

Title:

An Arctic sea ice concentration data record on a 6.25 km polar stereographic grid from three-years' Landsat-8 imagery

Authors:

Hee-Sung Jung, Sang-Moo Lee, Joo-Hong Kim, and Kyungsoo Lee

Main Overview:

This study focuses on the development on a new high-resolution dataset from Landsat 8 for retrieval of sea ice concentration. There is novelty in this approach compared with more traditional passive microwave methods, and whilst the notion of sea ice concentration retrieval from Optical/IR methods is not intrinsically new, the research presented here demonstrates work of significant value – it is very nice to see research that transitions away from the MODIS sensor (which is approaching end of life and Terra no longer has a regular orbit), and the amount of data processed at high resolution is very novel, and quite exciting! I wholeheartedly agree with the authors thoughtful and well-presented argument of the importance of an independent from PMW SIC dataset given in the introduction.

The code used in this study is also on zenodo, this is really nice to see! I particularly was appreciative of the 'requirements.txt' outlining the specific versions of packages used in this study, this is excellent to see from the perspective of reproducibility. The code is organised with a logical structure and is sufficiently commented and documented for publication.

The dataset is also published as linked on zenodo, which is a suitable archive. The data is in a typical netcdf format with sensible file names, the time parameter is logical with a clearly defined epoch in the metadata, fill values indicated, and it is intuitive and easy to open and plot this data, which looked reasonable when I plotted some data in the Central Arctic.

This is a very good and innovative concept which has been robustly executed. The methodology is sound. On the whole, the manuscript is thorough, but at this stage there are some gaps in justification of particular choices and insufficient detail in the some of the explanations, particularly in regards to the visual inspection component of cloud masking, as I have outlined in my Specific Comments. Therefore, it is my view that this manuscript should be published subject to minor revisions. The manuscript would benefit from some additional proofing, I outline minor typographical corrections where I see them but do not intend for this to be exhaustive.

Specific Comments:

General Introduction: There is, of course, already a 1km SIC dataset produced by the University of Bremen based on merging Optical/IR MODIS data with PWM methods. The value of your work is that is truly independent of PWM, and the UoB dataset isn't. Using MODIS in 2024 also has it's own problems. However, I think you need to write a paragraph on this dataset in the intro, and explain the niche of what your dataset does that this one doesn't. Certainly, without reference to this dataset I do not feel your argument can be considered complete.

Ludwig, V., G. Spreen, & L. T. Pedersen (2020). *Evaluation of a New Merged Sea-Ice Concentration Dataset at 1 km Resolution from Thermal Infrared and Passive Microwave Satellite Data in the Arctic*. *Remote Sens.* 12(19), 3183. [doi:10.3390/rs12193183](https://doi.org/10.3390/rs12193183) [Article (PDF file)]

Line 107: I think there is a better way that 'the rest of this paper' but I like seeing an outline at this point in the paper. I'd probably also do a line break before to enhance clarity.

Line 117: This is probably a good time to mention the 'pole-hole' or unimaged portion of the central Arctic owing to Landsat-8 inclination. You could also do a diagram to illustrate which bit of the Arctic can't be imaged, but I leave this to your judgment. Noted that you explain this in line 130 but it really is quite important for it to go as soon as the sensor is introduced in my opinion.

Line 122: I appreciate why Landsat-8 was chosen over Sentinel 2 given data coverage and robustness of cloud masking. However, I think you need to explicitly explain why you chose Landsat-8, as someone less familiar with Arctic sea ice research (and hasn't had the pleasure of trying to get Sen2Cor working over sea ice!) is going to be scratching their head wondering why you didn't choose Sentinel 2 (given the bands you use are overlapping, Landsat-8 band 5 and 6 approx. equal to S2 8a and 11, and S2 has a better resolution and repeat time). So, my advice here is to add in a couple of sentences explaining why you chose Landsat-8 and make an explicit reference to why you didn't choose Sentinel-2.

Line 127: What method of cloud mask is implemented here? Is it appropriate? Why did you choose 10%?

Section 2.4: I agree with the study choice to use Ice Charts for validation, and I think you have done a good job in conveying why this is the most appropriate choice.

Line 224: Great explanation!

Line 230: The phrase 'a visual inspection' is not exhaustive enough an explanation for this step. I want to know exactly what this process entailed, what images you looked at, what questions you asked, how you looked at them, how many images you looked at, whether anyone else looked at them. I think a diagram of one typical judgement you made would be assistive. I need the thought process here to be outlined to the extent that I could replicate this step and generate the labels myself purely based on what is on the page, and it feels too muddled at the moment for me to plainly understand your meaning.

Line 270: Are these Top of Atmosphere (TOA) or not? In my view it is OK to use TOA for Arctic studies as the aerosol profiles are very poorly constrained and the magnitude of the atmospheric component is small compared to the bright surface, but this should be made very clear. If these are not TOA what processing was applied, and what is the justification for appropriateness?

Line 355: This is robust error propagation and well justified.

Line 380: Thanks for making this clear, it's the right choice and good to see.

Line 461: This is actually quite similar to how MOD29 is derived, which is considered to be one of the better cloud masks over sea ice. I leave it to your judgment if you feel inclusion of that would strengthen or weaken your case, but I think it is robust in any case.

Section 5.2: There does seem to be a notable bias between the ice chart and your dataset at lower SIC%. Can you comment on why this might be?

Line 578-59: always lovely to see DOIs but these should be referenced properly as datasets and cited in text.

Section 5.3: At what point in the season is this Landsat-8 data or is it the whole dataset? Would a more specific analysis around peak melt pond presence perhaps be better? (It is perfectly fine if you do not agree with this but adjust the manuscript to justify what choice was made and why in any case).

Figures, Tables, and Equations:

I feel that your captions lack detail throughout. Figures and Tables should have terms used redefined so they can be interpretable independently of the text surrounding it. This will really help improve the clarity of the overall paper and should not be too arduous. You have done a very good job of defining terms within equations – I want to see the same clarity from the captions associated with your Tables and Figures.

Figure 1: Add reference to the mask you used to make this figure. Please also add in the projection and the software you used to make this map. I do not think a scale is necessary on a pan-Arctic map. I would also prefer to see your lat/lon labels appear in-front of the boxes, rather than behind, ie. The 70 deg is partially obscured in the Laptev sea. Please also reiterate the 'study period' with dates, the figure should be as self-contained as possible without reference to the text.

Figure 2: As with figure 1, add a citation for the regional mask and please also add in the projection and the software you used to make this map.

Figure 3: Really like to see diagrams like this, they're very helpful graphical illustrations of your work and improve readership comprehension. Good job!

Table 2: As with figures, tables should be as self-contained as possible. Please add a very brief explanation of C1, C2, C3, and C4 in your caption, as you have with Figure 5.

Figure 4: Redefine p_5 in your caption please.

Figure 5: Redefine p_5 in your caption please.

Figure 6: Please add a very brief explanation of C1, C2, C3, and C4 in your caption, as you have with Figure 5.

Figure 7: I think these plots warrant a scale bar, what do you think? I'd also prefer to see the dataset for the true colour images as a reference instead of a hyperlink to a USGS tool.

Figure 8: Define σ_{SIC} and p_5 in your caption please.

Table 4: Very quick definition of CFMask please.

Figure 9: Very nice to see error bars and the percentiles these confer to explicit. What R metric is this? Not clear if Pearson/Spearman/Coefficient of Determination or other, so make this explicit in your caption.

Figure 10: I'm not sure at all what this trendline represents, or if it has much physical meaning. Either add context to it in your caption or remove it.

Figure 11: Put the USGS source as a reference. Define specifically what R we are talking about. A scale bar would be nice if it is easy to do, but not essential. Define BT and NT.

Figure 12: What are the units for (b)? Add a definition for 'in' and 'out' in (d).