

## ***Interactive comment on “Validation metrics for ice edge position forecasts” by Arne Melsom et al.***

### **Anonymous Referee #2**

Received and published: 12 March 2019

#### 1. General comments

This article proposes to review and test existing published metrics for sea ice edge forecast verification. I would have mentioned in the title that this is a review or evaluation of a series of existing metrics. After an introduction that presents the state-of-the-art published results on sea ice edge validation, the section 2 provides a description of the metrics that are tested in this article. Then, evaluation and relative behaviour of these metrics is performed using basic synthetic cases (section 3), then 2 five day forecast synoptic cases in 2017 in the Arctic Ocean (section 4), then the full series of weekly five-day forecast over the full year 2017 (section 5). Discussion and recommendations are presented in section 6.

A supplementary information is provided that give additional description of the metrics, together with technical details.

C1

This is a technical science-based article that should be published. The topic is of high relevance for the sea ice forecast community. However this article is poorly written and organised in section 2, and should deserve a revision to better introduce every metrics in this section, merging with some part of the supplementary paper material, then have a proper annex for very technical aspects. The metrics evaluation and discussion are also lacking an assessment of the robustness of the metrics.

#### 2. Specific comments

2.1 Scientific significance The author are selecting the existing state-of-the-art metrics used by the sea ice forecast community. The overall metrics evaluation gives some guidance for any forecasting centre to perform similar tests and define adequate metrics. And the recommended metrics are based on well justified discussions.

2.2 Scientific quality The authors are choosing three kind of test-bed experiment. One "theoretical" and synthetic. Then two synoptic cases on the Arctic Ocean, at two different and relevant periods, with different sea ice distributions. Then one year (2017) of statistics established every week. All metrics are tested the same way, and the observation reference data set are ice charts, not used by the forecasting system assimilation mechanism. Only robustness of each metrics should be addressed in addition, either referring to existing published work, or discussed in this article. For instance, the 2017 statistics could be used to test the robustness of the metrics through MonteCarlo or boot-strapping approaches.

2.3 Presentation quality Section 2 needs to be extensively revised. The authors refer to some existing articles for each metrics presented: they should take example of these articles in the way metrics are scientifically, then technically presented. Because sea ice edge and sea ice edge displacement are complicated geometric features that need to be correctly explained. For instance, the authors refer to Roberts and Lean (2008) for the FSS definition. For instance this article says "The purpose of this verification method is to obtain a measure of how forecast skill varies with spatial scale in a

C2

way that can be intuitively understood by users and is also directly applicable for post-processing". We would appreciate some similar explanation in section 2. Hence, the Roberts and Lean (2008) article is very easy to read and understand, because there is a substantial effort to describe the metrics in their physical meaning, give some figures or schematics to explain the geometry, and then provide some mathematical definition: the ration of text versus equation is maybe 10 sentences for a given equation. In the present articles's section 2, there is approximately one equation after 1-2 sentences. The reader gets incomplete understanding of the essence of each metrics, in particular if the reader is not used to formal mathematical definitions (all geoscientists don't have strong mathematical background). However, there is this effort, with the supplementary document to provide more technical explanation. I strongly recommend to merge this document inside the section 2, and if needed, to separate some very technical aspect in an annex at the end of the article.

#### 2.4 Overall evaluation:

1. Does the paper address relevant scientific questions within the scope of OS? YES
2. Does the paper present novel concepts, ideas, tools, or data? YES in the sense that if recommend some adequate metrics, among all existing metrics, for sea-ice edge forecast verification
3. Are substantial conclusions reached? YES the recommendation are clear and straightforward
4. Are the scientific methods and assumptions valid and clearly outlined? YES the evaluation of the different kind of metrics is based on different case studies, and results are scientifically justified
5. Are the results sufficient to support the interpretations and conclusions? YES conclusions of the authors and proposition of the selected metrics are robust
6. Is the description of experiments and calculations sufficiently complete and precise

C3

to allow their reproduction by fellow scientists (traceability of results)? Partly, due to the fact that section 2 should provide more comprehensive description of each metrics, in order to facilitate their reproduction. . . unless reader directly refer to the reference article.

7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution? YES
8. Does the title clearly reflect the contents of the paper? NO the title should mention that this article provides an evaluation of several metrics, and eventually give some recommendations for sea ice edge forecast verification
9. Does the abstract provide a concise and complete summary? The abstract should be slightly improve to really describe the article content
10. Is the overall presentation well structured and clear? NO section 2 needs to be revisited
11. Is the language fluent and precise? NO, there are many shortcuts and some statements are difficult to understand, even if english is correct
12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? NO as mentioned above. Section 2 needs to be revisited
13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated? YES section 2, merged with the supplementary Information article
14. Are the number and quality of references appropriate? YES
15. Is the amount and quality of supplementary material appropriate? NO needs to be re-organized, with maybe a technical annex

#### 3. Technical and detailed review:

- 3.1 Abstract: Line 7: Sentence not clear, in particular with the confusing use of "con-

C4

centrated": "Such information is traditionally available as a set of metrics that provide a concentrated assessment of the information quality."

L14: "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 ?

3.2 Article:

L8 p2: Is Melsom et al. (2011) reference easily available ?

L9 p2: The reference Palerme et al (2019) is only submitted: not available for readers at this stage

L15 p2: 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.

L16 p2: Carriers et al., 2017 reference: TYPO, this is Tom Carrieres, as mentioned in the reference list page 16. . . fund in <https://www.cambridge.org/core/books/sea-ice-analysis-and-forecasting/B74BD33160B03EE1FA77CC9BB80E7DA7>

L16 p2: "integral quantities" of what, please clarify.

L24 p2: 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".

L26 p2: CMEMS forecast modelling tools are not limited to "circulation models" : biogeochemical models, wave models. . .

L28-30 p2: number of production centres: please update following what is presented at <http://marine.copernicus.eu/about-us/about-producers/>

C5

L5-7 p3: "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.

L16 p3: 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.

L18 p3: "grid properties". . . you mean here "grid characteristics" ? "properties" might be more general

L25 p3: In equation (1) please define the "logical AND" symbol that might not be known by all readers

L27 p3: "We also introduce the metric position of grid cell" confusing. Do you refer to the geographical coordinates in a given frame of the cell  $i,j$  ?

L29 p3: "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"

L30 p3: Why introducing "Ealt" when just above you have introduced "Eo" ?

L1 p4: still confusing: what to call the "reference product" ? M or O ?

L2 p4: 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

C6

could corresponds in the second product.

L3-5 p4: 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 ?

L9 p4: 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" .

L23-24 p4: Again, not sure this is useful.

L10-15 p4: The two metrics should be discussed: in practice what do they inform on ? In particular "A+ - A-"

L21 p5: Already mentioned above: the author is Tom Carrieres, not "Carriers"

L1-6 p6: 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

L7 p7: "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.

L1 p8: 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.

L25 p8: "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 as-

C7

assessment of the ice edge position in the forecasts from CMEMS ARC MFC". typo in "ÃSect.sÃ" Again the authors introduce here, too shortly, some conclusions obtained later on in this article. This is rather difficult to follow and confusing.

L29 p8: "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 ?

L6 p9: "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

L10 p9: "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

L23 p9: the CMEMS acronym is already provided.

L6 p10: Typo: overlaid.

L8 p10: "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.

L2 p11: 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.

L6-7 p11: "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.

L14 p11: I recommend to include section S1.1 into the main article.

L23-25: 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.

C8

L29 p11: 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 my PDF version)?

L29-30 p11: "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 .

L1-4 p12: "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).

L16 p12: "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.

L20-22 p12: Not clear to what these four group refers. . . high, low correlation between them ? Please explain.

L9 p14: 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).

L20 p14: 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.

C9

<http://dx.doi.org/10.5670/oceanog.2009.71>

3.3 Supplementary information review:

L6-10 p2: Here a diagram/figure showing the 2 rectangles, and their overlapping area

---

Interactive comment on Ocean Sci. Discuss., <https://doi.org/10.5194/os-2018-149>, 2019.

C10