Author Response to Reviewer #1

We thank the reviewer for their positive assessment and constructive suggestions. The comments have helped us improve the clarity and completeness of the manuscript. This is my first paper and it helped me improve my writing. The comments lead to additional validations. Comments from the other reviewer lead to a large restructuring of the paper. Hence, the manuscript has changed considerably. We present here a high-level overview of the changes before addressing each specific comment, in order to help the reviewer navigate the revised version. Changes have been highlighted in green.

The first important change is that the dataset has been given a name: SO_SLICE.

Original manuscript structure :

- **Sect. 1** — Introduction
- **Sect. 2** — Data
  - 2.1 Satellite data / missions
  - 2.2 Open ocean satellite data / CMEMS
  - 2.3 In-situ data for validation
- **Sect. 3** — Methods
  - 3.1 Altimetry above the open ocean: SSH equation, corrections, MSS, ADT
  - 3.2 Lead detection and retracking
  - 3.3 Editing (including MERIS validation mixed in here)
- **Sect. 4** — Results
  - 4.1 Product overview mixed with CMEMS comparison and seasonal climatology
  - 4.2 Validations.
- **Sect. 5** — Discussion and conclusions

Revised manuscript structure :

- **Sect. 1** — Introduction
- **Sect. 2** — Data and standard processing
  - 2.1 Missions + range-to-SLA short paragraph. Shortened, cites references instead.
  - 2.2 Ancillary and validation data. New: introduces *all* comparison datasets upfront: C3S (instead of CMEMS), Dragomir, Auger and TGs + BPRs.
- **Sect. 3** — Sea-ice altimetry processing
  - 3.1 Lead detection, retracking, and along-track editing
  - 3.2 Ocean/lead bias correction
  - 3.3 Inter-mission harmonization
  - 3.4 Optimal Interpolation. New, detailed subsection on OI inputs
  - 3.5 Derived variables
- **Sect. 4** — Results and validation
  - 4.1 Along-track qualitative validation. New: MERIS moved here from methods
  - 4.2 Product overview. New: Seasonal cycle has been dropped + spectral analysis and comparison with C3S to quantify SO_SLICE capabilities.

  - 4.3 Validation: Inter-satellite consistency (new diagnostics), formal error, comparison with in-situ observations, comparison with other altimetry products (new).
  - 4.4 Effective resolution of SO_SLICE.

Sect. 5 — Discussion and conclusions

Below, we address each comment individually. Reviewer comments are in bold italics, and our responses follow in regular text.

General Comments

1 - Many citations have parentheses when they should not have (l. 45, 46, 48, 51, …)

All citation formatting has been corrected throughout the manuscript.

2 - Please revise all the figures. Too often the labels for the axes, legend, colorbar are not readable because too small. Every letter of a plot should be readable, otherwise it should be removed. And often the images do not look high-resolution.

All figures have been regenerated at 300 dpi with larger labels and consistent font sizes throughout. We apologise for the poor resolution in the initial submission. We had put some care in producing the figures at the correct resolution, but the Copernicus preprint rendering process degraded them.

3 - The generated product needs a name. It would help the reader in this manuscript (instead of calling it "our/the product" in the text or just "Product" in the images). Also, it will be beneficial for future works using and referencing it.

This is a valuable remark. The product is now named SO_SLICE (Southern Ocean Sea Level In the ICE) and is referred to consistently throughout the manuscript and all figures.

4 - Check if the tables' caption has to be below or above the table.

Table captions have been moved above their respective tables, consistent with the layout of published ESSD papers.

5 - Use consistent layout for tables (linewidth of borders).

The table styles have been updated with consistent border widths throughout.

6 - In the methods section, it would be useful to have a recap of all the processing applied to the data as a diagram. What editing, debiasing techniques are applied to each product sequentially, otherwise it's hard to keep track.

We have added a processing pipeline diagram as the new Figure 3, which describes the full processing chain from L1b waveforms through to the final gridded product, indicating which steps are described in which section. It can be found in line 205.

7 - The altimetry product created is not characterized enough in terms of capabilities. The minimum information required is at least the effective spatial resolution (which I imagine will change in time and in the different regions due to coverage) and the temporal resolution. This is crucial information to understand what kind of processes can be resolved and studied.

We agree that this was a significant gap in the original manuscript. The maps we distribute are on a 25 km grid at daily to monthly frequencies, but this is not the effective resolution of SO_SLICE. We now describe the capabilities of the dataset in two ways.

First, we computed the spatial and temporal power spectral density (PSD) of SO_SLICE and compared it with the Copernicus reference product (Fig. 6, Sect. 4.2). This shows the scales at which energy is present within the altimetry products.

Second, we estimated the effective spatial and temporal resolution following the noise-to-signal ratio (NSR) protocol of Ballarotta et al. (2019). The effective spatial resolution ranges from approximately 120 km between 50°S and 65°S to 200 km south of 65°S, and does indeed vary regionally. The effective temporal resolution averages 37 days across in situ stations. Both are discussed in the new Sect. 4.4. In the paper, we present resolution estimation on the year 2019, where we have 3 satellites. The same diagnostic was done in 2011, with Envisat only maps and the results were similar.

Specific Comments

1 - 85: Cryosat-2 not Cryosat.

Corrected throughout. We now use "CryoSat-2" with a capital S, consistent with the ESA official spelling.

2 - Page 4. The table of the satellite missions should be Table 1.

Corrected.

3 - 116: Supplementary Table S2. Somewhere in the text or in the Supplementary table it would be interesting to add the distance of the in-situ stations from the coast. This would strengthen your comparison later, when you explain why some of the differences with the altimetry may arise.

We have added a distance-to-coast column to the supplementary table listing the in situ station characteristics (Table S1).

4 - Fig 2: Please add in the caption the meaning of the red dot of the bottom panels.

The red dot indicates the retracking point on the waveform. This has been added to the caption. (line 171-176).

5 - 157: the sentence is missing a verb.

We've removed this sentence from the original paragraph (now lines 124-135).

6 - 174: LaTeX rendering not working.

Corrected. lines 124-135

7 - 206–211: this explanation is too quick. Why is the extra-editing needed? What does the threshold remove? Why do you first perform an "extra-editing" on the data and then an "iterative editing" on the ocean data (only open ocean, no leads?)? Also add one sentence to briefly describe the "off-nadir hooking" phenomenon. Is this processing the same for all satellites? Please rephrase the full paragraph adding information and clarity.

The editing procedure has been completely rewritten with three explicitly numbered steps.

Step 1 (empirical threshold) applies to all lead data and removes lead measurements that go through the classification step, but for which we are not 100% confident. For example, we have a condition on sea ice concentration where sea ice concentration needs to be at least > 30% for us to accept the waveform as a lead.

Step 2 (iterative Lanczos filter) applies to open-ocean data only, where observations are dense enough to compute meaningful along-track statistics. It iteratively removes points that deviate from a low-pass filtered profile.

Step 3 (hooking filter) applies to lead data from LRM altimeters only (Envisat, SARAL). The hooking effect is now described in the text. This correction is not applied to SAR altimeters (CryoSat-2, Sentinel-3A), whose smaller footprint makes them much less susceptible to off-nadir contamination. (lines 218-230).

8 - 215: fewer by how much? What is the proportion of data removed? Is there a general statistic about the difference between leads and open-ocean points kept, compared to the initial amount?

We have added a supplementary figure (Fig. S3) showing the number of lead measurements per 100 km × 100 km × 1-month bin before and after editing, for each satellite mission. For Envisat, the median number of lead observations per bin decreases from approximately 300 to 35 after editing. Similar values are obtained for SARAL. In both cases, the hooking filter accounts for the largest fraction of removed points. CryoSat-2 and Sentinel-3A retain a higher fraction of lead observations, as the hooking correction is not applied to SAR altimeters. For open-ocean waveforms, approximately 85% are retained after editing for all missions. These statistics are briefly discussed in Sect. 3.1. (lines 232-238).

9 - What is the statistical occurrence of the limitations mentioned in paragraph 225–235 for Envisat, and other satellites if relevant? How many points are correctly classified as leads, how many are incorrectly classified as leads and how many are missing? How does a misinterpretation of a lead influence the results? This is important to judge the robustness of the method.

The limitations described in this paragraph (now Sect. 4.1) are relevant for all LRM altimeters (Envisat, SARAL). We have not ourselves quantified the fraction of correctly vs incorrectly classified leads. This has been done in part by Poisson et al. (2018), who compared waveform-based lead classifications against MERIS optical imagery for 42 Envisat along-track sections. In their Figure 8, they show scatter plots of two optical properties for leads, open ocean, and ice floes. The three classes are efficiently discriminated, although a small overlap exists between leads and ice floes. The cases we discuss in Sect. 4.1 (points 3 and 4 in Fig. 4) fall within this overlap region. Poisson et al. do not provide a single fraction of misclassified leads, but the overlap is visually small. A comprehensive quantification of the classification error rate is beyond the scope of this study. However, the impact of residual misclassifications on the final gridded fields is diluted by the OI, which combines hundreds of observations per grid point over each mapping window. The validation in Sect. 4.3–4.6 confirms that the product achieves comparable accuracy in ice-covered and open-ocean regions, indicating that classification errors do not significantly degrade the final fields.

10 - 229: what do you mean "at least for this example"? How do the results change for the other satellites?

What we meant is that this is a single illustrative example, not a systematic assessment. The classification methodology is the same for all missions (Poisson et al., 2018). While we did not repeat this qualitative illustration for the other satellites, the gridded product validation in Sect. 4.3–4.6 provides an integrated assessment across all missions. We have rephrased the text to avoid ambiguity. This paragraph refers to Sect. 4.1.

11 - Fig 3: Add the units for the time dimension in panel (a) and add the geographical coordinates in panel (c). Panel b is hard to interpret. Please make bigger. One suggestion could be to keep the horizontal panel (a) at the top of the figure, while (b) and (c) can be displaced in a second row next to each other. Make legend labels and markers bigger because they are hard to understand. Colorbar label and ticks must have bigger font.

Figure 3 (now Figure 4) has been redesigned following all of these suggestions. Panel (a) is at the top, with panels (b) and (c) side by side below. All labels, markers, legend entries, and colorbar ticks have been enlarged. Units have been added to the time dimension in panel (a) and geographical coordinates are now shown in panel (c). (lines 340).

12 - 257: Why is the additional debiasing step only performed for Envisat?

The additional debiasing step is applied to Envisat because its ocean/lead bias exhibits a very strong spatial pattern that cannot be treated as spatially homogeneous noise, unlike the other missions. We empirically found a correlation between the residual ocean/lead bias and the average peakiness of the neighbouring lead waveforms. This relationship is exploited to apply a spatially varying correction: the bias, which was initially spatially heterogeneous, is reduced to a level where it becomes more like a white noise and is comparable to the other missions. For CryoSat-2, SARAL, and Sentinel-3A, the ocean/lead bias is spatially more uniform and is adequately removed by the standard correction described in Sect. 3.2. (lines 242-257).

13 - 262: Why is CryoSat-2 used as a reference? You say it at l. 234, but should be better explained here.

CryoSat-2 is used as the inter-mission reference because it is the only mission with temporal overlap with all three other satellites (Envisat from 2010–2012, SARAL from 2013 onward, Sentinel-3A from 2016 onward). This makes it the natural pivot for harmonising the missions. This has been clarified in the relevant paragraph. (line 261).

14 - Add reference for EASE2 grid: https://doi.org/10.3390/ijgi1010032

The reference (Brodzik et al., 2012) has been added. (line 284).

15 - Section 3.6: rephrase. The interpolation is an important step of your methodology and needs more details. For example, you say that "the expected variance and noise estimation necessary for the interpolation also come from Auger et al. (2022a) and Veillard et al. (2024)" without mentioning why you need these variables and how they influence the interpolation. Mention the basics of the DUACS-DT2022 mapping and how this kind of interpolation could influence the results.

We have expanded Sect. 3.4 (previously 3.6) to explain the role of each OI input parameter. The text now reads: "The OI weights are determined by a space-time covariance function controlled by three inputs: the spatial and temporal correlation scales, the expected signal variance, and the measurement noise variance. The correlation scales define the space-time radius within which observations contribute to each grid point estimate. They control the smoothness of the output and set a lower bound on the resolved scales. The signal variance determines the expected amplitude of SLA variability. The noise variance controls how much the OI trusts individual observations relative to the prior." The prescribed correlation scale values are described and shown in supplementary figures (Figs. S3–S4). We also explain that the correlation scales and variance files come from Auger et al. (2022a) and Veillard et al. (2024), which were tuned for the subpolar Southern Ocean. This can be found in lines 270 – 279.

16 - 277: is this the effective spatial resolution of the product or the distance between grid points? What is the effective temporal resolution? What are the covered latitudes?

This is the distance between grid points (25 km), not the effective resolution. The effective spatial resolution (120–250 km depending on latitude) and effective temporal resolution (23–48 days, mean 37 days) are now discussed in the new Sect. 4.4. The product covers the Southern Ocean south of 50°S. We've switched from resolution to spatial grid cells for clarity (line 285).

17 - The comparison between your product and CMEMS is too quick. We need some spectra (at least in the supplementary material). You could do frequency-wavenumber spectra to know exactly what kind of structures the product resolves in space and time. At least in the open-ocean region, to provide more robust results for this comparison.

You should add an extra panel per row with the difference between CMEMS and "your product", to highlight where the major differences are.

The comparison with CMEMS has been substantially expanded, though we now use C3S as an altimetry product. We have added difference maps (Fig. 5c) and computed both spatial and temporal PSD for SO_SLICE and C3S, separately in the ACC and the subpolar Southern Ocean (Fig. 6). The spatial PSD shows close agreement in the ACC at wavelengths longer than 150 km, with SO_SLICE retaining slightly more energy at shorter wavelengths due to less aggressive along-track filtering. In the subpolar SO, SO_SLICE contains substantially more energy at all wavelengths, with a PSD ratio ranging from 5 at 1000 km to 17 at 50 km. The temporal PSD confirms consistency at periods longer than 90 days, with SO_SLICE retaining more energy at shorter periods (90–20 days), which we again attribute to differences in filtering. We chose to present spatial and temporal PSD separately rather than combined frequency-wavenumber spectra, as we found this more visual and more directly informative for the user who needs to know what scales are resolved. This can be found in section 4.2 (lines 365-280).

18 - Section 4.2.1: Frequency-wavenumber spectra of the SLA would be a good way to quantify the scales at which the satellites agree or differ. It is not enough to visually compare the images finding "good agreement in large-scale spatial patterns across missions". What is large and small scales? It is important to do this to be sure of what you are comparing and to add robustness to the final product. Generally throughout the entire paper, scales need to be quantified better. You can do it monthly, but it would be even better to do it at the highest temporal frequency that you have.

We now present the spatial spectral coherence between mission pairs (Sect. 4.3.1, Fig. 7a), computed on binned 75 km × 10-day maps. This directly quantifies the scales at which the satellites agree or differ: all mission pairs display coherence greater than 0.5 at wavelengths longer than 300 km. In the ACC, the 0.5 threshold is crossed at approximately 180 km for all pairs. In the subpolar SO, it varies from 170 km (SAR–SAR pair) to 304 km (CryoSat-2 vs Envisat). This is done at the 10-day time step, the highest temporal frequency available for the binned maps. We believe the spectral coherence is a more targeted diagnostic than frequency-wavenumber spectra for answering the specific question of inter-mission agreement, while being easier to interpret. (lines 406-414).

19 - Fig 6: "(b-c/e-g) Comparison of monthly SLA snapshots" — it is not the comparison but just the individual snapshots. Right? - Panel (c) should be CryoSat-2.

Correct. Fig 6 is now Fig7, however we have removed the maps from the analysis now.

Corrected in Fig 7. (line 439)

20 - 362: "at face value"? Are you sure this is the right expression you want to use here?

No, what we meant was simply that this metric may overestimate the true errors of the final product. We have rephrased: "These leave-one-out RMSE values represent upper bounds on the errors of the final product, since SO_SLICE integrates all four missions simultaneously rather than withholding one ». (line 476-479).

21 - Fig 7b: what season is this plot? Is it showing only winter or also other seasons? If other seasons are included, how do you treat the region south of your white contour in spring/summer/autumn? Is it included in the open-ocean distribution?

Both the map and the distributions show winter (JJA) RMSE values, which was not specified in the original caption. We have added this precision. The regional split is between the ACC and the subpolar Southern Ocean (defined by the MDT contour), applied uniformly regardless of season. (line 480).

22 - 376–380: can you provide the mathematical formulation of the formal error?

The formal error equation has been added to Sect. 3.4: $\sigma^2\_formal = \sigma^2\_signal - c_0^T(C + N)^{-1}c_0$, where σ^2_signal is the prescribed signal variance, C is the space-time covariance matrix of the observations, N is the noise covariance matrix, and c_0 is the covariance vector between the estimation grid point and the observation locations. (line 305).

23 - 389: (REF)

The placeholder reference has been removed. We apologise for the oversight.

24 - 393: what do you mean by "in addition to Envisat"?

This was a mistake. There are no Envisat data available in 2019. The sentence has been corrected to list only CryoSat-2, SARAL, and Sentinel-3A. (line 505).

25 - Fig 8: No need to write (top) and (bottom) in the caption if you have separate letters for the panels. Mention panel (d) in the text. Add info on the time scale on which the average is performed. Is it for the full product?

The caption has been revised: panel labels are now used exclusively, panel (d) is referenced in the text, and the averaging period (over the full period) is specified. (line 523-528).

26 - 459: Add OSISAF acronym or shortly explain what it is.

The acronym (Ocean and Sea Ice Satellite Application Facility) and the reference to the dataset have been added. Line 573.

27 - Tab 5: You don't mention in the text the change of the results when the sea-ice coverage exceeds 10%. How do you explain that most of the Pearson coefficients increase in the higher sea-ice scenario?

We do not have a fully convincing explanation for this. In some instances, such as for Myrtle B or C, the portion of the time series with sea-ice concentration > 10% is relatively short, so the higher correlation could be partly due to sampling variability. In most cases, the gain in correlation is lower than 0.1, so we are not confident it represents a definitive improvement. In the revised manuscript, Tables 4 and 5 now report metrics computed exclusively during ice-covered periods (sea-ice concentration > 10%), which is the most relevant condition for validating SLA retrievals under ice. This is to not confuse the reader.