

Response to referee

ilo and hsk

August 2025

We thank the reviewer Alek Petty for the review. Please find below reviewer comments provided in black and responses in blue.

1 Answer to review by Alek Petty - General review

I still find the framing around reference measurements as a little odd and still quite poorly defined. By your definition they are comparison datasets of similar quantities to the relevant satellite datasets (you should better list these out from the start) that happen to not be satellite derived. I'm surprised your definition doesn't include any mention of reliability, accuracy, higher-resolution, different sensing approaches etc? I personally think a better framing would be something like: 'A compilation and description of potential validation datasets [or reference measurements] for satellite-derived altimetry estimates of freeboard, snow depth and thickness.' Then state how/why you seek non-satellite data for whatever reason. A mouthful yes, but at least clearer? I also think you need to be much clearer about the fact the underlying datasets were not produced and disseminated towards the goal of being reference measurements for satellite sea ice altimetry validation, but you are using them for that purpose. Focusing on OIB which I am more familiar with, you are using OIB quick-looks and refereeing to them as reference measurements 'The OIB reference measurements used in this study...'. Which I just still think is quite odd as the goal of the quick-looks were really to just see how things look/capture the regional sea ice distribution and by no means to be considered as a long-term archived dataset for satellite validation.

We thank the reviewer for this constructive comment. We agree that the previous version of the introduction lacked a clear "red thread" and that the framing of reference measurements was not fully clarified. We have now re-organized the introduction to focus on the goal and purpose of the CCI SIT RRDP. Specifically, we frame our study as a compilation and description of potential validation datasets for satellite-derived altimetry estimates of freeboard, snow depth, and ice thickness.

We clarify that these datasets, which we refer to as reference measurements, were not inherently collected or post-processed with the aim of being used for

satellite evaluation, but we use them for this purpose because they provide independent, high-resolution, or complementary observations that are useful for evaluation of satellite-derived SIT CDRs. We now explicitly list the geophysical quantities included in the comparison datasets and describe their roles in validating satellite altimetry. We have also read through the manuscript and have critically removed the term reference measurement directly related to one type of measurement e.g., OIB, but use it more general for a collection of, for example, sea ice thickness measurements which can potentially be used for sea ice thickness CDR validation.

To acknowledge issues regarding selection biases and scaling, which are related to some of the datasets included in the CCI SIT RRDP such as OIB, when compared to satellite-derived SIT products, we refer to recent literature by Langsdale et al. 2025; Xu, Zhou, and Wang 2020.

I think the newly introduced opening introduction paragraphs do not flow very well. The line ‘Comprehensive validation of satellite-derived sea ice thickness products necessitates...’ makes more sense as a start-point I think so I recommend you build from there then introduce your concept of reference measurements/validation dataset cataloging and why you are doing this in a much clearer way.

We appreciate the reviewers input, and have re-organized the introduction accordingly. We now start with the suggested paragraph, followed by introducing the explicit goals of the CCI Sea Ice project and its RRDP.

I think you also need to be much clearer about the goals of the CCI RR project and include higher up in the intro as that’s the clear motivation.

We have included a clearer goal of the CCI RR in the second paragraph of the paper.

The CCI RR paragraph also too vaguely refers to satellite data but not specific missions or goals. Is ERS 1 and 2 part of the motivation/goal or is that just secondary? What is FDR4ALT and how does that fit in exactly? Overall, I appreciate the efforts towards a reframed introduction but I think it doesn’t read too well now, so I would really consider the author spends some time trying to improve this for clarity.

Thank you for your questions and suggestions. The primary objective of the CCI SIT project is to produce Climate Data Records (CDRs) of sea ice thickness (SIT) for the entire satellite-derived altimetry era (1993–present). Although ERS-1 was launched in 1991, usable sea ice thickness maps were only available from 1993, and this has now been clarified in the first paragraph of the introduction. The overarching aim of the ESA CCI Sea Ice project is therefore to produce CDRs of SIT from existing polar radar altimetry missions starting in 1993. The CCI SIT RRDP cover this same period.

The latest public available CCI SIT CDR do not include ERS-1 nor ERS-2 (only Envisat and CryoSat-2). The parallel ESA project FDR4ALT provides Fundamental Data Records of among others high-quality radar freeboards from historical missions, such as ERS-1 and ERS-2. To illustrate the coverage of CCI SIT RRDP, we have therefore included Fundamental Data Records of ERS-1 and ERS-2 radar freeboards produced within the FDR4ALT project.

We have reorganized and clarified this section, broken up dense sentences, and explicitly linked the datasets to the project aims to improve readability.

I don't have time to go through the manuscript in detail again, but here are some more specific comments:

I am quite surprised and worried about the impact from assessing median not mean quantities, especially with things like the 'average date', which seemed odd to me. I appreciate there can be good reason to assess median when distributions are highly skewed, but is that the case with all these? At the very least this needed more justification.

The median date provides a useful indicator of when, within a given month, the bulk of the data was collected. This helps users place greater weight on matchups closer to the median date and gives insight into when the data are most densely clustered. This is particularly important for datasets from airborne campaigns, where measurements often occur within a short time window.

When considering mean versus median quantities, it's important to note that changes from version 3 to version 4 of the manuscript reflect more than just a shift in statistical method. Several datasets were updated to include additional data—ASSIST, BGEP, IMB-CRREL, SB-AWI, and AEM-AWI (due to the addition of MOSAiC and Nansen Legacy data)—and an outlier filtering process was applied. These updates alone have a substantial effect on the results. Moreover, many datasets have distributions that deviate strongly from normality, both along an entire flight/ship track and within individual grid cells as shown in figures 1 - 5. BGEP and SCICEX, in particular, show a strongly skewed distribution where the choice between mean and median makes a large difference.

During re-processing for version 4, we also identified and corrected two errors:

ASSIST snow depths: In some cases, grid cells contained valid snow depth measurements but no valid sea ice thickness values. Upon review, these included multiple instances of zero-centimeter snow depth in areas without valid ice thickness values that appear dubious and contributed to skew the distribution. We now include snow depth only in grid cells where at least one valid ice thickness sample exists.

SB-AWI (northern hemisphere): The initial snow depth had been added inadvertently twice, which significantly affected bias estimates. This has been corrected.

Given these updates and corrections, a direct comparison of mean versus median between versions 3 and 4 does not isolate the effect of the statistical method. To address your concern about skewness, we examined histograms of measurements within a small set of grid cells (BGEP: 11 cells; OIB, AEM-AWI, IMB-CRREL: 16 cells each). The results, provided in the Appendix (although based on a very small random sample from the combined data), show:

- BGEP: Distributions are far from normal, and as in the main results, medians are generally smaller than means.

- SCICEX: Shows a similar pattern, with a pronounced difference between mean and median.
- OIB & AEM-AWI: Skewness is present, but effects are less consistent and smaller in magnitude.
- IMB-CRREL: Sample sizes are smaller, distributions are narrower, and mean–median differences vary in direction.

Overall, the data are not evenly distributed, and significant outliers—especially in mooring and airborne datasets—can arise from leads and/or large ridges. For this reason, we consider the median to be a more representative measure for a typical value in a grid cell, as it is less sensitive to outliers than the mean. To address this in the paper the following have been added to section 3.2: "Median and robust standard deviation were selected to reduce the influence of outliers and to better handle skewed distributions present in several datasets. This approach was preferred over the mean and ordinary standard deviation to obtain a typical representative value for each grid cell. Using the median date also indicates when the reference data are most densely distributed and enables matchup weighting based on temporal distance."

L32: if you have very good reference measurements of multiple quantities, then you could theoretically validate this concept too perhaps! Yes, if you have all reference measurements, you could in principle validate the concept of hydrostatic equilibrium - if this is what you mean. If you have three different measurands you can calculate the densities as was done by Jutila et al. (2022) from total thickness, total freeboard and snow depth.

L36: here is another example where you are directly referring to things as reference measurements even if that's not what they were intended to be and its again unclear what this means in my view? We agree. We have updated the text to accordingly.

L41: why particularly radar?

We agree that this might not necessarily be particularly for radars. During our re-organization of the introduction, the sentence still contains radar as we have a focus on radar altimetry-derived sea ice thicknesses in this paper. The term "particularly" has been removed.

L102: be clear about what temperature.

We have specified that included temperatures are 2 meter and surface temperature and that the temperatures of importance to the penetration depth are surface temperatures, along with temperatures within the snowpack.

L106: I think this is generally considered the pole hole not gap. Products already include interpolation over this (e.g. my IS2SITMOGR4 product) so this isn't a future idea exactly.

Thank you for sharing your product including interpolation over the pole hole. We were not aware of this, but only referred to the CCI project, where this is still a thing to do. We have updated the text accordingly. First of all we changed the "polar gap" to "pole hole" as suggested. Additionally, we updated

the text to read: "Reference measurements north of the satellite altimeter coverage i.e., the pole hole, have been included for evaluation of satellite products interpolated across the pole hole, or as reference measurements for models."

L418: 'All reference data in the CCI SIT RRDp are related with some degree of uncertainty..' what does this mean?

Here, we mean that all reference data in the CCI SIT RRDp have inherent measurement uncertainties, as no observation is perfectly accurate. This sentence is intended as an introduction to the subsequent section where the uncertainties are discussed for the different data products in the CCI SIT RRDp. To make the intended purpose clearer we have reformulated this to: "*All source reference data in the CCI SIT RRDp are associated with some level of uncertainty*"

Figure 13: confused by the two colorbars here and how they relate to the figure, needs fixing.

We agree, that the snow depth colorbar is not adding any value to the plot, which basically demonstrates the temporal as well as the spatial representativeness. We have updated the plot, by including another legend explaining the difference in footprints, and removed the snow depth color bar. The caption has been updated accordingly.

Figure 14 – really hard to see anything in that scatter plot, recommend distributions instead.

Figure 14 has been updated to include distributions. To make it easier to see the distribution of scatterplots in figures 6 to 9 these have also been updated to use distributions instead of a combined scatterplot.

2 Appendix

References

- Xu, S., L. Zhou, and B. Wang (2020). "Variability scaling and consistency in airborne and satellite altimetry measurements of Arctic sea ice". In: *The Cryosphere* 14.2, pp. 751–767. DOI: 10.5194/tc-14-751-2020. URL: <https://tc.copernicus.org/articles/14/751/2020/>.
- Langsdale, M. et al. (2025). "The Challenges and Limitations of Validating Satellite-Derived Datasets Using Independent Measurements: Lessons Learned from Essential Climate Variables". In: *SurvGeophys* 2025, pp. 1–38. DOI: 10.1007/s10712-025-09898-4. URL: <https://link.springer.com/content/pdf/10.1007/s10712-025-09898-4.pdf>.

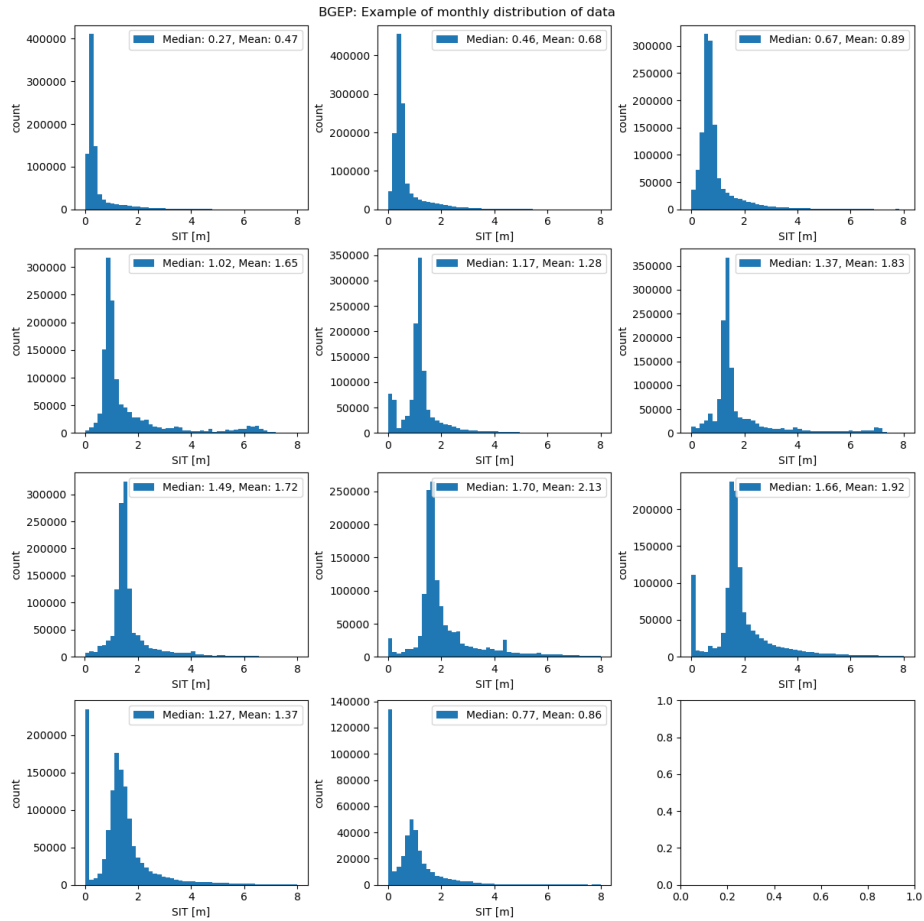


Figure 1: BGEP: An example of the distribution of values used to obtain monthly averages for 11 data points

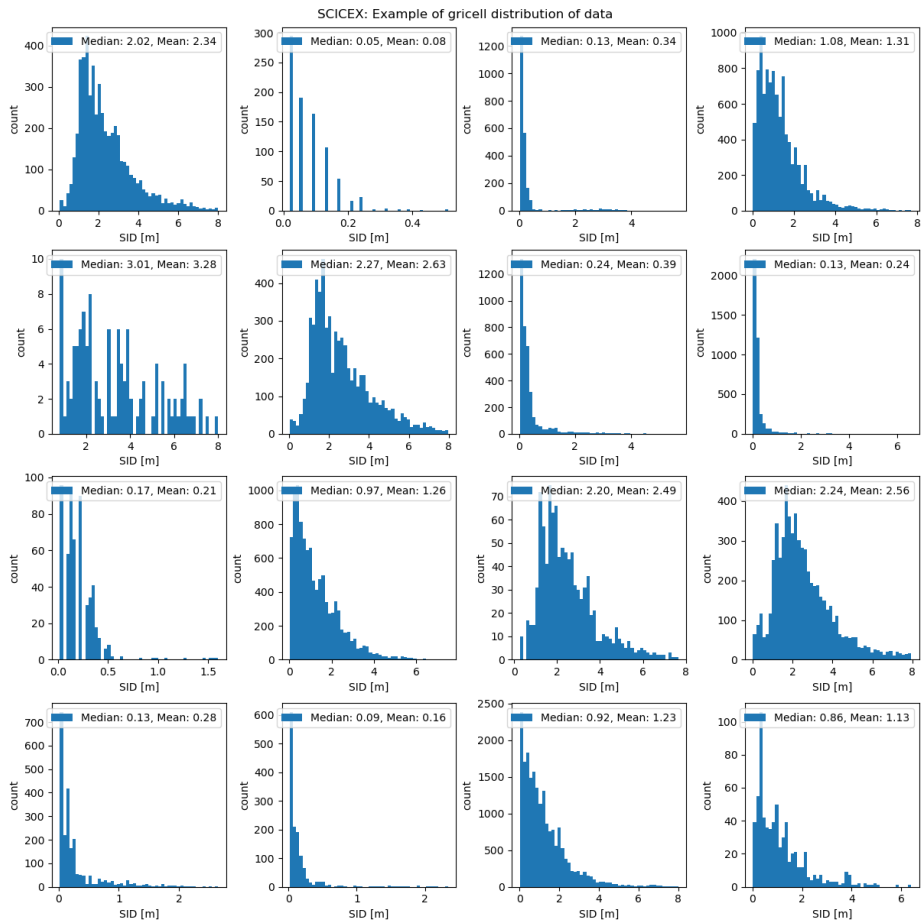


Figure 2: SCICEX: An example of the distribution of values for 16 grid cells

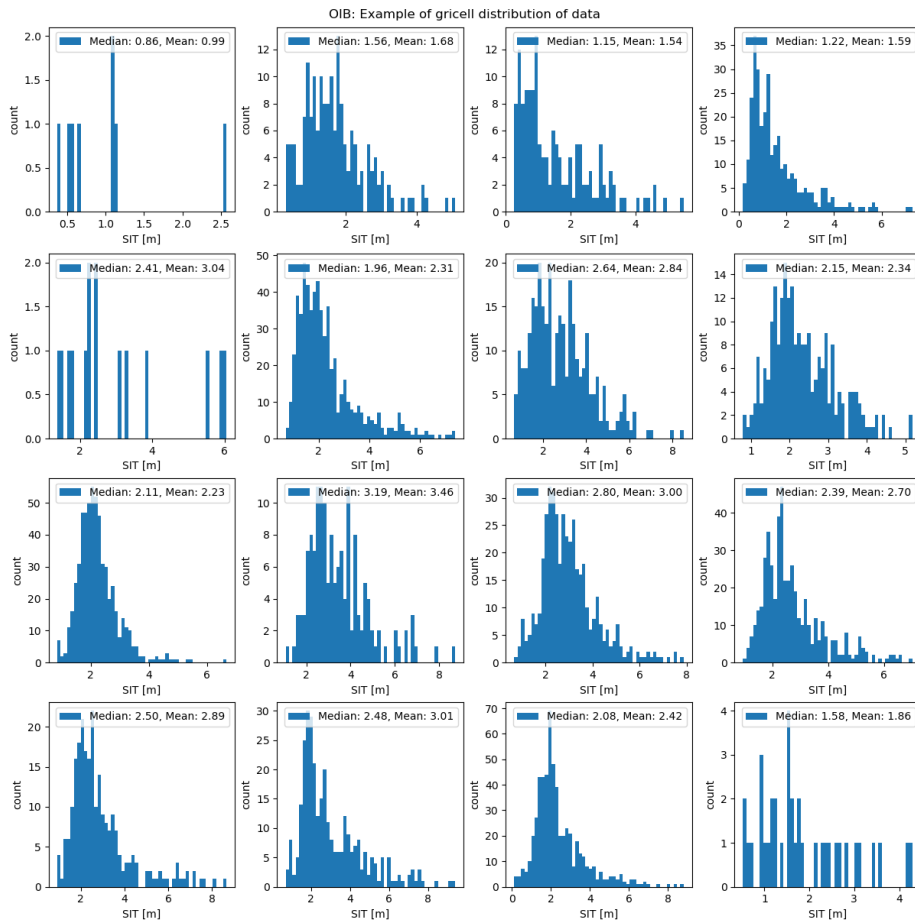


Figure 3: OIB: An example of the distribution of values for 16 grid cells

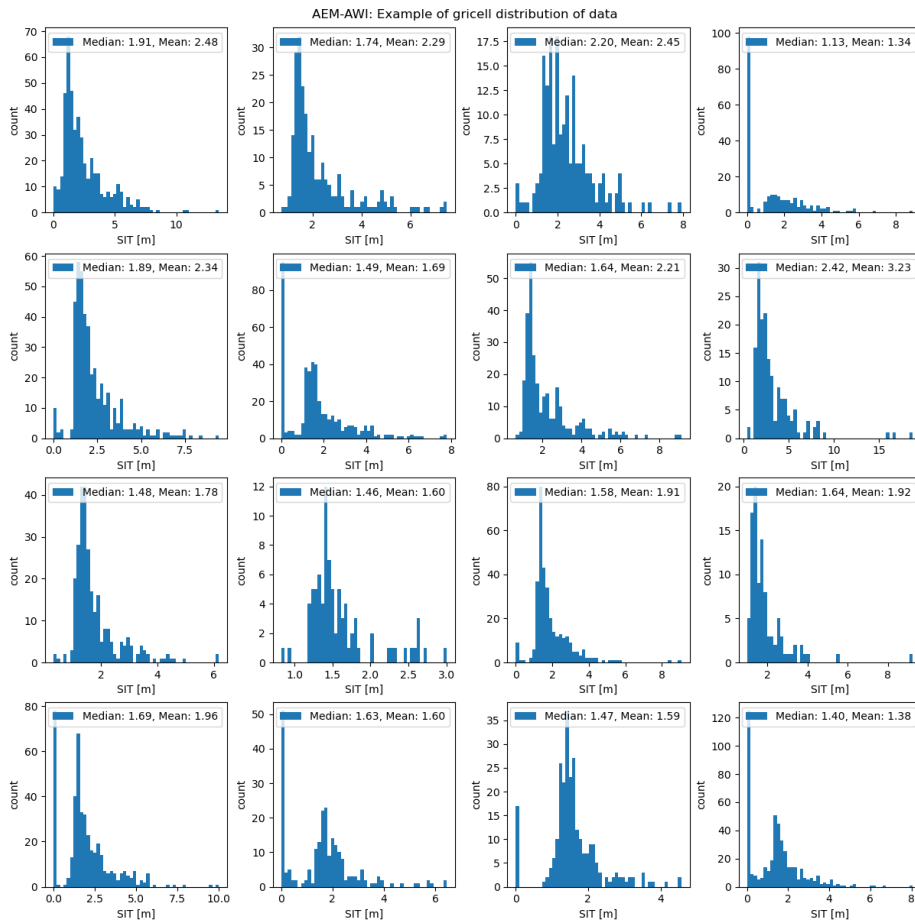


Figure 4: AEM-AWI: An example of the distribution of values for 16 grid cells

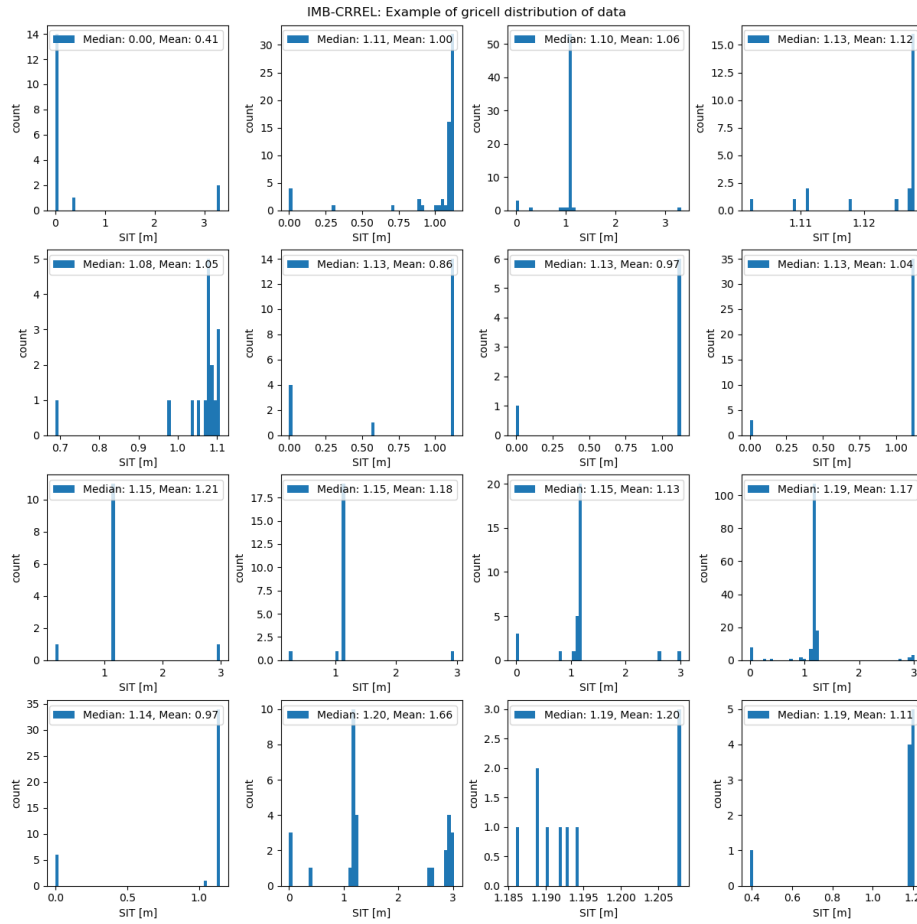


Figure 5: IMB-CRREL: An example of the distribution of values for 16 grid cells

Response to editor

ilo and hsk

August 2025

We thank the editor for the editorial comment. Please, find below editor comments provided in black and responses in blue.

Thank you for your revised submission. After careful consideration of the reviewers' comments and our editorial assessment, I am recommending major revision for your manuscript.

While your revisions have improved the manuscript and one reviewer has changed their initial recommendation from rejection, both reviewers and editors remain concerned about the current state of the manuscript. Minimal changes in response to the suggestions below are not likely to be sufficient to advance this paper to the next round.

The lack of clarity regarding the manuscript's purpose remains a significant issue (primary concern of Reviewer 1). To address this, you must:

- Reframe "reference measurements" throughout the manuscript

In a recently published manuscript by Langsdale et al. 2025, they use the term *reference data* in the same way as we use the term *reference measurement*. Citation from Langsdale et al. 2025: "Validation of satellite-derived essential climate variable (ECV) datasets requires comparison against independent measurements. These independent measurements, which include ground-based, airborne, and other non-satellite-based measurements, are typically the product of a different measurement system and may include some contribution from models. These *reference data* therefore have their own characteristic." Based on this we have kept the term *reference measurement* throughout our manuscript.

To clarify how we use it, we have made following changes; 1) updated the introduction with a clear definition in the second paragraph "We use the term *reference measurements* for such a compilation of independent non-satellite observations collected from airborne platforms, submarines, moorings, drifting buoys, and ships, which can potentially be used for evaluation of satellite-derived SIT CDRs." 2) We have read through the entire manuscript and have critically removed the term *reference measurement* directly related to one type of measurement e.g., OIB, but use it more general throughout the manuscript for a collection of for example non-satellite sea ice thickness or snow depth measurements, which can potentially be used for comparison of sea ice thickness CDR.

- Clarify that this is a catalogue and description of datasets that can support validation, but were not designed to do so

We have in the manuscript added the following statement in the third paragraph of the introduction to make this clear to the reader: *"... we acknowledge that the datasets in the CCI SIT RRDP were not inherently collected or post-processed with the aim of being used for satellite evaluation; thus, their use requires careful consideration of inherent biases, preferential sampling, and measurement uncertainties."*

- Explain that you have compiled datasets not originally produced for validation and have described known limitations and uncertainties in each dataset. This should include outlining how/why each dataset can serve as a reference (accuracy, reliability, resolution), what gaps it fills, and how it should (and should not) be used. Although you describe what the datasets are, how they were processed, and link to previous studies that describe the uncertainty of the original datasets, there is a lack of discussion on their accuracy, reliability, and resolution in the context of satellite validation, the stated primary intended purpose of the dataset.

To address this point we have made several changes to the manuscript, with some of them being mentioned above. We further highlight the limitations during our introduction of the quality flags Section 4.4, as well as in the SIT examples in Sections 7.6 and 7.7. We finally, updated the conclusion.

- Address the concern about OIB quick-looks not being designed as a long-term archived dataset for satellite validation.

We address this in the introduction *"... the datasets in the CCI SIT RRDP were not inherently collected or post-processed with the aim of being used for satellite evaluation; thus, their use requires careful consideration of inherent biases, preferential sampling, and measurement uncertainties."* We further address in Section 2.1 second paragraph that: *"Data after this are provided as quicklooks (Kurtz et al., 2016), which basically means that significantly less processing has been performed to provide a quick assessment of the collected data, leading to biases in the derived products, see Section 2.1.2."* where we describe the inherent biases of OIB QL. Finally, we acknowledge the limitations of OIB QL in the conclusion: *"Furthermore, care should be taken when using OIB QL products due to existing biases in snow depth processing, as this dataset were only meant to be processed for a quick assessment of the data and not designed as a long-term archived dataset for satellite inter-comparison."*

Two other key issues stand out in the reviewer's comments

- Please address Reviewer 1's technical concern about using median quantities in skewed distributions.

We here refer to our response to Reviewer 1

- Please address Reviewer 2’s comment around the N-ICE2015 HEM data. In your response to the first review, you mentioned that the N-ICE2015 data would be included; however, this was not done, and the references have been removed. Please either include the dataset or provide a clear justification for not including it.

We apologize for this - N-ICE data from HEM has now been included. Data from buoys lacked information on position of the buoys and were only partly processed. This data has therefore not been included.

I look forward to receiving your revised manuscript that addresses these fundamental concerns.

References

Langsdale, M. et al. (2025). “The Challenges and Limitations of Validating Satellite-Derived Datasets Using Independent Measurements: Lessons Learned from Essential Climate Variables”. In: *SurvGeophys* 2025, pp. 1–38. DOI: 10.1007/s10712-025-09898-4. URL: <https://link.springer.com/content/pdf/10.1007/s10712-025-09898-4.pdf>.

Response to referee

ilo and hsk

January 2025

Please find below reviewer comments provided in black and responses in blue.

1 Answer to review by anonymous - General review

It is in the response to the first reviewers comment the authors mentioned that data from N-ICE2015 would be included in the study. Though to the best of my knowledge this wasn't done and in fact references to that extensive field campaign has instead been removed from the manuscript. As HEM data was collected during this campaign should rather be included in this manuscript.

We apologize for not including the N-ICE2015 HEM data in the latest revision. We have now included the data set and updated the manuscript accordingly.