Response to essd-2024-264 RC1

- The referee's comments are in blue
- The authors' responses are shown in black

The article presents a data set of high resolution and high quality sea ice concentration maps derived from Landsat data that can be used to validate algorithms for ice concentration fields derived from other satellite data. The authors note that the data should find utility as a benchmark data set by which to judge the accuracy of passive microwave-derived ice concentration data. While the spatial extent of each classified Landsat image is quite small relative to a passive microwave ice concentration field, the large number of classified Landsat images should somewhat make up for this limitation.

The data and methods are new, while having been built upon established methods for classifying visible and near-IR data. Methods are described in detail with sufficient references. Error estimates and sources of error are given and discussed in the article. The data are accessible and can be downloaded from the linked Zenodo site. There is one NetCDF file for each of 12 regions. I plotted SIC fields from the Beaufort region file using Panoply, and displayed a number of image days at random. These looked reasonable, but I did not assess the data quality beyond that. I'll note that the summary on the Zenado site needs significant copy editing.

I rate the data set as excellent in terms of its uniqueness, usefulness, and completeness.

The article is well-structured, and overall, the presentation quality is very good, but in some places it lacks clarity. I have made comments under Technical Corrections that may help the authors improve clarity.

The authors sincerely appreciate your time and valuable comments which definitely led to a much better version of the manuscript. In revising the paper, we strove to take up your valuable suggestions and comments and incorporate them into the revision. Please check the authors' responses to the comments below.

Specific comments

Section 3.1 says that an addition step in cloud mask quality assessment was to visually compare the cloud mask array from each image with the corresponding true-color image. Please provide more information on this process, including who did the visual inspection and approximately how much time it took for each image. More that 15,000 images are a lot to visually screen, and this step should be described in more detail.

Thank you for pointing this out. The authors agree with your comment and the necessity of more detailed description of the visual inspection process. In the revised manuscript, details of the visual inspection process, including the exact steps taken in the visual inspection process, an example case to show how the visual inspection was executed, who did the inspection, and the time it took to perform the inspection, are added in the appendix. A sentence to point the readers to the appendix was also added in Section 3.1.

[Added] 'Further details of this visual screening step are provided in Appendix B.'

[Added] Appendix B: Visual Inspection for Cloud Mask Quality Control

'In this section, a step-by-step description of the process taken to perform the visual inspection of Landsat-8 scenes is presented. As defined in Section 3.1, each pixel in a Landsat-8 scene can be sorted into the following four categories depending on the state of the cloud mask for the pixel: False negative (FN; cloud pixel mistaken as clear pixel), false positive (FP; clear pixel mistaken as cloud pixel), true negative (TN; clear pixel identified as clear pixel), and true positive (TP; cloud pixel identified as cloud pixel). It is noted that the pixels with FN are used to calculate SICs while the pixels with FP are not, indicating that the presence of FN pixels can directly introduce errors in the calculated SIC value Therefore, visual inspection was performed very strictly to detect FN pixels.

Figure B1 outlines the steps taken to perform the visual inspection. The descriptions of each step are provided along with an example case of a Landsat-8 scene that is categorized into C1 during the section-wise inspection stage.

Step 1. Generating jpeg file of cloud mask (i.e., cloud mask image).

> For each Landsat-8 scene, a false-colour image with each pixel classified as ice (white pixels in Fig. B2b, d), open water (blue pixels in Fig. B2b, d), cloud (grey pixels in Fig. B2b, d), and fill value (black pixels in Fig. B2b, d) is constructed using the OpenCV module in Python. Ice and open water pixels are then differentiated using the method described in Section 3.2. Cloud pixels are classified by masking the medium confidence cloud, high confidence cirrus, cloud shadow, and dilated cloud pixels identified by the quality assessment band (i.e., the cloud mask array produced by CFMask).

Step 2. Comprehensive inspection of cloud mask quality.

> The cloud mask images generated in Step 1 are visually inspected against the true-colour image to identify sections populated with FN, FP, TN, or TP pixels. This is done in the following order: First, if no cloud pixels are observed from both the cloud mask image and the true-colour image (i.e., all pixels in the image are TN pixels), the scene is labelled as C4. Second, if any cluster of FN pixels is observed, the scene is labelled C1. Third, if any cluster of FP pixels is identified, the scene is labelled C2 and passed on to Step 3. If the clusters of cloud pixels in the cloud mask image are well corresponding to the position of clouds observed in the true-colour image (i.e., TP pixels), the scene is labelled C3 and passed on to Step 4 (Fig B2a, b).

Step 3. Comprehensive inspection of cloud mask quality for C2.

> For the scenes passed on to this step (i.e., scenes labelled C2 from Step 2), the cloud mask image is recreated using a higher confidence threshold (i.e., high confidence cloud, high confidence cirrus, cloud shadow, and dilated cloud pixels) for the quality assessment band. The new cloud mask image is visually inspected against the true-colour image, and if any cluster of FN pixels are observed, the confidence threshold for the quality assessment band is returned to its initial value (i.e., medium confidence cloud, high confidence cirrus, cloud shadow, and dilated cloud pixels). If the observed cluster of cloud pixels in the new cloud mask image is well corresponding to the position of clouds observed in the true-colour image, the higher confidence threshold is kept, and the scene is passed on to Step 4.

Step 4. Section-wise inspection of cloud mask quality.

In this step, the identified clusters of TP pixels are inspected in more detail. For each cluster of TP pixels observed, we zoom in (i.e. about 1000 × 1000 pixels; the full-size image is approximately 8000 × 8000 pixels) to the section of the cluster to check for the existence of FN pixels. If any FN pixels are found within the cluster, the scene is labelled C1 (Fig B2c, d).

An example of how a Landsat-8 scene may be categorized according to the process described in Fig B1 is presented using the case of a Landsat-8 scene acquired on Mar. 25, 2022, over the Barents Sea (Fig B2). First, from Step 2. (i.e., the comprehensive inspection step), visual inspection of the cloud mask image (Fig. B2b) against the true-colour image (Fig. B2a) shows that the position of clouds in the cloud mask array is generally well corresponding to those observed in the true-colour image. Therefore, at this step, this scene is labelled C3 and passed on to Step 4 as described above. Next, the section-wise inspection of the cloud mask quality is performed by zooming in to the cloud areas. This is illustrated in Fig. B2c and Fig. B2d, which is a zoomed in image of the area enclosed by the red rectangle in Fig. B2a and Fig. B2b. Inspection of this sub-section shows the presence of unmasked cloud shadow pixels, which results in the erroneous classification of ice as open water. Therefore, at this step, the label of this scene is changed to C1.

The visual inspection was done by HJ and it took approximately 5 – 10 minutes to inspect one Landsat-8 scene for cloud cover.'

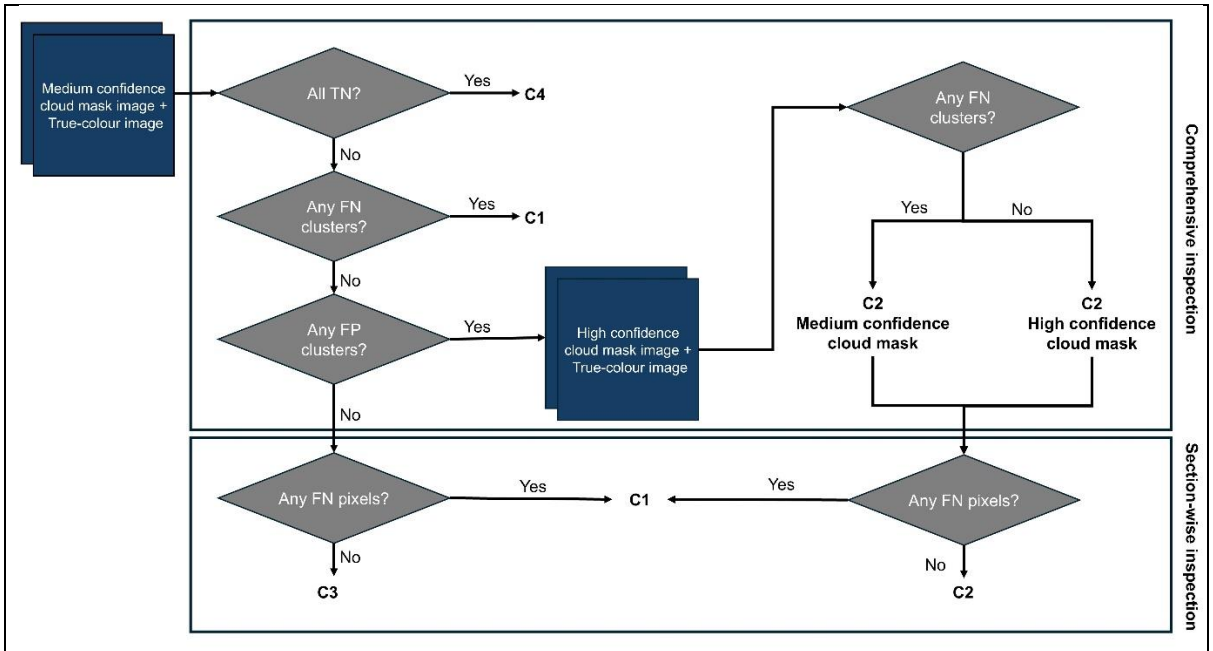

Figure B1: Processing pipeline of the visual inspection step. Each Landsat-8 image is labelled as C1 (i.e., underestimated cloud cover), C2 (i.e., overestimated cloud cover), C3 (i.e., correctly estimated cloud cover for cloudy sky), or C4 (i.e., correctly estimated cloud cover for clear sky) depending on the observed dominance of true negative (TN; clear pixels identified as clear pixels), false negative (FN; cloud pixels mistaken as clear pixels), false positive (FP; clear pixels mistaken as cloud pixels), and true positive (TP; cloud pixels identified as cloud pixels) pixels.

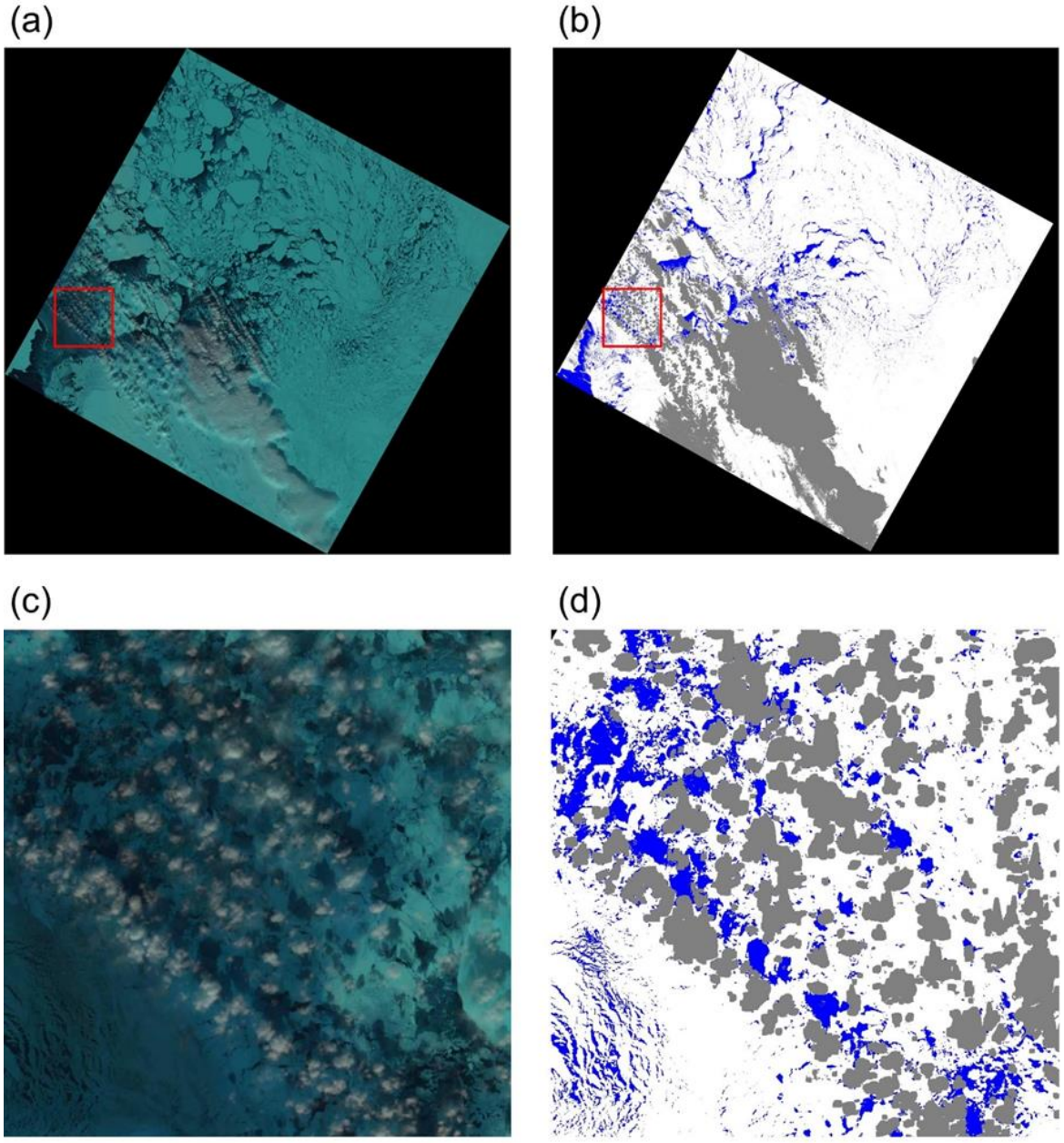

Figure B2: The case of a Landsat-8 scene classified as C1 (i.e., underestimated cloud cover) from the visual inspection step. Shown in the panels are (a) full size true-colour image, (b) full size cloud mask array, (c) true-colour image of the area enclosed by the red rectangle in (a) and (b), and (d) cloud mask array of the area enclosed by the red rectangle in (a) and (b). The blue, white, gray, and black pixels in (b) and (d) are open water, ice, cloud, and fill value pixels, respectively. The scene is from Mar. 25, 2022, over the Barents Sea. The true-colour image is obtained from Earth Resources Observation and Science (EROS) Center (2020).

Technical corrections or wording suggestions

The authors appreciate the technical corrections and wording suggestions. The authors fully agree with your suggestions in that they can help improve the clarity of the manuscript. Please see the changes made for each of your corrections/suggestions.

(Line 36) Remove "at least"

Thank you for your suggestion. In the revised manuscript, "at least" has been removed.

(Line 54) There have been developed various PMW SIC algorithms >> Various PMW SIC algorithms have been developed

In the revised manuscript, the suggestion has been incorporated.

(Line 72) However, there exist discrepancies >> However, discrepancies exist

In the revised manuscript, 'there exist discrepancies' is changed into 'However, discrepancies exist' by following your suggestion.

(Line 78) What is meant by "ice/water mixtures"? Note that the differences in algorithms isn't due to the presence of these things, but rather due to the differing sensitivity of the algorithms to these things, so it may be helpful to reword this sentence to reflect that.

Thank you for your catching of the rather ambiguous phrasing. "Ice/water mixture" was meant to reflect the state of sea ice with melt pond presence which can lead to changes in the emissivity of sea ice. The sentence has been changed to better clarify the meaning of "ice/water mixture" and to reflect that the differences in the retrievals are due to the differing sensitivity of the algorithms.

[Old] (Lines 78-79) '…due to the presence of melt ponds and ice/water mixtures, as well as a humid atmosphere…'

[New] '…due to the differing sensitivity of retrieval algorithms to the presence of melt pond and the associated emissivity change, as well as a humid atmosphere…'

(Line 88) It would be helpful add a sentence here that notes that you will use the Kern et al. data to validate your own data as described in Section 2.3.

Thank you for this helpful suggestion. A sentence mentioning the utility of the dataset by Kern et al. (2022) is added to the revised manuscript after line 88.

[Added] The dataset by Kern et al. (2022) is also utilized for validation of the produced Landsat-8 SIC in this study, and the results of the comparison are presented in Section 3.2.

(Line 107) The use of "sub-region" confused me. Here, and for most if not all other occurrences

throughout the paper, “region” would serve equally well. Consider changing “sub-region” to “region” throughout.

The authors fully agree with your opinion. In the revised manuscript, all occurrences of “sub-region” were replaced to “region”.

(Line 118) short-wave IR >> short-wave IR (SWIR)

Thank you for your catching. “(SWIR)” has been added in the revised manuscript.

(Line 122) Suggest you add “(used in the Normalized Difference Snow Index)” to tie this back to the abstract. So it would read “…SWRI band 6 (used in the Normalized Difference Snow Index)…”

Thank you for your suggestion. This is done in the revised manuscript as the following.

[Old] (Line 122) ‘…SWIR band 6 were used in this study.’

[New] ‘…SWIR band 6 (used in the Normalized Difference Snow Index) were used in this study.’

(Line 167-170) This isn’t clear. Does it mean that six of the scenes that Kern classified were used by the authors to validate their method? Or does it mean that six of the scenes that Kern classified are being offered to readers in the supplement, so that readers can evaluate the author’s method? I think it means the former. A re-written sentence might read something like this: “In order to evaluate the classification method suggested by our study [to distinguish it from “this study” used earlier for Kern’s study] we processed Landsat 8 reflectance from six clear-sky scenes that Kern (2021) had classified, and then compare results.” Then, point readers to where those comparisons can be found (Section 3.2?)

Thank you for the comment and suggestion. The sentence (Lines 167-170) in the original manuscript does indeed mean the former as you pointed out. In the revised manuscript, to clarify the unclear meaning of the sentence, this has been changed following your suggestion.

[Old] (Lines 167-170) ‘In order to evaluate the ice and water classification method (see Section 3.2) suggested by this study, 6 classified scenes (Kern, 2021) under clear sky condition of which scene location and time are provided in the supplements Fig. S1 and Table S1 and the corresponding Landsat-8 reflectivities were used.’

[New] ‘In order to evaluate the classification method suggested by our study we processed Landsat-8 reflectance from six clear-sky scenes that Kern (2021) had classified, and then compared results. The result of the comparison is presented in Section 3.2 and the location and time of the Landsat-8 scenes that were used in the evaluation are provided in the supplements Fig. S1 and Table S1.’

(Supplements) In the caption for Figure S1, it looks like “left of each panel” and “right of each

panel" are reversed. Also, where it says "…and the reference classification map (left of each panel) are provided", consider changing to " …and the reference classification map that our method produced (right of each panel) are shown",

The authors sincerely thank you for this catching. The caption in Figure S1 has been revised following your suggestion as the following.

[Old] …True-color images (right of each panel) and the reference classification map (left of each panel) are provided…

[New] …True-color images (left of each panel) and the reference classification map that our method produced (right of each panel) are provided…

(Line 264) Classification of a Landsat-8 pixel into ice and open water >> Classification of a Landsat-8 pixel as ice or open water

Thank you for this suggestion. The sentence has been changed following your suggestion in the revised manuscript.

(Line 276) Remove "in order"

Thank you. In the revised manuscript, "in order" has been removed.

(Line 282) Remove "steps"

Thank you for this suggestion. In the revised manuscript, "steps" has been removed.

(Line 301) are not consisted solely >> do not consist solely

Thank you for your suggestion. In the revised manuscript, 'are not consisted solely' is changed into 'do not consist solely'.

(Line 321) are not fully concentrated by Landsat-8 pixels >> are not entirely filled by Landsat-8 pixels

Thank you very much for your suggestion. 'are not fully concentrated by Landsat-8 pixels' is modified by 'are not entirely filled by Landsat-8 pixels' in the revised manuscript.

(Line 331) each twelve >> all twelve

Thank you for the correction. In the revised manuscript, your correction is incorporated.

(Line 366-368) As mentioned in section 3.3, Landsat-8 SIC can be largely deviated from actual

SIC if Landsat-8 measures partially-covered grid cell, in other words, SIC computed from partially-covered grid cells may not be representative of actual ice coverage over the entire grid cell >> As mentioned in section 3.3, SIC computed from partially-covered grid cells may not be representative of actual ice coverage over the entire grid cell

Thank you for your suggestion which led to the more concise phrasing. In the revised manuscript, the phrase has been changed by following your suggestion.

(Line 391) consider adding "along with the mean and standard deviation of sea ice concentration" after "…shown in Fig. 6"

Thank you for your suggestion. In the revised manuscript, your suggestion has been reflected as the following:

[Old] (Line 391) '…shown in Fig. 6 (see Table S9 in the supplementary for values).'

[New] '…shown in Fig. 6 along with the mean and standard deviation of SIC (see Table S9 in the supplementary for values).'

(Line 415) Should "..estimated over the pixels with such wrongly-masked pixel.." be "estimated for grid cells with such wrongly-masked pixels …" ?

The authors appreciate your catching. Your description is indeed more accurate than what was in the original manuscript. This has been changed following your suggestion in the revised manuscript.

[Old] (Line 415) '…estimated over the pixels with such wrongly-masked pixel…'

[New] '…estimated for grid cells with such wrongly-masked pixels…'

(Line 424) are well corresponding >> correspond well

Thank you for your suggestion. In the revised manuscript, 'correspond well' can be found rather than 'are well corresponding'.

(Line 443) sub-range >> range for line 443 and also in the figure caption

Thank you for your comment. The suggestion is fully incorporated in the revised manuscript.

(Line 445) The contribution of the two threshold variables to σSIC was found that ρ5 threshold explains most of … >> Still, the ρ5 threshold explains most of …

The authors thank you for your valuable suggestion which led to a more concise phrasing of the sentence. In the revised manuscript, your suggestion has been reflected.

(Line 446) In spite of the relatively high uncertainty in Landsat-8 SIC ranged from 20% to 80%,

>> In spite of the relatively high uncertainty in Landsat-8 SIC between 20% and 80%,

In the revised manuscript, 'In spite of the relatively high uncertainty in Landsat-8 SIC ranged from 20% to 80%,' has been changed into 'In spite of the relatively high uncertainty in Landsat-8 SIC between 20% and 80%,' by following your suggestion.

(Line 460) …category, sub-section with 100% cloud cover based on visual inspection, but less than 100% cloud cover from CFMask was selected. From the collected sub-sections, the ρ5 and NDSI values were collected … >> …category, those having 100% cloud cover based on visual inspection, but less than 100% cloud cover from CFMask were selected. From these images, the ρ5 and NDSI values were collected …

Thank you for your suggestion. The authors totally agree that the latter sentence better clarifies the process of testing the unmasked cloud pixels. In the revised manuscript, therefore, the following correction has been made:

[Old] (Lines 460-462) '…category, sub-section with 100% cloud cover based on visual inspection, but less than 100% cloud cover from CFMask was selected. From the collected sub-sections, the ρ5 and NDSI values were collected …'

[New] '…category, those having 100% cloud cover based on visual inspection, but less than 100% cloud cover from CFMask were selected. From these images, the ρ5 and NDSI values were collected …'

(Line 472) …and thus SICs produced … >> …and thus for SICs produced …

Thank you for your correction. This has been corrected in the revised manuscript.

(Line 482) "Chart" should be plural: "Charts"

Thank you for your correction. This has been corrected in the revised manuscript.

(Line 495) The spreads >>The spread

Thank you for your correction. This has been corrected in the revised manuscript.

(Line 498) "SIC from the ice chart was found to be positively biased to Landsat-8 SIC," Would it be more clear to say ""SIC from ice charts tends to be higher than that found using Landsat-8 SIC, "

Thank you for your suggestion. The authors agree that the meaning of the sentence is better clarified in the latter. Your suggestion has been incorporated in the revised manuscript.

(Line 528) should be "Although a few…"

Thank you for your correction. This has been corrected in the revised manuscript.

(Line 529) …because melt ponds are not easily discernible to open water, … >> …because melt ponds are not easily distinguished from open water, …

Thank you for your suggestion. This has been changed by following your suggestion in the revised manuscript.

(Line 533) has robustness >> is robust

Thank you for your suggestion. This has been changed by following your suggestion in the revised manuscript.

(Line 617) Comparison of Landsat-8 SIC against SIC retrievals from NASA Team (NT) and Bootstrap (BT) algorithms reveal >> Comparison of Landsat-8 SIC against SIC retrievals from NASA Team (NT) and Bootstrap (BT) algorithms for two cases reveal

Thank you for the detailed correction. What is shown in Section 5.4 is indeed limited to two cases of the PMW retrievals. This has been corrected by following your suggestion in the revised manuscript.

(Line 619) related with >> related to

Thank you for your correction. This has been corrected in the revised manuscript.