Comment on tc-2021-228
Anonymous Referee #1

Referee comment on "An Evaluation of Antarctic Sea Ice Thickness from the Global Ice-Ocean Modeling and Assimilation System based on In-situ and Satellite Observations" by Sutao Liao et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2021-228-RC1, 2021

Review of "An Evaluation of Antarctic Sea Ice Thickness from the Global Ice-Ocean Modeling and Assimilation System based on In-situ and Satellite Observations" by S. Liao et al.

The authors compare Antarctic sea-ice thickness (SIT) in the global sea-ice reanalysis GIMMAS with observational estimates from satellite radar altimetry and various in-situ observations. They find that anomalies and trend of Antarctic sea-ice volume (SIV) match quite well between model and observations. However, compared to satellite observations, the model simulates much lower mean SIV and SIT. Variability in the model also seems to be lower than observed, at least in some regions. Some hypothesis on the reasons for the discrepancies between model and observations are stated but not tested.

The study addresses an important gap in our knowledge of the Earth system: the thickness of Antarctic sea ice. It makes good use of novel satellite observations and the available in-situ observations and documents several intriguing facts such as the link between variability and mean state. It thus shows good potential. However, the manuscript is lacking maturity and substance. It stays purely descriptive and misses the chance to better understand discrepancies, which would really help the community to make progress with developing better observational and modelling products of Antarctic sea-ice thickness. There are also some methodological shortcomings when it comes to comparing model data with in-situ observations. I recommend publication of the manuscript after a major revision that addresses my concerns stated below.

Major comments:

1) p 5 to p6 l 165 and elsewhere: I appreciate the basic description of discrepancies between observed and modelled SIV and SIT, but more analysis on the reasons for the discrepancies is needed in order for the paper to make a substantial contribution to the field. As it stands, only hypotheses and speculations are offered (e.g. p 4 l 131f "These differences may lead to overestimation...", p 4 l 142f "... which may be due to model bias ...", p 4 l 149f "... may be partially attributed to the uncertainties of SIT retrieved...").

I would appreciate if the authors could provide at least some attempts to disentangle the
respective roles of observation errors, model errors and insufficient data assimilation methods. To this end, I would suggest to
1a) discuss sea-ice thickness in a model run /without/ assimilation of sea-ice concentration. A complementary approach would be to discuss the assimilation increments (here: nudging tendencies) for sea-ice concentration. Assuming that GIOMAS still uses the original PIOMAS nudging methods (Lindsay and Zhang, 2006), there will also be implicit increments on SIT and SIV. It would be good to also discuss the implicit increments on SIT and SIV that arise from SIC the nudging, as their impact on the mean state might be substantial.
1b) better quantify observational uncertainty. From the figures presented, it is evident that there are major discrepancies between ES and CS2-derived SIT. The same is likely true when comparing the various in-situ observations with the satellite records - can we have some scatter plots of these please? To clarify, I am not asking for an intercomparison of observational products - this would clearly be outside the scope of what the authors wanted to present. However, it is necessary to present the model-observation discrepancies in the context of the observational uncertainties, so there needs to be some quantitative discussion of these.
1c) use at least some of the in-situ observations (the ones who are judged most reliable) as ground truth and verify both satellite observations and GIOMAS against these. This would allow to draw some much-needed conclusions on whether it is the observational data sets or the model (or both) that need to be improved.

2) The discussion of trends (p 6 ll 166 - 175) is very interesting but disappointingly rudimentary. I consider Figure 5 a success for GIOMAS and a central result of the manuscript: it convincingly shows that year-to-year variability as well as trends in GIOMAS and the satellite record correspond quite well. This is an important message, but it needs to be backed up with more analysis (see minor comments). I would also suggest to dedicate a separate subsection to the trends.

3) Comparing monthly-mean model fields to in-situ observations at a specific point in space and time might introduce substantial errors, because you are not comparing "like with like". Is it true that the authors use monthly-mean fields from GIOMAS? If daily-mean fields are available, these really ought to be used to compare to in-situ data. The mean and variability of sparsely sampled instantaneous in-situ observations might not meaningfully represent the monthly mean and variability over a large-scale region like a model grid cell. Please specify and discuss how exactly you match in-situ observations to model fields and give some justification why you think the direct comparison is not misleading (see e.g. Janjic (2018) for an introduction/overview of the problem).

Minor comments:
- The manuscript would benefit from language editing, as there are various places with small grammatical errors and slightly inappropriate word choices. Nothing serious though. To give two examples: p 1 l 9f: "not very clear" is too colloquial and imprecise. p 2 l 34: "limited by the short of..." needs to be replaced by "limited by the lack of..."
- p 3 l 81: nudging of SIC is not state of the art anymore, it can introduce grave problems with SIT (see Tietsche et al. 2012). Please add some discussion on implicit changes to SIT when nudging SIC in your system.
- p 4 l 115 "grid of observations": in-situ data are not gridded. Please rephrase, e.g. "...take place in observation space, which means that GIOMAS was converted to the locations of the observations"
- p 4 l 119: You claim the difference ES - CS2 is "much less" than the variability, but 574 is almost half of 967! Can you please rephrase to acknowledge that there is considerable differences in CS2-ES SIV? I would like to see a figure with the time series of SIV anomalies as well.
Please be more precise on your method to "splice together" the two data sets. The problem with the manuscript that I describe above in major comment (1) is most evident in this paragraph. There is lots of speculation of where discrepancies may come from, but no further insight offered at all. Look at these phrases: "... may be due to model bias...", "... maybe caused by smaller freeboard of CS2 than ES2...", "... still disputed whether radar altimeter signals originate from the snow/ice or snow/air interface...". Without at least an attempt to decide whether discrepancies are due to deficiencies in GIOMAS or the satellite observations, it is difficult to draw any conclusion on where improvements are needed. Maybe it makes sense to treat (at least some of) the in-situ observations as the ground truth, and verify both model and satellite observations against them?

Can you please show a figure of Antarctic SIE, to clarify whether SIV anomalies and trends are mostly explained by SIE anomalies and trends, or there is some independence? A scatter plot with SIE/SIV anomalies would also be helpful.

The trend lines after 2013 are quite different. Can you please give the numbers for the trends in GIOMAS and the satellite record, and add to the discussion that GIOMAS seems to underestimate the trend seen in the satellite record?

The variability of SIT anomalies in satellite observations quite probably have a contribution from measurement noise, which the model should not try to imitate/simulate. Can you offer some comment or quantification on that point?

"... which is consistent with Fig. 4b."": This looks like a mistake. Figure 4b shows difference in mean not difference in variability. Please clarify and correct.

I agree there appears to be a relationship to some extent, but the authors might be over-simplifying this relationship. Look at ULS 232: GIOMAS underestimates variability by a factor of 3, but has no bias. To clarify this, can we please have a scatter plot of standard deviation ratio GIOMAS/satellite versus mean bias?

Can you please add a paragraph with the conclusions?

References:
