+ Abstract: "16 algorithmically-generated flux gates, which are continuous around Antarctica" Generally flux gates are per glacier or basin so this is confusing without reading the paper - which should not be required to understand the abstract. What about "16 algorithmically generated overlapping inter-comparable flux gates, which..." or something a bit longer to clarify that it isn't 16 outlets?

We have modified this sentence to: “We calculate ice flux through 16 algorithmically-generated flux gates across 998 ice sheet, glacier, ice stream and ice shelf drainage basins”. We also note than in Antarctica, there is precedent for referring to flux gates in the manner we have here (Gardner et al., 2018).

Gardner, A.S., Moholdt, G., Scambos, T., Fahnstock, M., Ligtenberg, S., Van Den Broeke, M. and Nilsson, J., 2018. Increased West Antarctic and unchanged East Antarctic ice discharge over the last 7 years. The Cryosphere, 12(2), pp.521-547.

+ Zenodo URL: Maybe best to give the top/latest/permanent DOI but mention "this paper written using vX of the data" ?

Done

+ Does the pole hole (L112) cover any flux gates? If so, what % or weighted %? For the remainder of the velocity gap filling exercise, what % is filled? At L89 you mention the pole hole and 6 % weighted discharge contribution but this is for CryoSat-2, not velocity products.

Yes, this is a good point. The Sentinel-1 pole hole does cover some portion of the flux gates, amounting to 6.2 % of the Antarctic-wide discharge. We have added a note to that affect on line 167: “The forward-filling of the velocity time-series is used on all flux gate pixels south of 81.8° during the Sentinel-1 era, which contribute 6.2 % to our Antarctic-wide discharge”. The ITS_LIVE and MEaSUREs velocity products are comprised of measurements from multiple satellites, so the pole hole for those aren’t clear cut.

+ You describe "we use our reference velocity map" at line 166 near the end of 2.3, but this is 2+ pages after describing how you create it. Clarify why you generate a reference map earlier. Perhaps even why before how.

We agree that this was not an ideal structure. We have added a sentence at the beginning of section 2.3: “Prior to constructing an ice velocity and discharge time-series, we generate a reference velocity grid in order to fill gaps in the time-series products.” Thanks for the suggestion.

+ L22: The Antarctic Ice Sheet is losing mass at an accelerating rate (Diener et al., 2021; Otosaka et al., 2023; Shepherd et al., 2019; Slater et al., 2021). But wasn't there mass gain last year?

Fair point. We have modified the wording of this statement to: “The rate of mass loss from the Antarctic Ice Sheet has accelerated since the early-1990s”.

+ Fig 3: black dots show something? I don't see black dots. I see white dots, red dots, and blue dots.

Apologies, they are black circles, not black dots. We have modified the figure caption accordingly.

+ L152: Maybe (?) replace "typically indicate" with "have" (you started sentence with "generally" which may be a synonym with "typcially" and therefore this could be repetitive).

Good spot. Done.

+ based on a robust linear fit <-- what does "robust" mean here?

'Robust' linear regression is less sensitive to outliers than standard linear regression (ordinary least-squares fitting). Robust regression uses iteratively reweighted least-squares to assign a weight to each data point, and it is less sensitive to large changes in small number of elements in the dataset. We chose to use a robust fit not an ordinary least-squares fit because we align the datasets before removing outliers, which we in turn chose to do to maximise the amount of time-series information available to identify, remove and replace outliers. We note on line 157 that the 'robust' fit is an iteratively re-weighted least-squares algorithm.

+ Outlier removal discussed 2x in separate paragraphs (L153, 158). Maybe join?. Split "gap filling" into new paragraph near L162. Are these gaps you create from outlier removal? Or other gaps due to assigning velocity products to time-stamps rather than time-spans?. Or both?

We have described this stage of the processing chronologically (remove gross outliers using the reference grid –> align all velocity sources -> then use full time-series information to remove remaining outliers -> then fill the gaps), which we think is clearer to the reader than the alternative, though of course one could make an argument alternative structures.

We have split the gap-filling into a new paragraph as suggested. The gap-filling is done on all gaps, whether or not they were originally present or if they were created by removing outliers. We don't think it's necessary to specify that here, given that we have already stated on line 150 that we create a gapless velocity time-series.

+ How do you calculate EPSG 3031 grid distortion? What is the mean/median/min/max?

We assume this refers to our section on flux gate pixel spacing, now on line 184. We use the methodology described in Numerical Weather and Climate Predication (Thomas Tomkins Warner, 2011, Cambridge University Press). $m = (1 + \sin(trueLat)) / (1 + \sin(lat))$, where m is the ratio of distance on a polar stereographic projection to distance on a sphere (where $m<1$, projected distances are shorter than those on a sphere), and trueLat is the reference latitude of the projection (-71 for EPSG 3031). We did not retain the un-corrected distances or the correction ratio for the flux gates, and we do not think it would be very helpful for the reader to provide summary statistics for the distortion given that we remove the effect of it.

+ Eq 6. Not sure I agree with the edits (changing variable names). The output of this function is the discharge error (a function of velocity error, thickness error, and probably others not accounted for (see other comments), but inputs are velocity error and thickness error. What is "velocity-induced discharge error" isn't this just the "velocity error"? It is also the "velocity component of the discharge error", but no need to make it overly complicated.

1) We think it is worth distinguishing between error in the velocity and thickness measurements vs the resulting errors in the grounding line discharge because (for example) the velocity-induced discharge error resulting from a given velocity error is dependent on the ice thickness. The data-user is probably most interested in the velocity-induced discharge error and the thickness-induced discharge error, rather than the errors in the underlying thickness and velocity data.
2) We now use “velocity component of the discharge error” (also for thickness) as suggested.

We have substantially revised the section describing the derivation of the discharge errors to clarify our approach.

+ It also isn't clear to me where the "uniformly distributed random numbers" come from. That sentence explains "from time-stamped velocity error". But at each gate pixel at each time, how many velocity products (or velocity error products) exist? You list 8 at lines 127 -- 144 but not all exist everywhere or at all times, so I don't see how a normal distribution can be generated from them.

At each pixel and each time we have one measurement of cross-gate velocity error (incorporating the error in the vx and vy components of the velocity) and one measurement of thickness error (incorporating the bed error, surface error, firn error and dH/dt error). We want to know how those velocity and thickness errors impact grounding line discharge. One approach (used in your Greenland discharge datasets) is to add/subtract those errors from your central velocity and thickness measurement, to get upper and lower bounds on the grounding line discharge, given the errors in thickness and ice velocity. This approach therefore provides the discharge error under the assumption that the thickness and velocity value in every pixel could differ from the central estimate by the respective errors in the thickness and velocity. Our approach here is very similar, but we reasoned that the probability of every flux gate pixel being simultaneously different from the central estimate by the full error value was very small. Therefore, in every pixel and at all times, we take the velocity and thickness errors and say “those are the largest the errors can be in those pixels and at those times”. Then, through 100 iterations, we randomly generate new thickness and velocity errors in every pixel and at all times, assuming those upper bounds from the original component errors. The random generation of errors assumes uniformly distributed random numbers, not normally distributed random numbers. This could potentially be improved if one had information about the form of the error in the underlying data, but that information is not always available. This gives us 100 discharge estimates in every pixel and at all times, which represents the spread in possible discharge measurements given the velocity and thickness errors. We actually end up with 2x100 discharge estimates (one only using the velocity errors and one using the thickness errors). We then take the standard deviation of those 2x100 measurements as the velocity and thickness components of the discharge errors. We then combine those components in quadrature to get the discharge error in each gate pixel and measurement epoch. The basin-integrated errors are just $((D_{max}-D) + (D-D_{min}))/2$. As noted above, we have substantially revised the section describing the derivation of the discharge errors to clarify our approach.

+ Do you need to updated Eqs. 7 and 10 V_{\sigma} and H_{\simga} to the new D^2_{vel_\sigma} and D^2_{H_\sigm} in Eq. 6 (although I don't agree with those variable names)?

These equations describe the underlying thickness and velocity error, not their impact on the discharge estimate, so we think those variable names are correct. We do not have an equation to describe the Monte-Carlo approach, but we think the text description is sufficiently clear.

+ Fig 1: Ice thickness increase of ~100 % between FB and FBAdj? This seems extreme and warrants discussion. Is "unrealistically thin" on L71 the only discussion of this? You add Fig. 1 panel E in response to reviewer comments. Mention Fig A3 here as it contains more samples.

We agree this is extreme. We do raise this in our section 4.2 “implications for mass budget estimates”, where we write: “Given that the magnitude of the thickness adjustments in FrankenBedAdj exceed 50 % in some places, and that these are typically in locations where SMB and firn air content estimates from different regional climate models disagree the most (such as basin E-Ep in East Antarctica), we think that factors (1) and (2) likely contribute the majority of differences between mass budget approaches at the spatial scale considered here”.

+ In your response to reviewers you repeatedly mention that FrankenBedAdj compensates for uncertainties in mass change (from altimetry), ice thickness, and SMB. What are the relative %? In the abstract you answer this with """The uncertainties in our discharge dataset primarily result from uncertain bed elevation and flux gate location, which account for much of difference between our results and previous studies.""" If it's not all ice thickness (assigned to bed elevation, as opposed to thinning), is *Bed* the correct name for this term? Or should it be assigned a more general 'scaling factor' term such as k or alpha? L341 mentions PIG has 1.5 % error due to gates, but SMB error is likely larger than that at PIG and everywhere (SMB models have errors >> 1.5 %).

Section 4.2 discusses this and related issues at length, but given that the uncertainties in SMB are unknown, we don’t think it’s possible to provide a definitive response to this and we don’t think it would be meaningful to do this at a coarse scale because they are location specific (depending on the local accumulation rate, density of bed elevation observations, surface slope, surrounding topography, to name just a few). Indeed, if we knew the relative contributions with confidence everywhere, then we would not need FrankenBedAdj.

These points are somewhat separate from the sentence in the abstract quoted here, which focuses on the contributions to error in the discharge dataset and the magnitude of those errors relative to differences between our discharge estimate and other datasets – we have modified the abstract text to better reflect that.

The quoted figure for PIG is the relative difference in discharge between flux gates. We provided that to provide the reader with a means to evaluate the uncertainty in the grounding line discharge associated with gate location, which in turn reflects the quality of the thickness, velocity and SMB data. It indicates the precision of the discharge estimate rather than the accuracy. We agree the SMB models likely have larger errors than 1.5 % - those errors are not well quantified and are certainly not available at every RCM pixel and time step – but that doesn't necessarily translate to a large spread in flux gate discharge estimates.

+ Fig 10: Seasonality appears pot-2016, presumably due to Sentinel and not a change in the system. Can you estimate annual discharge vs. discharge sampled only when pre-2016 observations occur (summer only?) and from this estimate if pre-2016 estimates are low or high, and/or adjust uncertainty?

We could certainly compare annually-averaged discharge estimates after with those before, but we do not think this would be very informative. The annual mosaics from MEaSUReS and ITS_LIVE are, as far as we can tell from the available documentation, derived from image pairs in both and winter (and some spanning full years, summer to summer, or winter to winter) using a range of sensors. As mentioned in the main text, there are also systematic differences between each of the velocity products, which may be due to aliasing seasonal signals, but could also be due to various choices in the product development.

Additional private note (visible to authors and reviewers only):
+ R2 raised some generally good concepts (e.g., "more detail, more error analysis"), but I understand why you found it challenging to reply. Still, I wonder if in revision you can better address any of their concerns. As one easy example, they question your comment

"""Given that Antarctic Ice Sheet mass changes estimated from gravimetry and altimetry agree much more closely with each other than either do with the single available input-output mass change estimate"""

To which you point to Otosaka 2023 Fig. 3 and agree that EAIS is the major source of disagreement. You could clarify that, and that WAIS has IO and Gravimetry in closer agreement that Grav v. Alt, rather than not addressing their comment in the revised text. There may be other places where you can address some of their comments, even if not all - and you did, for example removing 'best' comment near the end.

We have substantially revised this paragraph discussing the differences amongst mass change estimates in different locations, which we think helps to more completely address this comment from the reviewer.