Authors response to Anonymous Referee #1

We would like to thank the referee for the time spent reviewing the manuscript in depth. From our point of view, the comments helped us to significantly improve our manuscript, in particular by refocusing the central objective of the article as suggested by the reviewer. We respond (in blue) to each of the reviewer's comments (in black) below and propose a revised manuscript, hoping to meet the expectations.

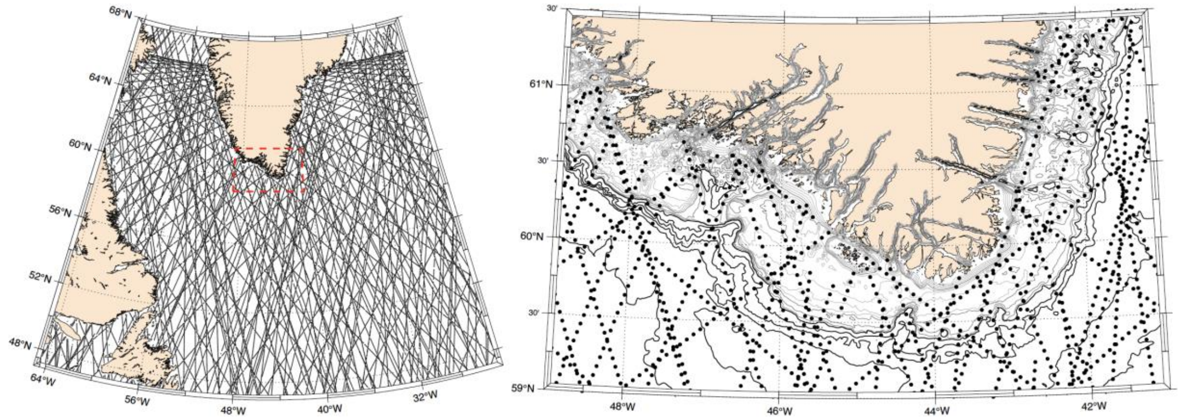
general comments

1. The authors propose a more comprehensive approach for evaluating and intercomparing different Eulerian gridded surface current products against Lagrangian drifting data. It consists in evaluating gridded Eulerian products in 3 independent steps: a direct comparison of model velocities to drifter velocities, a comparison in the Eulerian frame of reference by transforming Lagrangian drifting velocities onto a Eulerian grid, and finally a comparison in the Lagrangian framework through the generation of model trajectories from eulerian velocities. These approaches are not novel as suggested in the title as each one is widely applied in the community but not necessarily altogether in one same study as done here. A novelty however is the complementary assessment of products using along and across shore velocities.

We agree with the reviewer that the individual steps in the framework we outline are not novel. We have instead rewritten much of the paper to focus on the validation of the altimetry fields around the southern tip of Greenland. This region is particularly important due to its potential for shelf-basin exchange and the supply of fresh polar water masses to the subpolar North Atlantic.

2. The authors illustrate the approach with a case study around Greenland comparing two altimetry-derived surface velocity products (geostrophy and geostophy+ekman), and conclude that both products reproduce well the shelf circulation of the region without entering too much in the detail as to why. The tone of the paper is ambitious, yet the manuscript does not contain major advances, and I feel that the manuscript would require significant modification for reaching the objectives it sets itself.

We agree that the dual motivations for our original submission were perhaps too ambitious for a single manuscript – a fact that was brought up by both reviewers. We have instead focused the text around the validation of altimetry-derived surface currents around southern Greenland. The result that the velocity field in this region is well-represented by the altimetry does advance the field because it allows the time-dependency of the flow field to be quantitatively assessed. To address why the altimetry fields perform well on the South Greenland Shelf, we have added text in the introduction about the structure of the shelf bathymetry that is amenable to reliable SSH retrievals, as well as added a new figure (Fig. S1, copied below) .This figure shows the along-track data coverage over the 10-day period in which the drifters were deployed, demonstrating the considerable coverage of the satellites that feed into the gridded product.



Caption: Along-track data coverage of six satellite altimeters incorporated into the gridded product during the 10 day period from August 15-24, 2021. Red dashed line in the left panel outlines the region shown in the right panel. Bathymetry is shown in the right panel.

specific comments

3. A concluding sentence of the introduction highlights the model errors in the region as a justification for such study, yet no model have been included in the investigation, solely altimetry derived products. The regional analysis system TOPAZ4 was included in step1 as a reference point, but not carried across in the following steps which is regrettable. One main purpose of carrying out such intercomparison work is to assess the skill of operational systems. Previous studies have evaluated altimetry products with operationnel systems (e.g. Dagestad and Rohrs, 2019). It would therefore be very valuable to assess TOPAZ4 in all 3 steps.

The focus of the paper is to evaluate the altimetry-derived surface currents in the region and thus we do not go into details about discrepancies between numerical models. There are many different reasons for why models may disagree with one another, notably sub-grid parameterizations, forcing fields, bathymetry, among others. Given the need for a central focus of this paper, we have decided to omit the reanalysis product TOPAZ4 from the majority of our analysis. The one exception is in Fig. 5, TOPAZ4 serves as a reference in our assessment of Taylor Skill Scores. This is done purely to compare the skill scores rather than an intent to assess the TOPAZ model.

4. The manuscript would benefit from a more exhaustive review of work carried out on the topic (and in the region), so to highlight the state of the art, present their limitation to justify the proposed approach.

We thank the reviewer for pointing out this omission and have now included a more robust introduction that includes information about the work in the region.

5. For step1, the methodology suggests to select the closest gridded product point to compare with the observation (daily average). I disagree with this approach and would recommend to interpolate the product value at the point of observation. It is not clear what resolution TOPAZ4 is, and for an approach to be robust a methodology should also be provided when comparing products of different resolution.

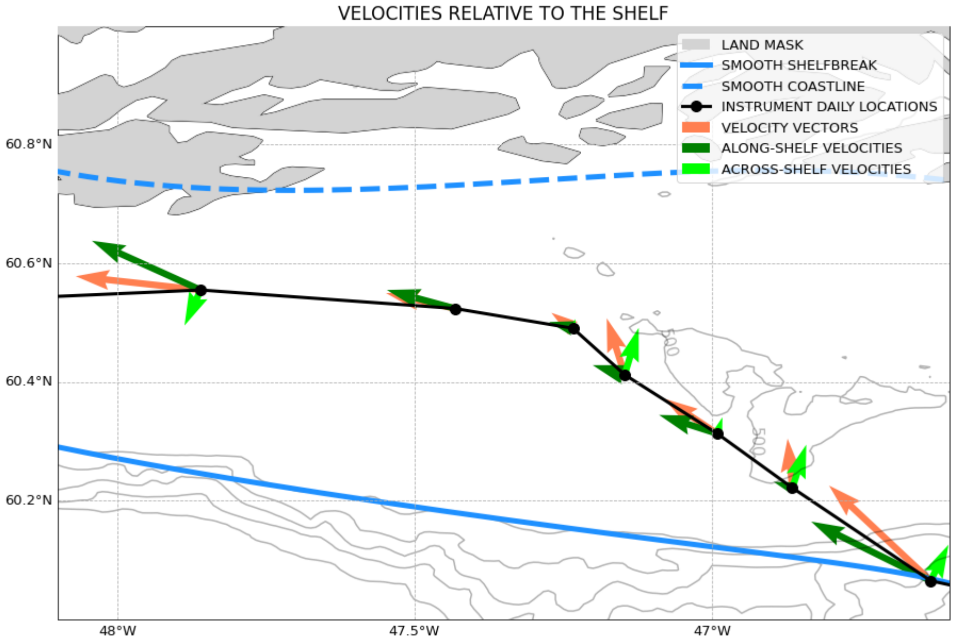
This is a valid concern that we had not considered, and thus we tested the sensitivity of our results to the interpolation methods and presented the results in section 3.1 lines 299-304:

"We tested the sensitivity of our results to the interpolation methods. When comparing the 'nearest neighbor' method with linear interpolation, the correlation improved in average by 0.02 the percent variance explained increased by 2%, but the normalized standard deviation and the Taylor Skill Score both decreased by 0.03 in average. Thus, interpolating reduced the variance in the gridded data while improving the correlations. The gridded altimetry is already a smoothed product from the along-track altimetry. We felt that retaining the natural variance is important. Given the similarity between the results, our conclusions from this first step are robust to this choice."

The resolution of TOPAZ4 is 1/4°, we thank the reviewer for pointing out this omission. We added this information in the section 2.2, line 105.

6. It is not clear how across and along shelf velocity are calculated, (this could be illustrated in figure1). The author mention smoothing the bathymetry, but without further detail (which bathymetry product, is the bathymetry common to all products?), how is the distance to the coast calculated? The authors introduce frequency spectrum analysis but do not carry it out in this study.

We have added text on lines 155-160 and equation 1 to further explain how we built the smooth shelfbreak and coastline and calculated the along and across shelf velocities. In addition, as suggested by the second reviewer as well in its comment 5, we zoomed out in the figure 1, which is now figure 2, and added the figure S2 in supplementary material (copied below) that covers the same area and helps to visualize the direction of along- and across-shelf velocities.



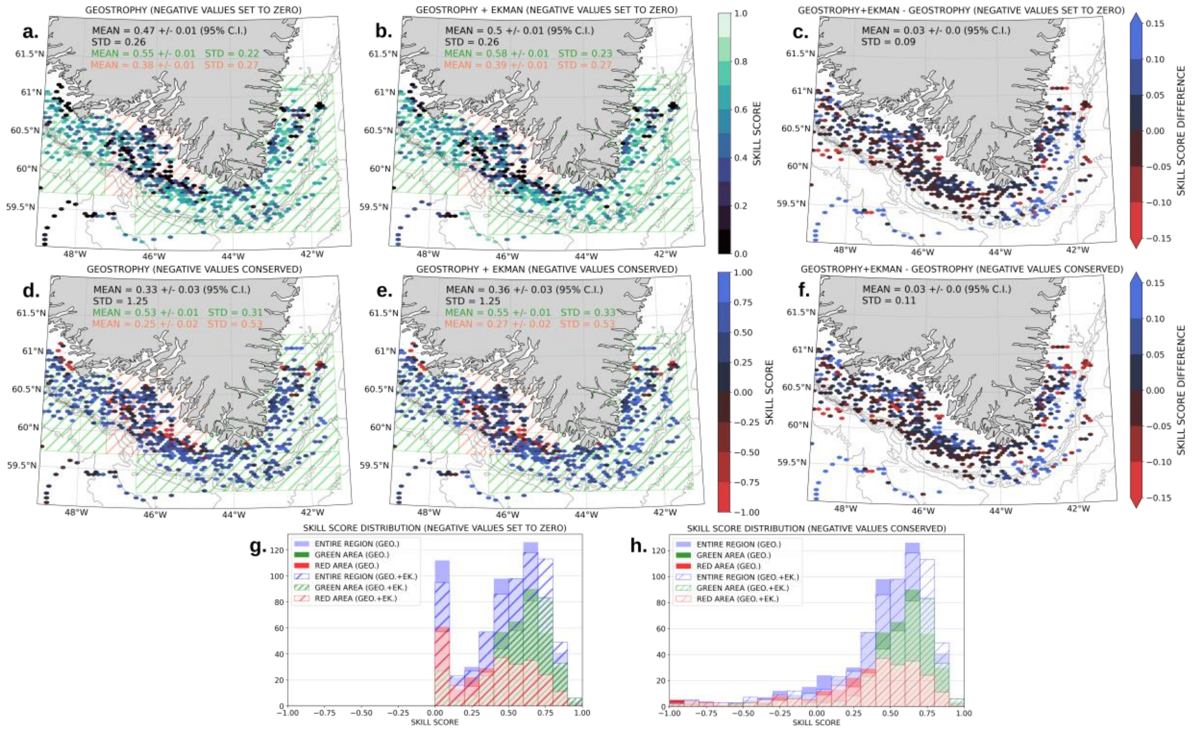
Comments for step2:

7. Lagrangian velocity is transformed onto Eulerian grid for comparison. The result is a map of average velocity for each grid cell where through drifter travelled for the period that the drifter covered. The spatial averaging is well explained and clear, what is not so clear is how the eulerian gridded product velocities have been selected to confront with each transformed lagrangian velocity. Let say if velocity values for a drifter travelling a given cell covers the temporal window not centered on the time window of the daily average gridded product, how is this addressed?

In this step, we build a time-mean surface flow from the drifters and compare it to the time-mean surface flow from altimetry. So we do not account for differences in when each drifter crossed through each grid cell.

8. A map of mean velocity vectors for each product, together with differences in magnitude and direction are proposed to evaluate the skill of products for the different subregions of the shelf. It would be interesting to measure second order statistics so to get a better feel of the product behaviour, and put uncertainties on the values (e.g. on skill score like Revelard et al. 2021).

We agree with this comment and thank the reviewer for this suggestion. We added in Fig. 9 (copied below) the standard deviation (STD) of scores over the different subregions (green and red hatched). In addition, we present in subplots g and h the distribution of scores for the different subregions, products and using the two metrics. These modifications help to better understand the different scores obtain for the two subregions. The region in green obtained only good scores and its STD is small while the region in red obtained more heterogeneous scores highlighted by a much larger STD. The interpretation of those distributions has been added on the manuscript on lines 393-400.



Comments for step3:

9. Statistics are done for the Liu skill score. A score that sets to 0 the negative values. Mean values of this skill score do not take into account the negative values, and may not be so robust when intercomparing models. Other studies have suggested alternative statistics for the skill score distribution such as the proportion of score > 0.6 (Revelard et al. 2021), which may be more informative.

We thank the reviewer for these suggestions and agree that the Liu skill score has some limitations, especially concerning the replacement of negative values with zero. However, we think that this skill score is still very informative because it allows us to directly compare our results with those obtained by Liu et al. . To investigate the sensitivity of our results to this skill score, we followed the reviewer's suggestion and reproduced the results using the metric derived in Revelard et al. (2021)

that conserves the negative values. In addition, we evaluated the proportion of negative values among our results. The figure 9 (copied above for comment 8) has been updated with the results from both metrics. Below is a copy of the paragraph interpreting these results, which has been added to the manuscript (lines 381-392):

The mean results decreased from 0.47 and 0.5 respectively for geostrophy and geostrophy+Ekman to 0.33 and 0.36, which remain good compared to skill scores obtained by Liu et al. on the shelf (0.35 and 0.41) without accounting negative values. The skill scores in the large shelf area, highlighted by green hatches, remain very good and only decrease from 0.55 and 0.58 to 0.53 and 0.55 with this new metric. The results in the red shelf area present larger decreases from 0.38 and 0.39 respectively for geostrophy and geostrophy+Ekman to 0.25 and 0.27 but it remains on the order of Liu's results without negative values. Values using this metric reported in Revelard et al. (2021) from the Ibiza channel were considerably lower, with some regions hitting -0.6. The proportion of negative scores, proposed in Révelard et al. (2021) is also a very informative metric to understand the full picture. In our case for geostrophy (geostrophy+Ekman), we obtained 13.1% (12.2%) of negative scores in the entire region, 7.3% (7.1%) in the red area and 3.4% (3.6%) in the green one. This low proportion of negative values explained the relatively small change of skill scores when accounting negative values, especially in the area of good consistency highlighted in green. The impact of Ekman contribution for this metric (Fig. 9 f) remain very close to the result obtained using the first metric (Fig. 9 c).

No information is given for the particle tracking software used.

We thank the reviewer for pointing this out. The particle tracking software used is Parcelsv2.0 (Delandmeter and Van Sebille, 2019, https://doi.org/10.5194/gmd-12-3571-2019). We added this information line 223.

Overall point for the case study:

10. This study investigates altimetry derived products for the Greenland shelf, with an assessment of Ekman component on improving the representation of shelf circulation. The study would benefit from a more in depth study of the processes. For example, evaluating the altimetry tracks, in particular their angle relative to meridional/zonal component, which could explain the better resolution of one component relative to the other. Relative to Ekman, questions are raised to explain why it may perform better than other regions. The Ekman component of the product could be assessed by investigating wind patterns of the regions, and see if a better parameterisation would improve results.

We thank the reviewer for these very interesting suggestions. Regarding the influence of altimetry tracks, we have added a figure of all the altimetry tracks used to calculate geostrophic velocities in the supplementary material (Fig. S1) and in response to comment 2. We agree that the angle of the tracks could have an impact on the ability to recover a particular velocity component. However, the figure shows that the tracks have different angles to the north-south directions, so it may not be straightforward to fully disentangle these relationships. Concerning the spatial differences in improvements due to Ekman contribution, there may be some very interesting connections to local wind regimes. However, investigating this question would require additional datasets and we believe that this is beyond the scope of the present ADSC evaluation study.

11. A paragraph is missing in the discussion wherein the 3 steps are brought together and their complementarity illustrated, as such it feels like 3 independent steps.

We agree with the reviewer and have added text in our conclusions section (lines 411-413) to explicitly say that it is really the combination of all three steps that helped us to build a comprehensive understanding of the consistency of ADSC around Cape Farewell.

12. The study does not address approaches raised in the community such as for example particle ensemble releases with diffusion terms to account uncertainties.

We explored the role of diffusion early on in our project, and two results led us to conclude that it was not worth considering. First, using only the surface drifter data, we calculated the Péclet number (added in Fig 1) and observed that the regions is largely dominated by advection. Secondly, we used the particle ensemble release methodology and observed large beaching of particles. We think this has to do with the combination of isotropic dispersion assumptions, a sharp land mask linked to the 1/4° resolution of the ADSC velocity fields, and the absence of mass conservation in ADSC Though we acknowledge there are methods to avoid beaching particles, they also introduce biases and uncertainties themselves and we were not convinced their benefits outweighed their costs. Given the results with the Péclet number, we decided to not use this method for this application.

13. I feel that the manuscript would be better pitched if the authors stated that they used a complementary approach to validate surface current (here geo and geo + ekman) illustrated with their Greenland case study, rather than presenting it as a new framework.

We agree with the reviewer and have significantly rewritten the text to address this concern.