

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2022-261-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-261**

Ellyn Enderlin (Referee)

Referee comment on "Mass balance of the Greenland and Antarctic ice sheets from 1992 to 2020" by Inès N. Otosaka et al., Earth Syst. Sci. Data Discuss.,
<https://doi.org/10.5194/essd-2022-261-RC2>, 2022

Summary

The paper describes the process by which 50 mass balance time series for the ice sheets were combined to produce a consensus estimate of ice sheet mass loss since the early 1990s and then summarizes the results and their implications. They find that the scientific community is in generally good agreement regarding rates of mass loss within and across the three different methodologies used for these estimates – altimetry, input-output, and gravimetry – with the largest disagreements for the East Antarctic Ice Sheet. The paper is concise and generally well-written with several summary tables and figures that aid the presentation. I appreciate the complex data wrangling that likely took place to produce this paper and thank the authors for producing an updated IMBIE dataset. I have a few major comments regarding the presentation of numbers throughout the text as well as some minor figure recommendations, as described below.

Major Comments

- At the end of the Data section, two different basin definitions are described but then it isn't clear how these are used in the analysis. The data all seem to be split according to GrIS, APIS, WAIS, and EAIS, not these smaller drainage basins. When are these different basins used? The basin use should be clarified in the Methods section.
- In the dataset descriptions and the results, you state that the input-output method provides annual temporal resolution but in the methods your explanation of the dataset integration describes all estimates as monthly. Are the input-output datasets monthly? Do they have regular temporal intervals? It would be helpful to add temporal sampling flag or some other indicator of temporal resolution to Table 1. Similarly, these datasets

are described as all relying on the same SMB model. That model should be explicitly stated since SMB is a tremendously important component of GrIS mass loss.

- Throughout the results, I was uncertain how to interpret some of the metrics presented as summaries for the datasets and their intercomparison. It seems like the maximum difference in datasets is often reported. Why is this used and not the median or the trend? Why report the average of the standard deviations of the datasets? For small sample sizes, the average may be highly skewed. Finally, what are the metrics presented for the aggregate datasets? Are they the mean +/- standard deviation? Is the standard deviation calculated using the standard deviations of the independent datasets or is it a metric of variability over time for the aggregate dataset?
- At the beginning of the discussion, the aggregate rates of mass loss are compared to trends in global sea level rise. In addition to their contribution to the trend, it would be helpful to know what fraction of annual sea level rise is driven by ice sheet mass loss.

Minor Comments

- lines 64-68: In the abstract you switch between stating mass change for GrIS as a positive mass loss number and for Antarctica as negative numbers to also indicate mass loss. Make sure you are consistent with sign convention throughout.
- line 99: I prefer the use of the Oxford coma in sentences because I think it makes them easier to read. It is apparently not favored by these authors and I normally accept that stylistic preference, but there are several instances in this paper where the additional coma would help with sentence flow. For example, I had to read this particular sentence a few times. I recommend it is changed to "...geophysical corrections, SMB models, or GIA models in ..."
- line 152: Instead of "1 input-output method estimate" you could say "the input-output method estimate"
- lines 125-144: I appreciate the summary of the methods and their strengths and weaknesses!
- line 190: How did you quantify linear model structural error?
- line 223: Why are you only reporting differences from 2007-2011? The previous sentence states the datasets have a much longer period of overlap.
- Figure 1: While I really like the idea of this table, I struggled to see the aggregate average (black) when there are a large number of gravimetry estimates (green). Consider changing the shade or saturation of the green color. The aggregate average also needs to be stated in the caption and the difference in y-axis scaling should be noted as well. This is a stylist preference but I recommend only plotting the y-axis labels once per side to reduce clutter.