Response to the review tc-2019-120-RC3 by reviewer #3, Francois Massonnet,
of Reviewing Satellite Passive Microwave Sea-Ice Concentration Data Set Intercomparison: Closed Ice and Ship-Based Observations by Kern et al.

Summary:
Sea ice concentration (SIC) datasets are essential for a number of polar applications. Evaluating their accuracy and precision is therefore necessary to gain confidence in the value of those datasets. This study performs an assessment of ten popular products obtained from satellite microwave radiometry, against independent data. The goal is to inform users of those datasets about the accuracy and precision of each dataset. A secondary goal of the study is to question current practices, like the truncation of SIC at 100% or the dangers of using threshold-dependent diagnostics (like sea ice extent, SIE) when weather filters are applied

Overall I found this study interesting and, being a user of several of the products presented here, I am now more alert about pitfalls that such products can have. The main strength of the manuscript is that it provides a systematic and thorough assessment of the products, with many maps/histograms/time series that users can refer to when they will be using the data themselves. I also appreciate the work done in the Appendix to document the various algorithms. Having this information in one single paper is really a plus for the scientific community and anyone susceptible to work with this data one day. The main weakness of the manuscript is that sections 3, 4, 5 are overly long and descriptive without interpretation of the results (which comes in section 6). To be honest, paragraphs like the one at line 613-621 are difficult to follow. So many numbers, statistics and product names are given that it is just not possible to process the information when reading. So, I feel that the text could improve in clarity if the authors chose to focus the description on a few key points rather than trying to exhaustively describe the figures each time.

So, the paper is worth publishing in a journal like the Cryosphere but I think it would deserve a haircut in the middle sections because this is where the authors will lose most readers, unfortunately. Drowning readers with too detailed information in the text has the danger that the most interesting bits (like the issues of truncation at 100%) go unnoticed, so I think it's worth the effort.

We are very grateful for the positive perception of our manuscript by the reviewer. We are in line with the main concern that the manuscript is overly long, which is shared by reviewer #2 and the editor. In order to substantially shorten the manuscript we

- avoid to describe the content of figures and tables more than once
- avoid to repeat numbers given in figures and tables as much as possible
- substantially condense the discussion and conclusion, e.g. by shortening the description (see first point) and moving elements of the interpretation from section 6 into the respective section 3 to 5.

We also removed Figures 14 and 16 as we thought that the information given therein is kind of redundant to what is shown in Figure 15 and 17 anyways.

On top of that, I do have a few general, minor and editorial comments that I hope the authors can address.

General comments and questions ———————————————-

- The manuscript revolves around key concepts of metrology like accuracy, precision, uncertainty, error. To my knowledge, the ocean community has tackled the issue of defining the meaning of each term, while the sea ice community has not yet done so. Several authors of the manuscript organized a fruitful workshop in October 2017 where those questions were addressed. The present paper represents a timely opportunity to introduce those concepts to the cryosphere readership, so I wonder if one additional section or paragraph could not be devoted to explaining those basic concepts. The work needed is modest but the impact can be high, especially in terms of citations of this paper later on.

We are grateful for this comment. However, the paper is long and we needed to add relevant information and explanations based on the reviewers' comments. Because of this and the fact that the issue with the uncertainties is covered in parts of the literature already, we will carefully need to examine during the review process if we can follow this suggestion, or whether we instead simply include a link to the relevant published literature.

- Having used the ASPeCt data myself, I know that this dataset, while very valuable, has to be taken with extreme caution. The reason is that (1) individual measurements are prone to large errors because they are estimated visually, (2) the data is not necessarily representative since the ships preferentially navigate in thin ice areas, so retrievals might be biased thin (for sea ice thickness) with possible adverse consequences on SIC as well (I suppose, biased low) (3) co-locating the data is challenging. Points (1) and (3) are covered in the manuscript (nice discussion at the head of Sect. 5, by the way), but the point (2) would deserve one or two lines of additional comments.

We agree. The largest impact from navigating in thin-ice areas will possibly be for sea-ice thickness, snow depth and deformation features – independent of season. Observations of the sea-ice concentration are less affected during winter when leads are mostly frozen. During summer and/or in the marginal ice zone with more open water, a slight under-estimation of the true sea-ice concentration is more likely. We added in Line 549: "*Ships tend to avoid thick and deformed sea ice. Ship-based observations therefore often represent the thinner ice categories and/or conditions encountered in leads or openings. These are frozen over with sea ice in winter but are open water in summer. Therefore, particularly during summer or episodes of warmer weather, the sea-ice concentrations from the small scale ship-based observations are likely lower than from the larger scale satellite microwave radiometry. According to Ivanova et al. (2015), microwave radiometry tends to underestimate sea-ice concentrations over very thin (< 15 cm) ice. This suggests that during winter, sea-ice concentrations from ship-based observations could be, contrary to summer, slightly higher than from satellite microwave radiometry. While usage of daily, along-ship track averages reduces these effect it cannot mitigate it.*"

- I like the idea of creating groups of products that share similarities, that's a novelty as far as I know. I was not entirely clear though if those groups were built based on March Arctic SIC (as introduced at line 345) or following a more global criterion? That is, are the groups defined based on expert a priori knowledge, or are they based on the SIC fields they produce? According to my understanding, the former approach was chosen. But then, are the groups stable when the hemisphere changes, and when the season changes? It would be useful to know for the users if products that are close together in winter in the Arctic are also close together for summer and in Antarctica, for example. If not, a few explanations would be welcome.

We are grateful for this comment. The motivation to create groups was to ease interpretation of the results. We kind of used a more global criterion than only taking the March Arctic SIC to assign groups. However, we cannot give strict, quantitative boundaries about which product belongs to which group and we see that this is a shortcoming. We therefore revised the groups based on the algorithmic background. Group I still contains the ESA-CCI – OSISAF products. Group II contains the CBT-like algorithms, which is CBT-SSMI; CBT-AMSRE and NOAA-CDR (the latter is a combination of CBT-SSMI and NT1-SSMI but it is dominated clearly by CBT-SSMI); note that we removed NT2-AMSRE from this group. NT1-SSMI and ASI-SSMI are assigned to group III. These algorithms represent a slightly different concept to retrieve the sea-ice concentration where the sea-ice concentration is mainly based on the brightness temperature polarization difference; these algorithms still use tie points. Finally, NT2-AMSRE is assigned to group IV; its concept to derive sea-ice concentrations via a look-up table and model atmospheres is fundamentally different from the other nine algorithms.
The changes we applied to the manuscript are too widespread to document these in the reply and we kindly ask the reviewer to take a look at the revised manuscript. However, we introduce the grouping now in Section 2.1 as follows:
"*We group the products according to their concept for sea-ice concentration retrieval (Table 2, column "Group"). Group I contains the ESA-CCI – OSI-450 products. Group II contains the CBT-like algorithms, which is CBT-SSMI; CBT-AMSRE and NOAA-CDR (the latter is a combination of CBT-SSMI and NT1-SSMI but is dominated clearly by CBT-SSMI). NT1-SSMI and ASI-SSMI are assigned to group III. These algorithms follow a different concept to retrieve the sea-ice concentration where the sea-ice concentration is mainly based on a brightness temperature polarization difference. Finally, NT2-AMSRE is assigned to group IV; its concept to derive sea-ice concentrations via a look-up table and model atmospheres is fundamentally different from the other nine algorithms.*"

- In the scatterplot analyses (Figs. 15 and 17), a surprising result is that the evaluation is strongly dependent on whether one bins the satellite data to the ASPeCt categories (red colors) or one bins the ASPeCt data to the satellite categories (blue). I could not find in the text a satisfactory explanation of why this is so, could the authors elaborate? Can the better agreement in the case of red colors be used as justification to favor one evaluation method over another?

We are grateful for this comment. It made us to think again about these results. We see the necessity to include an illustration (new Fig. 17 and new Table 7, see below) that explains our observations. In short: We don't think that one of the colored (red or blue) should be favored over the other but rather see this as an imprint of the various ice conditions encountered as we explain now in the manuscript. We added:

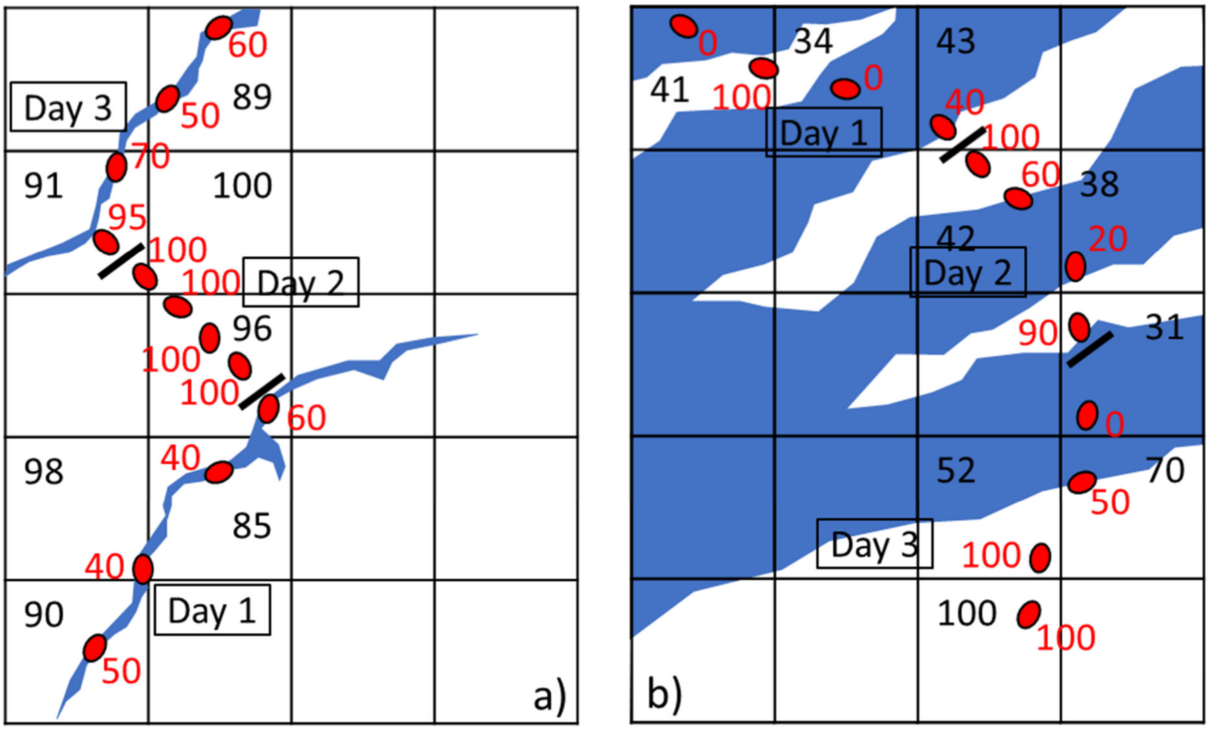

Fig. 17: Illustration of the representativity of ship-based observations (red ellipses and numbers) compared to gridded satellite observations (black grid and numbers) for a) close pack ice with leads and b) an open sea-ice cover in the marginal ice zone. Size of ellipses is in scale with the grid-cell size of 25 km by 25 km. Short black bars denote transitions between days. See also Table 7.

Table 7: Summary of hypothetic daily average sea-ice concentrations from Figure 17 for top: the pack ice case (a) and bottom the marginal ice zone (MIZ) (b). Left two columns: sea-ice concentrations as shown in Figure 17; middle two columns: illustration of the effect of sea-ice concentration under-estimation for thin ice; right two columns: illustration of the effect of sea-ice concentration over-estimation, e.g. thick ice with a wet snow cover (top), or of sea-ice concentration under-estimation, e.g. MIZ during end of summer (bottom).

Pack ice, a)	Lead is open water		Lead is thin ice; Sat underestimates by 20%		Lead is open water; Ice surface causes 5% over-estimation on 50% of the ice	
	Ship	Sat	Ship	Sat	Ship	Sat
Day 1	48	92	100	98	48	96
Day 2	100	95	100	99	100	98
Day 3	68	90	100	97	68	95
MIZ, b)			50% of sea ice is thin ice; Sat underestimates by 20%		50% of sea ice is soaked wet = open water for Sat	

	Ship	Sat	Ship	Sat	Ship	Sat
Day 1	35	40	35	36	35	20
Day 2	67	38	67	35	67	19
Day 3	63	63	63	57	63	44

"*With Figure 17 and Table 7 we aim to illustrate sources of the observed discrepancies between ship-based and satellite sea-ice concentrations. For the pack ice / lead case (Fig. 17 a), the variation of the ship-based observations depends strongly on the fraction of thin ice. For leads covered by open water satellite sea-ice concentrations tend to exceed ship-based concentrations. This could explain the banana-shaped distribution of data pairs for the SH (Fig. 16). For leads covered by thin ice, it is the other way round because ship-observations would reveal a total sea-ice concentration of 100% while most of the tested algorithms underestimate the true sea-ice concentration in presence of thin ice – the more the more thin ice is present and the thinner it is [Cavalieri, 1994; Comiso and Steffen, 2001; Ivanova et al., 2015]. This could explain the larger fraction of ship-based sea-ice concentrations near 100% as is particularly pronounced, e.g., for NT1-SSMI (Fig. 15 a) and, in general, the larger range of satellite versus ship-based sea-ice concentrations at comparably high concentrations. For the MIZ (Fig. 17 b), Table 7 bottom), it is more likely that most products provide smaller sea-ice concentrations than observed from a ship. During winter, a considerable fraction of the sea ice might be thin ice – causing under-estimation as stated above. During summer, a considerable fraction of the sea ice might be too wet to be recognizes as ice by satellite microwave radiometry – causing an under-estimation as well (see e.g. Worby and Comiso, 2004; Ozsoy-Cicek et al., 2009). In addition, low ice concentrations are often filtered by the weather filters applied. This results in a substantially larger range of ship-based sea-ice concentrations at comparably low satellite sea-ice concentrations. Furthermore, this also results in a larger fraction of low ice-concentration values for the satellite than the ship-based observations. This explains why mean ship-based sea-ice concentrations per binned satellite sea-ice concentration (blue symbols in Fig. 15 and 16) are shifted so much to the right compared to the red symbols; this is also evident from the larger fraction of data pairs below than above the 1-to-1 line for sea-ice concentrations below 60 to 80%.* "

Minor comments (starting with a line number) ——————————————————

81 Two studies are mentioned in which SIC was evaluated using independent data. It would be good to end that paragraph with a sentence stating the main limitations of those studies (Andersen 2007 is quite old I guess, Beitsch 2015 is only focusing on the Antarctic?) and why the present study is necessary.

We changed the part: "*… for the Arctic, and …*" to "*… for the high Arctic, and …*" and added in Line 83 after the last sentence: "*Both these studies each focused on one hemisphere only. The work of Andersen et al. (2017) is relatively old and is based on comparably old version of the algorithms and products. In the present paper, we inter-compare the newest available versions of the sea-ice concentration algorithm and products used in both studies, including three CDRs. We perform our evaluation for both hemispheres. Additionally, we take advantage of a recently published new calibration / validation data package (see Section 2.2).*"

88 The sentence "We focus on differences..." sounds like this paper will focus on winter conditions, and the summer sea ice concentrations will be analyzed in a subsequent paper. But the present manuscript covers in fact summer as well. Why would a second paper be needed then? Is it more the type of independent data for evaluation that will change?

In order to reduce potential confusion we changed L89-91 from:
"*… with winter-time near-100% reference sea-ice concentrations, and with a large suite of ship-based manual visual observations of the sea-ice conditions. The second contribution is going to focus on Arctic summer conditions, presenting and discussing results of an inter-comparison with sea-ice parameters derived from MODerate resolution Imaging Spectroradiometer (MODIS) satellite*

observations.” to “… *with near-100% reference sea-ice concentrations, and with a large suite of ship-based manual visual observations of the sea-ice conditions. The second paper is going to focus on an inter-comparison with sea-ice parameters derived from MODerate resolution Imaging Spectroradiometer (MODIS) satellite observations in the Arctic.*”

104 What is meant by "continuation"? I understand that the dataset production must be going on, but I am unsure.

We replaced “*continuation*” by “*sustained production*”.

146 I’m not entirely clear regarding the diagnostic used here (and in Fig. 1). From what I understand, only SIC in the range]0 %, 30%] is retained. Then, for each day of the month, one identifies what the 5% percentile of that SIC spatial distribution is (not all grid cells participate due to the restriction to the quoted interval). Finally, those 5% percentiles are averaged in time. Is that correct? Do not hesitate to spend one or two extra lines to avoid ambiguities in how the diagnostic is computed.

Thank you for your comment and suggestion regarding the text. We changed: “*the weather filters by computing time-series of the monthly mean 5% percentile of sea-ice concentrations of the range]0.0% to 30.0%], i.e. excluding the grid cells set by the open water filter to exactly 0.0%.*” to “*the weather filters. For this we focus on the sea-ice concentration interval]0.0%, 30.0%], i.e. exclude grid cells set to exactly 0.0% by the open water filter. Then, for each day of the month, we identify the 5% percentile of all gridded sea-ice concentrations falling into the above-mentioned interval. Subsequently, we average over those 5% percentiles.*”

147 Along the same lines, I’m not sure to understand why the diagnostic would not be impacted by secular sea ice trends and why stability is expected. The SIC distribution has been profoundly changing over the past decades (mean and higher-order moments), so why would percentiles not be impacted as well? In fact, it would be quite remarkable that a sea ice diagnostic, especially an Arctic one, would not exhibit trends over several decades. Clarifications would be welcome here.

With this diagnostic we aim at deciding if the “minimum detected sea-ice concentration value” is stable from one product to the other, between hemispheres, and across time (both multi-decadal and season). Although the sea ice concentration mean and higher order have been greatly impacted by climate change, we argue that the full range of sea-ice concentration values (between 0% and 100%) is present in the Arctic and Antarctic at any time, and that a product should detect this minimum value in a consistent way through time. This particular diagnosis was built through several iterations. For example, Lavergne et al. (2019) used the 1% percentile (of the same SIC range). We also tried the 0%-percentile (minimum value) but it was too noisy. We also tried to not cut the SIC values to <30% (thus take range (0-100%]), but then a seasonal cycle was (even more) apparent, with lower values of the diagnostic during summer than during winter. Because the 5%-percentile diagnostic still showed quite some seasonal cycle for some of the products, we went for plotting the March and September average values separately, so that sharp transitions in the long-term evolution of the diagnostic would appear better. Finally, the fact that the timing of the sharp transitions of our diagnostic coincide with transitions between sensor families (SMMR to SSM/I, SSM/I to SSMIS) for some products and not others, strongly suggest that they are the result of the underlying algorithm (specifically, the weather filters), and not the climate-driven evolution of the sea-ice cover.

We did not change the text.

151 As far as I know, the computation of SIA does not interfere with the choice of the 15% SIC threshold, so that statement only applies to SIE?

We are not sure what you referring to here. Either you mean that the 15% SIC threshold is not used for the computation of the SIA. Or you mean that in absolute terms it does not really matter whether one uses such a threshold or not. We'd say that it is quite common to use the 15% SIC threshold also for SIA (see e.g. Cavalieri and Parkinson, 2012; and Parkinson and Cavalieri, 2012, both in The Cryosphere 6(4)) but we note that others don't do it (Comiso et al, 2017, in J. Climate and J. Geophys. Research- Oceans); we think this is currently under debate.

In order to clarify this issue a bit better in our paper we changed in L150-152: "*… the 15% SIC value so that it does not interfere with the computation of sea-ice area (SIA) and extent (SIE) that use this threshold to define the sea-ice edge (e.g., Gloersen et al., 1992).*" to "*… the 15% SIC threshold that is commonly used in the computation of the sea-ice extent (SIE) (e.g. Gloersen et al., 1992; Meier et al., 2014, Comiso et al., 2017) but also of the sea-ice area (SIA) (e.g. Gloersen et al., 1992; Cavalieri and Parkinson, 2012; Ivanova et al., 2014).*"

178 The title of section 2.1.4 is "Distribution around 0% and 100%" but the section actually deals with near 100% concentration. This is a bit strange or I missed something.

Thank you for pointing this out. We changed the title of the section accordingly to "*Distribution around 100%*"

180 "sea-ice concentration on either side of 100%" is a sentence that could puzzle a few readers. I think I understand what it means, i.e. that the algorithms are designed in such a way that, in presence of noise, their output can yield above-100% values for sea ice concentration, and that some products apply truncation. But it would be good to write this down in one or two sentences. This is somewhat explained in the conclusion (line 943) but should appear earlier in the manuscript.

We are grateful for this comment because it is an important element of this manuscript. We provide this information now in the context of Section 2.1.4.

205 It would be good to know by how much the "near 100%" SIC values of the RRDP2 dataset deviate from 100% in the Areas of Interest. Are we talking about a maximum 1% deviation, 0.1%? Is this information available in the cited papers?

We changed "0% and 2%" in L208 to "0.4% and 1.5%" and replaced "*We only use AOIs for winter months, i.e., November through March (Arctic) and May through September (Antarctic). By doing so we ensure that small-scale openings, which may remain in an area of high ice concentration and net ice convergence, are frozen and thus sea-ice concentrations are close to 100%. We use only AOIs for which the average area reduction between day 1 and day 2 is between 0.4 and 1.5%.*" by: "*We cannot provide a definite uncertainty for this reference data set but for its production, we combine a suite of measures to ensure high precision and close-to-zero bias (high accuracy). The drift/convergence selection is based on convergence on two consecutive days of 1-day drift. During winter, i.e., November through March (Arctic) and May through September (Antarctic), this is assumed to ensure that all openings existing on day 0 (prior to the two convergence days) are closed by convergence or refrozen. The refreezing assumption is the reason why the quality of the product is higher during winter when openings rapidly refreeze whereas during summer some openings may still persist. There is no prior assumption of the initial ice concentration (on day 0), but the ice-drift product generally requires quite high concentrations for the 2-D cross correlation to work. Andersen et al. (2007) reported an accuracy of 2% for ice-water SAR image classification from ice analysts without additional drift / convergence information and a sea-ice concentration standard deviation of ~1% for own cold-season high-resolution high-quality SAR image classification. Based on our above-*

mentioned measures and the results of Andersen et al. (2007) we can state estimated values for precision: ~1% and accuracy: <0.5% for our reference data set."

288 What is the origin of the 98% value used for filling the pole hole? Why this choice?

We now motivate this choice in the manuscript by changing: "*… of 98%; this applies …*" to "*… of 98%. Andersen et al. (2007) found a mean sea-ice concentration of ~98% from a comparison of cold season passive microwave and synthetic aperture radar observations in the high Arctic. They noted a smaller value of ~95% for summer. Both values are confirmed by own work (Kern, 2018, http://icdc.cen.uni-hamburg.de/fileadmin/user_upload/ESA_Sea-Ice-ECV_Phase2/SICCI_Phase2_SIV-Retrieval_Report_v02.pdf, last accessed September 6, 2019). Using 98% instead of 95% during summer results in an overestimation of the SIA of about 10 000 km², a value small compared to other sources of biases for the SIA during summer. This filling is applied …*"

318 I find interesting the result that the spread for SIE is less than for SIA. This finding would warrant more discussion because it highlights that SIE, despite being certainly a less physical diagnostic than SIA, offers advantages as soon as one is not so much interested in errors above 15% SIC. Linking this finding to earlier papers by one of the co-authors (Notz et al., https://www.the-cryosphere.net/8/229/2014/tc-8-229-2014.html) would be a plus. In particular, what are the implications of Fig. 6b versus 6c when it comes to model evaluation? If I would like to evaluate an atmosphere forced ocean-sea ice simulation and focus on the iconic year 1996 and in September, for example, do I have an interest in using SIE instead of SIA? Or is it risky to use SIE because it could hide error compensations? I noticed that this is briefly touched in the paragraph at line 677, but it would deserve more discussion. In the current state, Fig. 6 could leave the feeling that the use of SIA should be discarded because it is less precise than SIE.

It is not new that SIE is less sensitive than SIA to SIC difference between products. At the same time, SIE contains less information than SIA. By definition one expects SIE to agree better across products than SIA, just because it by definition removes a lot of the uncertainty – actually for most of the sea-ice concentration range. Without having properly quantified the biases in the SIC during summer because of melt ponds it could be that SIE is a measure that could still be preferred over SIA in the Arctic, i.e. for three months. Nevertheless, it is not a helpful metric for a lot of reasons, including that in all publications, the mentioning of SIE without always giving the underlying grid is meaningless (as SIE depends on grid resolution). If we were about to give a recommendation than this would be that a comparison based on SIE should always only be a first (crude) step which needs to be followed by a comparison based on SIA and then SIC.
Note that we already had, and still have a paragraph in the discussion section (6.1.1) where we highlight recent findings (including the paper by Notz mentioned by you) and strongly recommend to use SIA instead of SIE.
We believe that the matrices we provide wherein we show the inter-product differences in SIC, SIA and SIE are quite informative into this direction as well. Sure, also these matrices makes one to favor SIE … but why? Just because all the products don't differ to much to each other? Is that a quantitative quality measure? We cannot force users to switch to SIA. All we can do is to show and discuss results in an as much as possible objective way. It is the SIC accuracy which comes first. And we do not hide problems neither with the known algorithms (too much SIC folding and truncation, weather filter issues) nor with the new algorithms (surface properties, too aggressive land-spill over filter). We tried to put proper formulations of these concerns and issues into the final version of the manuscript.

326 The deviations in SIC are computed as the difference of each observation and the ensemble mean. Given the small sample size, wouldn't it better to compute the difference with the median instead of the mean? The risk is that if there is an outlier, it will drag the mean towards itself and differences could not be so large. By choosing the median, one increases the probability to detect outliers (I think).

Thank you for this comment. We agree that the risk you described is not zero. Hence, to further optimize the manuscript we now replaced the differences to the mean by the differences to the median. As you will recognize, however, the change between these two approaches is not that large – but noticeable at places. Note that instead of Figures 7, 9, and 10, the respective figures now are Fig. 8, Fig. 10 and Fig. 11 due to a re-arrangement of figures.

344 I find interesting the idea of clustering the products in three groups. It would be useful to write the group name in each of the panels of Figs. 7, 9, 10, in order to visualize how this translates in terms of spatial maps.

We followed this suggestion and annotated the maps in question accordingly. See also our comment to the general comment about the assignment of groups. Note that we also annotated the Figures 15-18 (former 14 to 17) to ease interpretation of the results.

428 In the discussion of the near-100% SIC (this section and Fig. 12), it would be useful to tell the reader what a perfect product would look like. I assume, such a product would have all its power concentrated in the upper bin of Fig. 12 (almost like NT2-AMSRE), correct? In essence, I'm asking if the authors could interpret the results of Fig. 12 in terms of performance. From the current discussion in section 4.1, it is unclear if a few products are superior to others for that diagnostic.

We are hesitant to come up with a quantitative statement here. It is, on the one hand difficult to translate these histograms into overall performance, because algorithms that set all SIC above e.g. 95% to 100% would look perfect here, and the same for those overestimating the SIC before truncation. In a nutshell, to perform well at 100% SIC should ideally not be to the detriment of the rest of the SIC range – which would be the case however. We basically use this figure to illustrate how different SIC distributions near 100% are and to underline our hypothesis that it might be a good idea to look into the Gaussian fit. If you want to have a ranking, then we'd say that SICCI-50km is of superior quality compared to all other algorithms, because it allows a Gaussian fit, this fit AS WELL AS the originally retrieved SIC has quite a narrow distribution (good precision), and also exhibits one of the lowest overall biases with respect to the reference SIC (near zero bias, good accuracy).

435 I'm not sure to agree with the sentence "The distributions for these products look like as if one has taken the distribution of, for instance, OSI-450 but deleted the fractions of bins...". First, I don't quite see that from Fig. 12a. Second, by construction, the cumulative fractions have to sum to 1.0, so that products with smaller ranges have to compensate by higher frequencies. So they cannot simply be a truncated version of the other cumulative fractions, right?

Thank you for this comment. We i) made clear further up in the manuscript what we actually mean by "truncation" and ii) reformulated the whole description of this figure.

Editorial /style comments ——————————-
50 The meaning of "spatially and temporally resolved sea-ice concentration" is clear to me. Are there datasets that are not spatially or temporally resolved? Could you clarify, please?

We replaced "*spatially and temporally resolved*" by "*the*".

94 "It" grammatically refers to "input" while I think the authors mean "Section 2".

We replaced "*It*" by "*This section*".

167 I wonder if "hockey-stick" is appropriate here. To my knowledge, this term was initially used for describing the shape of mean temperatures over the past millennium, with a slight decrease until 1800 followed by sharp increase later on. The shapes in Fig. 1 are arguably not following the same logics.

We replaced “reveal a hockey-stick shaped cumulative distribution” by “reveal a cumulative distribution with a first increasing, later merely constant slope”

180 "the EUMETSAT-OSISAF – ESA-CCI products": could you name explicitly which products you refer to?

All of them. We added “*(see Table 2, group I)*”.

325 "Arctic sea-ice distribution" –> "Arctic sea-ice concentration distribution"

Changed as suggested.

344 Fig. 11 is referenced but Fig. 10 and 9. have not been referenced yet.

We split Figure 11 into an Arctic and an Antarctic part and now are able to refer to the Arctic part of this figure (new Fig. 9) before the Antarctic part (new Fig. 12).

369 "Antarctic sea-ice distribution" –> "Antarctic sea-ice concentration distribution"

Changed as suggested.

429 I would change the sentence "As expected the cumulative fraction increases towards 100% for all products" into ""As expected the cumulative fraction increases towards 1.0 for all products". The cumulative fractions are expressed as units in the figure.

Changed as suggested.

1286 Please add the units of frequencies in the middle column of Table 1.

We added “*[GHz]*” in the top cell of the middle column.

1357 Figure 1 labels in several panels (all but a and g) are too small to read, at least when printed on paper.

We enlarged Figures 1 and 2 to the maximum possible width. Readability of the labels should have improved therefore.

1367 The meaning of the grey lines in Fig. 3 is not clear to me. The figure caption mentions black symbols and lines. I suppose the former is the vertical dashed lines at 100%. Is the former the grey lines then? Please clarify.

This is an artifact due to not using a thick line when plotting this originally black line in Figure 3 and the respective figures in the appendix. We change the wording in the caption from:
“*Black symbols and lines show values cut off at 100%*” to “*Symbols connected and extended by grey lines show values truncated at 100%*”

1377 The colormap used for labeling the years in Fig. 5 might not be colorblind-friendly. Consider changing it.

We took up your suggestion and replaced the two greenish colors by two grey tones.

1397 The lower and upper limits of the color bars change depending on the diagnostic (SIE vs SIA) and the hemisphere. I think it would be better to use the same limits (-2 –2 or -3 – 3 million km2), this will then highlight the relative merits of using SIE or SIA, and also highlights that winter Antarctic spread is very large.

Thank you for this suggestion. Actually, I always like inter-comparably of such figures. We now use a range for SIA and SIE of -4.0 to +4.0 million square kilometers for both hemispheres. This applies to all figures of this kind in the main text as well as the appendices.