Responses to referee #3

Dear Dr. Daniel Price:

Thanks for your helpful comments to improve this manuscript.

Below, we repeat each comment and insert our replies in the text. All responses are in blue font for clarity of reading.

Qinghua Yang

On behalf of all the authors

Main Comments:

Point 1: Readers may benefit from a concise explanation of the general principles of sea ice thickness estimates from reanalysis products. It will be difficult for non-experts in this specific field to grasp the processes considered and limitations during the construction of a sea ice thickness reanalysis product. I appreciate this is the point of the references but it is often helpful to provide an insight as part of the text to assist the reader (supported by references)

Response: Good suggestion. We added some description:

Sea ice thickness is a prognostic variable in all ocean—sea ice models used to generate the reanalyses considered in this study. The use of a data assimilation scheme offers the possibility to provide revised estimates of sea ice thickness, by constraining the simulated model output with observations (ocean or sea ice, e.g., Sakov et al., 2012; Köhl, 2015; Mu et al., 2018).

Added reference:

Köhl, A., 2015: Evaluation of the GECCO2 ocean synthesis: transports of volume, heat and freshwater in the Atlantic. Q. J. Roy. Meteor. Soc., 141, 166-181.

Mu, L., M. Losch, Q. Yang, R. Ricker, S. N. Losa, and L. Nerger, 2018: Arctic-wide sea ice thickness estimates from combining satellite remote sensing data and a dynamic ice-ocean model with data assimilation during the CryoSat-2 period. J. Geophys. Res.-Oceans, 123, 7763-7780.

Sakov, P., F. Counillon, L. Bertino, K. A. Lisaeter, P. R. Oke, and A. Korablev, 2012: TOPAZ4: An ocean-sea ice data assimilation system for the North Atlantic and Arctic. Ocean Sci., 8, 633-656.

Point 2: I appreciate it is sometimes difficult to fit all the relevant information into the limited word count of an abstract, but I think the reader (and work) would benefit from some sort of quantified reporting in the abstract. Terms like 'well reproduce' are somewhat subjective, is there a way to effectively provide a quantitative measure in the abstract of how well these reanalysis perform compared to one another and the observations? i.e. report the key results in a quantitative manner. This could in some way be related to a 'score' suggested in point 5 below

Response: Good suggestion. A quantitatively description of the comparison results are necessary to know the performance of all four reanalyses. After a serious thinking, we tend to present values of root-mean-square error (RMSE) and correlation coefficient (CC), which are more objective instead of "score" ranking. It is noted that the CC with ULS means the temporal correlation between four reanalyses and ULS, while the CC with ICESat-1 means the spatial correlation because they are calculated by yearly mean SIT fields. Our results (Table 5) show that the SOSE has the highest CC of 0.77 and lowest RMSE of 0.72 m, when compared with

ULS ice thickness. All RMSEs are less than 0.9 m and all CCs are more than 0.4. Compared with ICESat-1, NEMO-EnKF has the highest CC of 0.54 and lowest RMSE of 0.44 m. CCs of the other three reanalyses are less than 0.3 and GIOMAS almost no spatial relation with ICESat-1.

By the way, we will add the RMSE anc CC results in the new abstract. But we cannot conclude the performance of reanalyses only by their RMSEs or CCs. First, the time coverage of ICESat-1 is quite limited. Second, the spatial representation of ULS data sets is very sparse.

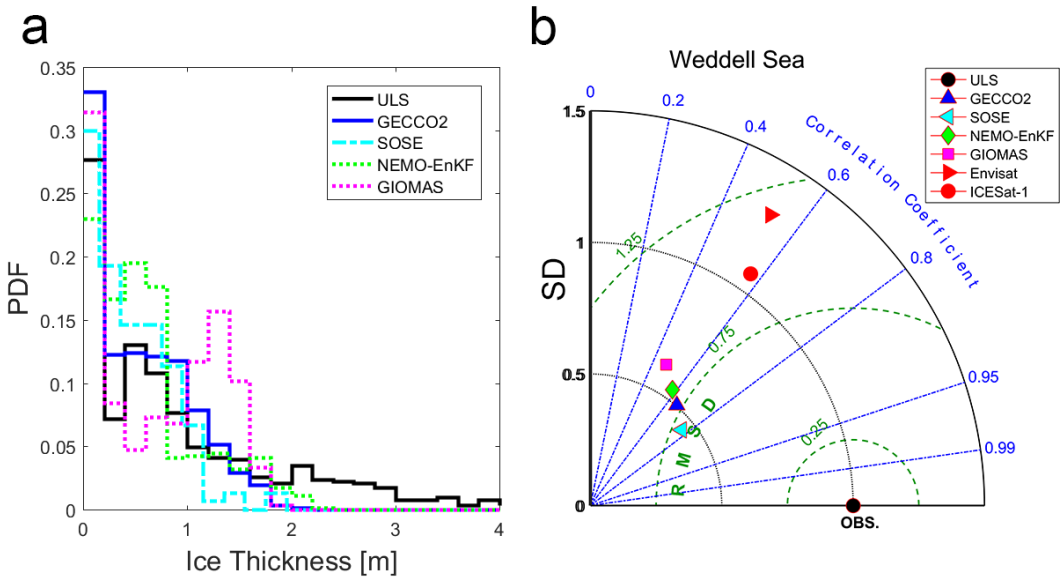
Table 5. Statistics of four reanalyses with respect to ULS and ICESat-1.

	GECCO2	SOSE	NEMO-EnKF	GIOMAS
ULS(RMSE)	0.77	0.72	0.82	0.89
ULS(CC)	0.65	0.77	0.58	0.47
ICESat-1(RMSE)	0.55	0.51	0.44	0.47
ICESat-1(CC)	0.19	0.26	0.54	0.03

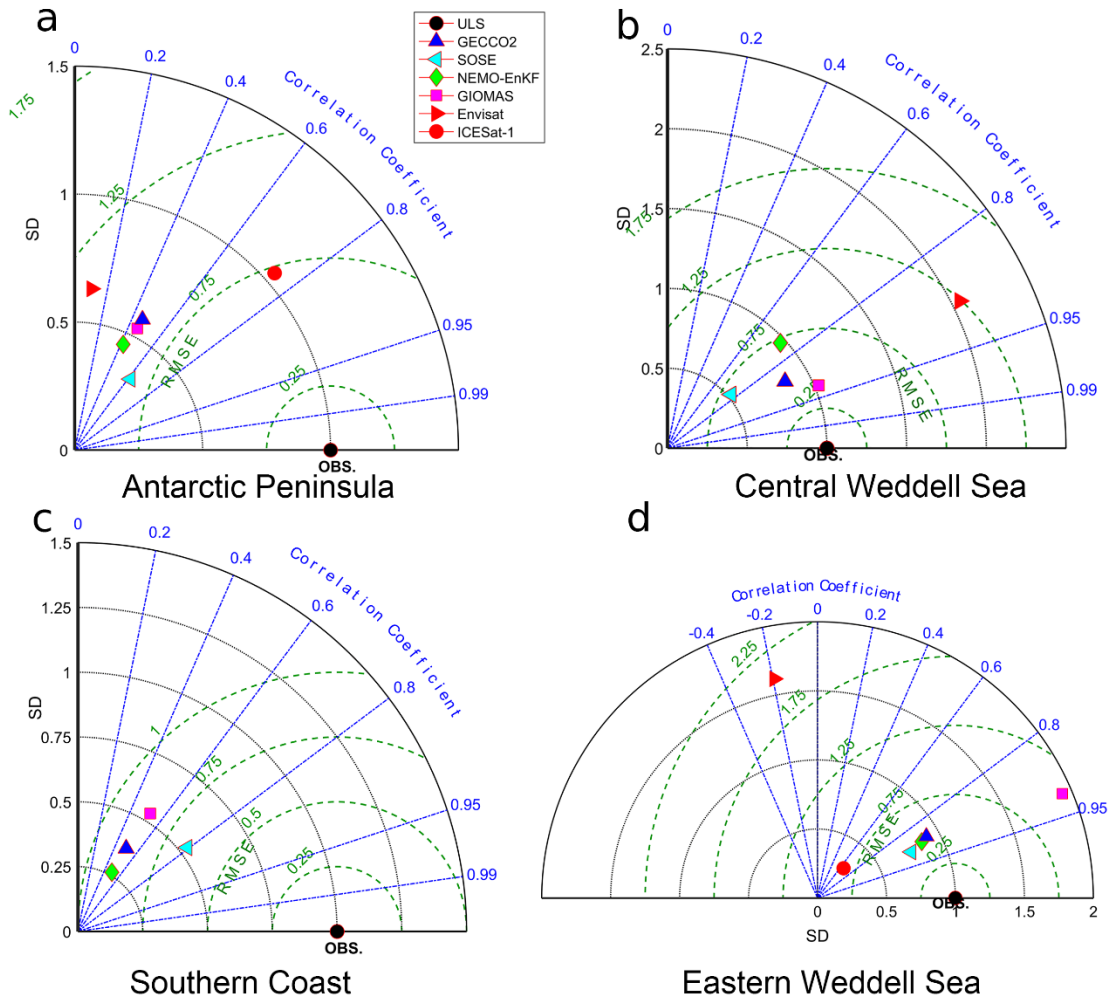
Point 3: Although the manuscript is written well and results are well displayed, it would be useful to maintain the colour coding of each reanalysis/observational datasets throughout the figures to avoid confusion

Response:

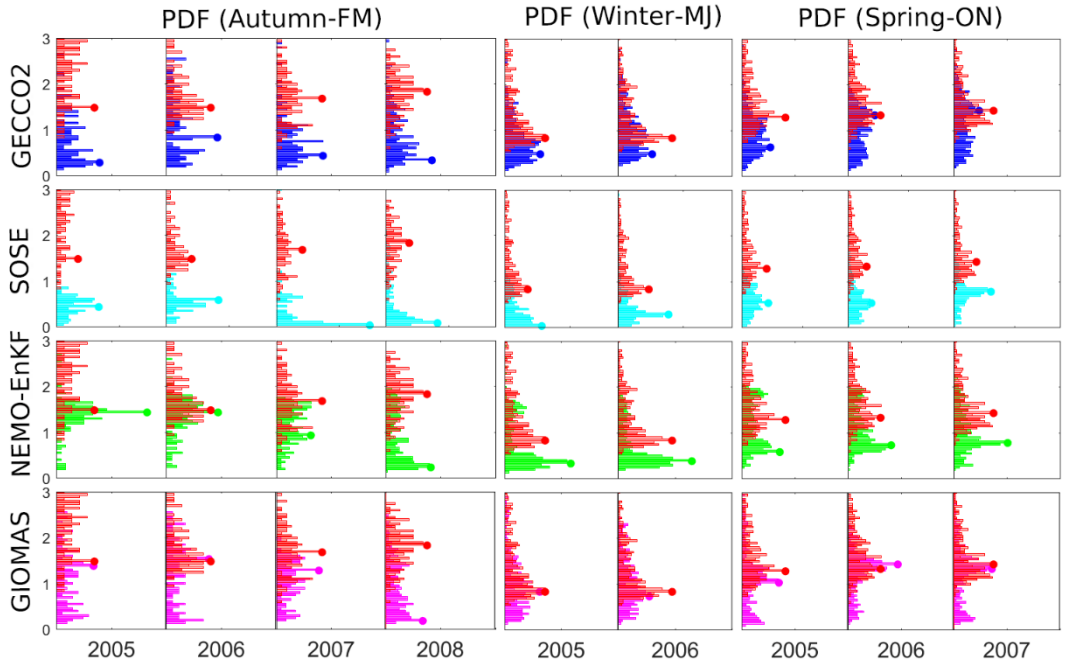
Corrected. In the new version, the ULS/ASPeCt data is colored in black and ICESat-1/Envisat is colored in red. Please see the new Figure 2 and 3. Besides, these color codes are applied to ITDs in the new Figure 5.



New Figure 2: a) Probability density distributions (PDF) of monthly sea-ice thickness from ULS and four reanalyses at the 13 ULS locations of the Weddell Sea. b) Normalized Taylor diagram for monthly sea ice thickness of four reanalyses as well as Envisat and ICESat-1 with respect to the sea-ice thickness from upward-looking sonar from 1990 to 2008 in the Weddell Sea. The green dashed lines indicate the normalized root-mean-square error (RMSE).



New Figure 3: Same as Figure 1b, but for the four sub-regions: a) Antarctic Peninsula, b) central Weddell Sea, c) Southern Coast, and d) eastern Weddell Sea.



New Figure 5: The variation of monthly ice thickness distribution from GECCO2 (blue), SOSE (cyan), NEMO-EnKF (green), GIOMAS (pink) and ICESat-1 (red) in Autumn-FM (left), Winter-MJ (middle) and Spring-ON (right). The colored dots represent the modal ice thickness. In order to make the histogram plots readable, different reanalyses has different x range.

Point 4: It is clear that the reanalyses underestimate the sea ice thickness distribution when compared to ICESat-1 but maybe some attention should be given to the description of this comparison in the results. "ICESat-1 thickness is much thicker than that of the reanalyses except GIOMAS in Spring-ON" – is this entirely accurate? GECCO2 also has two instances where the modes are similar when compared to ICESat-1, in Spring-ON (2007) the mode is higher in GECCO2, Spring-ON (2006) it appears to be the same. NEMO-EnkF also has two examples in Autumn-FM where the modes are higher (2006) and similar (2005). Is GIOMAS in Spring-ON really that notable? I am a little confused by the plots in Figure 5 from visual inspection - why does the ICESat-1 thickness distribution change in the same season and year for the SOSEcomparison? i.e. the ICESat-1 distributions seem to stay the same in the PDFs in the same year/season for the other plots but the distribution is shifted in the SOSE plots. Is this to do with some different sampling from different ge ographical coverage of the reanalyses? In addition, why is 2007/2008 omitted for Winter-MJ and 2008 for SpringON? Was this decision driven by a lack of ICESat-1 data for comparison? It is stated in the text (L264) that the 'we compare sea-ice thickness from four renanalyses: : :with that from ICESat-1 for the period from 2005 to 2008' but this does not appear to be the case in the corresponding figure. In Table 2 the ICESat-1 measurement periods are described and 2007 (Winter-MJ) does not have a '-' indicating the data is absent but instead 'Winter-MJ' is written. Also 2004 is shown but does not appear to be part of the described analysis, is there a reason for this? My main point related to this section of the study is that there seem to be some discrepancies in how the data is described and how it is presented in figures. This needs to be looked at and all data and descriptions must be consistent

Response: We are very sorry as we made a mistake in the old Table 2. In fact, there are no ICESat-1 measurements in the 2007 winter-MJ. We only compare four reanalyses with ICESat-1 from 2005 to 2008 to fit for SOSE time coverage as well as to be comparable with Figure 6. In order to make the histogram plots in Figure 5 more readable, we use a different x range for histogram plots of different reanalyses. The key point of Figure 5 is the location of modal ice thickness, which is the positions of colored circle dots. In addition, the descriptions of the differences between reanalyses and ICESat-1 were improved in the new version based on the new ranking results and the new monthly averaged approach.

Point 5: I would expect that the community would look to evaluations like this to understand what reanalyses could be useful for supporting their own work. As the manuscript currently reads, it is difficult to digest and really understand the limitations of each of the reanalyses. It may provide some clarity and assist the readers understanding of the results to have a table with the key parameters the authors are trying to evaluate (including but not limited to thickness, relationship between mean/mode, min/max thickness accuracy, spatial accuracy, sea ice growth/seasonal evolution of thickness, open ocean vs. coastal regimes, ice motion –divergence/convergence) and a score evaluating how well they have performed. This is not an explicit request, but just a suggestion for the authors to consider in order to improve the communication of important information from this work.

Response: Thanks and please see the response in Point 2.

Minor Comments:

1 L28: 'crucial component of the Earth system', understand what authors mean but perhaps more

specific ‘climate system’ for example.

Response: Corrected.

2 L89: add ‘a’ between ‘introduce’ and ‘sea-ice’.

Response: Corrected.

3 Section 2.1 ~ L110 and L130: should a spatial resolution be reported for GECCO2/GIOMAS as is given for the other reanalyses? I see they are in Table 1 but why report some resolutions in the text and not others?

Response: We added some description on the resolution of GECCO2 and GIOMAS in the new version.

4 L118 and L124: ‘â ¸A˘ r’ used in one instance and ‘degrees’ in another, perhaps adopt one standard

Response: we now use “°” in Line 124.

5 L136: to be absolutely accurate perhaps reword ‘(the part above the sea level)’ to ‘(combined ice and snow height above local sea level)’.

Response: Corrected.

6 L140: ‘suggested by Worby’? Is a complete reference available?

Response: Corrected. The complete reference is Kern et al. (2016) in Line 141. In the new version, we deleted “suggested by Worby” in Line 140.

7 L149: I understand that the limitations of radar altimeters are not the focus of this study but the complexity of the technique/its limitations in the Antarctic are understated by these few sentences. Perhaps include reference to other studies highlighting this to provide the reader with some context if they require it. This takes me to another point, it doesn’t appear CS-2 is used in the analysis, why is it described in the data section?

Response: Corrected. On the one hand, we added “the uncertainties of the radar altimeter can result from the inaccuracy snow-ice interface and snow-ice formation (Willatt et al., 2010), and also the surface type mixing and surface roughness (Schwegmann et al., 2016; Paul et al., 2018; Tilling et al., 2019).” On the other hand, we will delete the description about CS2 in the Antarctic since we have not used them.

Ref:

1.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 Trans. Geosci. Remote Sens., 48, 365–372, https://doi.org/10.1109/TGRS.2009.2028237, 2010.

2.Schwegmann, S., Rinne, E., Ricker, R., Hendricks, S., and Helm, V.: About the consistency between Envisat and CryoSat-2 radar freeboard retrieval over Antarctic sea ice, The Cryosphere, 10, 1415–1425, https://doi.org/10.5194/tc-10-1415-2016, 2016.

3.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, 2018.

4.Tilling, R., Ridout, A., and Shepherd, A.: Assessing the Impact of Lead and Floe Sampling on Arctic Sea Ice Thickness Estimates from Envisat and CryoSat-2, J. Geophys. Res., 124,

7473–7485,

8 L149: More accurate to say 'the radar altimeter is expected to' (and then provide relevant references) as opposed to 'the radar altimeter can measure'

Response: Corrected.

9 L159: Use acronym 'ULS' once it is provided and throughout manuscript use acronyms/abbreviations once they are supplied e.g. L199 'Antarctic Peninsula' to 'AP' as it is shortened on L195

Response: Corrected. In the new version, we use "ULS" instead of "upward looking sonar" in Line 82, 97, 158, 159, 192 and 193. We use "AP" instead of "Antarctic Peninsula" in Line 195, 199, 220, 236, 292, 338 and 384. We use "CWS" instead of "central Weddell Sea" in Line 200, 222, 239, 242, 292 and 381. We use "EWS" instead of "eastern Weddell Sea" in Line 199, 239, 241, 243, 258, 296, 317, 340, 345 and 379. We use "SC" instead of "southern coast" in Line 200 and 226.

10 L164: I don't think 'skilful' is appropriate here, perhaps 'accurate' or 'approximates thickness well' or something similar

Response: Corrected.

11 L167: What are these uncertainties? 5 cm/8cm/18 cm etc? Are they a spread around the mean +/- 5 cm or direct positive deviations from other reference measurements? If so are there references for these expected accuracies?

Response: Following Behrendt et al. (2013), the accuracy of the ice draft is ±5 cm in the freezing/melting seasons and ±12 cm in winter. Then, based on the linear regression function (Eq. 1 in Line 162) between ice draft and ice thickness, the accuracy of ULS ice thickness is ± 8 cm in the freezing/melting seasons and ±18 cm in winter.

12 Figure 1: Standard deviation is abbreviated to SD in the figure but to STD in the text, these should be consistent

Response: Corrected.

13 Figure 3: caption – Capitalise 'southern coast'.

Response: Corrected and the new caption will be change to: "Figure 3: Same as Figure 1b, but for the four sub-regions: a) Antarctic Peninsula, b) central Weddell Sea, c) Southern Coast, and d) eastern Weddell Sea."

14 Figure 5: Thickness (m) is not actually labelled on the y-axis. Insert '(red)' after second mention of ICESat-1 in the caption.

Response: Corrected.

15 L290: 'this means that the reanalyses may not well represent coastal processes' – what do the authors specifically mean here in reference to sea ice? Dynamics and convergence against the coast? Interaction with the coastline? Inaccurate bathymetry or coastal currents? Some of the concluding statements are a little vague. I think the study would benefit from being more specific and shed light on the limitations that need to be addressed.

Response: Exactly as what the reviewer pointed, these processes could all contribute to the underestimation, however, it is rather difficult to distinguish. For example, we do not know whether each model smooths the bathymetry, and actually how they manipulate the bathymetry can dramatically change the coastal current and sea ice. The only thing clear now is that it

should be lack of accurate sea ice dynamics rather than thermodynamics. We now refined the sentence to be: " this means that the reanalyses may not well represent the coastal sea ice dynamical processes.".

16 L290: Why is SOSE not included in the spatially averaged differences here?

Response: Sorry, we forgot to list the bias of SOSE. The new description should be "The spatial averaged differences between models and ICESat-1 are -1.30 m (GECCO2), 1.42 m (SOSE), -0.63 m (NEMO-EnKF) and -0.75 m (GIOMAS), respectively".

17 L325: 'primary' to 'prime'.

Response: Corrected.

18 L362: insert 'satellite' before 'altimeters'

Response: Corrected.

19 L369: 'still' before 'been'.

Response: Corrected.

20 L388: 'improve' to 'improving'

Response: Corrected.

21 L388: 'assimilate' to 'assimilating'.

Response: Corrected.

22 Figure 6: What time period is this data comparison for? Are they seasonal averages

for all years?

Response: The time period is from 2005 to 2008. They cover the irregular months listed in Table 2.

23 In acknowledgements: ICESat-1 data is provided by NASA and NSIDC not ESA. REFERENCED EXISTING REANALYSES

Response: Corrected.