values"

In this paper we do not argue that snow depth or SWE are underestimated in an absolute sense by comparison to W99. Instead we argue that the rate of decline (which is important for trends) is not properly represented by W99. As such we do not feel that we can add this line.

Line 457: "EM" -> Has EM been explained already ...?

This abbreviation had not been previously defined and we have replaced it with "electromagnetic"

Line 489: "... truncated" -> Please re-phrase. It is not the radar altimetry time series that is truncated. There is a region where simply no measurements could be taken.

We have reworded this to:

"at best limited to the CryoSat-2 era"

Line 511: "has longer to" -> "has longer time to"

We've made this change as suggested.

Comments for the supplementary material:

Line 19: Please refer to Figure S1 here. Otherwise it is completely unclear how you ended up with the expression in Eq. (S8). You might want to replace the "=" by an "is approximated by ... as illustrated in Figure S1."

We have rewritten L21 to read:

"This linearity is visualised in Fig. (S1) and allows the second term..."

Line 23: "reformulated as" -> Which value is used for rho\_water?

We have rewritten L23 as follows:

"This can be reformulated by setting  $\rho_w = 1023 \text{ kgm}^{-3}$  as follows:"

Figure S6, caption, line 3: "five" -> "four":

We wrote:

"A persistent, positive correlation exists in the Central Arctic and the East Siberian Sea in the last five months of winter."

It's possible that we're misinterpreting the reviewer's comment here, but we believe this should read "five". Here we reproduce the two regions in question. The five months of December, January, February, March and April all exhibit positive correlations.



Figure S10, caption "... SnowModel-LG data." -> I suggest to add something like ... "expressed as total sea ice area of all grid cells falling into a specific SIT bin." In addition: What is the bin size? What the bin borders? Since it includes the region "Central Arctic" this figure is based on years 2010-2018 only, correct?

We have added the suggested clarification and have included the bin size (which is 5 cm). The reviewer is correct that this is carried out in the CS2 period (2010-2018) and we have also added this.

Figure S11: There are two identical panels denoted (b).

This has now been fixed.

Please make a note in the caption about the differences in the y-axis range.

We have rescaled the plots so all regions have the same y-axis scales with the exception of the Central Arctic. We have noted this exception in the caption.

Please make a note about the error bars. i.e. what these represent.

The error bars represent 1 standard deviation either side of the mean of the timeseries. We have now added this information to the caption.

Figure S11, caption: "The SnowModel-LG contribution ..." -> While this statement is undoubtly correct also the aggregated marginal seas start low and end high. So why not also commenting on those regions?

We have now commented on this.

Again, I note that you need to provide the information whether all regions but the Central Arctic are based on the full period 2002-2018 or whether indeed all regions only used data from 2010-2018.

We have now included this in the caption.

Figure S13: Again the notion about the smaller time period covered for region Central Arctic is missing.

We have now noted this in the caption.

## **TCD** Discussion References

- Barrett, A. P., Stroeve, J. C., and Serreze, M. C.: Arctic Ocean Precipitation From Atmospheric Reanalyses and Comparisons With North Pole Drifting Station Records, Journal of Geophysical Research: Oceans, 125, https://doi.org/10.1029/2019JC015415, URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2019JC015415, 2020.
- Belter, H. J., Krumpen, T., Hendricks, S., Hoelemann, J., Janout, M. A., Ricker, R., and Haas, C.: Satellite-based sea ice thickness changes in the Laptev Sea from 2002 to 2017: comparison to mooring observations, The Cryosphere, 14, 2189–2203, https://doi.org/10.5194/tc-14-2189-2020, URL https://tc.copernicus.org/articles/14/2189/2020/, 2020.
- Boisvert, L. N., Webster, M. A., Petty, A. A., Markus, T., Bromwich, D. H., and Cullather, R. I.: Intercomparison of precipitation estimates over the Arctic ocean and its peripheral seas from reanalyses, Journal of Climate, 31, 8441–8462, https://doi.org/10.1175/JCLI-D-18-0125.1, URL http://journals.ametsoc.org/doi/10.1175/JCLI-D-18-0125.1, 2018.
- Cabaj, A., Kushner, P., Fletcher, C., Howell, S., and Petty, A.: Constraining Reanalysis Snowfall Over the Arctic Ocean Using CloudSat Observations, Geophysical Research Letters, 47, https://doi.org/10.1029/2019GL086426, URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL086426, 2020.
- ESA: Sea Ice Climate Change Initiative: Phase 2 D4.1 Product Validation & Intercomparison Report (PVIR), Tech. rep., 2018.
- Giles, K. A., Laxon, S. W., and Worby, A. P.: Antarctic sea ice elevation from satellite radar altimetry, Geophysical Research Letters, 35, L03 503, https://doi.org/10.1029/2007GL031572, URL http://doi.wiley.com/10.1029/2007GL031572, 2008.
- Hendricks, S. and Ricker, R.: Product User Guide & Algorithm Specification: AWI CryoSat-2 Sea Ice Thickness (version 2.2), 2019.
- Kwok, R. and Cunningham, G. F.: Variability of arctic sea ice thickness and volume from CryoSat-2, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373, https://doi.org/ 10.1098/rsta.2014.0157, 2015.
- Landy, J. C., Petty, A. A., Tsamados, M., and Stroeve, J. C.: Sea ice roughness overlooked as a key source of uncertainty in CryoSat-2 ice freeboard retrievals, Journal of Geophysical Research: Oceans, 44, 1–36, https://doi.org/10.1029/2019jc015820, 2020.
- Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S., and Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and volume, Geophysical Research Letters, 40, 732–737, https://doi.org/10.1002/grl.50193, 2013.
- Li, M., Ke, C., Shen, X., Cheng, B., and Li, H.: Investigation of the Arctic Sea ice volume from 2002 to 2018 using multi-source data, International Journal of Climatology, p. joc.6972, https://doi.org/10.1002/joc.6972, URL https://onlinelibrary.wiley.com/doi/10.1002/joc.6972, 2020a.
- Li, Z., Zhao, J., Su, J., Li, C., Cheng, B., Hui, F., Yang, Q., and Shi, L.: Spatial and temporal variations in the extent and thickness of arctic landfast ice, Remote Sensing, 12, 64, https://doi.org/10.3390/RS12010064, URL www.mdpi.com/journal/remotesensing, 2020b.
- Merkouriadi, I., Liston, G. E., Graham, R. M., and Granskog, M. A.: Quantifying the Potential for Snow-Ice Formation in the Arctic Ocean, Geophysical Research Letters, 47, no, https://doi.org/10.1029/2019GL085020, URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085020, 2020.
- Nandan, V., Geldsetzer, T., Yackel, J., Mahmud, M., Scharien, R., Howell, S., King, J., Ricker, R., and Else, B.: Effect of Snow Salinity on CryoSat-2 Arctic First-Year Sea Ice Freeboard Measurements, Geophysical Research Letters, 44, 419–10, https://doi.org/10.1002/2017GL074506, URL http://doi.wiley.com/10.1002/2017GL074506, 2017.
- Nandan, V., Scharien, R. K., Geldsetzer, T., Kwok, R., Yackel, J. J., Mahmud, M. S., Rosel, A., Tonboe, R., Granskog, M., Willatt, R., Stroeve, J., Nomura, D., and Frey, M.: Snow Property Controls on Modeled Ku-Band Altimeter Estimates of First-Year Sea Ice Thickness: Case Studies from the Canadian and Norwegian

Arctic, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 1082–1096, https://doi.org/10.1109/JSTARS.2020.2966432, 2020.

- Perovich, D. K., Longacre, J., Barber, D. G., Maffione, R. A., Cota, G. F., Mobley, C. D., Gow, A. J., Onstott, R. G., Grenfell, T. C., Scott Pegau, W., Landry, M., and Roesler, C. S.: Field observations of the electromagnetic properties of first-year sea ice, IEEE Transactions on Geoscience and Remote Sensing, 36, 1705–1715, https://doi.org/10.1109/36.718639, 1998.
- Petty, A. A., Holland, M. M., Bailey, D. A., and Kurtz, N. T.: Warm Arctic, Increased Winter Sea Ice Growth?, Geophysical Research Letters, 45, 922–12, https://doi.org/10.1029/2018GL079223, URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2018GL079223, 2018a.
- 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, 4577–4602, https://doi.org/ 10.5194/gmd-11-4577-2018, 2018b.
- Radionov, F. and Fetterer, V. F.: Environmental Working Group Arctic Meteorology and Climate Atlas, Version 1, https://doi.org/10.7265/N5MS3QNJ, 2000.
- Rösel, A., Itkin, P., King, J., Divine, D., Wang, C., Granskog, M. A., Krumpen, T., and Gerland, S.: Thin Sea Ice, Thick Snow, and Widespread Negative Freeboard Observed During N-ICE2015 North of Svalbard, Journal of Geophysical Research: Oceans, 123, 1156–1176, https://doi.org/10.1002/2017JC012865, 2018.
- Sallila, H., Farrell, S. L., McCurry, J., and Rinne, E.: Assessment of contemporary satellite sea ice thickness products for Arctic sea ice, The Cryosphere, 13, 1187–1213, https://doi.org/10.5194/tc-13-1187-2019, URL https://www.the-cryosphere.net/13/1187/2019/, 2019.
- Stroeve, J. and Notz, D.: Changing state of Arctic sea ice across all seasons, Environmental Research Letters, 13, 103 001, https://doi.org/10.1088/1748-9326/aade56, URL https://doi.org/10.1088/1748-9326/aade56, 2018.
- Stroeve, J., Nandan, V., Willatt, R., Tonboe, R., Hendricks, S., Ricker, R., Mead, J., Mallett, R., Huntemann, M., Itkin, P., Schneebeli, M., Krampe, D., Spreen, G., Wilkinson, J., Matero, I., Hoppmann, M., and Tsamados, M.: Surface-based Ku- and Ka-band polarimetric radar for sea ice studies, The Cryosphere, 14, 4405–4426, https://doi.org/10.5194/tc-14-4405-2020, URL https://tc.copernicus.org/articles/14/4405/2020/, 2020.
- Tilling, R. L., Ridout, A., Shepherd, A., and Wingham, D. J.: Increased Arctic sea ice volume after anomalously low melting in 2013, Nature Geoscience, 8, 643–646, https://doi.org/10.1038/ngeo2489, 2015.
- Tschudi, M. A., Meier, W. N., and Scott Stewart, J.: An enhancement to sea ice motion and age products at the National Snow and Ice Data Center (NSIDC), Cryosphere, 14, 1519–1536, https://doi.org/10.5194/tc-14-1519-2020, 2020.
- 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, https://doi.org/10.1175/1520-0442(1999)012j1814:SDOASI¿2.0.CO;2, 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, URL http://doi.wiley.com/10.1002/2014JC009985, 2014.
- Willatt, R., Laxon, S., Giles, K., Cullen, R., Haas, C., and Helm, V.: Ku-band radar penetration into snow cover on Arctic sea ice using airborne data, Annals of Glaciology, 52, 197–205, https://doi.org/ 10.3189/172756411795931589, 2011.
- Willatt, R. C., Giles, K. A., Laxon, S. W., Stone-Drake, L., and Worby, A. P.: Field investigations of Ku-band radar penetration into snow cover on antarctic sea ice, IEEE Transactions on Geoscience and Remote Sensing, 48, 365– 372, https://doi.org/10.1109/TGRS.2009.2028237, URL http://ieeexplore.ieee.org/document/5282596/, 2010.

# Faster decline and higher variability in the sea ice thickness of the marginal Arctic seas when accounting for dynamic snow cover

Robbie D.C. Mallett <sup>1</sup>, Julienne C. Stroeve <sup>1,2,3</sup>, Michel Tsamados <sup>1</sup>, Jack C. Landy <sup>4</sup>, Rosemary Willatt <sup>1</sup>, Vishnu Nandan <sup>3</sup>, and Glen E. Liston <sup>5</sup>

<sup>1</sup>Centre for Polar Observation and Modelling, Earth Sciences, University College London, London, UK

<sup>2</sup>National Snow and Ice Data Center, University of Colorado, Boulder, CO, USA

<sup>3</sup>Centre for Earth Observation Science, University of Manitoba, Winnipeg, Canada

<sup>4</sup>School of Geographical Sciences, University of Bristol, Bristol, UK

<sup>5</sup>Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

**Correspondence:** Robbie Mallett (robbie.mallett.17@ucl.ac.uk)

**Abstract.** Mean sea ice thickness is a sensitive indicator of Arctic climate change and in long-term decline despite significant interannual variability. Current thickness estimations from satellite radar altimeters employ a snow climatology for converting range measurements to sea ice thickness, but this introduces unrealistically low interannual variability and trends. When the sea ice thickness in the period 2002-2018 is calculated using new snow data with more realistic variability and trends, we find

- 5 mean sea ice thickness in four of the seven marginal seas to be declining between 60-100% faster than when calculated with the conventional climatology. When analysed as an aggregate area, the mean sea ice thickness in the marginal seas is in statistically significant decline for six of seven winter months. This is observed despite a 76% increase in interannual variability between the methods in the same time period. On a seasonal timescale we find that snow data exerts an increasingly strong control on thickness variability over the growth season, contributing 46% in October but 70% by April. Higher variability and faster
- 10 decline in the sea ice thickness of the marginal seas has wide implications for our understanding of the polar climate system and our predictions for its change.

### 1 Introduction

15

Sea ice cover moderates the exchange of moisture, heat and momentum between the atmosphere and the polar oceans, influencing regional ecosystems, hemispheric weather patterns and global climate. Sea ice thickness (SIT) is a key characteristic of the sea ice cover, as thicker sea ice weakens the coupling between the ocean and atmosphere systems.

Thicker sea ice is more thermally insulating and limits heat transfer from the ocean to the atmosphere in winter and consequent thermodynamic growth (Petty et al., 2018a). SIT also exerts control on sea ice dynamics and rheology (Tsamados et al., 2013; Vella and Wettlaufer, 2008). The thickness of sea ice during snow accumulation also dictates whether the sea ice surface drops below the waterline, potentially increasing thermodynamic sea ice growth through the formation of snow-ice

20 (Rösel et al., 2018). The impact of the end-of-winter SIT distribution persists into the melt season with thick sea ice decreasing the transmission of solar radiation to the surface ocean and reducing the potential for in- and under-ice primary productivity

(Mundy et al., 2005; Katlein et al., 2015). Finally, thick sea ice is far more likely to survive the melt season, increasing the average age of Arctic sea ice. Correct assimilation of ice thickness into models therefore offers opportunities for prediction of the sea ice state on seasonal timescales (Chevallier and Salas-Mélia, 2012; Blockley and Peterson, 2018; Schröder et al., 2019).

- The annual sea ice thickness distribution is highly spatially variable, with a cover of thick multi-year ice in the Central Arctic and a thinner, more seasonally variable cover of first year ice in the marginal seas. Regional sea ice thickness distributions are often characterised by the mean thickness,  $\overline{SIT}$ . As well as being a key parameter for the processes described above, the value can be multiplied by the sea ice area to produce the sea ice volume, one of the most sensitive indicators of Arctic climate change (Schweiger et al., 2019).
  - 30 While continuous and consistent monitoring of Pan-Arctic SIT has not been achieved on a multi-decadal timescale, a combination of different techniques has suggested a significant decline in thickness since 1950 (Kwok, 2018; Stroeve and Notz, 2018). Satellite altimeters using both radar and lidar have provided a valuable record of changing sea ice thickness, but have often been limited for various reasons. Some have been limited spatially by their orbital inclination (e.g. the European Remote Sensing (ERS) satellites, Envisat, AltiKa and Sentinel radar altimeters have operated up to only 81.5 degrees north), and
  - 35 others in temporal coverage (e.g. ICESat was operated in 'campaign mode' rather than providing continuous coverage). Two satellite altimeters currently offer continuous and meaningfully Pan-Arctic monitoring of the Arctic sea ice: the ICESat-2 and CryoSat-2 altimeters. ICESat-2 has been in operation since September 2018 and so far has documented only two winters of sea ice thickness (Kwok et al., 2020).
  - Although the launch of the CryoSat-2 radar altimeter (henceforth CS2) in 2010 allowed significant advances in understanding
    the spatial distribution and interannual variability of Pan-Arctic SIT (Laxon et al., 2013), a statistically significant decreasing trend within the CS2 observational period has not been detected for the Arctic as a whole. The lack of certainty regarding any trend in SIT is in part due to the various uncertainties associated with SIT retrieval from radar altimetry (Ricker et al., 2014; Zygmuntowska et al., 2014). Major contributors to these uncertainties are the relatively large footprint of a radar pulse when compared to laser altimetry, the variable density of sea ice, retracking of radar returns from rough sea ice, and the need for an *a priori* snow depth and density distribution (Kern et al., 2015; Landy et al., 2020).

The impact of snow-depth uncertainty on SIT retrievals was recently included by the IPCC in a list of 'Key Knowledge Gaps and Uncertainties' (Meredith et al., 2019). More specifically, Bunzel et al. (2018) found snow to have a strong influence on the interannual variability of SIT and consequent detection of thickness trends. Here we investigate the impact of a new, Pan-Arctic snow depth and density data set (SnowModel-LG; Liston et al., 2020; Stroeve et al., 2020) on trends and variability

50 in regional  $\overline{SIT}$  when used in place of the traditional, climatological data set (Warren et al., 1999). We show that traditional calculations of  $\overline{SIT}$  omit significant interannual variability due to their reliance on a snow climatology, and we quantify this omission. We also show that sea ice is likely thinning at a faster rate in some marginal seas than previously thought, because the snow water equivalent on the sea ice is declining too.

#### 1.1 The Role of Snow in Radar-Altimetry Derived Sea Ice Thickness Retrievals

- 55 Satellite radar altimetry involves the emission of radar pulses from a satellite and the subsequent detection of their backscatter. The time difference between emission and detection ('time of flight') corresponds to the distance traveled and thus the height of the transmitter above the scattering surface. Radar altimeters of different frequencies have been carried on board several earth observation satellites such as ERS-1/2, Envisat, AltiKa, CryoSat-2 and Sentinel-3A/B (Quartly et al., 2019). We now quantify the role of snow cover in conventional sea ice thickness estimation, before revealing and explaining the effects of previously
- 60 unincorporated trends and variability.

The Ku-band radar waves emitted from CryoSat-2 are generally assumed in mainstream SIT products to scatter from the snow/sea-ice interface (Kurtz et al., 2014; Tilling et al., 2018; Hendricks and Ricker, 2019; Landy et al., 2020). The difference in radar ranging (derived from time-of-flight) between areas of open water and areas of sea ice is known as the 'radar freeboard',  $f_r$ . The height of the sea ice surface above the waterline is referred to as the sea ice freeboard,  $f_i$ . This is extracted from the radar freeboard through (a) assuming that the primary scattering horizon corresponds to the sea ice surface, and (b) accounting for the slower radar wave propagation through the snow cover above the sea ice surface (Armitage and Ridout, 2015; Mallett et al., 2020). The sea ice freeboard can then be converted to sea ice thickness by considering the floe's hydrostatic equilibrium

given the sea ice density and weight of overlying snow. In the simplified case of bare sea ice, we would calculate:

$$SIT_{bare} = f_r \frac{\rho_w}{\rho_w - \rho_i} \tag{1}$$

70

65

Where  $\rho_w$  is the density of seawater and  $\rho_i$  the density of sea ice. In order to adjust the above equation for the presence of overlying snow, the twin effects of the snow's weight and the snow's delaying influence on radar pulse propagation must be taken into effect. These three influences on SIT (the radar freeboard measurement, the pulse propagation delay and the freeboard depression from snow weight) can therefore be expressed as three terms (see supplementary information) in the following way:

75 
$$SIT = h_r \frac{\rho_w}{\rho_w - \rho_i} + h_s \frac{\rho_w}{\rho_w - \rho_i} \left[\frac{c}{c_s} - 1\right] + h_s \frac{\rho_s}{\rho_w - \rho_i}$$
(2)

In this manuscript we introduce a simple method for combining the second and third terms of the above equation into a single term that is proportional to the snow water equivalent (Sect. 3.1). This helps to easily separate the influences of snow data and radar freeboard measurements on the determination of sea ice thickness. Specifically, we compare the impact of two snow products on regional trends and variability in sea ice thickness. These products are the snow climatology given by Warren

80 et al. (1999) and the output of SnowModel-LG (Liston et al., 2020; Stroeve et al., 2020).

#### 2 Data Description

#### 2.1 Regional Mask

We define six regions of the Arctic Basin using the mask from Stroeve et al. (2014) which is gridded onto a 25 km resolution EASE grid (Brodzik et al., 2012, Fig. 1). We define the 'marginal seas' of the Arctic Basin as the color coded areas of Fig. (1)
excluding the Central Arctic. All constituent regions of the 'marginal seas' grouping lie within the coverage of Envisat barring a negligible portion of the Laptev Sea.



**Figure 1.** The definitions of the marginal Arctic seas used in this paper, from Stroeve et al. (2014). Two black, concentric circles indicate the latitudinal limits of the CryoSat-2 (inner circle;  $88^{\circ}N$ ) and Envisat (outer circle;  $82.5^{\circ}N$ ) missions.

#### 2.2 Radar Freeboard Data

To examine the impact of snow products on Enivsat/CryoSat-2 thickness retrievals, we used radar freeboard data from the ESA Sea Ice Climate Change Initiative (Hendricks et al., 2018). This data is available from October in the winter of 2002/03 until April in the winter of 2016/17. This product was chosen for two main reasons: (a) it provides a consistent record for both the Envisat and CS2 missions (Paul et al., 2018), and (b) it is publicly available for download. CS2 carries a delay-Doppler

90

altimeter that significantly enhances along-track resolution by creating a synthetic aperture. For this reason as well as its higher latitudinal limit, we used CS2 radar freeboard measurements over Envisat during the period when the missions overlapped (November 2010 - March 2012). To create a radar freeboard product that is consistent between the Envisat and CS2 missions,

95 Envisat returns are retracked using a variable threshold retracking algorithm. This variable threshold is calculated from the strength of the surface backscatter and the width of the leading edge of the return waveform such that the inter-mission bias is minimised (Paul et al., 2018). The results are comprehensively analysed in the Product Validation & Intercomparison Report (ESA, 2018). One key finding of this report is that while Envisat radar freeboards are calculated so as to match CS2 freeboards during the period of overlap over the whole Arctic basin, there are biases over ice types. In particular, Envisat ice freeboards

(not radar freeboards) are biased 2-3 cm low (relative to CS2) in areas dominated by MYI, and 2-3 cm high in areas dominated

100

While the ESA CCI data are only available from the CCI website until the winter of 2016/17, the CryoSat-2 radar freeboards in this data are identical to the CS2 radar freeboard product of the Alfred Wegener Institute (Hendricks and Ricker, 2019, this was manually confirmed). We were therefore able to extend our radar freeboard timeseries through the winter of 2017/18,

#### 105

All radar freeboard data used in this study are supplied on a 25 km EASE grid (Brodzik et al., 2012), the same as that of SnowModel-LG.

#### 2.3 The Warren Climatology (W99)

by FYI. We discuss the implications of these biases in Sect. 5.3.

which is when our snow data from SnowModel-LG (see below) ends.

The most commonly used radar-altimetry SIT products use algorithms developed by the Centre for Polar Observation and 110 Modelling, the Alfred Wegener Institute and the NASA Goddard Space Flight Centre (Tilling et al., 2018; Hendricks and Ricker, 2019; Kurtz et al., 2014). Another commonly used but not publicly available product is from the NASA Jet Propulsion Laboratory (Kwok and Cunningham, 2015). All four groups utilize modified forms of the snow climatology assembled by Warren et al. (1999) from the observations of Soviet drifting stations between 1954 and 1991 (henceforth referred to as W99). While the consistent use of W99 for sea ice thickness calculation is convenient for intercomparison of products (e.g. Sallila

115 et al., 2019; Landy et al., 2020), the data have a number of drawbacks. This work is centered around two key issues with the use of W99 for SIT retrieval: inadequate representation of interannual variability and trends.

The Warren Climatology includes quadratic fits for every month of snow water equivalent and snow depth. We projected these fits over the  $361 \times 361$  EASE grid (for combination with our radar freeboard data and comparison with SnowModel-LG) to create SWE and depth distributions across the Arctic basin as defined in Sect. (2.1).

#### 120 2.3.1 Drifting Station Coverage Illustration

At this point it is instructive to briefly illustrate the coverage of the drifting stations from which W99 was compiled. We analysed position and snow depth data from the twenty-eight drifting stations that contributed to W99 (Fig. 2a). It is clear that the vast majority of these operated in the Central Arctic or in the East Siberian Sea, with very little sampling done in most other marginal seas. But while these tracks illustrate the movements of the drifting stations, it is important to note that the stations


**Figure 2.** (a) tracks of Soviet drifting stations 3 - 31. (b) Number of days in each region in each month that snow stake measurements were taken.

- 125 were not always collecting snow data which would contribute to the W99 climatology. To assess the spatial distribution of snow sampling, we cross-referenced the position data with days on which the drifting stations recorded the snow depth at their measuring stakes. We then calculated the number of 'measurement-days' in each region-month combination (Fig. 2b). We note that when two drifting stations were operating at the same day, we count this as two distinct days (as they were rarely so close together so as to collect redundant data).
- 130 This reveals that no snow measurements were taken in the Barents and Kara Seas, and none in the Laptev Sea for four of the seven winter months. While 'snow-line' transect data also contributed to W99 (and indeed was used in preference to stake data where possible), we find that snow-line data was overwhelmingly collected on days where stake-data was also collected.

Figure 2 illustrates that the quadratic fits of W99 are not appropriate for use in several of the marginals seas. However we note that a number of authors have still used the climatology for sea ice thickness retrievals in these regions, often in the course

135 of sea ice volume calculations (e.g. Laxon et al., 2003, 2013; Tilling et al., 2015, 2018; Sallila et al., 2019; Li et al., 2020a; Belter et al., 2020; Li et al., 2020b). We therefore consider these regions in this manuscript, but with the understanding that mW99 is likely not representative of the snow conditions.

#### 2.4 The modified Warren Climatology (mW99)

The W99 climatology is by definition invariant from year to year, and was implemented in this way by Laxon et al. (2003) and
Giles et al. (2008a) to estimate sea ice thickness using ERS 1 & 2. When implemented like this, the amount of snow on sea ice exhibits no interannual variability.

The implementation of W99 was then modified by Laxon et al. (2013) based on the results of Operation IceBridge flights which showed reduced snow depth over first year ice (FYI; Kurtz and Farrell, 2011). This implementation, known as 'mW99',

consists of halving snow depths over first year ice with snow density kept constant. Because the areal fraction and spatial

145 distribution of FYI changes from year to year, this modification introduces a small degree of interannual variability into the contribution of snow data to sea ice thickness. We investigate this in Sect. (4.1.1).

#### 2.5 Ice Type Data

155

Sea ice type data is required to modify W99 and create mW99. One popular product for this (e.g. Tilling et al., 2018; Hendricks and Ricker, 2019) is an operational product from the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF, www.osi-saf.org). However, this data series begins in March 2005. This is after our study begins (in October 2002).

A similar product exists, published by the Copernicus Climate Data Store (CDS, www.cds.climate.copernicus.eu; Aaboe, 2020). This product's underlying algorithm is adopted from the OSI SAF processing chain, but has been modified to produce a consistent record compatible with reanalysis. Furthermore, the CDS product only assimilates information from 'passive' satellite radiometers, whereas the OSI SAF operational product assimilates additional data from 'active' scatterometers. Despite these differences, a brief comparison of the products reveals a high degree of similarity.

It would be possible to use the CDS product prior to the beginning of the OSISAF product in 2005, but this approach raises issues surrounding the continuity of the products across the 2005 transition. Since our investigation focusses on trends and

variability, we prioritise a consistent record and opt to use the CDS ice type product for the entirety of our study.

Both ice type products occasionally include pixels of ambiguously classified ice. We implemented a very simple interpolation strategy to classify these points while creating our mW99 data, although they are rarely present in winter within the regions analysed in this paper. Where the ambiguous pixels are generally surrounded by a given ice type then they are classified as the surrounding type. In the case where the ambiguous classification is on the boundary between the two types, the snow depth was not divided by two.

#### 2.6 SnowModel-LG

165 To investigate the impact of variability and trends in snow cover on regional sea ice thickness we use the results of SnowModel-LG (Liston et al., 2020; Stroeve et al., 2020). SnowModel-LG is a Lagrangian model for snow accumulation over sea ice; the model is capable of assimilating meteorological data from different atmospheric reanalyses (see below) and combines them with sea ice motion vectors to generate pan-Arctic snow depth and density distributions. The sea ice motion vectors used were from the Polar Pathfinder dataset at weekly time resolution (Tschudi et al., 2020). SnowModel-LG exhibits more significant interannual variability than mW99 in its output because it reflects year to year variations in weather and sea ice dynamics.

SnowModel-LG includes a relatively advanced degree of physics in its modelling of winter snow accumulation. The model creates and merges layers based on precipitation and snowpack metamorphism. The effects of sublimation, depth-hoar formation and wind-packing are included. However, the effects of snow loss to leads by wind and extra snow accumulation due to sea ice roughness are not included. Furthermore, the heat flux to the snow is not sensitive to the thickness of the underlying sea 175 ice.

SnowModel-LG creates a snow distribution based on reanalysis data, and the accuracy of this snow data is unlikely to exceed the accuracy of the input. There is significant spread in the representation of the actual distribution of relevant meteorological parameters by atmospheric reanalyses (Boisvert et al., 2018; Barrett et al., 2020). The results of SnowModel-LG therefore depend on the reanalysis data set used. However, the data product used has been tuned to match Operation Ice Bridge derived

- 180 snow depths during spring time, and snow depth differences between the reanalysis products were found to be less than 5 cm (Stroeve et al., 2020). We note that the vast majority are over the Beaufort Sea and the Greenlandic side of the Central Arctic, which is generally covered by multivear ice. It is therefore conceivable that the scaling factors would be different if FYI were better sampled by OIB. The time-averaged regional differences between SnowModel-LG runs forced by ERA5 and MERRA2 reanalysis data are shown in Fig. (S3). The SnowModel-LG data used in this study are generated from the average
  - of SnowModel-LG runs forced by the two reanalysis products. The SnowModel-LG data is provided on the same 25 km EASE 185 grid as the ESA-CCI radar freeboards described above at daily time resolution. We averaged this daily product to produce monthly gridded data for combination with the monthly radar freeboard data.

#### 2.7 NASA Eulerian Snow on Sea Ice Model (NESOSIM)

To support and broaden the impact of our findings, we repeat our analyses with NESOSIM data from 2002-2015 (Petty 190 et al., 2018b). NESOSIM data is released on a  $100 \times 100$  km grid which was interpolated to the  $25 \times 25$ km EASE grid of the SnowModel-LG and radar freeboard data. NESOSIM runs in a Eulerian framework and like SnowModel-LG can assimilate precipitation data from a variety of reanalyses data. In contrast with SnowModel-LG's multi-layered scheme, NESOSIM uses a two-layer snow scheme to represent depth-hoar and wind-packed layers. To define these layers, it assimilates surface winds and temperature profiles from reanalysis. Wind-blown snow loss is parameterised to leads using daily sea ice concentration fields (Comiso, 2000, updated 2017).

```
195
```

In this study we use data from a NESOSIM run initialised on the 15th August for each year. The initial depth was produced by a 'near-surface air-temperature-based scaling of the August W99 snow depth climatology'. This is a linear scaling based on the duration of the preceeding summer melt season. Snow density was initialised using the August snow-line observations of Soviet NP drifting stations 25, 26, 30 and 31. Data from the most recent publicly available drifting stations were chosen to maximise their relevance in a changing climate.

## 3 Methods

215

#### 3.1 Contributions to thickness determination from snow and radar freeboard data

We now identify that the height correction due to slower radar pulse propagation in snow scales in almost direct proportion to the total mass of penetrated snow (m<sub>s</sub>; Fig. S1). As such, it can be easily combined with the change to the floe's hydrostatic
equilibrium from snow loading (also linearly dependent on m<sub>s</sub>) to make one transformation to modify Eq. (2) such that:

$$SIT = f_r \frac{\rho_w}{\rho_w - \rho_i} + m_s \frac{\rho_w}{\rho_w - \rho_i} \times 1.81 \times 10^{-3}$$
(3)

Physically, the first term of Eq. (3) corresponds to the SIT were the sea ice known to have no snow cover. The second term is the additional sea ice thickness that is inferred from knowledge of the overlying snow cover. SIT has been decomposed into linearly independent contributions from radar-freeboard data and snow data. This allows the contributions of the two data components to SIT to be assessed independently. A derivation of the  $1.81 \times 10^{-3}$  coefficient is available in the supplementary material.

We highlight here that our expression in Eq. (3) of the contribution of snow data to SIT determination solely in terms of snow mass is technically convenient for using W99 to estimate sea ice thickness, as quadratic fits of density (unlike depth and snow water equivalent) are not publicly available for all months. This has led to the required density values often being set to a constant value or 'backed out' by dividing the published SWE distributions by the depth distributions.

Eq. (3) and its factor of  $1.81 \times 10^{-3}$  allow the simple expression of the theoretical change to the radar freeboard under rapid snow accumulation or removal. Making  $f_r$  the subject of the equation and assuming SIT constant we find:

$$\frac{\partial f_r}{\partial m_s} = -1.81 \times 10^{-3} \quad (m/kgm^{-2}) \tag{4}$$

We stress that the above equation assumes total radar penetration of overlying snow, an assumption discussed in Sect. (5.3.1). 220 As well as allowing independent analysis of the radar and snow data contributions to SIT at a point, the linearly independent nature of Eq. (3) in terms of  $f_r$  and  $m_s$  allows for a simple calculation of the average SIT in a region ( $\overline{SIT}$ ) as:

$$\overline{SIT} = \overline{RF} + \overline{Snow} \tag{5}$$

Where  $\overline{RF}$  and  $\overline{Snow}$  represent the spatial averages of the first and second terms of Eq. (3).

#### 3.2 Assessing Snow Trends and Variability at a point

In Sect. (4.1) we briefly compare the statistics for trends and variability at drifting stations published in Warren et al. (1999) with those introduced by mW99 and SnowModel-LG at a given point. We carry out this analysis to establish that the mW99 variability and trends at a given point (chosen as pixels on a 25x25 km EASE grid) are considerably smaller than those observed at drifting stations.

The monthly interannual variability (IAV) values published in Warren et al. (1999) are calculated as the standard deviation of the snow depths at drifting stations when compared to the climatology at the position of the stations. The IAV values at 230 a point-like drifting station in a region will therefore naturally be higher than the IAV of the region's spatial-mean. As such, to compare IAV values from point-like drifting stations to mW99, we calculate the IAV at individual ice-covered points on a  $25 \times 25$  km equal-area grid (Brodzik et al., 2012). These are all positive values, which we then average for comparison with the drifting stations. By regionally averaging the IAV values of many points rather than calculating the IAV of regional averages, 235 we replicate the statistics of the point-like drifting stations.

However, the main part of this paper does not focus on trends and variability at a point (as measured by drifting stations), but instead investigates trends and variability in  $\overline{Snow}$  and  $\overline{SIT}$  at the regional scale (Sections 4.2 & 4.3). This variability is significantly lower than the typical variability at a point, as many local anomalies from climatology within a region are averaged out in the calculation of single, area-mean values which form a timeseries for each region.

#### 240 3.3 Assessing Regional Interannual Variability

Sect. (4.2) of this paper focuses on the interannual variability in regional  $\overline{SIT}$  which (treating  $\overline{RF}$  and  $\overline{Snow}$  as random, dependent variables) can be expressed thus:

$$\sigma_{\overline{SIT}}^2 = \sigma_{\overline{RF}}^2 + \sigma_{\overline{Snow}}^2 + 2\operatorname{Cov}(\overline{RF}, \overline{Snow}) \tag{6}$$

Where the final term represents the covariance between spatially averaged radar freeboard and snow contributions. This covariance term can be expressed as  $2r \times \sigma_{\overline{Snow}} \times \sigma_{\overline{RF}}$ , where r is the dimensionless correlation-coefficient between the variables 245 and ranges between -1 and 1. To further explain this term, if years of high  $\overline{RF}$  are correlated with high  $\overline{Snow}$ , then the covariance term will be high and interannual variability in  $\overline{SIT}$  will be amplified. If mean snow depths are anti-correlated with mean radar freeboard across the years, interannual variability in  $\overline{SIT}$  will be reduced.

 $\overline{SIT}$ ,  $\overline{RF}$  &  $\overline{Snow}$  were calculated where any valid grid points existed on the 25x25 km EASE grid. Because of this, 250 no average values were computed in the Kara Sea in October 2009 or 2012. Furthermore, no October values were generally available in the Barents Sea after 2008 (with the exception of 2011 and 2014). The impact of this on our resulting analysis is clearly visible in the top left panel of Fig. (10). We do not exclude the Barents Sea in October from our analysis because of the low number of valid points, but we do highlight the undersampling issue here. We continue to consider it because we do not find statistically significant declining trends with the data we have, so essentially we are reporting a null result. Our 255 calculations of interannual variability in this month is inherently adjusted for the small sample size, but we nonetheless urge caution in interpretation of the values. The number of grid points available for averaging in each region in each month are shown in Fig. (S2).

The three terms on the right hand side of Eq. (6) correspond to the three unique terms of the covariance matrix of the two terms of Eq. (5). The main-diagonal elements of this  $2 \times 2$  matrix correspond to the variance of the snow contribution and the radar freeboard contribution to sea ice thickness, terms one and two of Eq. (6). The off-diagonal elements are identical and

260 sum to form the third term of Eq. (6). We calculated this matrix for each region in each month to investigate the sources of regional interannual variability in  $\overline{SIT}$  for the time period under consideration (2002-2018). The Central Arctic region is not sufficiently well observed by the Envisat radar altimeter (see Fig. 1), so the covariance matrix for the region was only calculated for the CS2 period (2010-2018).

265

In some cases a natural degree of covariance is introduced between the regional  $\overline{Snow}$  and  $\overline{RF}$  timeseries because they both display a decreasing trend. This 'false-variance' would not be present were the system in a steady state. As such, we detrended the regional timeseries prior to calculation of the covariance matrix. We found that doing this significantly decreased the value of the covariance term in Eq. (6).

We consider the relative contributions of these three terms to  $\sigma_{\overline{SIT}}^2$  in calculations involving mW99 and SnowModel-LG (Sect. 4.2). In light of these results, we then re-assess the statistical significance of regional trends in  $\overline{SIT}$  using SnowModel-

LG.

275

Detection of temporal trends in  $\overline{SIT}$  is critically dependent on accurate characterisation of  $\sigma_{\overline{SIT}}^2$ . This is because conventional tests for trend exploit the known probability of a system with no trend generating the data at hand through variability alone (Chandler and Scott, 2011, p. 61). In this paper we argue that the  $\sigma_{\overline{Snow}}^2$  term of Eq. (6) has been systematically underestimated through the use of a quasi-climatological snow data set (mW99). As an alternative to this we use the results of SnowModel-LG, a snow accumulation model that incorporates interannual changes in precipitation amount, freeze-up timing

## 3.4 Assessing Regional Temporal Trends

and sea ice distribution.

In Sect (4.3) we examine temporal trends in regional  $\overline{SIT}$  for each month of the growth-season (October - April), and decompose the results by sea ice type. It is stressed that these regional trends are each the trend of a single timeseries of spatially averaged thickness values, rather than the average of many trends in sea ice thickness at various pixels in a region. Regional trends were deemed statistically significant if they passed a two-tailed hypothesis test with p-value less than 0.05, with a null hypothesis of no trend. Trends were calculated for regional  $\overline{SIT}$  over the Envisat-CS2 period (2002-2018) for all regions apart from the Central Arctic for which only CS2 data was available. We assess the relationship of these trends in  $\overline{SIT}$  to trends in  $\overline{RF}$  and  $\overline{Snow}$  (Fig S19).

In Sect. (4.1.2) we show that basin-wide average snow depth and SWE is decreasing in SnowModel-LG in most months, but only in October for mW99. We point out here that (under the paradigm of total radar wave penetration of snow on sea ice) under-accounting for potential reductions in SWE may partially mask a decline in sea ice thickness, as reductions in radar freeboards are partially compensated by reductions in snow depths. From Eq. (5):

$$290 \quad \frac{\partial(\overline{SIT})}{\partial t} = \frac{\partial(\overline{RF})}{\partial t} + \frac{\partial(\overline{Snow})}{\partial t} \tag{7}$$

## 4 Results

#### 4.1 Comparison of point-trends and point-variability

#### 4.1.1 Low interannual variability in mW99 compared to drifting stations and SnowModel-LG

295

How does the variability in mW99 and SnowModel-LG at a given point compare to the values recorded at Soviet drifting stations published by Warren et al. (1999)? These values for interannual variability are not currently used in sea ice thickness retrievals (although they do contribute to uncertainty estimates in the ESA-CCI sea ice thickness product). Nonetheless, they offer a benchmark against which to evaluate the variability induced by mW99 at a given location.

Using the method described in Sect. 3.2 we find that the snow variability at a point from mW99 (Fig. 3, blue bars) is on average about 50% of the values recorded at the drifting stations (Fig. 3, green bars). By comparison, SnowModel-LG snow depth variability at a given point is significantly higher, ranging from  $\sim$ 75% of the drifting station values in October to  $\sim$ 115%

300

by the end of winter.

We present this analysis of the point-like snow variability to illustrate that mW99 does not introduce enough variability at a given point to match that observed at drifting stations from year to year. Furthermore, the variability that does exist is confined to a distinct band of the Arctic Ocean (Fig. 4). This band represents areas where the sea ice type is not typically either FYI or

305 MYI. Instead it is either switching between the two, or it is an area where FYI has replaced MYI during the period of analysis. In areas where sea ice type is temporally unchanging, snow variability is not present. This has implications at the regional scale as marginal seas with a consistent sea ice type experience unrealistically low  $\sigma_{\overline{Snow}}$  in the mW99 scheme.



Figure 3. Interannual variability (2002-2018) in snow depth from mW99 and SnowModel-LG compared to the values given in Table 1 of Warren et al. (1999).



**Figure 4.** mW99 snow depth variability at each EASE grid point over the 2002-2018 period. This is calculated by generating a timeseries of snow depth at each point and then calculating the standard deviation of that timeseries. High variability is displayed in a band where sea ice type typically fluctuates from year to year. IAV is zero in areas that do not exhibit sea ice type variability, introducing unphysically low variability in SIT.

#### 4.1.2 Lack of Temporal Trends in mW99 compared to SnowModel-LG and in-situ data

Weak trends exist at some points in the mW99 Arctic snow distribution due to the shifting distribution and abundance of first

- year ice in the Arctic. In this section we briefly address their size, sign and veracity, leaving regional analysis until Sect. (4.2). 310 Values for SWE and depth trends measured by individual drifting stations are given in W99, but the values are not statistically significant for any of the winter months, and as such are not displayed here. We instead compare the point-trends at all color coded regions of Fig. 1 from mW99 and SnowModel-LG (Fig. 5).
- We find that when we average the point-trends at a basin-wide scale, the only statistically significant trend (at the 5% level) 315 for mW99 snow depth is a positive one for the month of October (+0.11 cm/yr; Fig. 5). This increasing trend in snow depth is in part due to the diminishing area of October FYI relative to that of MYI (Fig. S4), and in part due to the retreat of the October sea ice into the Central Arctic where W99 exhibits higher snow depths and SWE. The increasing October areal dominance of MYI is in part driven by delayed Arctic freeze-up (Markus et al., 2009; Stroeve et al., 2014). The area of sea ice over which the W99 climatology is halved in October is therefore shrinking, and basin-wide mean snow depths in mW99 are increasing.
- 320 Trends in sea ice type fraction for each winter month are displayed in Fig. (S4), and monthly timeseries for mW99 SWE are displayed in Fig. (S5).

Unlike mW99, SnowModel-LG exhibits statistically significant, negative point-trends for the later five of the seven winter months (when averaged at a basin-wide scale). We identify two processes as responsible for this decreasing trend: the MYI area is shrinking, so a smaller MYI sea ice area is present during during the high snowfall months of September and October

- 325 (Boisvert et al., 2018); also freeze-up commences later, so a lower FYI area is available in these months and more precipitation falls directly into the ocean. Webster et al. (2014) observed a -0.29cm/yr trend in Western Arctic spring snow depths using both airborne and *in situ* sources. This airborne contributions to this statistic included data over both sea ice types, and the in-situ contributions included data from individual Soviet drifting stations from the Western Arctic. The statistic compares well with the behaviour of SnowModel-LG (-0.27 cm/yr March; -0.31 cm/yr April), but is considerably beyond that of the non-statistically significant trends of W99 and mW99.
- 330

What might the effects of this decline be on SIT at regional scales and larger? In terms of Eq. (7), models and observations indicate that  $\partial(\overline{Snow})/\partial t$  is negative on long timescales (Webster et al., 2014; Warren et al., 1999; Stroeve et al., 2020). However, the use of mW99 effectively sets  $\partial(\overline{Snow})/\partial t$  to zero, and to a positive value in October. This has the effect of biasing  $\partial(\overline{SIT})/\partial t$  high (and towards zero). Section 4.3 examines the effect of using SWE data with a more realistic decline on regional  $\overline{SIT}$  trends; this is mediated by the effects of higher interannual variability, which is examined in Sect. (4.2).



Figure 5. Basin-wide spatial average of point-like trends in (a) snow depth and (b) SWE, from mW99 and SnowModel-LG. Calculated for the Envisat-CS2 period (2002-2018). Significance values (in %) are given at the base of each bar. Only October trends for mW99 are significant at the 5% level, whereas significant negative trends exist in SnowModel-LG for December - April.

#### 4.2 **Realistic SWE Interannual Variability Enhances Regional SIT Interannual Variability**

Having illustrated the deficiency of point-trends and point-variability in mW99, we now move on to the impact of snow data on SIT at the regional scale.

340

- We calculate the interannual variability of detrended timeseries of the snow contribution to the thickness determination (Snow) from mW99 and SnowModel-LG. We display some of these results in Fig. (6). We did this for every winter month (Oct-Apr) and for in each region defined in Fig. (1). SnowModel-LG data produce more variable timeseries of *Snow* (i.e. higher values of  $\sigma_{\overline{Snow}}^2$ ; c.f. Eq. 6). This is the case for all months, in all regions. For snow in the Kara Sea, mW99 introduces almost four times less interannual variability into SIT via  $\overline{Snow}$  than SnowModel-LG in the April timeseries. This analysis is further broken down by sea ice type in Figs S7 and S8.
- Having shown that SnowModel-LG's contribution to  $\overline{SIT}$  is more variable than mW99, how does this increased variability 345 propagate into sea ice thickness variability itself  $(\sigma_{\overline{SUT}}^2)$ ? To answer this question, we must examine the way in which the snow contribution to SIT combines with data from satellite radar freeboard measurements. Having calculated the  $\sigma_{\overline{Snow}}^2$  term of Eq. 6 (displayed in Fig. 6), we now turn to the  $2 \operatorname{Cov}(\overline{RF}, \overline{Snow})$  term. To assess this we calculate the magnitude and statistical significance of correlations between the detrended  $\overline{RF}$  and  $\overline{Snow}$  contributions to  $\overline{SIT}$  in individual years, regions and months.
- 350

To do this we calculated a monthly timeseries of  $\overline{RF}$  and  $\overline{Snow}$  for each region over the time-periods (2002-2018, with the Central Arctic being 2010-2018). Because we considered eight regions and seven months, this led to to 56 pairs of timeseries for  $\overline{RF}$  and  $\overline{Snow}$ . We then detrended each of them. We then calculated the correlation between each of the pairs of detrended timeseries. We note here that the correlation between the timeseries is dependent on their relative position to a linear regression.



**Figure 6.** Detrended timeseries of spatially averaged snow contributions to sea ice thickness  $(\overline{Snow})$  by region from W99 (blue) and SnowModel-LG (red). Standard deviation values are displayed for SnowModel-LG (lower left, red), and mW99 (lower right, blue). All regions are plotted in Supplementary Fig. (S6)

355 These correlation statistics are thus independent of the absolute magnitude of the values, their units, or any linear scaling of the axes. We therefore choose to present the correlations in Fig (7) without axes and scaled to the rectangular panels, so as to best show the relative positions of the points without extraneous numerical information.

We find statistically significant correlations between  $\overline{Snow}$  and  $\overline{RF}$  to generally range between 0.6 - 0.85 (Fig. 7). All statistically significant correlations were positive ones, and this was also the case when individual sea ice types were considered

- 360 for each region. When all sea ice types were considered, the Laptev and East Siberian seas exhibited statistically significant trends in five and six of the seven growth-season months respectively. The Barents Sea and the Beaufort Sea both exhibited one month of correlation, and the Central Arctic Region exhibited no months of correlation the reasons for this are discussed in Sect. (5.4). When analysed as a single, large region, the 'Marginal Seas' area exhibits correlations in four of the seven months analysed, with the strength of these correlations increasing over the season.
- 365

We continued this analysis by breaking down the regions by sea ice type. The area of the Central Arctic sea ice covered with first year ice exhibits strong correlations (all above 0.8) in the later five months of the winter (Fig. S9).

When considering correlations over multi-year ice (MYI), the 'Marginal Seas' grouping exhibits correlations in the first four growth-season months (Fig. S10). The MYI fraction Central Arctic, Chukchi and Barents Seas exhibited no correlations. We note that this analysis is relatively sensitive to the detrending process. When performed without detrending, statistically



Detrended SnowModel-LG Snow Contribution to SIT

**Figure 7.** Covariability of contributions to sea ice thickness from radar freeboard and SnowModel-LG derived snow components over all sea ice types. Plots are colored with magenta when a a statistically significant correlation is present between the contributions (p>0.95). Analogous plots are displayed for the FYI and MYI components of the regions in Figs S9 & S10.

370 significant correlations are noticeably more common. This is because  $\overline{Snow}$  and  $\overline{RF}$  are both in decline in some areas, which introduces an inherent correlation from the trend.

Having identified and quantified regions and months of significant covariance between  $\overline{Snow}$  and  $\overline{RF}$  (Fig. 7), we are in a position to fully answer the question of how the increased variability of SnowModel-LG over mW99 (shown in Fig. 6) ultimately impacts  $\sigma_{\overline{SIT}}^2$ . We plot the three contributing components to  $\sigma_{\overline{SIT}}^2$  for each region in each winter month (Fig. 8). We note that in the case of negative covariability between  $\overline{Snow}$  and  $\overline{RF}$ , it is possible for  $\sigma_{\overline{Snow}}^2 + \sigma_{\overline{RF}}^2$  to be larger than  $\sigma_{\overline{SIT}}^2$ .

375

This is not problematic because  $\sigma_{\overline{Snow}}^2 + \sigma_{\overline{RF}}^2$  does not represent a real quantity when the variables are not independent. In the marginal seas  $\sigma_{\overline{Snow}}^2$  overtakes  $\sigma_{\overline{RF}}^2$  to become the main constituent of  $\sigma_{\overline{SIT}}^2$  by end of the growth season (Fig. 8). This is particularly driven by the behaviour of the Beaufort and East Siberian Seas, where this relationship is clearly visible.



**Figure 8.** Constituent parts of  $\sigma_{\overline{SIT}}^2$  of different regions. Bars represent the variance ( $\sigma^2$ ) of  $\overline{RF}$  and  $\overline{Snow}$  and the covariance between the two. (a) illustrates the absolute variance contributions (b) illustrates their relative contributions. The variance of  $\overline{Snow}$  in mW99 is indicated in panel (a) by a superimposed black bar. Snow contributes significantly more variability in the late winter than radar freeboard in most of the marginal seas.

In the Central Arctic  $\sigma_{\overline{RF}}^2$  just remains the dominant component of  $\sigma_{\overline{SIT}}^2$ , throughout the cold season although  $\sigma_{\overline{Snow}}^2$  plays an increasing role as the season progresses.

Covariance between  $\overline{RF}$  and  $\overline{Snow}$  makes relatively constant contributions to  $\sigma_{\overline{SIT}}^2$  of the 'marginal seas' grouping in comparison to the other two components, but analysis of this grouping conceals more significant variation at the scale of the individual group members. The covariability term of Eq. (6) makes a larger contribution than radar freeboard variability itself at times, for example in the Kara and East Siberian seas at the end-of-winter, and for the Chukchi Sea in February and March.

385 For the Central Arctic, the covariability term generally makes less of a contribution to total SIT variability than radar freeboard or snow variability individually, and is negative in the first two months of winter. We note that the covariability is almost always positive in the marginal seas with the exception of December in the Kara and Chuckchi Seas.

Finally, we directly compare the variability of  $\overline{SIT}$  itself, when calculated using SnowModel-LG and mW99. We conduct this exercise in both absolute terms (Fig. 9a) and as a fraction of the regional mean thickness (Fig. 9b).

- Calculation of regional SIT with SnowModel-LG reveals higher variability in all marginal seas of the Arctic basin in all months. When the marginal seas are analysed as a contiguous entity, the standard deviation is 0.09 m with mW99 and 0.16 m with SnowModel-LG. This represents an increase in  $\overline{SIT}$  variability of 77%. For the Central Arctic this figure is considerably smaller, at 25%. When the individual marginal seas are considered, the largest increase was the Kara Sea (138%) and the smallest was the Beaufort Sea (35%).
- One key aspect of interannual variability is how it compares to typical values. When IAV is expressed as a percentage of the regional mean thickness, the Barents Sea exhibits the largest increase when calculated with SnowModel-LG: the standard deviation (as a percentage of mean thickness) increases from 15% to 25%. When variability is viewed in this way, the increase in the Central Arctic is small (7.9% to 9.4%). Variability as a fraction of mean thickness is also highest in the Barents Sea when calculated with SnowModel-LG whereas with mW99 this designation would go to the Beaufort Sea. When analysed as
- 400 one area, variability (as a fraction of mean thickness) in the marginal seas transitions from being 7.5% of the mean thickness to 13.8% when calculated with SnowModel-LG.

We also note that MYI exhibits more thickness variability than FYI (both absolutely and relative to the sea ice type's mean thickness) in all the marginal seas (Fig. S11). For the marginal seas as a single group, MYI is roughly twice as variable in absolute terms. This is not the case in the Central Arctic, where the thickness variability of the individual sea ice types is highly similar (with FYI IAV slightly larger when calculated relative to regional mean thickness).



**Figure 9.** Standard deviation in sea ice thickness over the period 2002 - 2018 except for the Central Arctic: 2010-2018 (a) calculated in absolute terms (b) calculated as a percentage of the regional mean thickness over the period. Mean growth-season values shown with dashed lines. The individual detrended regional timeseries from which this figure is synthesised are available in Fig. (S12).

## 4.3 New and faster thickness declines in the marginal seas

As well as exhibiting higher interannual variability than mW99, SnowModel-LG  $\overline{Snow}$  values decline over time in most regions due to decreasing SWE values year-on-year. Here we examine the aggregate contribution of a more variable but declining  $\overline{Snow}$  timeseries in determining the magnitude and significance of trends in  $\overline{SIT}$ .

We first assess regions where  $\overline{SIT}$  was already in statistically significant decline when calculated with mW99. This is the case for all months in the Laptev and Kara seas, and four of seven months in the Chukchi and Barents sea. The rate of decline in these regions grew significantly when calculated with SnowModel-LG data (Fig. 10; green panels). Relative to the decline-rate calculated with mW99, this represents average increases of 62% in the Laptev sea, 81% in the Kara Sea, and 102% in

the Barents Sea. The largest increase in an already statistically-significant decline was in the Chukchi Sea in April, where the

decline-rate increased by a factor of 2.1. When analysed as an aggregated area and with mW99, the total 'Marginal Seas' area 415 exhibits a statistically significant negative trend in November, December, January and April. The East Siberian Sea is the only region to have a month of decline when calculated with mW99 but not with SnowModel-LG.

We now turn our attention to new trends that stem from the use of SnowModel-LG over mW99 (Fig. 10; red panels). Our analysis reveals a new, statistically significant  $\overline{SIT}$  decline in the Chukchi Sea in October (taking the number of months with a decline in SIT to five). Perhaps more significantly, the aggregated Marginal Seas region exhibits two new months of 420 statistically significant declining  $\overline{SIT}$  in October and February, taking the total number of declining months to six. No months in any marginal sea exhibited a statistically significant increasing trend in  $\overline{SIT}$  (with either snow data set).

The Central Arctic region exhibits a statistically significant thickening October trend with both snow data sets (10 cm/yr and 9 cm/yr with SnowModel-LG and mW99). The region exhibits an additional month of increase in November when calculated with SnowModel-LG (7 cm/yr).

425

We also analyse these regional declines as a percentage of the regional mean sea ice thickness in the observational period (2002-2018). We observe the average growth-season thinning to increase from 21% per decade to 42% per decade in the Barents Sea, 39% to 56% per decade in the Kara Sea, and 24% to 40% per decade in the Laptev Sea when using SnowModel-LG instead of mW99. Five of the seven growth-season months in the Chukchi Sea exhibit a decline with SnowModel-LG of (on

average) 44% per decade. This is much more than that of the four significant months observable with mW99 (25% per decade). 430 We find the Marginal Seas (when considered as a contiguous, aggregated group) to be losing 30% of its mean thickness per decade in the six statistically significant months when SIT is calculated using SnowModel-LG (as opposed to mW99).

We further analyse these declining trends by sea ice type. This reveals the aggregate trends in the marginal seas to be broadly driven by thickness decline in FYI rather than MYI. We note that the FYI sea ice cover in the Kara and Laptev seas is

435 in statistically significant decline with either snow product in all months. The FYI cover in the Barents Sea is also in decline for six of the seven winter months when calculated with SnowModel-LG. We find that (when analysed with SnowModel-LG) if any month in a specific marginal sea is in 'all types' decline, its first year ice is also statistically significantly declining.

#### 4.4 Changes to the sea ice thickness distribution and seasonal growth

440

We now consider differences in the spatial sea ice thickness distribution introduced by a snow product with IAV. Because mW99 has low spatial variability in its SWE fields (the quadratic fits are relatively flat), it produces a more sharply peaked and narrow SIT distribution with lower probabilities of thinner or thicker sea ice in the months January - April. The SIT distribution also exhibits some degree of bimodality due to the halving scheme. This bimodality is to a large degree represented in the SnowModel-LG histograms - an encouraging result (Fig. S13).

The regional, seasonal growth rate is also similar when comparing calculations with SnowModel-LG and mW99 (Fig. S14).

These rates were calculated over the period 2002-2018 with the exception of the Central Arctic which was restricted to the 445 period 2010-2018. Among the most salient differences are the much smoother seasonal evolution of snow cover in the Barents Sea from SnowModel-LG and the decline in SWE from March to April in the Kara, Laptev and Beaufort seas with mW99



**Figure 10.** Regional  $\overline{SIT}$  timeseries calculated using mW99 and SnowModel-LG. Note different y-axis scale for Central Arctic and East Siberian Sea. Panels featuring a statistically significant trend in sea ice thickness when calculated both mW99 & SnowModel-LG framed with green. Red frames indicate where trend is only significant when calculated with SnowModel-LG. Blue frames indicate where a statistically significant increase is detected with mW99, but not with SnowModel-LG. Where trends are statistically significant, trend lines are superimposed.



**Figure 11.** Sea ice thickness trends in the four marginal seas that exhibited robust trends in several winter months in the period 2002-2018. Average winter trend (calculated only from statistically significant months) from each snow product shown with dashed lines. Data points are only shown where a statistically significant trend is present for that month and for the relevant snow data.

(compared to a continued increase with SnowModel-LG). In the East Siberian and Laptev seas there is clearly a slightly lower seasonal growth rate when calculated with mW99, and this is also true to a lesser extent in the Chukchi Sea.

#### 450 5 Discussion

#### 5.1 Sensitivity of Findings to Choice of Snow Product

#### 5.1.1 Choice of Climatology - Combining AMSR2 with mW99

The most recent sea ice thickness product from the Alfred Wegener Institute (Hendricks and Ricker, 2019) makes use of a new snow climatology, generated by the merging of W99 with snow depth data derived from the AMSR2 passive microwave record. This is then applied with a halving scheme based on sea ice type in a similar way to mW99 (but with the AMSR2 component not halved). This likely improves the absolute accuracy of snow depths (and thus sea ice thickness), but does not resolve the issues discussed in this paper involving trends and variability. The modified AMSR2/W99 climatology functions in a very similar way to mW99 - a weak IAV is introduced in areas of interannually fluctuating sea ice type. Any trends will be the result of trends in the relative dominance of sea ice type. This was discussed in Sect. 4.1.2 and illustrated in Fig. S4: sea ice type trends are only significant in October and January, where they are weak.

#### 5.1.2 Choice of Reanalysis Forcing for SnowModel-LG

Barrett et al. (2020) reviewed precipitation data from various reanalysis products over the Arctic Ocean using records from the Soviet drifting stations, and found the magnitude of interannual variability to be similar. They further broke these data down to the regional scale using the same regional definitions in this paper, and found that this similarity persisted. Boisvert et al.

- 465 (2018) conducted a similar analysis with drifting ice mass balance buoys, and found the interannual variability of the data sets to also be similar (although the authors found larger discrepancies in magnitude). These differences in magnitude however cannot be physical (as there is only one Arctic), and Cabaj et al. (2020) were able to bring precipitation estimates into better alignment using CloudSat data with a scaling approach. However this scaling approach preserved the interannual variability of the data sets, which Barrett et al. (2020) and Boisvert et al. (2018) found to be in comparatively good agreement. To investigate
- 470 how this variability propagates into *Snow* variability, we calculate *Snow* timeseries from SnowModel-LG runs forced by both MERRA-2 and ERA-5 data and find their variability to be very similar (Fig. S15).

With regard to trends, we find that the two different reanalysis forcings generally introduce minimal differences in the SIT trends (Fig. S16). We do however find that small differences in SWE cause the  $\overline{Snow}$  contribution of the MERRA-2 SnowModel-LG run to exhibit statistically significant decline in regions and months where the ERA-5 run does not (with only

475 a small change to the p-value). Analysis of the absolute  $\overline{Snow}$  timeseries reveals them to be otherwise similar (Fig. S17). We take these clear similarities as evidence that our findings are in principle robust to the choice of atmospheric reanalysis.

#### 5.1.3 Choice of Model - Comparison with NESOSIM

480

Some uncertainty is introduced into the spatial distribution of snow in a given year by SnowModel-LG snow parameterisations and simplifications, such as the lack of snow loss to leads. We therefore repeat our analyses with 2002-2015 data from the NASA Eulerian Snow On Sea Ice Model (NESOSIM; Petty et al., 2018b).

We find that doing this increases the relative importance of snow variability to sea ice thickness variability (Fig. S18). We also observe that the NESOSIM calculations are considerably more similar to those done with SnowModel-LG than with mW99. NESOSIM replicates the increasingly dominant  $\sigma_{\overline{Snow}}^2$  contribution to  $\sigma_{\overline{SIT}}^2$  over the winter in the Marginal seas, and also replicates the higher contribution of  $\sigma_{\overline{RF}}^2$  in the Central Arctic compared to both the individual and aggregated marginal seas. Striking means the neutrino provides the Vern Sec. Furthermore, the neutrino provides the marginal sease for the Vern Sec.

- 485 seas. Striking resemblances are seen for the Kara Sea and the East Siberian Sea. Furthermore, the negative covariances for November in the Barents Sea and December in the Chukchi are replicated (albeit with significantly greater magnitude in the Barents Sea). NESOSIM also replicates the negative covariances in October and November in the Central Arctic, but also introduces negative covariance in December (unlike SnowModel-LG).
- Because the NESOSIM data is only publicly available from 2002-2015, any underlying trends in the SIT timeseries are more challenging to detect because of the shorter observational period (by comparison to regions where all relevant data is available from 2002-2018). On the other hand, the calculated interannual variability is not reduced by the shorter timeseries, further obscuring any potential underlying trends. But despite these differences, both snow data sets produce statistically significant decline in all months in the Laptev Sea. NESOSIM reproduces six of the seven months of decline in the Kara Sea shown by SnowModel-LG, and three of the five in the Marginal Seas.
- 495 Further inspection of the individual data points across all regions and months reveals good agreement in regional SIT when calculated with either SnowModel-LG or NESOSIM we take this as evidence that our findings concerning trends and variability over the longer 2002-2018 period are robust to the choice of reanalysis-accumulation model.

#### 5.2 Study Limitations

#### 5.2.1 Statistical Treatment

500 We have assumed in calculating single figures for variances that the interannual variability of the systems at hand is timestationary. It is unclear whether this is the case, as the timeseries are limited in length and time-resolution and thus offer limited scope to test for stationarity. Furthermore we only tested for linear trends, when trends may in fact be non-linear. However, a visual inspection of Fig. (10) implies that this approximation is adequate on a qualitative level. Our trend tests also were twotailed, with the null hypothesis that there was no trend. We could have formulated an alternate test where our null hypothesis was that the trend was positive. This would have given a higher number of statistically significant instances of negative trends, but we deemed this inappropriate as one of the regions (the Central Arctic) does exhibit significant positive trends with the two-tailed test.

#### 5.3 Inter-Mission Bias between Envisat and CryoSat-2

An extensive validation exercise for the merged products indicated that although Envisat radar freeboards match well with

510 CS2 freeboards in the Arctic overall, some biases do exist over specific ice types (ESA, 2018). In particular, analysis of the inter-mission overlap period indicates that Envisat freeboards were biased low (relative to CS2) in areas dominated by MYI, and high in areas dominated by FYI.

We first make the point that this will have a relatively minimal effect on our findings regarding interannual variability, as  $\overline{Snow}$  is unaffected by this and  $\sigma_{\overline{RF}}^2$  is likely relatively independent of the absolute magnitude of  $\overline{RF}$ .

515 With regard to trends, if Envisat radar freeboards (and thus  $\overline{RF}$ ) are in fact biased high over FYI between 2002-2010 (relative to CS2), then the total trend in many regions dominated by FYI could potentially be smaller than calculated in this manuscript. We do however add that our findings regarding the impact of declining  $\overline{Snow}$  is unaffected by any inter-mission bias in  $\overline{RF}$ . Because the trend in  $\overline{SIT}$  is determined by both  $\overline{Snow} \& \overline{RF}$ , the trend in  $\overline{SIT}$  will always be more negative when calculated with downward trending data for  $\overline{Snow}$ .

#### 520 5.3.1 The Effects of Incomplete Radar Penetration of the Snowpack

This investigation has been carried out within the paradigm of total Ku-band radar wave penetration of the snow cover (as suggested by Beaven et al. (1995)), however some in situ investigations have cast doubt on this. The issue was highlighted in an Antarctic context by Giles et al. (2008b) for ERS radar freeboards, and it was shown subsequently that significant morphological features in the snowpack (e.g. depth hoar, wet snow or crusts) enhanced radar scattering from within the snowpack
(Willatt et al., 2010). For the Arctic, Willatt et al. (2011) found that airborne Ku-band radar backscatter in the Bay of Bothnia was returned from nearer the snow-ice than snow-air interface in only 25% of cases when the temperature was close to freezing, the figure increasing to 80% at lower temperatures. Nandan et al. (2017) observed that the presence of brine in the base of the snowpack can raise the scattering horizon by several centimeters. However, these investigations were often (but not exclusively) carried out at the end of the winter season or in the Sub-Arctic, when warmer temperatures may have increased the snow's brine volume fraction and diurnal forcing can drive rapid snow metamorphism. Both of these factors will be less prevalent in the colder months of winter. This analysis is therefore carried out using the imperfect historical assumption present in publicly available sea ice products (that of total penetration).

What would the effects of incomplete penetration of the snowpack be on our findings? As the height of the primary radar scattering horizon rises through the snow, the altimeter operation transitions from that of a radar altimeter to that of a lidar

- altimeter. Knowledge of overlying snow contributes positively to the inference of SIT in the case of a radar altimeter (i.e. the coefficient of  $m_s$  term of Eq. 3 is positive). However, the influence of overlying snow on lidar-based SIT estimates is negative (i.e. the presence of more snow for a given measured radar freeboard implies less underlying sea ice). As the scattering horizon rises through the snowpack, the SIT contribution of snow therefore decreases, reaches zero (in the top half of the snowpack, the exact location depending on snow density) and proceeds to negative values. The result of potential incomplete penetration
- 540 for our study is that the magnitude of the reported trend and variance underestimations is diminished. Were our investigation based on a similarly long timeseries of lidar freeboards combined with a snow climatology, one of our conclusions would be that diminishing snow cover is leading to *overestimation* of rates of decline in the marginal seas.

We finally note the potentially confounding influence of negative freeboard in regions such as the Atlantic sector of the Central Arctic region and the Barents Sea. In the case of high snowfall and low sea ice thickness, the sea ice surface can be depressed to the waterline or below. Beyond this point Eq. (5) no longer functions. The prevalence of negative freeboards has been studied by Rösel et al. (2018) and Merkouriadi et al. (2020), but has yet to be incorporated into any radar-altimetry based

sea ice thickness retrievals. This situation can be driven by storm tracks entering the Arctic from the Atlantic (but also the Bering Strait). These intrusions of warm air can also drive snow grain metamorphism, which may well affect radar penetration through the snowpack.

## 550 5.4 The Impact of Enhanced Variability from SnowModel-LG

When used instead of mW99, SnowModel-LG data increases the interannual variability of  $\overline{SIT}$  in the marginal seas by more than 50%. The main way that this occurs is though increasing  $\sigma_{\overline{Snow}}^2$  values (Fig. 6). The second and less significant way that  $\sigma_{\overline{SIT}}^2$  is increased is through some positive correlations between  $\overline{Snow}$  and  $\overline{RF}$  values for individual months in some regions (Fig. 7). Because the two timeseries are positively correlated in some cases,  $\sigma_{\overline{SIT}}^2$  is increased; for the Marginal Seas region this covariance term makes up around 15-20% of  $\sigma_{\overline{SIT}}^2$  (Fig. 8).

While values for interannual variability are given in W99, it was previously impossible to apply those values to either a given year or to fulfil Eq. (6). SnowModel-LG offers similar variability to the SWE statistics given in W99 (Fig. 3), and can generate a yearly timeseries of values. Furthermore it can be combined with radar freeboard data to generate all terms of Eq. (6) for a direct calculation of  $\sigma_{\overline{SUT}}^2$ .

- 560 Comparing our IAV values to the literature is challenging due to differences in the area over which other authors have calculated IAV values. Haas (2004) investigated the interannual variability of an area within the Transpolar Drift in the Central Arctic and Northern Barents Sea, and found a 0.73 m standard deviation. This is considerably higher than the values determined in this study, although this data was collected by electromagnetic sounding in late summer over a ten year period that does not overlap with this analysis. Laxon et al. (2003) defined a 'region of coverage', which essentially consisted of the marginal seas
- 565 considered in this analysis with the addition of some areas of the Canadian Archipelago and the Greenland sea. The authors found a variability of 0.24 m using W99 in this region of coverage over an eight year timescale. Unlike Haas (2004), this value is lower than our findings using either mW99 or SnowModel-LG. Similar to Haas (2004), the time period is considerably shorter and the geographical area is not identical. Finally, Rothrock et al. (2008) found interannual variability in SIT to be 0.46 m over a twenty-five year period (1975-2000), using submarine records from a variety of Arctic regions. It is likely that the
- 570 values in these studies differ due to the unequal spatial extent over which the IAV was calculated; averaging over a larger area reduces the IAV due to the averaging out of local anomalies.

#### 5.5 The Impact of New and Steeper Trends in Mean Sea Ice Thickness

The replacement of multiyear ice with first year ice has been documented to be reducing Arctic-mean SWE on sea ice in spring (Webster et al., 2014). However, progressively later freeze-ups in the Arctic are also likely driving a reduction in mean SWE in

575 the early cold-season. This is because sea ice covers a relatively smaller area in the high precipitation months of September and October. When the sea ice area then expands with the progression of the growth-season, the newer sea ice has not been exposed to this snowfall. This mechanism is not accounted for in mW99, and as such snow depths do not decrease at a statistically significant level in any month. In this study we have assessed how these negative trends in  $\overline{Snow}$  propagate through into trends in  $\overline{SIT}$ . In every area where

- a statistically significant decline in radar freeboards is observed, a statistically significant decline in SnowModel-LG SWE is also observed (Fig. S19). In addition to this, SnowModel-LG also exhibits  $\overline{Snow}$  decline in other months in the Beaufort and Barents Sea. As such, reductions in  $\overline{Snow}$  usually act in concert with observed reductions in  $\overline{RF}$ , amplifying decline in  $\overline{SIT}$ . This relationship is illustrated by the fact that several months in several regions do not exhibit either a statistically significant decline in  $\overline{RF}$  or  $\overline{Snow}$  (Fig. S19), but despite this they do exhibit decline in  $\overline{SIT}$  (Fig. 10). We note here that this 'co-decline'
- in  $\overline{Snow}$  and  $\overline{RF}$  is separate to the covariability presented in Sect. 4.2 and Fig. 7, as that was calculated from detrended data. Because SnowModel-LG data features a steeper decline in  $\overline{Snow}$  than mW99, a steeper decline is observed in the  $\overline{SIT}$  of several regions. However, SnowModel-LG  $\overline{Snow}$  contribution to  $\overline{SIT}$  also exhibits significantly more variability, which acts to reduce statistical significance of  $\overline{SIT}$  trends. Despite this compensating effect, the statistical significance of trends in  $\overline{SIT}$  were generally greater than those calculated using mW99. Furthermore, statistically significant trends emerged in new months and new regions.

Kwok and Rothrock (2009) analysed 42 years of submarine records and the five year ICESat record. However, it is challenging to draw comparison with our results, as trends were gleaned from submarine track crossings and by comparing the thickness difference between the period of submarine observation and that of ICESat observations. Difficulty in comparison is further compounded by differences in regional designation and the area of the submarine data release (which is generally confined to the Central Arctic region where the radar altimetry timeseries is at best limited to the CryoSat-2 era). This is also

595

the case for the updated analysis of Kwok (2018), who seasonally adjusted mean thickness values to match crossover points in submarine tracks in time and space. Our findings of enhanced interannual variability and steeper decline have implications for Arctic stakeholders and the de-

ployment of human infrastructure. The marginal seas are heavily used for the shipping of goods along the Northern Sea Route
in summer (Eguíluz et al., 2016) and provide the setting for potential extraction of natural resources (Petrick et al., 2017).
Furthermore, the season during which vessels may traverse the Northern Sea Route is lengthening. Higher variability in sea ice thickness may pose a challenge to the planning of this seasonal travel, particularly with regard to the need for ice-strengthened escorts for conventional vessels (Melia et al., 2017; Cariou et al., 2019). The enhancement of declining trends where they exist is perhaps of benefit these industries.

#### 605 5.6 The interannual relationship between radar freeboard and snow depth

We finally consider the physical mechanisms behind positive or non-significant correlations between  $\overline{Snow}$  and  $\overline{RF}$  displayed in Fig. (7). Assuming total radar penetration of the snow cover, as snow accumulates on sea ice it should lower the local radar freeboard by a distance on the order of half its accumulated height (Eq. 4). This lowering is a result of physical depression of the sea ice surface and an increase in the radar ranging due to slower radar wave propagation in snow (in approximately

610 a 60:40 ratio). Over short time scales (days to weeks), this would result in a negative correlation between local snow depth and local radar freeboard. This corresponds to a negative covariability term in Eq. (5) and is represented by purple bars in Fig. (8). Negative values are generally not seen, with the exception of October and November in the Central Arctic, November in the Barents Sea and December in the Chukchi and Kara seas. Furthermore, snow is a highly insulating material and its accumulation limits sea ice thermodynamic growth. This would also bring about a negative correlation between snow depth and radar freeboard, lagged over a period of weeks.

615

The lack of negative correlations between  $\overline{RF}$  and  $\overline{Snow}$  from year to year is likely indicative of the timescale of our analysis. If present, the negative correlation implied by Eq. (4) and the mechanisms above must only be present on shorter timescales (e.g. days). So what drives the positive correlations between  $\overline{RF}$  and  $\overline{Snow}$  where they exist? One driver over FYI is likely sea ice age. Sea ice formed at the beginning of the season has a longer time to (a) grow thicker, and (b) accumulate

620 snow. Both variables are therefore likely controlled by regional freeze-up timing, explaining the correlation. The combined evolution of  $\overline{Snow}$  and  $\overline{RF}$  anomalies as a function of regional freeze-up timings is likely to be the subject of future study. The relationship between MYI radar freeboards and accumulated SWE may also form an avenue for further study.

#### 6 Summary

In this paper we used a novel approximation for the slowing of radar waves in snow to decompose the conventional method for estimating sea ice thickness into two contributions: one originating from radar freeboard data (from satellite altimeters), the other from snow data of varying provenance.

This allowed a regional assessment of the conventional impact of snow on variability and trends in sea ice thickness. We then used a new snow data set (from SnowModel-LG) with a more realistic magnitude of interannual variability and trends to calculate the regional sea ice thickness timeseries.

630 We found that interannual variability in average sea ice thickness ( $\sigma_{SIT}^2$ ) of the marginal seas was increased by more than 50% by accounting for variability in the snow cover. On a seasonal timescale we find that variability in the snow cover makes an increasing contribution to the total variability of inferred sea ice thickness, increasing from around 20% in October to more than 70% in April.

We also observed that the trends in SnowModel-LG data propagated through to the SIT timeseries, amplifying decline in regions where it was already significant, and introducing significant decline where it did not previously exist. This occurred in spite of the compensating effect of enhanced interannual variability.

*Author contributions.* JCS, JCL, MT and RW proposed and conceptualised the study. VN and GEL provided extensive feedback on manuscript and GEL provided the SnowModel-LG data. RDCM carried out the main analysis. All authors contributed to the write-up.

Code and data availability. The code used for all analysis and visualisation was written in Python 3.6 and is available at
 github.com/robbiemallett/SnowModel-LG\_SIT\_Impacts. The radar freeboard data from Envisat and CryoSat-2 is available from the ESA
 CCI initiative at climate.esa.int/en/odp/#/project. The NESOSIM snow data is available from the NASA Cryospheric Sciences Laboratory

website at earth.gsfc.nasa.gov/cryo/data/nasa-eulerian-snow-sea-ice-model-nesosim. It is anticipated that the SnowModel-LG data will be hosted in a persistent data repository in the near future. Code and data last accessed 2020/9/20.

Competing interests. The authors declare no competing interests.

Acknowledgements. This work was funded primarily by the London Natural Environmental Research Council Doctoral Training Partnership grant (NE/L002485/1). JCL acknowledges support from the European Space Agency Living Planet Fellowship 'Arctic-SummIT' under grant ESA/4000125582/18/I-NS and the Natural Environmental Research Council Project 'Diatom-ARCTIC' under Grant NE/R012849/1. MT acknowledges support from the European Space Agency by project 'Polarice' under grant ESA/AO/1-9132/17/NL/MP, project 'CryoSat + Antarctica' under Grant ESA AO/1-9156/17/I-BG and project 'Polar + Snow' under Grant ESA AO/1-10061/19/I-EF. JS and MT also acknowledge support from the Natural Environment Research Council (grant no. NE/S002510/1). JCS and GEL acknowledge support from NASA grant 15-CRYO2015-0019 /NNX16AK85G.

## References

- Aaboe, S.: Copernicus Climate Data Records Sea Ice Edge and Sea Ice Type Product User Guide and Specification, Tech. rep., https://doi.org/10.24381/cds.29c46d83, 2020.
- 655 Armitage, T. W. and Ridout, A. L.: Arctic sea ice freeboard from AltiKa and comparison with CryoSat-2 and Operation IceBridge, Geophysical Research Letters, 42, 6724–6731, https://doi.org/10.1002/2015GL064823, 2015.
  - Barrett, A. P., Stroeve, J. C., and Serreze, M. C.: Arctic Ocean Precipitation From Atmospheric Reanalyses and Comparisons With North Pole Drifting Station Records, Journal of Geophysical Research: Oceans, 125, https://doi.org/10.1029/2019JC015415, https://onlinelibrary. wiley.com/doi/abs/10.1029/2019JC015415, 2020.
- 660 Beaven, S. G., Lockhart, G. L., Gogineni, S. P., Hosseinmostafa, A. R., Jezek, K., Gow, A. J., Perovich, D. K., Fung, A. K., and Tjuatja, S.: Laboratory measurements of radar backscatter from bare and snow-covered saline ice sheets, International Journal of Remote Sensing, 16, 851–876, https://doi.org/10.1080/01431169508954448, 1995.
- Belter, H. J., Krumpen, T., Hendricks, S., Hoelemann, J., Janout, M. A., Ricker, R., and Haas, C.: Satellite-based sea ice thickness changes in the Laptev Sea from 2002 to 2017: comparison to mooring observations, The Cryosphere, 14, 2189–2203, https://doi.org/10.5194/tc-14-2189-2020, https://tc.copernicus.org/articles/14/2189/2020/, 2020.
  - Blockley, E. W. and Peterson, K. A.: Improving Met Office seasonal predictions of Arctic sea ice using assimilation of CryoSat-2 thickness, The Cryosphere, 12, 3419–3438, https://doi.org/10.5194/tc-12-3419-2018, https://www.the-cryosphere.net/12/3419/2018/, 2018.
    - Boisvert, L. N., Webster, M. A., Petty, A. A., Markus, T., Bromwich, D. H., and Cullather, R. I.: Intercomparison of precipitation estimates over the Arctic ocean and its peripheral seas from reanalyses, Journal of Climate, 31, 8441–8462, https://doi.org/10.1175/JCLI-D-18-0125.1, http://journals.ametsoc.org/doi/10.1175/JCLI-D-18-0125.1, 2018.
- 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, https://doi.org/10.3390/ijgi1010032, http://www.mdpi. com/2220-9964/1/1/32, 2012.
- Bunzel, F., Notz, D., and Pedersen, L. T.: Retrievals of Arctic Sea-Ice Volume and Its Trend Significantly Affected by Interannual Snow
   Variability, Geophysical Research Letters, 45, 751–11, https://doi.org/10.1029/2018GL078867, 2018.
  - Cabaj, A., Kushner, P., Fletcher, C., Howell, S., and Petty, A.: Constraining Reanalysis Snowfall Over the Arctic Ocean Using CloudSat Observations, Geophysical Research Letters, 47, https://doi.org/10.1029/2019GL086426, https://onlinelibrary.wiley.com/doi/abs/10.1029/ 2019GL086426, 2020.
    - Cariou, P., Cheaitou, A., Faury, O., and Hamdan, S.: The feasibility of Arctic container shipping: the economic and environmental im-
- 680 pacts of ice thickness, Maritime Economics and Logistics, pp. 1–17, https://doi.org/10.1057/s41278-019-00145-3, https://doi.org/10.1057/ s41278-019-00145-3, 2019.
  - Chandler, R. E. and Scott, E. M.: Statistical Methods for Trend Detection and Analysis in the Environmental Sciences, https://doi.org/10.1002/9781119991571, 2011.
  - Chevallier, M. and Salas-Mélia, D.: The Role of Sea Ice Thickness Distribution in the Arctic Sea Ice Potential Predictability: A Diagnostic
- 685 Approach with a Coupled GCM, Journal of Climate, 25, 3025–3038, https://doi.org/10.1175/JCLI-D-11-00209.1, http://journals.ametsoc. org/doi/10.1175/JCLI-D-11-00209.1, 2012.
  - Comiso, J.: Bootstrap Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I, National Snow and Ice Data Center, Boulder, Colorado. Digital media, 2000.

Eguíluz, V. M., Fernández-Gracia, J., Irigoien, X., and Duarte, C. M.: A quantitative assessment of Arctic shipping in 2010-2014, Scientific Reports, 6, 1–6, https://doi.org/10.1038/srep30682, www.nature.com/scientificreports, 2016.

ESA: Sea Ice Climate Change Initiative: Phase 2 D4.1 Product Validation & Intercomparison Report (PVIR), Tech. rep., 2018.

690

- 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, L22 502, https://doi.org/10.1029/2008GL035710, http://doi.wiley.com/10.1029/2008GL035710, 2008a.
- Giles, K. A., Laxon, S. W., and Worby, A. P.: Antarctic sea ice elevation from satellite radar altimetry, Geophysical Research Letters, 35,
   L03 503, https://doi.org/10.1029/2007GL031572, http://doi.wiley.com/10.1029/2007GL031572, 2008b.
- Haas, C.: Late-summer sea ice thickness variability in the Arctic Transpolar Drift 1991-2001 derived from ground-based electromagnetic sounding, Geophysical Research Letters, 31, n/a–n/a, https://doi.org/10.1029/2003GL019394, http://doi.wiley.com/10.1029/ 2003GL019394, 2004.
  - Hendricks, S. and Ricker, R.: Product User Guide & Algorithm Specification: AWI CryoSat-2 Sea Ice Thickness (version 2.2), 2019.
- 700 Hendricks, S., Paul, S., and Rinne, E.: Northern hemisphere sea ice thickness from the CryoSat-2 satellite on a monthly grid (L3C), v2.0, https://doi.org/10.5285/ff79d140824f42dd92b204b4f1e9e7c2, 2018.
- Katlein, C., Arndt, S., Nicolaus, M., Perovich, D. K., Jakuba, M. V., Suman, S., Elliott, S., Whitcomb, L. L., McFarland, C. J., Gerdes, R., Boetius, A., and German, C. R.: Influence of ice thickness and surface properties on light transmission through <scp>A</scp> rctic sea ice, Journal of Geophysical Research: Oceans, 120, 5932–5944, https://doi.org/10.1002/2015JC010914, https://onlinelibrary.wiley.com/doi/abs/10.1002/2015JC010914, 2015.
  - Kern, S., Khvorostovsky, K., Skourup, H., Rinne, E., Parsakhoo, Z. S., Djepa, V., Wadhams, P., and Sandven, S.: The impact of snow depth, snow density and ice density on sea ice thickness retrieval from satellite radar altimetry: Results from the ESA-CCI Sea Ice ECV Project Round Robin Exercise, Cryosphere, 9, 37–52, https://doi.org/10.5194/tc-9-37-2015, 2015.

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

710 38, n/a–n/a, https://doi.org/10.1029/2011GL049216, http://doi.wiley.com/10.1029/2011GL049216, 2011.

- Kurtz, N. T., Galin, N., and Studinger, M.: An improved CryoSat-2 sea ice freeboard retrieval algorithm through the use of waveform fitting, Cryosphere, 8, 1217–1237, https://doi.org/10.5194/tc-8-1217-2014, https://www.the-cryosphere.net/8/1217/2014/, 2014.
  - Kwok, R.: Arctic sea ice thickness, volume, and multiyear ice coverage: Losses and coupled variability (1958-2018), Environmental Research Letters, 13, 105 005, https://doi.org/10.1088/1748-9326/aae3ec, https://doi.org/10.1088/1748-9326/aae3ec, 2018.
- 715 Kwok, R. and Cunningham, G. F.: Variability of arctic sea ice thickness and volume from CryoSat-2, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373, https://doi.org/10.1098/rsta.2014.0157, 2015.
  - Kwok, R. and Rothrock, D. A.: Decline in Arctic sea ice thickness from submarine and ICESat records: 1958-2008, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL039035, 2009.
- Kwok, R., Kacimi, S., Webster, M. A., Kurtz, N. T., and 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, 1–19, https://doi.org/10.1029/2019JC016008, 2020.
  - Landy, J. C., Petty, A. A., Tsamados, M., and Stroeve, J. C.: Sea ice roughness overlooked as a key source of uncertainty in CryoSat-2 ice freeboard retrievals, Journal of Geophysical Research: Oceans, 44, 1–36, https://doi.org/10.1029/2019jc015820, 2020.
    - Laxon, S., Peacock, H., and Smith, D.: High interannual variability of sea ice thickness in the Arctic region, Nature, 425, 947–950, https://doi.org/10.1038/nature02050, 2003.

- 725 Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S., and Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and volume, Geophysical Research Letters, 40, 732-737, https://doi.org/10.1002/grl.50193, 2013.
  - Li, M., Ke, C., Shen, X., Cheng, B., and Li, H.: Investigation of the Arctic Sea ice volume from 2002 to 2018 using multi-source data, International Journal of Climatology, p. joc.6972, https://doi.org/10.1002/joc.6972, https://onlinelibrary.wiley.com/doi/10.1002/joc.6972, 2020a.
  - Li, Z., Zhao, J., Su, J., Li, C., Cheng, B., Hui, F., Yang, Q., and Shi, L.: Spatial and temporal variations in the extent and thickness of arctic landfast ice, Remote Sensing, 12, 64, https://doi.org/10.3390/RS12010064, www.mdpi.com/iournal/remotesensing, 2020b.
  - 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:
- 735 Oceans, https://doi.org/10.1029/2019jc015913, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019JC015913https://agupubs. onlinelibrary.wiley.com/doi/abs/10.1029/2019JC015913https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019JC015913, 2020.
  - Mallett, R. D., Lawrence, I. R., Stroeve, J. C., Landy, J. C., and Tsamados, M.: Brief communication: Conventional assumptions involving the speed of radar waves in snow introduce systematic underestimates to sea ice thickness and seasonal growth rate estimates, Cryosphere, 14, 251–260, https://doi.org/10.5194/tc-14-251-2020, 2020.
- 740 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, https://doi.org/10.1029/2009JC005436, 2009.
  - Melia, N., Haines, K., Hawkins, E., and Day, J. J.: Towards seasonal Arctic shipping route predictions, Environmental Research Letters, 12, 084 005, https://doi.org/10.1088/1748-9326/aa7a60, https://doi.org/10.1088/1748-9326/aa7a60, 2017.
  - Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., Kofinas, G., Mackintosh, A., Melbourne-Thomas,
- 745 J., Muelbert, M., Ottersen, G., Pritchard, H., and Schuur, E.: Polar Regions, in: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, edited by Portner, H.-O., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegria, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., and Weyer, N., pp. 203–320, IPCC, 2019.
  - Merkouriadi, I., Liston, G. E., Graham, R. M., and Granskog, M. A.: Quantifying the Potential for Snow-Ice Formation in the Arctic Ocean, Geophysical Research Letters, 47, no, https://doi.org/10.1029/2019GL085020, https://onlinelibrary.wiley.com/doi/abs/10.1029/ 2019GL085020, 2020.
- 750

- Mundy, C. J., Barber, D. G., and Michel, C.: Variability of snow and ice thermal, physical and optical properties pertinent to sea ice algae biomass during spring, Journal of Marine Systems, 58, 107-120, https://doi.org/10.1016/j.jmarsys.2005.07.003, https://linkinghub. elsevier.com/retrieve/pii/S0924796305001417, 2005.
- Nandan, V., Geldsetzer, T., Yackel, J., Mahmud, M., Scharien, R., Howell, S., King, J., Ricker, R., and Else, B.: Effect of 755 Snow Salinity on CryoSat-2 Arctic First-Year Sea Ice Freeboard Measurements, Geophysical Research Letters, 44, 419-10,
- https://doi.org/10.1002/2017GL074506, http://doi.wiley.com/10.1002/2017GL074506, 2017.
  - Paul, S., Hendricks, S., Ricker, R., Kern, S., and Rinne, E.: Empirical parametrization of Envisat freeboard retrieval of Arctic and Antarctic sea ice based on CryoSat-2: progress in the ESA Climate Change Initiative, The Cryosphere, 12, 2437–2460, https://doi.org/10.5194/tc-12-2437-2018, https://tc.copernicus.org/articles/12/2437/2018/, 2018.
- 760 Petrick, S., Riemann-Campe, K., Hoog, S., Growitsch, C., Schwind, H., Gerdes, R., and Rehdanz, K.: Climate change, future Arctic Sea ice, and the competitiveness of European Arctic offshore oil and gas production on world markets, Ambio, 46, 410-422, https://doi.org/10.1007/s13280-017-0957-z, 2017.

Petty, A. A., Holland, M. M., Bailey, D. A., and Kurtz, N. T.: Warm Arctic, Increased Winter Sea Ice Growth?, Geophysical Research Letters, 45, 922–12, https://doi.org/10.1029/2018GL079223, https://onlinelibrary.wiley.com/doi/abs/10.1029/2018GL079223, 2018a.

- 765 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, 4577–4602, https://doi.org/10.5194/gmd-11-4577-2018, 2018b.
  - Quartly, G. D., Rinne, E., Passaro, M., Andersen, O. B., Dinardo, S., Fleury, S., Guillot, A., Hendricks, S., Kurekin, A. A., Müller, F. L., Ricker, R., Skourup, H., and Tsamados, M.: Retrieving sea level and freeboard in the Arctic: A review of current radar altimetry methodologies and future perspectives, https://doi.org/10.3390/RS11070881, 2019.
- 770 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, Cryosphere, 8, 1607–1622, https://doi.org/10.5194/tc-8-1607-2014, https://www.the-cryosphere.net/ 8/1607/2014/, 2014.
  - Rösel, A., Itkin, P., King, J., Divine, D., Wang, C., Granskog, M. A., Krumpen, T., and Gerland, S.: Thin Sea Ice, Thick Snow, and Widespread Negative Freeboard Observed During N-ICE2015 North of Svalbard, Journal of Geophysical Research: Oceans, 123, 1156– 1176, https://doi.org/10.1002/2017JC012865, 2018.
- Rothrock, D. A., Percival, D. B., and Wensnahan, M.: The decline in arctic sea-ice thickness: Separating the spatial, annual, and interannual variability in a quarter century of submarine data, Journal of Geophysical Research: Oceans, 113, https://doi.org/10.1029/2007JC004252, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2007JC004252https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2007JC004252, 2007JC004252, 2007JC004252, 2007JC004252, 2008.
- 780 Sallila, H., Farrell, S. L., McCurry, J., and Rinne, E.: Assessment of contemporary satellite sea ice thickness products for Arctic sea ice, The Cryosphere, 13, 1187–1213, https://doi.org/10.5194/tc-13-1187-2019, https://www.the-cryosphere.net/13/1187/2019/, 2019.
  - 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, https://doi.org/10.5194/tc-13-125-2019, https://www.the-cryosphere.net/13/125/2019/, 2019.
  - Schweiger, A. J., Wood, K. R., and Zhang, J.: Arctic Sea Ice volume variability over 1901-2010: A model-based reconstruction, Journal of

785 Climate, 32, 4731–4752, https://doi.org/10.1175/JCLI-D-19-0008.1, 2019.

775

- Stroeve, J. and Notz, D.: Changing state of Arctic sea ice across all seasons, Environmental Research Letters, 13, 103 001, https://doi.org/10.1088/1748-9326/aade56, https://doi.org/10.1088/1748-9326/aade56, 2018.
- Stroeve, J., Liston, G. E., Buzzard, S., Zhou, L., Mallett, R., Barrett, A., Tschudi, M., Tsamados, M., Itkin, P., and Stewart, J. S.: A Lagrangian Snow-Evolution System for Sea Ice Applications (SnowModel-LG): Part II Analyses, Journal of Geophysical Research: Oceans, https://doi.org/10.1029/2019JC015900, https://onlinelibrary.wiley.com/doi/abs/10.1029/2019JC015900, 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, 1216–1225, https://doi.org/10.1002/2013GL058951, http://doi.wiley.com/10.1002/2013GL058951, 2014.
  - Tilling, R. L., Ridout, A., Shepherd, A., and Wingham, D. J.: Increased Arctic sea ice volume after anomalously low melting in 2013, Nature Geoscience, 8, 643–646, https://doi.org/10.1038/ngeo2489, 2015.
- 795 Tilling, R. L., Ridout, A., and Shepherd, A.: Estimating Arctic sea ice thickness and volume using CryoSat-2 radar altimeter data, Advances in Space Research, 62, 1203–1225, https://doi.org/10.1016/j.asr.2017.10.051, https://doi.org/10.1016/j.asr.2017.10.051, 2018.
  - Tsamados, M., Feltham, D. L., and Wilchinsky, A. V.: Impact of a new anisotropic rheology on simulations of Arctic Sea ice, Journal of Geophysical Research: Oceans, 118, 91–107, https://doi.org/10.1029/2012JC007990, 2013.

Tschudi, M. A., Meier, W. N., and Scott Stewart, J.: An enhancement to sea ice motion and age products at the National Snow and Ice Data

800 Center (NSIDC), Cryosphere, 14, 1519–1536, https://doi.org/10.5194/tc-14-1519-2020, 2020.

- Vella, D. and Wettlaufer, J. S.: Explaining the patterns formed by ice floe interactions, Journal of Geophysical Research, 113, C11011, https://doi.org/10.1029/2008JC004781, http://doi.wiley.com/10.1029/2008JC004781, 2008.
- 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, https://doi.org/10.1175/1520-0442(1999)012<1814:SDOASI>2.0.CO;2, 1999.
- 805 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, http://doi.wiley.com/10.1002/2014JC009985, 2014.
  - Willatt, R., Laxon, S., Giles, K., Cullen, R., Haas, C., and Helm, V.: Ku-band radar penetration into snow cover on Arctic sea ice using airborne data, Annals of Glaciology, 52, 197–205, https://doi.org/10.3189/172756411795931589, 2011.
- Willatt, R. C., Giles, K. A., Laxon, S. W., Stone-Drake, L., and Worby, A. P.: Field investigations of Ku-band radar penetration into snow cover
  on antarctic sea ice, IEEE Transactions on Geoscience and Remote Sensing, 48, 365–372, https://doi.org/10.1109/TGRS.2009.2028237, http://ieeexplore.ieee.org/document/5282596/, 2010.
  - Zygmuntowska, M., Rampal, P., Ivanova, N., and Smedsrud, L. H.: Uncertainties in Arctic sea ice thickness and volume: New estimates and implications for trends, Cryosphere, 8, 705–720, https://doi.org/10.5194/tc-8-705-2014, https://www.the-cryosphere.net/8/705/2014/, 2014.

# Supplement to: Faster decline and higher variability in the sea ice thickness of the marginal Arctic seas when accounting for dynamic snow cover

Robbie D.C. Mallett <sup>1</sup>, Julienne C. Stroeve <sup>1,2,3</sup>, Michel Tsamados <sup>1</sup>, Jack C. Landy <sup>4</sup>, Rosemary Willatt <sup>1</sup>, Vishnu Nandan <sup>3</sup>, and Glen E. Liston <sup>5</sup>

<sup>1</sup>Centre for Polar Observation and Modelling, Earth Sciences, University College London, London, UK
 <sup>2</sup>National Snow and Ice Data Center, University of Colorado, Boulder, CO, USA
 <sup>3</sup>Centre for Earth Observation Science, University of Manitoba, Winnipeg, Canada
 <sup>4</sup>School of Geographical Sciences, University of Bristol, Bristol, UK
 <sup>5</sup>Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

## S1 Snow's impact on conventional sea ice thickness retrievals can be characterised solely by its snow water equivalent

Using the expression of ice freeboard from Armitage and Ridout (2015):

And using the expression of the propagation correction from Tilling et al. (2018):

5 
$$h_i = h_r + h_s(c/c_s - 1)$$
 (S2)

Where  $h_s$  is snow depth, c is the speed of light in free space and  $c_s$  is the speed of light in snow. Numerous empirical expressions for  $c_s$  exist, in this work we use the expression for the permittivity of dry snow from Mätzler (2006):

$$\epsilon_{ds} = \left(1 + 0.5194\rho_s\right)^3 \tag{S3}$$

Relating the radar wave speed to the permittivity using  $c_s = c/\sqrt{\epsilon}$  (Ulaby and Long, 2014):

10 
$$c_s = c (1 + 0.5194 \rho_s)^{-3/2}$$
 (S4)

The conversion of  $h_i$  to SIT then invokes the floe's hydrostatic equilibrium and Archimedes' principle. Like the freeboard correction for slower radar pulse propagation in snow, this operation requires a priori knowledge of the depth and density of the snow cover.

$$SIT = h_i \frac{\rho_w}{\rho_w - \rho_i} + h_s \frac{\rho_s}{\rho_w - \rho_i}$$
 (Tilling et al. 2018) (S5)

15 Separating  $h_i$  into its  $h_r$  and  $\delta h_{prop}$  components using Eq. (S2), we can express SIT for a given ice type as a linear combination of the radar freeboard and snow properties.

$$SIT = h_r \frac{\rho_w}{\rho_w - \rho_i} + h_s \frac{\rho_w}{\rho_w - \rho_i} \left[\frac{c}{c_s} - 1\right] + h_s \frac{\rho_s}{\rho_w - \rho_i}$$
(S6)

$$SIT = h_r \frac{\rho_w}{\rho_w - \rho_i} + h_s \frac{\rho_w}{\rho_w - \rho_i} \left( \left[ \frac{c}{c_s} - 1 \right] + \frac{\rho_s}{\rho_w} \right)$$
(S7)

The equation  $y = c/c_s - 1$  where  $c_s$  is a function of  $\rho_s$  as in Eq (S4) is highly linear as a function of  $\rho_s$  as follows:

20 
$$\frac{c}{c_s} - 1 = 8.36 \times 10^{-4} \times \rho_s$$
 (S8)

This linearity is visualised in Fig. (S1) and allows the second term in Eq. (S7) to be written to a close approximation:

$$SIT = h_r \frac{\rho_w}{\rho_w - \rho_i} + m_s \frac{\rho_w}{\rho_w - \rho_i} \left( (8.36 \times 10^{-4}) + \frac{1}{\rho_w} \right)$$
(S9)

Where  $m_s$  represents the mass of snow per unit area. This can be reformulated by setting  $\rho_w = 1023 \text{ kgm}^{-3}$  as:

$$SIT = h_r \frac{\rho_w}{\rho_w - \rho_i} + m_s \frac{\rho_w}{\rho_w - \rho_i} \times 1.81 \times 10^{-3}$$
(S10)



**Figure S1.** Value of the propagation factor used to convert radar freeboard to ice freeboard, plotted as a function of snow density. This function is highly linear and is approximated as such in this work. The factor is multiplied by the snow depth to generate the total correction.



**Figure S2.** The number of  $\overline{RF}$  25×25 km data points in each region for each month. We were not able to compute  $\overline{RF}$  in the Kara Sea for October 2009 or 2012. Nor were we able to calculate it in the Barents Sea in October after 2008 (with the exception of 2011 and 2014).



Figure S3. (a) difference in snow depth in SnowModel-LG when driven by ERA5 and Merra2 reanalysis data at each 25x25 km pixel on the EASE grid averaged over the period 2002-2018. (b) time average of absolute differences in SnowModel-LG when driven by ERA5 and Merra2 reanalysis data. We note that (b) is not the absolute value of (a), but instead the time-average of the absolute values of monthly differences.



**Figure S4.** Basinwide trends in first year ice extent as a fraction of total extent from 2003-2018. Statistically significant trends exist in October (declining) and January (increasing). When trends of any significance are considered, all months show positive slopes barring October, which shows distinct decline. The October trend is due to later freeze-ups, the other positive trends fit in with established trends of increasing FYI dominance. Shaded regions represent the 95% confidence level for the linear regression.



Figure S5. Basinwide trends in mW99 SWE fields from 2003-2018. A statistically significant trend only exists in October, where SWE is increasing due to the increasing dominance of MYI in the month due to later freeze-ups. Shaded regions represent the 95% confidence level for the linear regression.


**Figure S6.** Detrended timeseries of spatially averaged snow contributions to sea ice thickness (Snow) by region from W99 (blue) and SnowModel-LG (red). Standard deviation values are displayed for SnowModel-LG (lower left, red), and mW99 (lower right, blue)



Figure S7. Detrended timeseries of spatially averaged snow contribution to sea ice thickness ( $\overline{Snow}$ ) from W99 (blue) and SnowModel-LG (red) over first year ice. SnowModel-LG is significantly more variable from year to year than W99, which only varies due to shifting dominance of ice types. This increased variability propagates through to sea ice thickness, but is moderated by its covariance with radar freeboard variability. The standard deviations of the two timeseries are displayed in the lower corners of each panel.



**Figure S8.** Detrended timeseries of spatially averaged snow contribution to sea ice thickness  $(\overline{Snow})$  from W99 (blue) and SnowModel-LG (red) **over multiyear ice** (MYI). SnowModel-LG is significantly more variable from year to year than W99, which only varies due to shifting dominance of ice types. This increased variability propagates through to sea ice thickness, but is moderated by its covariance with radar freeboard variability. A substantial number of data points are missing from some panels - these absences reflect months where no MYI is present in the relevant region. The standard deviations of the two timeseries are displayed in the lower corners of each panel.



Detrended SnowModel-LG Snow Contribution to FYI SIT

**Figure S9.** FYI correlations between radar freeboard and snow contributions to sea ice thickness, where the snow contribution is calculated using SnowModel-LG. All statistically significant correlations are positive (i.e. years with more snow exhibit higher radar freeboards). A persistent, positive correlation exists in the Central Arctic and the East Siberian Sea in the last five months of winter. The Barents and Kara Seas both exhibit significant correlations in the last two months of winter. The Baufort sea exhibits no months of statistically significant correlation between radar freeboard and snow contributions.



Detrended SnowModel-LG Snow Contribution to MYI SIT

**Figure S10.** MYI correlations between radar freeboard and snow contributions to sea ice thickness, where the snow contribution is calculated using SnowModel-LG. Fewer correlations exist for MYI than for FYI. The Central Arctic and Chukchi Sea exhibit no correlations between snow and radar freeboard contributions.



**Figure S11.** Regional IAV displayed by ice type. MYI represented by orange points, FYI represented by purple. When averaging over the growth season in a given region, MYI is more variable in all the marginal seas.



Figure S12. Detrended timeseries of spatially averaged sea ice thickness  $(\overline{SIT})$  by region from W99 (blue) and SnowModel-LG (red) for all ice types. Standard deviation values are displayed for SnowModel-LG (lower left, red), and mW99 (lower right, blue).



**Figure S13.** 2010-2018 basin-wide sea ice thickness distribution calculated using both mW99 and SnowModel-LG data expressed as total sea ice area of all grid cells falling into a specific SIT bin. Bin size is 5 cm. Shaded areas represent the area constituted by the Central Arctic.



**Figure S14.** Seasonal evolution of (a) snow thickness and (b) sea ice thickness by region. All regions calculated over 2002-2018 with the exception of the Central Arctic, which is 2010-2018. Note different y-axis scales for Central Arctic panels. 'Error bars' represent the one standard-deviation range either side of the mean value for the timeseries. The SnowModel-LG contribution starts lower but ends higher in the Central Arctic, the region that dominates Pan-Arctic statistics. This is also true for the Marginal Seas grouping, but not necessarily true for the individual constituent regions. This corresponds to faster thickness increase than would be calculated with W99.



**Figure S15.** Interannual variability of SnowModel-LG contribution to  $\sigma_{SIT}^2$  ( $\sigma_{Snow}^2$ ) when forced by two different reanalysis data sets. MERRA2 (orange) and ERA5 (green) produce very similar variability.



**Figure S16.** Trends in sea ice thickness (2002-2018) by region, when calculated using SnowModel-LG runs using two different sources of reanalysis (ERA5, Purple; MERRA2, Orange). Panels are framed with green where statistically significant trends exist independent of reanalysis choice. Purple (orange) frames represent month/region pairs where statistically significant trends are only present with ERA5 (MERRA2). Slope values are given where significant in the lower corners. All significant trends in the marginal seas are negative, all significant trends in the Central Arctic are positive. In the Central Arctic, two of the four statistically significant increasing trends are only evident with ERA5 reanalysis. In the Marginal Seas, the decline in some months is only statistically significant with MERRA2.



**Figure S17.** Trends in snow contribution to sea ice thickness ( $\overline{Snow}$ ; 2002-2018) by region, when calculated using SnowModel-LG runs using two different sources of reanalysis (ERA5, Purple; MERRA2, Orange). Panels are framed with green where statistically significant trends exist independent of reanalysis choice. Purple (orange) frames represent month/region pairs where statistically significant trends are only present with ERA5 (MERRA2). Slope values are given where significant in the lower corners.



**Figure S18.** Interannual variability of NESOSIM data's contribution to SIT, shown as (a) absolute contribution to SIT variability, and (b) relative contribution. Variability from snow is of a similar magnitude to that of SnowModel-LG, although regional differences exist between the corresponding plots, particularly in the Barents Sea. As well as differences in the snow accumulation scheme, the two data sets differ in spatial resolution and the timespan over which they are analysed.



**Figure S19.** Timeseries of the thickness contributions of radar freeboards ( $\overline{RF}$ ) and snow ( $\overline{Snow}$ ) over all ice types. Orange framed boxes indicate statistically significant decline in both  $\overline{RF}$  and  $\overline{Snow}$ . The red framed box indicates statistically significant decline in  $\overline{Snow}$  only. No boxes feature a statistically significant decline in  $\overline{RF}$  without a concomitant decline in  $\overline{Snow}$ . All statistically significant trends in both  $\overline{Snow}$  and  $\overline{RF}$  are negative.

## 25 References

Armitage, T. W. and Ridout, A. L.: Arctic sea ice freeboard from AltiKa and comparison with CryoSat-2 and Operation IceBridge, Geophysical Research Letters, 42, 6724–6731, https://doi.org/10.1002/2015GL064823, 2015.

Mätzler, C.: Thermal microwave radiation: applications for remote sensing, vol. 52, Iet, 2006.

Tilling, R. L., Ridout, A., and Shepherd, A.: Estimating Arctic sea ice thickness and volume using CryoSat-2 radar altimeter data, Advances

- in Space Research, 62, 1203–1225, https://doi.org/10.1016/j.asr.2017.10.051, https://doi.org/10.1016/j.asr.2017.10.051, 2018.
  Ulaby, F. and Long, D.: Microwave Radar and Radiometric Remote Sensing, The University of Michigan Press,
  - https://doi.org/10.3998/0472119356, http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=4537961, 2014.