Response to referee

ilo and hsk

January 2025

1 Answer to review by anonymous - General review

The manuscript aims to combine a range of "reference" dataset containing sea ice thickness, sea ice freeboard, snow depth and sea ice draft measurements. It's commendable the amount of work that has gone into compiling this dataset. However, the dataset and the description of it requires some additional work. Specifics has already been pointed out by Reviewer 1 and I have opted to not repeat those also here.

We appreciate the reviewer's time and feedback, which have helped identify areas for improvement. We acknowledge the concerns raised and will make the necessary revisions to improve clarity, consistency, and accuracy. Below are our detailed responses to each comment. Please, see also our replies to reviewer #1

2 Comments

A concern for the quality of the reference dataset is the introduction of unknown information into the reference dataset, e.g., by adding a timestamp to the AEM-AWI when a time is missing. The adding of a time stamp (effectively assigning the data additional information) is then listed as a very minor pre-processing. Similarly, is the submarine dataset manipulated by adding dates to the dataset, the pre-processing level is then listed as 1-3. Minor pre-processing in my opinion would be removing clearly erogenous data points in a QC of the data, not adding information. A definition of the authors interpretation of the pre-processing flags is therefore needed in the manuscript. For section 8 should it be clearly stated that at times NaN have been replaced with arbitrary time stamps.

We will clarify the use of pre-processing flags to ensure sharper and more precise definitions and we propose to add this information in a table for better overview. Regarding the AEM-AWI data, the additional information provided is a timestamp within the given day, meaning the maximum error introduced is 24 hours. Given that the data is later averaged on a monthly scale, the impact of this addition is minimal. However, we acknowledge the need to explicitly state this in the manuscript to avoid confusion.

The IMB-CRREL data is listed as the most important data in this study (L337-338) and has also been listed as a dataset where major pre-processing is needed, and that the data should be used with care (L342). Assigning the dataset as the most important and at the same time the most unreliable raises questions to the usefulness of this entire reference dataset.

Agreed. However, the statement on L337-338 appear misleading. The intended meaning was that mass balance data (snow depth and sea ice thickness) are more relevant to our study than temperature measurements (available in the buoy data). We will revise the wording accordingly.

In section 3 the levels 0-3 are listed as pre-processing flags, and in section 4 levels 0-3 are listed as uncertainty flags. Using the same numbers/listings may cause confusion, a recommendation is therefore to use different numbers for the different types of levels.

Thank you for raising this point, we acknowledge that using the same numerical values (0-3) for both pre-processing and uncertainty flags may be confusing. However, as they are provided as two separate variables, they should be straightforward to implement and separate from one another, when using the data. This is a common use of flags (e.g., ICESat-2 implements several quality flags with values ranging from 0-2 or more). We will clarify in the text, with use of acronyms (QF = Quality Flag, UF = Uncertainty Flag), which flag the specific value is specified for.

How was the length of the different sensor time series selected? E.g., why is the Fram Strait wide mooring data after 2018 not included in the study (L224), why are IMB buoys only until 2015 included (L242), and why are only ASSIST data until 2021 included (L278)? On L238, why is the updated version of the data not used in this study? The updated dataset is from 2022 which is now 3 years ago.

We thank the reviewer for this comment. In terms of the Fram Strait wide mooring data, the data source from which we have gotten the data only contains data until 2018-08-31 (Sumata, Divine, and Steur 2021). All IMB data from after 2015 lacks measurements of snow depth and sea ice thickness. The latest deployed buoy was the 2016A buoy in the Beaufort Sea but this buoy, unfortunately, contains no measurements of snow depth and sea ice thickness (Perovich, Richter-Menge, and Polashenski 2022). We propose to extend time-series, such as the ASSIST data, and include data of similar types, such as HEM from MOSAiC, Nansen and N-ICE, but not include entirely new data types (see also responses to reviewer #1).

Deformed ice appears to be defined as ice >3m thick (L441). Meaning that large parts of MYI will fall into the deformed ice category and many FYI areas, incl. rubble fields, will be classified as level with this definition. Should perhaps a different thickness range have been used here? Rough ice is used in other places (e.g. L462), what is the overall sea ice classifications used in this work? Should rough be the same as deformed? For clarity please include a sea ice type definition early on in the manuscript, I understand that there may be a large number of ice types and that there will be significant overlap between different ice types but it's good to be able to see the definition in one place.

This is a great recommendation. We acknowledge that the use of rough, deformed and thick sea ice interchangeably is confusing and this will be revised accordingly.

Section 4.2.4. This data is based on visual observation and is very dependent on the experience of the person making the observations, yet this dataset has been given an uncertainty flag of 1. Whereas data that are independent on human errors such as the airborne data has been given an uncertainty flag of 3. Measured and quantified errors have therefore been given a higher degree of uncertainty than those who is not easy (impossible?) to assess errors. E.g. how common is it for the estimates for the ship data to be made by an experienced observer? In addition is there a statement on L367-369 where the ship-observations show the opposing trend to what is expected, is the reason behind this the data quality? On L669 the SIT for ship observations is described by the authors as dubious. Should the ship-based observations perhaps therefore be ranked a level 3 like the airborne measurements? Section 7.5. It is great to see uncertainty in the data being discussed, though a clear and consistent definition is needed.

These are valid concerns. In terms of the ship data, we believe that the opposing trend, e.g. that the sea ice thickness measured by ships tends to be lower than the sea ice thickness measured by satellites, is mainly caused by a bias in ship routes toward thinner ice. Regarding your point about the uncertainty flagging, we have currently used uncertainty flags as a way to quantify, the level of variability and information that we have in our uncertainty, but as you point out we lack to quantify information such as representation error and human interpretation in our flagging. We will look into making updated uncertainty flag categories, to take into account these. However, differences in data sources make it challenging to establish a fully consistent flagging method across all datasets without introducing a degree of subjectivity. We encourage suggestions to specific datasets, which you believe should have different uncertainty or preprocessing flags.

Section 7.1. Flight measurements may at times avoid certain ice types, such as thin/young ice, and open water areas which will affect the sea ice distribution in these datasets. The deployment of IMBs on pre-dominantly stable (thicker) sea ice is brought up it would be useful to also discuss the un-representativeness of the ice types in the air-borne data. This may help explain the discrepancies discussed on L638-640. The sea ice in the brackish Baltic Sea has a very low salinity, in areas equivalent to fresh ice. What is the error uncertainty associated with this ice type in the reference dataset?

Thank you for raising this. We were not aware of the fact that flight measurements may at times avoid certain ice types, and have not encountered this when considering flight lines below satellite orbits where the main aim is to capture the same ice conditions as observed by the satellites. We have since been made aware that single-engine helicopters may avoid flying at low level over water. However, missing open water is not expected to have a large impact since the CCI sea ice thickness is the "mean thickness of the ice-covered fraction of the grid cell area". Nevertheless, this is an interesting point, that we will look

further into and we will make sure to raise this point in the manuscript when we discuss representativeness.

A "reference" dataset such as the one mentioned could be used by other research groups outside the altimeter community. The down sampling to 25 km NH (50 km for SH) grid cell of the reference coordinates, therefore, makes the compiled dataset less useful. Others within the altimeter community may also want to perform this down sampling in different mathematical ways, potentially rendering the dataset less used than if the original resolution of the reference data is kept. The effect of this down sampling should also be discussed with the uncertainty assessment in mind, i.e. what uncertainty is introduced by the down sampling. A quantification of the uncertainties and errors introduced to the data should be discussed and quantified in section 7.6.

As we have made the processing pipeline (code on GitHub available Ida Olsen and Henriette Skourup 2025), where it is relatively easy for the user to change time-space averaging to their needs. We will at this stage not be able to quantify the errors introduced by the time-space averaging as this is ongoing research, see also general comments to reviewer #1. However, we propose to include additional information in the paper about how the code can be modified to be used for other purposes, to make it more straightforward for readers to obtain the reference data in the desired temporal and spatial resolutions. Publishing the data in its original resolution amounts to redistributing the source data and this is not in the scope of this work. However, links to all source data are provided in table 2, making it straightforward for potential users to obtain the used source data in is original resolutions.

3 Minor comments

L4-5. What is meant with format matching? The file type, the data type?

We acknowledge that this is not very well formulated. We here refer to time-space averaging. We will update the text accordingly.

L47-48. The issues with combining in-situ/airborne/drone based etc validation data to any remote sensing sensor is challenging due to the differences in temporal and spatial resolution. The snow depth is not unique in this regard, and it is unclear why this parameter is listed as uniquely challenging.

Here, we justify why we include snow depth reference measurements in the RRDP, even though it might not be the prime measurand or even derived from the dedicated satellite altimetry mission. Specifically, snow depth plays a pivotal role when converting into sea ice thickness from satellite (or airborne) altimetry, and therefore a comparison with reference observations of other aspects of the sea ice (its thickness or draft) will rely on the snow depth product used in this conversion. However, we have been encouraged by reviewer #1 to update the introduction with more focus on the reference observations. During this process, we will update these sentences for clarity.

L138. Why is section 8 listed before 3,4... etc?

Thank you for bringing our attention to this. This will, of course, be adjusted to keep the sections in chronological order.

Table 1. ASSIST data originates not only from ice breakers, but also ice capable ships etc. Later in the text the terms ship (e.g. Figure 2) or support vessels are used, it would be good to be consistent throughout the manuscript. Ship would suffice.

Thank you for raising this point, we will use ship throughout the manuscript for consistency.

Figure 2. How was dominate data source assigned? >8%?

If you refer to the numbers in bold between campaign/provider and geophysical variable in Figure 2, then these numbers were highlighted as they show the primary contributor (whichever is highest) to each of the four main geophysical variables (SD, SIT, FRB and SID). We will clarify this in the caption.

Introduce the acronyms the first time they are used, e.g. now AEM is used many times not explaining that is stands for Airborne Electro Magnetic soundings (?) EM is explained on L184.

Thank you for pointing this out. We will ensure that acronyms are described when first mentioned.

L337-338. What makes this the most important data for this study?

As previously mentioned, this is a mistake in the wording and it will be updated.

L451-452. What is the uncertainty for the NPI data?

The uncertainty for NPI data is described in lines 453 to 458 and is stated in Table 4.

L467. Consider using months instead of summer and winter to allow for easier interpretation of the time period for the SH and NH.

Thank you for the suggestion. We will implement this approach.

L476. Is depth = draft depth?

Yes it is. We will here correct "depth" to "draft depths" when necessary.

References

Sumata, Hiroshi, Dmitry Divine, and Laura de Steur (2021). *Monthly mean sea ice draft from the Fram Strait Arctic Outflow Observatory since 1990.* DOI: 10.21334/npolar.2021.5b717274. URL: https://doi.org/10.21334/npolar.2021.5b717274.

Perovich, D., J Richter-Menge, and C Polashenski (2022). *Observing and understanding climate change: Monitoring the mass balance, motion, and thickness of Arctic sea ice.* URL: http://imb-crrel-dartmouth.org.

Ida Olsen and Henriette Skourup (Feb. 2025). *ESA-CCI-RRDP-code.* Version v1.0. DOI: https://doi.org/10.5281/zenodo.14808969. URL: https://github.com/Ida2750/ESA-CCI-RRDP-code.