Authors' response to Referee Comment 2

We are grateful for the referee's careful consideration of our manuscript and provision of a large number of comments which we find very helpful for the present revision of our manuscript, particularly for Sect. 2.

The referee has a number of suggestion of expanding the main article, *e.g.* (i) by moving material from the Supplementary Information to Sect. 2, (ii) by assessing robustness of metrics through MonteCarlo or boot-strapping approaches in Sect. 5, and (iii) by including more results from a comparison between the microwave product and ice charts in Sect. 5. Our general response is that expansions of the main article should not include material that for a large part becomes distractions from the topic, which is an evaluation of metrics for sea ice edge position forecasts. Following this guideline, we have chosen to comply with the referee's advice concerning (i) and (ii). Item (iii) is also addressed, but additional results are given in the Supplementary Information.

We are advised to change the title so that it includes references to evaluation of several metrics, and subsequent provision of recommendations for sea ice edge forecast verification. We disagree. The title should not be a long sentence, but provide enough information that the attention of an interested reader would be caught from a contents listing or from a web search lookup. We believe that our title serves this purpose. The fact that evaluations and recommendations are given follows implicitely from the title as is. The abstract has been rewritten slightly, following the relevant detailed comments made by the referee.

Please find our detailed responses to all specific comments below, and note that while we have followed the referee's advise on most of the items, there are a few items upon which we have not acted. Initial page and line numbers below (in bold) and comments (in italics) are repeated from the referee's document. This is followed by our response (in regular font) and, when relevant, reference to where changes can be found in the revised submission (in italic bold).

P1L7

Sentence not clear, in particular with the confusing use of 'concentrated':
"Such information is traditionally available as a set of metrics that provide a concentrated assessment of the information quality."

Here, 'concentrated' referred to the fact that a metric is a single number that provides a condensed assessment of a two-dimensional field. Since this is basically the nature of a metric, we have taken the referee's advise an removed this word in the revised document. *P1L7-8*

P1L14

"These metrics are analyzed in synthetic examples, in selected cases of actual forecasts, and for a full year of weekly forecast bulletins" This sentence is also confusing: are analyses performed separately for 1) synthetic examples ; 2) few real cases; 3) a full year of weekly forecast? Or only one kind of analyses on selected forecast among 1 year of weekly bulletins in some synthetic cases?

The sentence in question has been rewritten to make its content more clear. *P1L14-15*

P2L8

Is Melsom et al. (2011) reference easily available?

We have added a web reference from which Melsom et al. (2011) is available. Furthermore, we note that Melsom et al. (2011) was cited by one of the references in the present study (Goessling et al., 2016, GRL). *P18L29*

P2L9

The reference Palerme et al (2019) is only submitted: not available for readers at this stage

Referee statements to the submission to GRL of Palerme et al. have been provided. The editor concludes that the manuscript "may be suitable for publication after minor revisions". A revised manuscript was submitted to GRL earlier this week. We have not been able to find a policy statement regarding

when a citation is acceptable, so we leave it as is until we are advised differently by the editor or the technical editor.

P2L15

In these two sentences, are you mentioning statistics of the sea ice extent per se, *or statistics of erroneous determination by forecasting centres of the sea ice extent quantity? This is confusing, also the introduction of 'contingency table' made need some more detailed explanation for non-expert reader.*

Model *vs.* observation contingency tables provides results for the sea ice extent for each of the two product, as well as for the sea ice extent mismatch between the product. However, details regarding contingency tables are not appropiate in the Introduction section. Accordingly, we have added some sentences to explain this matter in Sect. 2.2. **P6L7-11**

P2L16 + P5L21

Carriers et al., *2017 reference: TYPO, this is Tom Carrieres, as mentioned in the reference list page 16... found in*
https://www.cambridge.org/core/books/sea-ice-analysis-and-forecasting/B74BD33160B03EE1FA77CC9BB80E7DA7
+ *Already mentioned above: the author is Tom Carrieres, not 'Carriers'*

The reference has been corrected. The duplicity was due to using the LaTex citation feature. **P2L17, P6L8**

P2L16

'integral quantities' of what, please clarify.

The integral quantity here is the sea ice extent. We have rewritten the sentence to make this clear. **P2L15**

P2L24

Not sure that the CMEMS, funded by European Commission DG Grow as part of the Copernicus Program can be defined as a 'pan-European project'.

We have substituted 'pan-European project' with the description given by the EU Copernicus Programme. **P2L29**

P2L26

CMEMS forecast modelling tools are not limited to "circulation models" : biogeochemical models, wave models...

We have added other model systems to the list, as suggested by the referee. **P2L31-32**

P2L28-30

number of production centres: please update following what is presented at http://marine.copernicus.eu/about-us/about-producers/

The number of CMEMS centers listed in the text has been updated. **P3L1**

P3L5-7

"As we demonstrate in this study, the assessment of quality of the forecasted ice edge position is highly sensitive to the definition of metrics, and to some degree uncertainty due to differences in observational products. The amount of available data is not a limiting factor in this context" This sentence is a concluding statement that should not appear this way in the introduction of this article.

We have rewritten these sentences along the lines suggested by the referee. **P3L11-13**

P3L16

Please rephrase. You mean 'between' model and observed quantities. And 'eg' looks not adequate here: this is not an example among many... It is your purpose to investigate discrepancies between Model and Observed estimates of sea ice edge position.

We have replaced 'in' by 'between' (**P3L22**). Further, the referee implies that our analysis is limited to comparisons between model results on one hand and observations on the other. This is incorrect.

Metrics like the ones we examine are also used when comparing results for ice edge position between different observational products, which is what we do in Sects. 5 and S2 where we compare ice charts with a microwave product.

P3L18

'grid properties'... you mean here 'grid characteristics'? 'properties' might be more general

We have rewritten 'grid properties' as 'grid cell quantities'. ***P3L24***

P3L25

In equation (1) please define the 'logical AND' symbol that might not be known by all readers

A statement on the symbol $\wedge$ used for logical AND has been added after Eq. (1). ***P4L4***

P3L27

"We also introduce the metric position of grid cell" confusing. Do you refer to the geographical coordinates in a given frame of the cell i,j?

We have rewritten 'metric position' as 'coordinate position'. This is not the geographical longitude, latitude position, but the coordinate position from origo in a projection plane. ***P4L9***

P3L29 + P3L30 + P4L1 + P4L9

"Next, for each edge grid cell in each product, we find the distance to the nearest edge grid in the alternative product." Again confusing. Why not saying ... for each grid cell in the model product, we find the distance in the observed product, or vice-versa? You have just defined above O and M, and it is not clear to what refers 'alternative'
+ Why introducing 'Ealt' when just above you have introduced 'Eo'?
+ still confusing: what to call the 'reference product'? M or O?
+ Here the confusion mentioned above clearly appears: Equation 4,5,6 contain reference to 'M' and 'O' while reader can believe that 'Ealt' was 'O'.

Following these suggestions and comments, we have removed references to the 'alternative product' and 'reference product' and rewritten Sect. 2.1 accordingly. ***P4L8-16***

P4L2

Equation (2) looks like the Euclidian distance between a given ice edge position between the 'alt' product (not clear as mentioned above) and the 'reference product' (also not clear) QUESTION: how are associated the ice edge cells between the 2 compared products? I assume that for a given cell in the first products, several cells could corresponds in the second product.

A statement on the symbol $\forall$ used for the FOR ALL operator has been added after Eq. (2). $\min$ is the minimum function (applied to all distances to all grid cells in the second product). ***P4L11***

P4L3-5 + P4L23-24

Not clear if separating situation with/without considering ocean/land boundaries need to be discussed by providing equation (3), similar to equation (2). Maybe just including the ocean/land node point when presenting the detailed explanation on the way this metrics is computed might be sufficient?
+ Again, not sure this is useful.

A good number of the referee's comments and suggestions ask for more information, and we think rightly so in most cases. However, here the referee asks us to omit information as removal of Eq. (3) is recommended, and then the way that the resulting metrics are introduced after Eq. (7) is criticized. We find that keeping Eq. (3) is an approach that is more in line with the general level of detail in the manuscript, even more so for the present revision than for the initial submission. By keeping Eq. (3) we find that sufficient information is provided regarding the separation between the metrics defined by Eq. (4)-(7), thus these are not repeated for the hatted metrics counterpart. Hence, no action has been taken in response to these items.

P5L10-15 [1]

The two metrics should be discussed: in practice what do they inform on? In particular 'A^+ - A^-'

---

[1]Erroneously listed as P4L10-15 in the referee statement

Here, A^+ and A^- expresses mismatching of the sea ice extent between model and observations. We have added a sentence at the end of the relevant paragraph to make this clear. ***P5L19-20***

P6L1-6

For the sake of simplicity, some diagrams could have been provided, summarising the different configurations of grid cell with/without ice edge and the way the length is determined

To demonstrate how the ice edge length is determined, we have added a schematic figure and updated the text accordingly. ***Fig. 1, P6L18-24***

P7L7 + P8L1

"Next, we introduce the coarse grid ice edge fraction for a neighbourhood with an extent of n grid cells as" This definition deserve much more explanation, because this is key-definition to understand equations 17 to 20. "with an extent of n grid cells" is not clear to me, and I imagine for many readers, unless reading the Roberts and Lean (2008), what I have done the shortcut of the present text. Please, give more comprehensive definition before your equations.

+ It is unfortunate that the supplement explanations are not directly introduced in the article: this is the way Roberts and Lean (2008) proceeded to give shape of their explanation and equations. This should be done in the present article.

We include information that was previously provided as Sect. S1.2, now in the main article in Sect. 2.3. This includes a figure (revision of Fig. S2 in the original submission) where the concept of neighbourhood size is exemplified. We believe that this reorganization of text and a figure makes the presentation easier to comprehend for readers who are new to the FSS score. ***P7L24-P9L14, Fig. 2***

P8L25

"We will demonstrate in Sect.s 4 and 5 below that differences which are qualitatively similar to the Modified case are important to leading order for the quality assessment of the ice edge position in the forecasts from CMEMS ARC MFC". typo in '[gibberish]' Again the authors introduce here, too shortly, some conclusions obtained later on in this article. This is rather difficult to follow and confusing.

The sentence in question has been rewritten to point to the subsequent discussion in Sect.s 4 and 5, without stating a conclusion. We cannot find the typo that the referee indicates, likely because the quote on the pdf file with the referee statement appears as gibberish. However, in the event that there is a typo, we are confident that the technical editor will spot it, in the event that our manuscript is accepted for publication. ***P10L32-34***

P8L29

"and the main purpose of this document is to present metrics for the separation in this set of lines" Again very difficult to understand. Document ? This particular example of Fig 1? The full article? lines... the ice edge lines? a line of discussion?

The "document" refers to the entire paper. The sentence in question has been rewritten to better reflect our ambition. ***P10L10-11***

P9L6

"From experience, we know that discrepancies where sea ice emerges or disappears at a distance from other ice covered regions arise from time to time" Not clear. Please explain and/or re-phrase

To make clear which experience we refer to, we have added "in an operational sea ice forecasting service" at the end of the sentence in question. ***P10L2-3***

P9L10

"Since an additional discrepancy between the observations and model results has been introduced at a large distance, this change is according to our expectations". Not clear. Please explain and/or re-phrase

The discrepancy we refer to is the one that is described in the paragraph in question, and also in the section in question, as displayed on Fig.s 3 and 4. To make this clear, we have rewritten "Since an additional discrepancy" as "Since the additional discrepancy". ***P10L15***

P9L23

the CMEMS acronym is already provided.

We have retained the acronym (CMEMS) only in the present revision. **P11L2**

P10L6

Typo: overlaid.

Corrected. **P11L18**

P10L8

"In order to explore how sea ice edge metrics from actual forecasts and observations are affected by changing conditions" .. Not clear to what refers 'conditions'. Please explain and/or re-phrase.

Regarding the referee's comment that it is not clear what 'conditions' we refer to, we disagree. The type of conditions we have in mind is stated in the same sentence that the referee only partly cites: "...contrasts of the type that was examined in Sect. 3". No change has been made.

P11L2

Figure 4 horizontal axes: problem with the time labels on my PDF version. And labels (a) and (b) do not appear in my PDF version.

There was an error in the compilation of the document that gave rise to the Fig. 4 labeling issues. Fig. 5 had the same problems. We provided corrections in the *Interactive discussion* on 04 Jan 2019, see item 'AC1'.

P11L6-7

"which reveal that the sea ice extent is larger in the ice chart product than in the model product." Also mentioning that this brings the negative values of fig 4b.

We now mention the relation to negative values in Fig. 6b. **P12L23**

P11L14

I recommend to include section S1.1 into the main article.

All information that is relevant for the recommended metrics should be explained in the main article. However, our conclusion in Sect. 6.3 is that we don't recommend any of the $\widehat{D^{IIEE}}$ metrics to be included in operational validation of the sea ice edge position. Hence, on balance our preference is to keep the original organization where details on the $\widehat{D^{IIEE}}$ metrics' definitions are given in the Supplementary Information document.

P11L23-25 [2]

these statistics of comparison between ice concentration assimilated product and ice charts should be added to Table 3,4, wherever they can appear... This would be more readable.

The purpose of this study is to examine the results for metrics when two products are compared. To keep this focus, we disagree that including results from a third product in tables in the main article. Nevertheless, we wish to provide the reader with some additional results that can shed light on the underlying reasons for discrepancies. So, rather than making any changes in the main article, we add a section (Sect. S2), a figure (Fig. S3), and two tables (Tables S1-S2) in the supplementary information, so that details related to mismatching of the assimilated microwave data and ice charts are available. **S-P3L13-28; Fig. S3; Table S1-S2**

P11L29

Figure 5: In my PDF version, label (a) and (b) are mission in the figures, and it should be more readable to add x- and y-axis label titles... Also some x-axis label numbers are missing (only 1, 2, 5). What happens in both figures for lead-time days 2 to 5 ? Why curves are dashed lines and x-ticks missing (in may PDF version)?

Regarding the labeling issues, we refer to our reply **P11L2** above. Dashed lines are used to indicate results that bridge days with no data (ice charts are not produced on Saturdays and Sundays; see

---

[2]The page number is missing in the referee's report

P9L27 in the original submission, P11L5 in this revision). An explanation has been added in the figure caption. ***Fig. 7***

P11L29-30

"We also note that results for the two metrics in group 2 nearly overlap at all lead times" referring here to curves blue and red would be more readable.

We now include a reference to the two curves in questions as blue and red, as suggested by the referee. ***P13L17***

P12L1-4

"The FSS scores reveal that useful forecasts with a five day lead time are obtained at a scale of about 90x90 km, when the FSS reaches a value of 0.5 (which is criterion recommended by Skok and Roberts (2016)). When comparing with the microwave data, the FSS is well above 0.5 for a neighbourhood extent n = 5 (not shown), corresponding to useful data at a scale of approximately 60x60 km." Here It would have been interesting, with the 2017 comparison, to show the asymptotic behaviour of FSS discussed in Roberts and Lean (2008). It is also interesting to notice the higher resolution quality of the ice concentration (60km useful scales) compared to model results (90km useful scales).

We have moved the comparison between FSS results for the model product and the microwave product to the paragraph where changes as a function of lead time are discussed. Note that the comparison is now restricted to the period from January to mid-May. We have also include a figure that displays the FSS score and the asymptote values as defined by Roberts and Lean (2008). ***P13L18-23; Fig. 8***

P12L16

"by systematically computing the correlation coefficients between all possible sets of two displacement metrics" This definition is not clear. Here some more explanation of equation would be useful.

We have rewritten the sentence, and we have also added some more detail in the text on the next lines. We now refer to this analysis as "systematically computing the correlation coefficients between all possible combinations of displacement metrics time series pairs". ***P14L16-17***

P12L20-22

Not clear to what these four group refers... high, low correlation between them ? Please explain.

We now explicitly state which bounds we have used to separate large positive and large negative correlation coefficients from the intermediate and low coefficient values. The absolute values of correlation coefficient meets this criterion for metric pairs inside each of the four groups, as stated in our original submission. ***P14L20-22***

P14L9

this is the first time robustness of the metrics is discussed. As mentioned in the general review comments, there is a lack in this article of robustness assessment of the different metrics (eg, using bootstrap methodology over the 2017 data set).

We have followed the referee's suggestion, and now include results from bootstrapping in Tables 3 and 4. ***Tables 3, 4***

P14L20

Sea Ice metrics computed on specific areas was already presented in the GODAE validation article: Hernandez, F., and Coauthors, 2009: Validation and intercomparison studies within GODAE. Oceanography Magazine, 22, 128-143. http://dx.doi.org/10.5670/oceanog.2009.71

We have included the reference to Hernandez et al. (2009) in the present revision. ***P16L13; P18L17-19***

S-P2L6-10

Here a diagram/figure showing the 2 rectangles, and their overlapping area

A schematic diagram displaying a sample configuration with rectangular IIEE areas has been included in the present version of the Supplementary Information document. ***Fig. S2***