Authors' response to the OS Editorial Board

Dear Sir, Madame

We have now completed the revision of our manuscript
Validation metrics for ice edge position forecasts
which we hereby submit in its final form to Copernicus Publications - Ocean Science; Special Issue: "The Copernicus Marine Environment Monitoring Service (CMEMS): scientific advances".

We have completed revising our original submission, based on comments and suggestions by the referees. Our detailed and itemized responses to both referee statements follow, starting on the next page. Their efforts helped us revise our original submission in a way that we find was highly beneficiary to the quality of our work. We are profoundly grateful for the referees' efforts. Note that a modest additional reorganization and editing was performed on our own initiative, notably moving a section from P11L3-10 to P12L3-10 (referring to the mark-up document).

On behalf of all authors,
Arne Melsom

Authors' response to Referee Comment 1

We would like to thank the referee for taking the time to carefully read our manuscript and provide a large number of suggestions and comments which we find very useful for the present revision of our manuscript.

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 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 mark-up version of the revised submission (in italic bold).

P3L29 and elsewhere

often "grid(s)" is used when "grid cell(s)" is meant. Also, some- times "nodes" is used instead. I recommend to use "grid cell(s)" consistently (where that is meant, of course). This also holds for the Supplement.

Where applicable we have replaced "node(s)" and "grid(s)" with "grid cell(s)" (also in the Supplementary Information document). An example where "grid" was not modified is when referring to a "stereographic grid".

P4Eq7

I suggest to make it explicit that d_o and d_m are not single scalars but sets, if I am not mistaken, by writing the right-hand-side as "$\max(\max(d_o), \max(d_m))$".

Even though the contents of Eq. 7 is not affected, we have elected to follow the referee's suggestion and modified the equation as recommended. ***P5L16***

P5Eq8

It seems that statements like "$a^+ = 0$ elsewhere" and "$a^- = 0$ elsewhere" are missing in the upper and lower equation, respectively.

The referee is correct, and Eq. 8 has been rewritten accordingly. ***P5L25-P6L1***

P7Eq17

I am somewhat irritated by this equation. For example, when I substitute i_k^n (bottom left) into the upper equation, the first term in the brackets becomes $1 + k \cdot (n + 1)$, which doesnt seem to make much sense. Isnt i_k^n supposed to stay the same when the sums are evaluated, that is, should the indices be different?

The referee's irritation regarding Eq. 17 is highly justified. We have taken two actions related to this issue. First, in our original manuscript I as defined by Eq. 16 is a grid cell quantity. Since all other quantities for the grid cell level are in lower case, this was unfortunate. Hence, we have replaced I by λ in the present revision. Second, the reviewer rightly rejects the use of *e.g.* i_k^n in Eq. 17, the correct here is i^n. The equation has been corrected accordingly. ***P8L25-26***

P8L3-9

It might be OK not to repeat the algorithm for the FSS displacement, but at least a qualitative description of how that quantity is derived from the FSS should be provided.

We have rewritten Sect. 2.3 to provide more general information on the FSS metric in the first paragraphs, and we also provide an approximate expression for the relation between FSS values and FSS displacement lengths in the final paragraph. ***P8L7-13, P10L15-19***

P9L16-17

"the resulting displacement metrics are also reduced substantially from the Reference case to the Modified case, due to the added ice areas proximity to land."; Is this sentence really saying what its supposed to say? After all, they are still increasing, only much less.

The referee is correct, and the sentence in question have been rewritten accordingly. ***P11L31-32***

P10L4-7

It seems worth mentioning that the Hausdorff-type metrics do not require remapping, although it seems OK to do it in this study to ensure consistency. This could also be mentioned in the discussion part

The referee is correct that Hausdorff-type metrics do not require remapping. The main contrasts between our approach and that of some other investigations is that we treat the ice edge as being composed of grid cells, rather than one-dimensional curves. We have added a paragraph on this topic (the second paragraph in Sect. 2). Moreover, while it is possible to define displacement metrics also for sets of grid cells given on different resolutions and projections, there are then complications related to representativeness that we find to be somewhat beyond the scope of the present study. ***P3L29-31***

P10L23-32

Here I was surprised that the relation between $\widehat{D^{IE}}$ and D^{IE} is not mentioned, and also not the relation between $\widehat{D^{IIEE}}$ and D^{IIEE}. Likewise, its worth to highlight already that D^{IIEE} and $\widehat{D^{IE}}$ are very similar. You elaborate on this only in the next section, and I think this is an interesting outcome that gives confidence about the robustness of these two metrics which are technically derived in quite different ways.

As suggested here by the referee we have added a section (second to last in Sect. 4) where results for various metrics from the two forecasts are mentioned. ***P13L19-24***

P11L22-26

What can be concluded from the comparison of the two different observational products? Can this help to understand the relatively large errors that are present already in the initial states? It would be good to comment on this.

The referee is correct about the impact of the contrasts between the assimilated microwave data and the ice chart data used for validation. We have added a sentence about this in the paragraph in question, and also in the following section. However, we refer to initial differences as 'deviations' rather than 'errors' since the two observational products in question have their separate strengths and weaknesses, so the 'truth' is not known. Finally, additional results from the comparison between the two observational products are now given in the Supplementary Information. ***P14L18-21,P14L29-P15L2, Sect. S2, Tables S1,S2, Fig. S3***

P12L8

"This was to be expected"; Actually, I would not have expected such a close match, given the considerably different approach to derive these two metrics.

We admit that the expected relationship between displacement metrics should have been explained more carefully. In the present revision we have included a discussion of idealized cases in the beginning of Sect. 6 which should shed light on this topic. ***P15L9-15***

P12L18

"50 such pairs" -> "50 out of 105 pairs" (correct?)

Yes, it's 50 out of 105 pairs. This is stated explicitly in the revised manuscript. ***P16L3-4***

P14L5-6

Regarding the maps, these would be examples of past performance rather some kinds of averages, which I wouldnt know how that should work, right? Or maps showing the errors for the latest previous forecasts (making use of the slow decorrelation)?

Our recommendation is due to the latter, *i.e.* the long decorrelation time scale. In order to explain this better, we have rearranged Sect. 6.3 and rewritten the sentence in question. ***P17L21-23, P18L12-15***

P14L14-18

I have difficulties to understand this paragraph on the usefulness of providing FSS in addition. I suggest to either explain a bit more, or to remove this paragraph.

The sentence concerning steepness of 0.5-crossing was not documented, and may thus be incorrect.

This sentence has been removed. ***P17L32-P18L2*** However, the application of FSS for examination of systems with different resolutions is at the core of this metric, and has been described thoroughly in papers that we cite, see *e.g.* Roberts and Lean. This is also stated in the Introduction section of the present manuscript. Based on suggestions from another referee the presentation of the FSS metric has been re-arranged in this revision.

P15L1-3

Is the Palerme et al. paper published now? Its not ideal to base an important final recommendation partly on a not-yet-published paper.

Palerme et al. is not yet published, but a revised manuscript based on a 'minor revision' recommendation has been submitted. However, we disagree that our recommendation is partly based on this study. Palerme et al. was mentioned here for context. Nevertheless, we have moved this sentence to the Introduction section, where it fits nicely in a paragraph where relevant literature is listed. ***P2L20-21, P18L27-29***

P15L3-4

"We have shown that the deterioration in the forecast quality is moderate for these lead times"; Again, I think there should be some discussion on why there is such a relatively large initial error (which is partly responsible for this slow initial error growth, I would say).

A discussion on the impact of initial errors, or rather deviations, is provided in Sect. 5 in the revised manuscript, see our reply to item **P11L22-26** above.

Figure2

Is A^- and A^+ the wrong way around here? Shouldnt A^+ be the part where the model/forecast has too much ice?

The referee asks if there is an error in the color shading in Fig. 2 in our original submission, and we have indeed made the mistake that the referee has spotted. We are very grateful that the referee pointed us to our mistake. In the revised manuscript the error has been corrected. We can add that we double-checked Fig. 5 (Fig. 3 in the original submission), and found that this did *not* contain the same mistake. ***Fig. 4***

Figure5

A statement on the units of the y-axis is missing.

The units of the y-axis is now given in the caption. ***Fig. 7***

EqS2-S4

It appears strange to me to use the areas (a^{ia}) as weights when averaging over the different segments the edge consists of. Wouldnt it make more sense to use the lengths l as weights? In case of S3, and neglecting A_0, this would yield simply $D... = \sum a / \sum l$. Also, for the same reason, the term A_0 seems a bit arbitrary: this one would converge to zero for increasing resolution, right? I am also suspecting that this awkward weighting is the reason why the hat-versions of D^{IIEE} are by such a large factor larger than those without hat.

The application of area weights was introduced in order to highlight effects of the geometry of IIEE areas, as stated in Sect. 2.2.2. With the referee's suggestion (*e.g.* $\sum a^{ia} / \sum l^{ia}$) the metrics would essentially give the same information as the original D^{IIEE} metrics: consider the fraction $\widehat{D^{IIEE}_{AVG}} / D^{IIEE}_{AVG}$ in the three idealized cases we present. For $\nu = 1/4$ the resulting fractions are 1.5, 1.7 and 1.35, respectively. Adopting the referee's suggestion we find the set of corresponding values to be 1.38, 1.36 and 1.36. For $\nu = 4$ the resulting fractions are 3, 2.5 and 2.3, while the referee's alternative yields fractions of 1.17, 1.13, 1.13. Moreover, the term A_0 is not arbitrary: in the case of two identical ice edges, dropping A_0 will lead to an ill-defined value for $\widehat{D^{IIEE}_{AVG}}$ since with no A_0 it becomes $0/0$. Note that A_0 is the area of all grid cells where the products overlap, the sentence in question has been rewritten to make this clear. ***S-P1L20***

Technical corrections

All of these items have been corrected according to the referee's advice.

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 mark-up version of 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. ***P1L8***

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 rephrased to make its content more clear. ***P1L15***

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). ***P20L29***

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. ***P6L19-24***

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. Note that these changes are not highlighted in the mark-up document since they resulted from latex citation code rather than an explicit typographical mistake. ***P2L17, P6L19***

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. ***P2L17-18***

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. ***P2L29-30***

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. ***P3L12-15***

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' (***P3L24***). 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'. ***P3L26-27***

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). ***P4L6***

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. ***P4L7-8,15-16***

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. ***P4L10-27***

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). ***P4L15***

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. ***P6L7-8***

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, P7L6-12***

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. ***P8L16-P10L14, Fig. 2***, red text in ***S-P3L13-P4L23***

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. ***P11L8-10, P12L8-10***

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. ***P11L12-13***

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. ***P11L22***

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". ***P11L25***

P9L23

the CMEMS acronym is already provided.

We have retained the acronym (CMEMS) only in the present revision. ***P12L12-13***

P10L6

Typo: overlaid.

Corrected. ***P12L28***

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. ***P14L1***

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. Blue text in ***S-P3L14-P425; 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 to item **P11L2** above. Dashed lines are used to indicate results that bridge days with no data (ice charts are not produced on Saturdays and Sundays;

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

see P9L27 in the original submission, P12L16 in the mark-up revision). An explanation has been added in the figure caption. ***Fig. 7 caption on P32***

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. ***P14L27***

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. (The latexdiff software has split a section in two.) Note that the comparison is now restricted to the period from January to mid-May, which reduces the useful scale. We have also include a figure that displays the FSS score and the asymptote values as defined by Roberts and Lean (2008). ***P14L29-P15L5; 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". ***P16L2***

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. ***P16L7-8***

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. ***P18L6; P20L17-19***

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***

Validation metrics for ice edge position forecasts

Arne Melsom[1], Cyril Palerme[1], and Malte Müller[1]

[1]Norwegian Meteorological Institute

Correspondence: Arne Melsom (arne.melsom@met.no)

Abstract.

The ice edge is a simple quantity in the form of a line that can be derived from a spatially varying sea ice concentration field. Due to its long history and relevance for operations in the Arctic, the position of the ice edge should be an essential element in any system that is designed to monitor or provide forecasts for the physical state of the Arctic Ocean and adjacent ocean regions.

~~Like for all components~~ Users of monitoring and forecast products ~~, users need to complement information about the ice edge position with~~ for sea ice must be provided with complementary information of the expected accuracy of the data or model results. Such information is traditionally available as a set of metrics that provide ~~a concentrated~~ an assessment of the information quality. In this study we provide a survey of metrics that are presently included in the product quality assessment of the CMEMS Arctic Marine Forecasting Center sea-ice edge position forecast. We show that when ice edge results from different products are compared, mismatching results for polynya and local freezing at the coasts of continents and archipelagos have a large impact on the quality assessment. Such situations, which occur regularly in the products we examine, have not ~~previously~~ properly been acknowledged when a set of metrics for the quality of ice edge position results have been constructed.

We examine the quality of ice edge forecasts using a total of ~~17~~ 15 metrics for the ice edge position. These metrics are analyzed in synthetic examples, as well as in selected cases of actual forecasts, and finally for a full year of weekly forecast bulletins. Using necessity and simplicity of information as a guideline, we recommend using a set of four metrics that sheds light on the various aspects of product quality that we consider.

Moreover, any user is expected to be interested in a limited part of the geographical domain, so metrics derived as domain-wide integrated quantities may be of limited value. Consequently, we recommend that metrics are also made available for an appropriate set of subdomains. Furthermore, we find that the metrics' decorrelation time scales are much longer than the present forecast range. Hence our final recommendation is to include depictions of gridded mismatching of ice edge positions using maps for the integrated ice edge error.

1 Introduction

The ice edge location is a primary source of information for safe navigation in ice infested waters. The retreating sea ice in the Arctic Ocean has given rise to increased naval traffic in the region. The navigation distance from Northern Europe to the Far East is about 40% shorter using the northern sea route when compared to the ~~distance~~ length of the southern route *via* the Suez Canal. Hence, commercial shipping is becoming viable from an economic perspective due to the changing physical conditions (Ho, 2010; Schøyen and Bråthen, 2011). Our motivation is to provide the increasing number of operators in the Arctic region with easily comprehensible and robust information about the quality of relevant forecasts.

Basic computations of ice edge displacement in operational sea ice forecasts relative to observational products have *e.g.* been performed by Posey et al. (2015) and Melsom et al. (2011). Results for the ice edge position from seasonal ensemble forecasts have been examined by Zampieri et al. (2018) and Palerme et al. (2019). Dukhovskoy et al. (2015) examined five metrics for ice edge displacement, and based on sensitivity ~~test~~ tests for scale, rotation, translation, and noise, their recommendation is to apply the Modified Hausdorff Distance~~(defined by their Eq. (11))~~.

Model results for sea ice concentration are frequently examined by presenting differences from corresponding observations, or results from other models, as shaded contours on maps, see *e.g.* Johnson et al. (2007), Arzel et al. (2006). In these and other studies, results for sea ice are often quantified by simple statistics for integrated quantities, notably sea ice extent (Massonnet et al., 2012). Statistics for sea ice extent is one of ~~several~~ the quantities that can be derived from contingency tables for sea ice concentration categories (Carrieres et al., 2017). A sophisticated approach to examinations of ~~integral quantities~~ results for sea ice extent has been proposed by Goessling et al. (2016) who introduced the integrated ice edge error (IIEE) as an objective score for differences in the position of the ice edge. An extension relevant for ensemble predictions was recently published (Goessling and Jung, 2018). Using this extension, Palerme et al. (2019) find that ECMWF SEAS5 seasonal forecasts Johnson et al. (2019) that are initialized between April and September are more skillful than climatology for forecast ranges of 6-12 weeks.

The fractions skill score (FSS) metric was developed for small scale features in forecast ~~system~~systems, originally applied to convective precipitation in weather forecasting (Roberts and Lean, 2008). ~~The~~ One purpose of the FSS is to provide an objective analysis of how the forecast skill changes as a function of horizontal scales, which is potentially relevant for skill assessments of the ice edge position. The FSS was designed for features whose spatiotemporal evolution can't be forecasted exactly but rather in a statistical sense.

The present examination of validation metrics for the ice edge position has been performed with an aim of improving the information of product quality for users of products available from the Copernicus Marine Environment Monitoring Service (CMEMS)~~is a pan-European project~~. CMEMS is the marine component of the European Union's Earth Observation Programme. CMEMS has been set up to meet today's climate and marine challenges by providing the public with observational multiyear and near real-time products, and reanalyses and forecasts from ~~circulation models~~ocean circulation models, sea ice models, wave models and biogeochemical models. The information is integrated into an open and free catalog of products that is available from http://marine.copernicus.eu/.

CMEMS is presently organized as ~~11~~ 15 production centers, ~~four~~ eight of which process observational data from satellite and *in situ* platforms, and the remaining seven centers run and process results from ~~ocean circulation~~ numerical models. These groups of centers are referred to as *thematic data assembly centers* (TACs) and *monitoring and forecast centers* (MFCs), respectively.

One of the TACs is dedicated to observations of sea ice, mainly based on data from satellite-born instruments. Furthermore, three of the MFCs' model systems have their ocean circulation model coupled to sea ice models. These are the centers responsible for forecasts and reanalyses in the Baltic Sea (BAL MFC), the Arctic Ocean (ARC MFC), and the global oceans (GLO MFC). Sea ice can also occur in the Black Sea, but the relevant forecast center (BS MFC) presently has no sea ice product.

Information about the product quality is available for all CMEMS model products, provided ~~by~~ as statistics for a variety of metrics which are calculated by comparing results with observational products. Relevant data for sea ice concentration and the position of the ice edge is available from satellite-born instruments. ~~As we demonstrate in this study , the assessment of quality of the~~ In this study we assess the quality of forecasted ice edge ~~position is highly sensitive to the definition of metrics, and to some degree uncertainty~~ positions using a large number of metrics. The sensitivity of the metrics due to differences in observational products ~~. The amount of available data is not a limiting factor in this context~~is also considered.

The present examination is organized as follows. In Sect. 2 we introduce the metrics used in our analysis: ice edge displacement metrics in 2.1, IIEE and derived metrics in 2.2, and FSS metrics in 2.3. Next, an idealized situation which is constructed to shed light on situations which ~~leads~~ lead to large differences between model results and observations is explored in Sect. 3. This issue is investigated in the context of sea ice forecasts from CMEMS ARC MFC in Sect. 4, where results for two forecast bulletins with different error characteristics are presented. Then, results for a full year of sea ice forecasts ~~is~~ are given in Sect. 5. These results are discussed in Sect. 6, and our examination concludes with a recommended best practice for validation of sea ice edge forecasts in 6.3.

2 Definition of metrics

We consider metrics for offsets in ice edge position ~~in~~ between two gridded products, *e.g.* with one product derived from observations and with the other from simulation results from a numerical coupled sea ice-ocean circulation model. In this section, the two products are referred to as O and M, respectively. Below we ~~will associate grid properties~~ associate grid cell quantities by lower-case indices, and integral properties by upper-case indices. Analogously, we separate ~~grid scale metrics~~ Euclidean grid cell distance values and integral ~~metrics~~ distance metrics' values by denoting these as d and D, respectively.

Note that in our approach, ice edges are associated with areas due to their composition of sets of grid cells, rather than curves. The definitions that lead to edge displacement metrics below do not directly apply to one-dimensional curves. Several displacement metrics between pairs of curves are given by Dukhovskoy et al. (2015).

2.1 Ice edge displacement metrics

~~The first step in the algorithm to compute displacement metrics~~ In order to compute ice edge displacement metrics the first step is to find the grid cells which ~~constitutes~~ constitute the ice edge in the gridded observations as well as in the model product. Let c be the sea ice concentration, and let c_e be the sea ice concentration value that defines the ice edge (usually set to 0.15). Then, we take the ice edge to be constituted by the grid cells $[i, j]$ that meet the condition

$$c[i,j] \geq c_e \quad \wedge \quad \min\left(c[i-1,j], c[i+1,j], c[i,j-1], c[i,j+1]\right) < c_e \tag{1}$$

where $\wedge$ is the logical AND operator. Let E be the ice edge. Ice edges E_O and E_M then correspond to the set of grid cells e_o and e_m that are returned by this algorithm step when applied to products O and M, respectively. We also introduce the ~~metric~~ coordinate position of grid cell $[i, j]$ as $[x, y]$, and let N_O be the number of edge grid cells in product O, and N_M be the number of cells in product M.

Next, for each edge grid cell in each product, we find the distance to the nearest edge grid ~~in the alternative product. Hence, if the set of edge grids in the alternative product is E_{alt}, and each of the cells in E_{alt} is denoted e_{alt}~~cell in the other product. Consider first the distance from an ice edge grid cell $[i_m^1, j_m^1]$ in the model product at the coordinate position $[x_m^1, y_m^1]$. Then, the displacement ~~from an~~ of the observed ice edge from this grid cell becomes

$$d_m^1 = \min\left(\forall e_o \in E_O : \left[(x_o - x_m^1)^2 + (y_o - y_m^1)^2\right]^{1/2}\right) \tag{2}$$

where $\forall$ is the FOR ALL operator and $[x_o, y_o]$ is the coordinate position of an ice edge grid cell ~~$[i, j]$ in the reference productat the metric position $[x, y]$ can be written~~

~~$$d = \min\left(\forall e_{alt} \in E_{alt} : \left[(x_{alt} - x)^2 + (y_{alt} - y)^2\right]^{1/2}\right)$$~~

in the observed product.

A variant is to consider any land/ocean boundary ~~node~~ grid cell as included in the ~~alternative product~~observed sea ice edge. When adopting this variation we refer to the ~~alternative product as $\hat{E}_{alt}$~~observational product as $\hat{E}_O$, constituted by grid cells ~~$\hat{e}_{alt}$~~$\hat{e}_o$. We note that ~~$E_{alt} \subset \hat{E}_{alt}$~~$E_O \in \hat{E}_O$. The corresponding displacement ~~from an~~ becomes

$$\hat{d}_m^1 = \min\left(\forall \hat{e}_o \in \hat{E}_O : \left[(\hat{x}_o - x_m^1)^2 + (\hat{y}_o - y_m^1)^2\right]^{1/2}\right) \tag{3}$$

We compute the displacement d_o^1 of a model ice edge from an ice edge grid cell in the ~~reference product at the metric position $[x, y]$ is then~~

~~$$\hat{d} = \min\left(\forall \hat{e}_{alt} \in \hat{E}_{alt} : \left[(\hat{x}_{alt} - x)^2 + (\hat{y}_{alt} - y)^2\right]^{1/2}\right)$$~~

observational product analogously. This is also done for $\hat{d}_o^1$ after E_m has been expanded to $\hat{E}_m$ by including all land/ocean boundary grid cells.

~~Now,~~ We can now define a set of symmetric ice edge position metrics ~~can be~~ expressed as functions of the edge ~~displacement in Eq. 2~~displacements. Here, a symmetric metric is a parameter whose value is independent of whether it is the observations or the model product that is the base of the analysis. We introduce four such metrics here, based on results for d_m and d_o.

1. The root-mean-squared ice edge displacement:

$$D^{IE}_{RMS} = \frac{1}{2}\left[\left(\frac{1}{N_O}\sum_{n=1}^{N_O}(d^n_o)^2\right)^{1/2} + \left(\frac{1}{N_M}\sum_{n=1}^{N_M}(d^n_m)^2\right)^{1/2}\right] \tag{4}$$

2. The average ice edge displacement:

$$D^{IE}_{AVG} = \frac{1}{2}\left[\frac{1}{N_O}\sum_{n=1}^{N_O} d^n_o + \frac{1}{N_M}\sum_{n=1}^{N_M} d^n_m\right] \tag{5}$$

3. The ice edge displacement bias, here defined as positive when the ice edge in the model product is on the open ocean side of the ice edge in the observational product:

$$\Delta^{IE} = \frac{1}{2}\left[\frac{1}{N_O}\sum_{n=1}^{N_O}\frac{c_m[i^n_o, j^n_o] - c_e}{\|c_m[i^n_o, j^n_o] - c_e\|}d^n_o + \frac{1}{N_M}\sum_{n=1}^{N_M}\frac{c_e - c_o[i^n_m, j^n_m]}{\|c_e - c_o[i^n_m, j^n_m]\|}d^n_m\right] \tag{6}$$

where $\|x\|$ is the absolute value of x, and c_o, c_m are the sea ice concentrations in the observations and model, respectively. Also, $[i_o, j_o]$ and $[i_m, j_m]$ denotes ice edge ~~grids~~ grid cells in the observations and model, respectively. One may construct situations where a denominator in Eq. 6 becomes 0. In reality, such cases will be very rare, and most of the time this will occur when edge grid cells in the two products overlap, *i.e.*, $d^n = 0$. In these cases, we set the fraction to 0.

4. The extreme ice edge displacement, also known as the Hausdorff distance:

$$D^{IE}_{H} = \max\left(\max(d_o), \max(d_m)\right) \tag{7}$$

where d_o, d_m are the full sets of gridded displacements as given by Eq. 3.

Finally, ~~displacements in the alternative product variant~~ substituting displacements d in Eq.s 4–7 by $\hat{d}$ as given by Eq. 3 ~~give~~ gives rise to a set of ~~displacement metrics analogous to those given above by Eq.s 4–7. We denote these metrics as hatted variables and~~ supplementary metrics $\widehat{D^{IE}_{RMS}}$, $\widehat{D^{IE}_{AVG}}$, $\widehat{\Delta^{IE}}$, and $\widehat{D^{IE}_{H}}$. We note that *e.g.* $\widehat{D}^{IE}_{RMS} \leq D^{IE}_{RMS}$.

2.2 IIEE metrics

Recently, the integrated ice edge error (IIEE) has been suggested as an alternative approach to quantifying the offsets between two ice edges (Goessling et al., 2016). The IIEE is computed from the area between the ice edges in the two products. For a gridded product with a grid cell size a, set

$a^+ = a$ for grid cells where $c_m > c_e \ \wedge \ c_o < c_e$

$a^- = a$ for grid cells where $c_o > c_e \ \wedge \ c_m < c_e$

$$a^+ = \begin{cases} a & \text{for grid cells where } c_m > c_e \ \wedge \ c_o < c_e \\ 0 & \text{elsewhere} \end{cases}$$
$$a^- = \begin{cases} a & \text{for grid cells where } c_o > c_e \ \wedge \ c_m < c_e \\ 0 & \text{elsewhere} \end{cases} \tag{8}$$

Then, the area where the ice edge position in the model product is on the open ocean side of the observed ice edge is

$$A^+ = \sum_A a^+ \tag{9}$$

whereas the complementary situation with the observed ice edge on the open ocean side of the model edge covers the area

$$A^- = \sum_A a^- \tag{10}$$

(an illustrated example is provided in Sect. 3). The ice edge is here the perimeter of the sea ice extent area. Thus, A^+ is the area where the ice extent in the model results overshoots the ice extent in the observations, and *vice versa* for A^-.

Two area metrics can then be constructed, as given by Goessling et al. (2016).

1. The integral score:

$$A^{IIEE} = A^+ + A^- \tag{11}$$

2. The bias score:

$$\alpha^{IIEE} = A^+ - A^- \tag{12}$$

Note that Goessling et al. (2016) also ~~introduces~~ introduced additional area metrics which are not considered here.

The IIEE metrics defined in Goessling et al. (2016) are all provided for areas of sea ice, while no displacement metrics are introduced. Here, IIEE-based displacement metrics are derived by dividing the IIEE areas by an IIEE characteristic length scale. Below, we introduce two definitions of such a length scale.

Summary statistics in the form of a contingency table ~~provides~~ provide versatile information for validation of sea ice concentration results (Carrieres et al., 2017). ~~Here, we note that the~~ After categories have been defined by a set of ranges in sea ice concentration, table cells will give areas with category match-ups. Here it is essential to have the sea ice concentration value that defines the ice edge as a value that separates two categories. The sea ice extent for each product is then found as the sum of the relevant rows and columns, respectively. The differences in sea ice extent (quantities A^+ and A^- ~~may easily be derived from such tables~~) emerge from adding the areas in cells that corresponds to categories on different sides of the ice edge in the two products.

2.2.1 Edge length based IIEE displacement metrics

In order to provide scores that have the same dimension as those produced by the ice edge displacement metrics in Sect. 2.1, we here introduce metrics that arise when dividing the area metrics ~~above~~ given by Eq.s 11, 12 with the ice edge length. Presently, the ice edge is given as a set of grid cells that were identified from Eq. 1. For simplicity we consider the case where the resolution in both horizontal directions is constant and equal, and write the grid cell size as s.

~~Next~~Consider the schematic example provided in Fig. 1. When calculating the length of the ice edge, we must account for the presence of diagonal edge ~~nodes~~grid cells. This is performed by looping all edge ~~nodes~~ grid cells e and counting the number of $[i, j]$ edge ~~node~~ grid cell neighbours (*i.e.*, among [i-1,j], [i+1,j], [i,j-1], [i,j+1]) in the same product. If there are two or more neighbours, the edge grid cell contributes with a length $l^e = s$ (edge grid cells e_c, e_d in Fig. 1). If there are no such neighbours, the edge length is set to the length of the diagonal, *i.e.*, $l^e = \sqrt{2}s$ (edge grid cell e_a). If there is exactly one such edge neighbour, the contribution becomes $l^e = 0.5 \cdot (s + \sqrt{2}s)$ (edge grid cells e_b, e_e). Note that by this definition "open ended" edge ~~nodes~~ grid cells (*e.g.* adjacent to land; e_a, e_e) will contribute with a diagonal representation towards the open end.

The ice edge length in the observational product becomes

$$L_O = \sum_{e \text{ in } E_O} l_o^e \tag{13}$$

and the corresponding length in the model product is given analogously.

Two length metrics can now be derived from the corresponding area metrics.

1. The IIEE average displacement:

$$D_{AVG}^{IIEE} = \frac{2}{L_O + L_M} A^{IIEE} \tag{14}$$

2. The IIEE bias:

$$\Delta^{IIEE} = \frac{2}{L_O + L_M} \alpha^{IIEE} \tag{15}$$

Note that if there are no overlapping ice edge ~~nodes~~ grid cells in the two products and if no IIEE area is bounded by dry ~~nodes~~ grid cells or an open boundary, the length scale used for derivation of the displacement metrics given by Eq.s 14 and 15 is half the circumference of the IIEE areas.

2.2.2 Separation based IIEE displacement metrics

An alternative to the application of the scaling length $(L_O + L_M)/2$ in Sect. 2.2.1 is introduced in Sect. S1 ~~.1~~ in the Supplementary Information document. The alternative expression for the scaling length is solely dependent on the geometry of the IIEE areas. We then derive a supplementary set of displacement metrics that is analogous to the ~~D^{IIEE}~~ D^{IE} metrics defined by Eq.s 4-7.

The definitions of metrics in Sect. S1 ~~.I take dry nodes~~ take dry grid cells adjacent to IIEE areas into account, which the scaling length definition in Sect. 2.2.1 does not. Hence, we adopt here the hatted notation as introduced in Sect. 2.1. The resulting displacement metrics defined in Sect. S1 ~~.I~~ are thus denoted as $\widehat{D^{IIEE}_{RMS}}$, $\widehat{D^{IIEE}_{AVG}}$, $\widehat{D^{IIEE}_{MAX}}$, and $\widehat{\Delta^{IIEE}}$.

2.3 Fractions skill score

We next consider the fractions skill score (FSS), as introduced by Roberts and Lean (2008). This metric was defined with the purpose of providing information on the impact of ~~errors~~ differences on small scales that can appear in results from high resolution observations and models. ~~In the present case of the position of the sea ice edge , the~~ The FSS is computed for binary results, such as gridded hits and misses due to a criterion, from a pair of products (usually observations and model results). Values for FSS provide information of how the two products compare as a function of resolution. Representation on different resolutions are computed by integration onto coarser (larger) grid cells, and the binary results on the original grid become hit fractions on coarser grids. The FSS reaches its maximum value of 1 at resolution(s) where the representation of the two products are identical, and has a minimum value of 0 when no grid cells have overlapping non-zero values.

In the present context, we define hits as grid cells which are part of the ice edge as defined by Eq. 1, in both products. The probability of a ~~grid-by-grid~~ grid cell-by-grid cell match up of the edge positions is expected to be reduced when the resolution is enhanced.

The presentation of FSS in this section is largely based on the Roberts and Lean (2008) article, adapted to representation of lines of grid cells rather than areas. We provide a relevant schematic example as Fig. 2, and we use this to illustrate some of the quantities that are introduced below.

Recall from Sect. 2.1 that we identified the sets of N_O and N_M grid cells e_o and e_m that constitute the ice edges E_O and E_M in products O and M, respectively. We construct a binary gridded representation of the ice edge in product O as

$$I\lambda_o[i,j] = \begin{cases} 1 & \forall e_o \in E_O \\ 0 & \text{elsewhere} \end{cases} \tag{16}$$

so that ~~$\sum I_o = N_O$~~ $\sum \lambda_o = N_O$. The corresponding binary representation of the edge in product M, ~~I_m~~ λ_m, is defined analogously. Next, for product O we introduce the coarse grid cell ice edge fraction for a neighbourhood with an extent of n grid cells as

$$I_o^n[i^n, j^n] = \frac{1}{n^2} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} I_o\left[i_k^n + k - \frac{n-1}{2}, j_l^n + l - \frac{n-1}{2}\right],$$
$$i_k^n = \frac{n+1}{2} + k \cdot n \quad , \quad j_l^n = \frac{n+1}{2} + l \cdot n$$

$$\lambda_o^n[i^n, j^n] = \frac{1}{n^2} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \lambda_o\left[i^n + k - \frac{n-1}{2}, \; j^n + l - \frac{n-1}{2}\right] \tag{17}$$

where n is an odd number. Again, we define ~~I_m^n~~ λ_m^n analogously, and we note that ~~$I_o = I_o^1$~~ $\lambda_o = \lambda_o^1$. In the example in Fig. 2, a neighbourhood extent of 3 grid cells is indicated by the thick grid lines and for this case, we find

$$\lambda_O^{n=3} = \frac{1}{9}\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \quad ; \quad \lambda_M^{n=3} = \frac{1}{9}\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix} \tag{18}$$

The mean square edge fraction error for a neighbourhood extent of n ~~grids becomes~~ grid cells becomes

$$\mathrm{MSE}^n = \frac{1}{N_x^n N_y^n} \sum_{i^n=1}^{N_x^n} \sum_{j^n=1}^{N_y^n} \left[I\lambda_m^n[i^n, j^n] - I\lambda_o^n[i^n, j^n] \right]^2 \tag{19}$$

where N_x^n, N_y^n are the number of the neighbourhood extent n ~~grids~~ grid cells in the x and y directions, respectively. Following Roberts and Lean (2008) we introduce a reference MSE value as the largest possible with the present extent of the edge ~~nodes~~ grid cells

$$\mathrm{MSE}_{\mathrm{ref}}^n = \frac{1}{N_x^n N_y^n} \min\left\{ \left[\sum_{i^n=1}^{N_x^n} \sum_{j^n=1}^{N_y^n} \lambda_o^n[i^n, j^n]^2 + \sum_{i^n=1}^{N_x^n} \sum_{j^n=1}^{N_y^n} \lambda_m^n[i^n, j^n]^2 \right], \right.$$
$$\left. \left[\sum_{i^n=1}^{N_x^n} \sum_{j^n=1}^{N_y^n} \left(1 - \lambda_o^n[i^n, j^n]\right)^2 + \sum_{i^n=1}^{N_x^n} \sum_{j^n=1}^{N_y^n} \left(1 - \lambda_m^n[i^n, j^n]\right)^2 \right] \right\} \tag{20}$$

This expression is a worst case arrangement of hits and misses that takes into account ~~e.g.~~ situations where hits outnumber misses. This is a modification of the corresponding definition in Roberts and Lean (2008) whose Eq. 7 allowed for situations with $\mathrm{MSE}_{\mathrm{ref}}^n$ exceeding 1.

For the skill score with the original 6×6 grid in Fig. 2 we have $\mathrm{MSE}^{n=1} = 6/6^2$ and $\mathrm{MSE}_{\mathrm{ref}}^{n=1} = 12/6^2$, while for the $n = 3$ neighbourhood displayed by the thick grid lines we have $\mathrm{MSE}^{n=3} = 2/(2 \cdot 9)^2$ and $\mathrm{MSE}_{\mathrm{ref}}^{n=3} = 9/(2 \cdot 9)^2$.

Now, the resolution-dependent fractions skill score is introduced as

$$\mathrm{FSS}^n = 1 - \frac{\mathrm{MSE}^n}{\mathrm{MSE}_{\mathrm{ref}}^n} \tag{21}$$

which has a value of 1 for a perfect forecast for neighbourhood extent n (~~$I_m^n = I_o^n \ \forall \ i^n, j^n \ \Rightarrow \ \mathrm{MSE}^n = 0$~~ $\lambda_m^n = \lambda_o^n \ \forall \ i^n, j^n \ \Rightarrow \ \mathrm{MSE}^n$ and a value of 0 when ~~$I_m^n \cdot I_o^n = 0 \ \forall \ i^n, j^n$~~ $\lambda_m^n \cdot \lambda_o^n = 0 \ \forall \ i^n, j^n$ ($\Rightarrow \ \mathrm{MSE}^n = \mathrm{MSE}_{\mathrm{ref}}^n$). Note that invoking the modified definition of $\mathrm{MSE}_{\mathrm{ref}}^n$ in Eq. 20 makes the FSS^n metric symmetric in the sense that reversing the definition of hits and misses does not affect the FSS^n score.

~~Illustrative examples for interpretations of FSS metrics for convective precipitation are provided in Roberts and Lean (2008) .Here, we consider gridded representations of lines rather than areas. Consequently, we have chosen to provide a relevant schematic example to supplement the cases introduced in Roberts and Lean (2008). Our example is given in Section S1.2 in~~ For the sample case in Fig. 2 we then find that $\mathrm{FSS}^{n=1} = 1/2$ and for the ~~Supplementary Information document~~ $n = 3$ neighbourhood displayed by the thick grid lines we have $\mathrm{FSS}^{n=3} = 7/9 \approx 0.78$.

Moreover, we note from Eq.s 19-21 that the FSS score will not change if we introduce a set of additional grid cells where neither product has an ice edge, provided that non-events dominate events (*i.e.*, the first term in Eq. 20 is used, here: that the number of nodes without an ice edge is larger than the number of edge nodes). This observation has consequences for two different aspects in the present study.

First, when modeling the ocean, dry nodes are usually not considered to be part of the computational domain, and are assigned a special value in numerical results. When integrating over a neighbourhood $n > 1$ one option would be to discard the grid cells that are dry in the original representation. We will then be left with a result which has a non-constant neighbourhood size where n^2 if dry nodes are not present, and $< n^2$ for neighbourhoods where dry nodes are present. Here, we choose to avoid the problem of non-constant neighbourhood sizes by adopting $\lambda_o = \lambda_m = 0$ for dry grid cells.

Second, the grid for $n = 3$ indicated by thick lines in Fig. 2 is only one of nine possible configurations. Since the FSS results are not affected by additional grid cells where neither product has an ice edge, we can expand the original domain by adding a padding region of $n - 1$ grid cells. In the case of $n = 3$ all configurations are attained by shifting the neighbourhood by 0, 1 and 2 original grid cells in both directions. The average FSS score from all of the configurations will be used henceforth in this article, since the alternative is a set of results that will depended on an arbitrary configuration subset choice.

As an expansion of the FSS metrics, Skok and Roberts (2018) introduced the FSS displacement~~. The~~, which we will refer to as D^{FSS}. An initial estimate for D^{FSS} is derived by first determining for which neighbourhood size the FSS exceeds 0.5. The full algorithm for computing this displacement metric is given at the end of ~~their article~~Skok and Roberts (2018), and is not repeated here. ~~We will refer to this displacement quantity as~~ In most cases D^{FSS} will become about half of the minimum metric neighbourhood size at which the FSS exceeds 0.5. The reliability of D^{FSS} decreases when the frequencies are biased (Skok and Roberts, 2018). Here, this translates to differences in the number of ice edge ~~grids~~ grid cells in observations and in the forecast. In the present study we implement a reduction of the product with the longest ice edge by randomly removing ice edge ~~grids~~ grid cells from this product. Thus, an unbiased version of the two ~~grids~~ grid cells is used when computing D^{FSS}. The random removal of ~~nodes~~ grid cells is repeated a number of times, and the average value of the resulting displacements is taken to represent the D^{FSS}.

3 Ice edge metrics in two synthetic cases

In order to illustrate the various sea ice metrics and to examine how the results for these metrics compare, we have constructed a set of synthetic distributions of sea ice concentrations. The distributions will serve as representing observations and model results, respectively. The sea ice concentration distributions are introduced on a 200×200 grid, and they are displayed in Fig. 3.

We take the sea ice concentration field in Fig. 3a to represent a reference observation. One aspect of interest here is the effect on the validation scores when ice is introduced or removed locally in one product, but not in the other. In order to accentuate such conditions, we supplement the reference observation with modified observation as displayed in panel b. A corresponding model result is given as shown in Fig. 3c.

We denote the comparison of the reference observation and model results as the *Reference case*, while the comparison of the modified observation and model results is referred to as the *Modified case*.

~~A digression which is relevant here is that we have not included the Modified Hausdorff Distance, which was recommended by Dukhovskoy et al. (2015), in our analysis. In our formulation, this quantity is the maximum of the two terms in the bracket in Eq. 5, and will generally exhibit similar results to D^{IE}_{AVG} but with larger magnitudes. While the sensitivity study in Dukhovskoy et al. (2015) is rich in detail, changes like contrasts between the *Reference case* and the *Modified case* are not considered. In their study of results from seasonal forecasts, Palerme et al. (2019) conclude that results for the Modified Hausdorff Distance are sensitive to differences with similar qualitative aspects as those discussed in this section. 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.~~

The ice edges (0.15 concentration isolines) as given by Eq. 1 are displayed as colored lines in Fig. 3. Edges from synthetic observations have been added in Fig. 3c~~, and the~~ . The main purpose of this ~~document~~ article is to present metrics for the separation in ~~this set~~ such sets of lines.

Now consider the areas between the ice edges, from which the IIEE metrics are computed. The regions corresponding to the definitions in Eq.s 9 and 10 are shown in pink and red in Fig. 4.

The results for the various displacement metrics that were defined in ~~section~~Sect. 2 are given in Table 1. First, we note that in the *Reference case*, all D^{IE} and D^{IIEE} scores have similar values (with the expected exception of the maximum displacement score D^{IE}_{H} which has a larger value than the other D^{IE} scores by design). Also, Δ^{IE} and Δ^{IIEE} are of similar magnitudes in the *Reference case*.

For the modified case, we assume that the bottom boundary is adjacent to land. This is relevant for the hatted ice edge displacement metrics. 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 in an operational sea ice forecasting service. An example will be presented in Sect. 4. We find that the value of the D^{IE} ice edge displacement metrics given by Eq.s 4, 5 and 7 increase from the *Reference case* to the *Modified case* by a factor of about 2-5 even though a fairly modest area with additional sea ice has been introduced in the latter case. Since ~~an~~ the additional discrepancy between the observations and model results has been introduced at a large distance, this change is according to our expectations.

Even though an additional discrepancy has been ~~introduces~~ introduced in the *Modified case*, its shape and size is such that with the exception of bias metrics all IIEE displacement metrics increase by a very modest degree in these synthetic examples. In conclusion, we find that the deterioration according to scores for the *Modified case* is much larger for the D^{IE} ice edge displacement metrics than for the IIEE metrics since the latter do not explicitly depend on the displacement between the pair of ice edges. Moreover, we note that if the ice edge displacement is defined by Eq. 3 the resulting $\widehat{D^{IE}}$ displacement ~~metrics are also reduced substantially~~ increase only by a marginal fraction from the *Reference case* to the *Modified case*, due to the added ice area's proximity to land.

Finally, we note from Table 2 that the fractions skill score is only moderately reduced when additional observed sea ice is introduced locally in the *Modified case*, and the FSS displacement also increases modestly (Table 1, D^{FSS}). The changes in

the IIEE area scores provide a quantification of the change in ice extent when ~~substituing~~ substituting the *Reference case* with the *Modified case*.

A digression which is relevant here is that we have not included the Modified Hausdorff Distance, which was recommended by Dukhovskoy et al. (2015), in our analysis. In our formulation, this quantity is the maximum of the two terms in the bracket in Eq. 5, and will generally exhibit similar results to D^{IE}_{AVG} but with larger magnitudes. While the sensitivity study in Dukhovskoy et al. (2015) is rich in detail, changes like contrasts between the *Reference case* and the *Modified case* are not considered. In their study of results from seasonal forecasts, Palerme et al. (2019) conclude that results for the Modified Hausdorff Distance are sensitive to differences with similar qualitative aspects as those discussed in this section. In Sect.s 4 and 5 below we will examine if differences which are qualitatively similar to the *Modified case* have an effect on the quality assessment of the ice edge position in the forecasts from CMEMS ARC MFC.

4 Ice edge metrics for two forecasts

We compare model results with observations which both are products that are distributed by ~~the Copernicus Marine Environment Monitoring Service (CMEMS)~~CMEMS. The observational product is the Arctic Ocean Sea Ice Concentration Charts *Svalbard* which is a multi-sensor product that uses data from Synthetic Aperture Radar (SAR) instruments as its primary source of information (WMO, 2017). This product covers the northern Nordic Seas, the Barents Sea and adjacent ocean regions. It is available on working days as mean values on a 1 km stereographic grid and will be referred to as the ice chart data hereafter.

Model results are taken from the Arctic Ocean Physics Analysis And Forecast product. Assimilation of sea ice concentration is implemented by use of microwave data, while no SAR data are assimilated. ~~This~~ The model product will from here on be referred to as the ARC model product. In our investigation we will consider daily mean fields of sea ice concentration, which presently are distributed on a 12.5 km stereographic grid. We restrict this study to the forecasts from the Thursday bulletins, which are available with a forecast range of ten days. The microwave data that are assimilated are available as the Ocean and Sea Ice Satellite Application Facility northern hemisphere product (Breivik et al., 2001), which is available from the CMEMS catalogue. The ~~assmilation~~ assimilation was performed three days prior to the ~~Thurday~~ Thursday bulletins. The main topic of this investigation is to provide an independent assessment of the quality of results for the ice edge, and not to assess the impact of assimilation. Thus, we compare results with ice chart data rather than with the microwave data.

Prior to performing the analysis both products are regridded. The ice chart product is aggregated onto a 13 km grid, while the ARC model product is interpolated onto the same grid ~~. (The~~ (the axes of the two CMEMS products, both available on polar stereographic grids, are rotated differently~~.)~~). The land-sea masks of the two regridded products are ~~overlain~~ overlaid so that the geographical extent of the two regridded products is identical.

In order to explore how sea ice edge metrics from actual forecasts and observations are affected by changing conditions, we here examine two cases that illustrate contrasts of the type that was examined in Sect. 3. The two cases that are chosen are the day 5 ARC forecast products issued on 2017-03-30 and 2017-05-25. The quality of the forecasted ice edge positions will be assessed by comparing the model results with the ice edge position in the ice chart data on the respective forecast valid dates.

The positions of the ice edges on these two dates according to model and observations are shown by displaying the IIEE fields in Fig. 5a and b.

For the situation on 2017-05-29 (panel b) we notice that there are large discrepancies in the position of the ice edge in several locations: a polynya to the northwest of Greenland is open in the model, but not in the observations; there is a region along the coast in the Barents Sea where the model ice edge has retreated from the coast in the southern Kara Sea while the entire Kara Sea is frozen over in the ice chart; there remains some ice along the coast in the southeastern Barents Sea in the ice chart but not in the model. These objects are indicated by labels in Fig. 5. Note also that polynyas have opened around Franz Josef Land (FJL), but since these are seen in both products this region doesn't affect the displacement metrics to the same degree as the other discrepancies that are mentioned here.

In contrast, the situation on 2017-04-03 (panel a) has notable offsets along the sea ice edge, but polynyas and mismatching results in coastal regions play a much smaller role than on 2017-05-29.

Results for the various displacement metrics are given in Table 3. As was seen in the results for the synthetic cases in Sect. 3, the scores that deviate substantially between the two forecasts are for the D^{IE} ice edge displacement metrics and for Δ^{IE}. The inflated values for the 2017-05-29 forecast when compared to the results for the 2017-04-03 forecast can largely be attributed to the ice edges associated with the IIEE features that are labeled in Fig. 5b. Furthermore, we note that the values for $\widehat{D^{IIEE}_{AVG}}$ and $\widehat{D^{IIEE}_{RMS}}$ are larger than those for the corresponding $\widehat{D^{IE}}$ metrics by a factor of 1.5-2. This contrast, which is much larger than in the synthetic case (Table 1), can be attributed to the fact that the individual IIEE features in the synthetic cases were few and regular. In the forecasts there is a large number of IIEE features with irregular shapes.

Furthermore, we find that the $\widehat{D^{IE}}$ metrics change only very modestly from 2017-04-03 to 2017-05-29 due to the proximity to the coast for the features that are labeled in Fig. 5b, in contrast to the results for D^{IE}. We also note that the definitions for the displacement metrics that are derived from the IIEE lead to values for $\widehat{D^{IIEE}_{AVG}}$ that are about twice as large as the corresponding D^{IIEE}_{AVG} values. Finally, we observe that for each of the two forecasts that are examined here, the relative difference between D^{IIEE}_{AVG} and $\widehat{D^{IE}_{AVG}}$ is only about 10% or less. The relationships between the various displacement metrics are examined based on results from a full year of weekly forecast bulletins in the next section.

From the results for supplementary metrics in Table 4 we note that the FSS values are only slightly lower for the 2017-05-29 forecast than for the 2017-04-03 forecast, even though this forecast performs much poorer when diagnosed with the D^{IE} ice edge displacement metrics.

5 Ice edge position metrics for 2017

The comparison of model results and observations in Sect. 4 ~~have~~ has been performed for all weekly forecast bulletins from 2017. The results for mean displacement metrics and biases for the 5-day forecasts are displayed in Fig. 6. We note that there is a seasonal variation in all metrics with large deviations during the months that lead up to the sea ice minimum in mid-September. We will refer to the period from the start of July to mid-September as the pre-minimum. A substantial part of the pre-minimum discrepancies is explained by the biases, which reveal that the sea ice extent is larger in the ice chart product than

in the model product. The smaller extent in the model product gives rise to negative values in Fig. 6b. Annual average values for the various displacement metrics are given in ~~the bottom rows (~~rows *All 5-day forecasts* ~~)~~ of Tables 3 and 4.

Furthermore, we note that the curves in Fig. 6 can be separated into two groups:

1. D^{IE}_{AVG}, $\widehat{D^{IIEE}_{AVG}}$ and D^{FSS}
2. $\widehat{D^{IE}_{AVG}}$ and D^{IIEE}_{AVG}

Group 1 metrics generally have larger values than group 2 metrics. This is expected since *e.g.* $\widehat{D^{IE}_{AVG}} \leq D^{IE}_{AVG}$ by definition, notably the different impact on these two metrics when the displacements occur in the vicinity of land or islands. Moreover, we demonstrated in Sect. S1 ~~.1~~ that the definition of $\widehat{D^{IIEE}_{AVG}}$ in group 1 leads to values that are larger than the D^{IIEE}_{AVG} metric in group 2.

Interestingly, we find that there is a contrast in the results between the two metrics groups during the pre-minimum: the deterioration exhibited in the evolution of group 1 metrics are larger than the corresponding deterioration for group 2 metrics in absolute terms. When we inspect the results from the two cases presented in Sect. 4, Table 3 reveals that the group 2 metrics have the lowest values in both cases. However, the separation into two distinct groups of metrics does not apply. We note that these two cases (indicated by vertical lines in Fig. 6) precede the July to mid-September pre-minimum during which the separation between the groups is most striking.

We have supplemented this analysis by a comparison between the microwave product that is assimilated by the model, and the ice charts. The deviations between these two observational products reveal similar peaks during the pre-minimum, *e.g.* with values for D^{IE}_{AVG} and Δ^{IE} in ranges of about 60 - 120 km and -40 - -120 km, respectively (~~not shown). Moreover, the~~ see Sect. S2 in the Supplementary Information document for details). Hence, the pre-minimum peaks that are seen in Fig. 6 can at least to some degree be attributed to assimilation of an observational product that deviates from the ice charts during the pre-minimum season. The correlation coefficients for the time series of D^{IE}_{AVG} for the 5-day forecasts *vs.* ice charts (black line in Fig. 6a) and the time series of D^{IE}_{AVG} for microwave data *vs.* ice charts is 0.89. The corresponding correlation coefficient for Δ^{IE} is 0.92.

Next, we have examined how the quality of the ice edge forecasts changes as a function of lead time. In order to limit the impact of the strong seasonal signal that is evident from Fig. 6, we have restricted this part of the analysis to the period from January to mid-May. The deterioration of the forecast quality that can be inferred from Fig. 7 is very weak. We also note that results for the two metrics in group 2 (blue and red curves in Fig. 6a) nearly overlap at all lead times, and are also lower in magnitude than the group 1 metrics at all lead times, as expected.

~~Finally, from the results in Table 4 we note that the model has a tendency to have a lower sea ice extent than the ice chart, as more than 70% of the IIEE areal misrepresentation is due to such conditions. This tendency is a confirmation of the negative bias values reported in Table 3. The FSS scores~~ The FSS scores for the same period are depicted as a function of resolution in Fig. 8, for model forecasts issued with a five day lead time, as well as for the microwave data. These results reveal that useful forecasts with a five day lead time are obtained at a scale of about ~~90x90~~60x60 km, when the FSS reaches a value of 0.5 (which is a criterion recommended by Skok and Roberts (2016)). When comparing ~~with~~ the microwave data with ice charts, the FSS

is well above 0.5 for a neighbourhood extent ~~$n = 5$ (not shown)~~$n = 3$, corresponding to useful data at a scale of approximately ~~60x60~~40x40 km if ice chart data are taken as truth.

Finally, from the results in Table 4 we note that the model has a tendency to have a lower sea ice extent than the ice chart, as more than 70% of the IIEE areal misrepresentation is due to such conditions. This tendency is a confirmation of the negative bias values reported in Table 3.

6 Discussion

6.1 ~~Reducing the set of displacement metrics~~

Our investigation of the results for the ice edge in the 2017 forecast bulletins in Sect. 5 revealed that the metrics $\widehat{D^{IE}_{AVG}}$ and D^{IIEE}_{AVG} nearly overlap, and this is also the case for $\widehat{\Delta^{IE}}$ and Δ^{IIEE}. ~~This was to be expected, since including the coastal nodes in the search for closest nodes in the alternative product's ice edge (Sect. 2.1) is very similar to a coastal restriction of IIEE areas (Sect. 2.2).~~ These similarities can to some degree be understood from the following simplified cases: consider first a situation where one ice edge is shifted by a constant distance from the other, *i.e.*, they are parallel lines. Then, all of the average displacement metrics will be nearly identical, and this will also be the case for the displacement bias metrics. This is an idealized description for cases similar to the forecast for 2017-04-03 (Fig. 5a) where D^{IE}_{AVG} is only moderately larger than $\widehat{D^{IE}_{AVG}}$ (Table 3).

Next, consider a situation where a part of one ice edge is shifted from the other, and the remaining part is due to discrepancies with coastal ice cover in one product, but not in the other. When the length of boundaries between IIEE areas and adjacent dry grid cells is much shorter than the ice edge length, the impact of disregarding coastal segments in Eq. 13 is small. Then, nearly identical displacement metrics values will again be the result for *e.g.* $\widehat{D^{IE}_{AVG}}$ and D^{IIEE}_{AVG} by the same argument as above since the coastline will have taken on the role as an ice edge, or IIEE area limit. However, the value for D^{IE}_{AVG} will inflate in this situation. These differences in displacement metrics will be further accentuated when such coastal discrepancies are separated geographically from the remaining ice edges as *e.g.* is seen with the labeled features in Fig. 5b, and $D^{IE}_{AVG} \gg \widehat{D^{IE}_{AVG}}$ (Table 3).

The main exception to the two types of situations described above occurs when polynyas form in the open ocean, away from the continental coasts as well as the Arctic islands. However, such cases rarely arise in the set of results that are investigated here. ~~Also, the length of boundaries between IIEE areas and adjacent dry nodes is much shorter than the ice edge length in cases examined here, so the impact of disregarding coastal segments in Eq. 13 is small~~

6.1 Reducing the set of displacement metrics

The expected relationship between displacement metrics, conceptually described above, is confirmed by the results in Sect. 5. Hence, with the present configuration of validation domain and the results from model and observation, one in each of these two metrics pairs $\widehat{D^{IE}_{AVG}}$, D^{IIEE}_{AVG} and $\widehat{\Delta^{IE}}$, Δ^{IIEE} can be disregarded. Of the two approaches, we find adopting D^{IIEE}_{AVG} and Δ^{IIEE} to be the more intuitive and simpler choice (but admittedly this preference is somewhat subjective).

We can take this analysis one step forward, by systematically computing the correlation coefficients between all possible ~~sets of two displacement metrics~~ combinations of displacement metrics time series pairs. If we perform such an analysis for all 2017 forecasts and list the pairs whose correlation value is outside the range [-0.85, 0.85], 50 such pairs from a total of ~~50 such~~ 105 pairs become listed. However, an influential seasonal cycle in the metrics, evident from the strong bias during the pre-minimum, has a sizable impact on the correlation results. If we instead restrict the analysis to the months prior to the pre-minimum, and retain the criterion that pairs with correlation outside [-0.85, 0.85] is of interest, we find that 13 of the proposed 15 metrics can be divided into four groups inside which metrics have large positive ~~or negative~~ (> 0.85) or large negative (< -0.85) correlation coefficients. These groups are

1. All three D^{IE} metrics
2. D^{IIEE}_{AVG}, D^{FSS}, $\widehat{D^{IE}_{AVG}}$
3. $\widehat{\Delta^{IE}}$, Δ^{IIEE}, $\widehat{\Delta^{IIEE}}$
4. $\widehat{D^{IE}_{RMS}}$, $\widehat{D^{IIEE}_{AVG}}$, $\widehat{D^{IIEE}_{RMS}}$, $\widehat{D^{IIEE}_{MAX}}$

The two remaining displacement metrics are Δ^{IE} and $\widehat{D^{IE}_{H}}$.

Note also that the Hausdorff/maximum metrics are at times subject to large fluctuations depending on presence or absence of outliers. This was also noted in the investigation of skill metrics for sea ice model results by Dukhovskoy et al. (2015). Hence, a case can be made for disregarding the Hausdorff/maximum metrics.

6.2 Relative ice edge metrics

From the synthetic cases that were analyzed in Sect. 3, we note that the penalty for local freezing in one product but not in the other is much smaller for the IIEE-based displacement metric D^{IIEE}_{AVG} than for the ice edge displacement metric D^{IE}_{AVG}. We therefore introduce two combined, relative metrics:

$$r_{AVG} = \frac{D^{IE}_{AVG}}{D^{IIEE}_{AVG}} \tag{22}$$

$$\widehat{r_{AVG}} = \frac{D^{IE}_{AVG}}{\widehat{D^{IE}_{AVG}}} \tag{23}$$

These derived metrics will *e.g.* increase in magnitude as local freezing are seen in the observational product and not in model results since the common ~~nominator~~ numerator D^{IE}_{AVG} will inflate. Then, if the model eventually becomes able to represent the local freezing, the metrics will decrease. For the synthetic cases we investigated in Sect. 3 we find $r_{AVG} = 1.03$ and $\widehat{r_{AVG}} = 1$ in the *Reference case*. In the *Modified case* we have $r_{AVG} = 1.82$ and $\widehat{r_{AVG}} = 1.90$. The corresponding set of ratios for the two forecasts that were examined in Sect. 4 are $r_{AVG} = 1.21$ and $\widehat{r_{AVG}} = 1.14$ on 2017-04-03, and $r_{AVG} = 2.89$ and $\widehat{r_{AVG}} = 3.17$ on 2017-05-29.

We started this discussion by noting that results for the two metrics which are the denominators in Eq. 22 and 23 nearly overlap. Hence, the curves in Fig. 9a also nearly overlap. However, this is not the case for the 5-day forecast for 2017-09-11, indicated by the rightmost vertical line in Fig. 9a. This outlier in the context of the metrics ratios can be explained by examination of the IIEE areas, for which the results in the Fram strait is shown in Fig. 9b. We can infer that there is a complex shape of a large part of the ice edge in the observational product (the red ~~grids~~ grid cells that have a blue neighbour) which is at some distance from the model ice edge. This inflates the edge-integrated metric $\widehat{D^{IE}_{AVG}}$ much more than the area-derived D^{IIEE}_{AVG}, and consequently $\widehat{r_{AVG}}$ (2.18) is significantly smaller than r_{AVG} (2.94) in this case.

6.3 Recommendation

Our recommendations regarding a set of metrics to use for assessing the quality of ice edge forecasts are made from a preference of simplicity and necessity. By simplicity we have in ~~minds~~ mind metrics which are simple, not convoluted, in their implementation, and also have an intuitive interpretation. By necessity we have in mind a set of metrics for which each value provides useful information that is supplementary to the other values, and not overlapping.

From the analysis of validation results from a full calendar year that was presented in Sect. 5, and the subsequent discussion in 6.1 above, we recommend that validation results for ice edge displacement are provided for a set of three metrics:

1. D^{IE}_{AVG}
2. D^{IIEE}_{AVG}
3. Δ^{IIEE}_{AVG}

Here, 1. and 2. give a high and a low bound for the expected displacement error for the position of the ice edge, respectively. The bias metric 3. provides information about whether the ice edge should be expected before or after a user reaches the forecasted position of the ice edge.

~~Frequently, users of forecast products are interested in the results for a small portion of the full domain. Hence, when possible validation results should be provided as easily accessible representations on maps. In the present context, we strongly recommend to supplement the above set of metrics with maps showing the distribution of IIEE areas (as *e.g.* Fig. 5).~~

Moreover, while no new metrics are involved, we also encourage displaying results for

4. r_{AVG}

since time series for this quantity provides information on the robustness of the metrics results that can be easily presented as a line plot. In situations with large values of this fraction a user should be aware that the quality of the forecasted ice edge position is sensitive to how the displacement error is formulated. Note that of the two formulations in Eq. 22 and 23, our preference is the former since the episodic high impact of a complex ice edge makes interpretation of the latter less intuitive in the present context.

~~This ends our recommendation for a basic set of ice edge displacement metrics. Nevertheless, more advanced users may also benefit from access to results for the FSS as a function of neighbourhood size. Here, steep 0.5-crossings indicate low~~

~~domain-wide variability in the local displacement, since such situations arise when results by neighbourhood extent tend toward geographic homogeneity. The FSS will also be highly relevant when performance changes due to increased resolution in model system upgrades are evaluated.~~

Another useful supplement when the pan-Arctic ice edge is considered is metrics statistics that are computed for sectors or sub-domains. IN CMEMS ARC MFC, we have adopted the Global Ocean Data Assimilation Experiment ~~(GODAE; Bell et al., 2015)~~ (GODAE; Bell et al., 2015; Hernandez et al., 2009) definitions of Arctic region when comparing forecasts to microwave observations. The GODAE Arctic regions are displayed in Fig. S3 in the Supplementary Information document. An alternative definition of Arctic sectors was adopted by Posey et al. (2015) in their quantification of the sea ice edge displacement.

Obviously, in a context of forecasting, validation results will always be available after the fact only. However, recent validation results are more often than not also relevant for a future period. We apply an auto-correlation crossing at e^{-1} to define the decorrelation time scale. Then, we find that the decorrelation time scales of the metrics 1.-4. above are 6-7 weeks.

Frequently, users of forecast products are interested in the results for a small portion of the full domain. Hence, when possible validation results should be provided as easily accessible representations on maps. Hence, taking advantage of the long decorrelation time scale we recommend to supplement the above set of metrics with maps showing the distribution of IIEE areas (as *e.g.* Fig. 5).

This ends our recommendation for a basic set of ice edge displacement metrics. Nevertheless, more advanced users may also benefit from access to results for the FSS as a function of neighbourhood size: the FSS will also be highly relevant when performance changes due to increased resolution in model system upgrades are evaluated.

The above set of recommendations are based on an examination of results covering one year, for a specific forecast system and a specific observational product. While we believe that such an analysis is relevant for other sets of forecasts and observational products, each configuration should be checked separately, if resources are available. Issues like domain size (*e.g.* pan-Arctic *vs.* regional) and resolution (representation of archipelagos and straits) can conceivably affect the characteristics of the forecast quality.

We end this study by noting that the travel time for commercial shipping between ports in Northwestern Europe and the Far East is about 20-30 days with speeds in the range 10-15 knots (5-7.5 m/s) (Schøyen and Bråthen, 2011). Adding a few days for advanced decision making of sea route, and subtracting some days for sailing time in ice free conditions at the end of the leg, forecast lead times of ~~uo~~ up to 20-30 day period is expected to be required in this context. ~~Palerme et al. (2019) find that the seasonal forecasts that are initialized during April – September are more skillful than climatology for forecast ranges of 6-12 weeks.~~ Presently, CMEMS forecasts are available for lead times up to 10 days. We have shown that the deterioration in the forecast quality is moderate for these lead times (Fig. 7). Since maritime safety is one of the four core CMEMS areas of benefits, our final recommendation is to double the forecast lead time range of the CMEMS forecasting systems.

Data availability. All observational data that are used in this study is available from the CMEMS catalogue. The ice chart data and their documentation are available as product SEAICE_ARC_SEAICE_L4_NRT_OBSERVATIONS_011_002 from

http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=SEAICE_ARC_SEAICE_L4_NRT_OBSERVATIONS_011_002, and the microwave data and their documentation are available as product SEAICE_GLO_SEAICE_L4_NRT_OBSERVATIONS_011_001 from

http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=SEAICE_GLO_SEAICE_L4_NRT_OBSERVATIONS_011_001.

The CMEMS ARC forecasts (product ARCTIC_ANALYSIS_FORECAST_PHYS_002_001_a) are also distributed from the CMEMS catalogue, but the forecasts are overwritten on a weekly basis by results from a delayed-mode ensemble simulation that is used for data assimilation purposes. The forecasts that are analyzed in this investigation is however publicly available from

http://thredds.met.no/thredds/myocean/ARC-MFC/myoceanv2-class1-arctic.html.

Author contributions. Melsom performed the analysis and wrote the article. Based on results from the analysis Palerme provided Fig.s 6 and 7, the remaining figures were provided by Melsom. Palerme and Müller contributed in discussions and provided comments and suggestions that significantly improved the presentation of the present study.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We would like to express our gratitude to two anonymous referees whose comments and suggestions significantly improved our manuscript. We are also indebted to the model development team at the Nansen Environmental and Remote Sensing Center, and the provision of ice chart from the Norwegian Ice Service at the Norweian Meteorological Institute. This study has been performed on behalf of the the Copernicus Marine Environmental and Monitoring Service under Mercator Océan contract no. 2015/S 009-011301. Additional support was available from the Nansen Legacy project and the Salienseas project which are funded by the Norwegian Research Council under contracts no. 276730 and 276223, respectively. This is a contribution to the Year of Polar Prediction (YOPP), a flagship activity of the Polar Prediction Project (PPP), initiated by the World Weather Research Programme (WWRP) of the World Meteorological ~~Organisation~~ Organization (WMO). Fig.s ~~3~~1-5, 8-9, S1-3,5 were made using the NCAR Command Language (NCL, 2017).

References

Arzel, O., Fichefet, T., and Goosse, H.: Sea ice evolution over the 20th and 21st centuries as simulated by current AOGCMs, Ocean Mod., 12 (3-4), 401-415, doi:10.1016/j.ocemod.2005.08.002, 2006.

Bell, M.-J., Schiller, A, Le Traon, P.-Y., Smith, N. R., Dombrowsky, E, Wilmer-Becker, K.: An introduction to GODAE OceanView, J. Op. Oceanogr., 8 (S1), 2-11, doi:10.1080/1755876X.2015.1022041, 2015.

Breivik, L.-A., Eastwood, S, Godøy, Ø, Schyberg, H, Andersen, S, Tonboe, R. T.: Sea Ice Products for EUMETSAT Satellite Application Facility, Can. J. Rem. Sens., 27 (5), 403-410, doi:10.1080/07038992.2001.10854883, 2001.

Carrieres, T., Casati, B., Caya, A., Posey, P., Metzger, E. J., Melsom, A., Sigmond, M., Kharin, V., and Dupont, F.: System evaluation, in: Sea Ice Analysis and Forecasting, edited by: Carrieres T., Buehner M., Lemieux J. F. and Pedersen, L. T., Cambridge University Press, doi:10.1017/9781108277600, 2017.

Dukhovskoy, D. S., Ubnoske, J., Blanchard-Wrigglesworth, E. , Hiester, H. R., and Proshutinsky, A.: Skill metrics for evaluation and comparison of sea ice models, J. Geophys. Res. Oceans, 120, 5910-5931, doi:10.1002/2015JC010989, 2015.

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T.: Predictability of the Arctic sea ice edge, Geophys. Res. Lett., 43, 1642–1650, doi:10.1002/2015GL067232, 2016.

Goessling, H. F., and Jung, T.: A probabilistic verification score for contours: Methodology and application to Arctic ice-edge forecast, Q. J. R. Meteorol. Soc., 144, 735–743, doi:10.1002/qj.3242, 2018.

Hernandez, F., Bertino, L., Brassington, G., Chassignet, E., Cummings, J., Davidson, F., Drévillon, M., Garric, G., Kamachi, M., Lellouche, J.-M., Mahdon, R, Martin, M. J., Ratsimandresey, A. and Regnier, C.: Validation and Intercomparison Studies Within GODAE, Oceanography, 22(3), 128-143, doi:10.5670/oceanog.2009.71, 2009.

Ho, J., The implications of Arctic sea ice decline on shipping, Mar. Policy, 34, 713-715, doi:10.1016/j.marpol.2009.10.009, 2010.

Johnson, M., Gaffigan, S., Hunke, E., and Gerdes, R.: A comparison of Arctic Ocean sea ice concentration among the coordinated AOMIP model experiments. J. Geophys. Res., 112, C04S11, doi:10.1029/2006JC003690, 2007.

Johnson, S.J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D, Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B., M.: SEAS5: the new ECMWF seasonal forecast system. Geosci. Model Dev., 12, 1087-1117, doi:10.5194/gmd-12-1087-2019, 2019.

Massonnet, F., Fichefet, T., Goosse, H., Bitz, C. M., Philippon-Berthier, G., Holland, M. M., and Barriat, P.-Y.: Constraining projections of summer Arctic sea ice, The Cryosphere, 6, 1383–1394, doi:10.5194/tc-6-1383-2012, 2012.

Melsom, A., Simonsen, M., and Bertino L.: MyOcean Project Scientific Validation Report (ScVR) for V1 real-time forecasts, Tech. Rep., met.no, Oslo, Norway, 21 pp. Retrieved from http://cmems.met.no/ARC-MFC/Validation/validationReport01.pdf

The NCAR Command Language (Version 6.4.0) [Software], Boulder, Colorado: UCAR/NCAR/CISL/TDD, doi:10.5065/D6WD3XH5, 2017.

Palerme, C., Müller, M., and Melsom, A.: An intercomparison of skill scores for evaluating the sea ice edge position in seasonal forecasts, Geophys. Res. Lett., submitted.

Posey, P. G.., Metzger, E. J., Wallcraft, A. J., Hebert, D. A., Allard, R. A., Smedstad, O. M., Phelps, M. W., Fetterer, F., Stewart, J. S., Meier, W. N., and Helfrich, S. R.: Improving Arctic sea ice edge forecasts by assimilating high horizontal resolution sea ice concentration data into the US Navy's ice forecast system, Cryosphere, 9, 1735–1745, doi:10.5194/tc-9-1735-2015, 2015.

Roberts, N. M., and Lean, H. W.: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events, Mon. Weather Rev., 136, 78–97, doi:10.1175/2007MWR2123.1, 2008.

Schøyen, H., and Bråthen, S.: The Northern Sea Route versus the Suez Canal: cases from bulk shipping, J. Transp. Geogr., 19, 977-983, doi:10.1016/j.jtrangeo.2011.03.003, 2011.

Skok, G., and Roberts, N. M.: Analysis of Fractions Skill Score properties for random precipitation fields and ECMWF forecasts, Q. J. R. Met. Soc., 142, 2599-2610, doi:10.1002/qj.2849, 2016.

Skok, G., and Roberts, N. M.: Estimating the displacement in precipitation forecasts using the Fractions Skill Score, Q. J. R. Met. Soc., 144, 414-425, doi:10.1002/qj.3212, 2018.

WMO: Sea-Ice Information Services in the World, WMO No. 574, World Meteorological Organization, 103pp, 2017.

Zampieri, L., Goessling, H. F., and Jung, T.: Bright Prospects for Arctic Sea Ice Prediction on Subseasonal Time Scales, Geophys. Res. Lett., 45, 9731-9738, doi:10.1029/2018GL079394, 2018.

	Ice edge displacement metrics							
	D^{IE}_{AVG}	D^{IE}_{RMS}	D^{IE}_{H}	$\widehat{D^{IE}_{AVG}}$	$\widehat{D^{IE}_{RMS}}$	$\widehat{D^{IE}_{H}}$	Δ^{IE}	$\widehat{\Delta^{IE}}$
Reference case	9.1	10.6	20	9.1	10.6	20	0.24	0.24
Modified case	17.5	27.4	112	9.2	10.7	20	-9.1	-0.8

	FSS	IIEE displacement metrics						
	D^{FSS}	D^{IIEE}_{AVG}	$\widehat{D^{IIEE}_{AVG}}$	$\widehat{D^{IIEE}_{RMS}}$	$\widehat{D^{IIEE}_{MAX}}$	Δ^{IIEE}	$\widehat{\Delta^{IIEE}}$
Reference case	8.8	8.8	10.4	10.5	10.6	0.17	0.21
Modified case	9.8	9.6	11.0	11.1	13.4	-1.7	-2.3

Table 1. Results for the various displacement metrics defined in Sect. 2. Vertical lines are introduced to separate non-negative displacement metrics from signed bias metrics, and the FSS metric from IIEE metrics. The *Reference case* and the *Modified case* refer to the observational sea ice concentrations that are displayed in Fig. 3a and b, respectively. All values are given in non-dimensional grid units. Note that in the *Reference case*, all boundaries are considered open, and so the ice edge displacement metrics are unaffected when computing the hatted variables. Note also that in the *Modified case*, the bottom boundary was treated as adjacent to a closed (land) boundary.

	IIEE area metrics		Fractions skill score		
	A^{IIEE}	α^{IIEE}	$n=3$	$n=7$	$n=11$
Reference case	2002	38	0.14	0.26	0.37
Modified case	2470	-430	0.12	0.24	0.34

Table 2. Supplementary metric scores. IIEE area scores are given in non-dimensional grid units. The fractions skill scores is computed by Eq. 21.

	Ice edge displacement metrics							
	D^{IE}_{AVG}	D^{IE}_{RMS}	D^{IE}_{H}	$\widehat{D^{IE}_{AVG}}$	$\widehat{D^{IE}_{RMS}}$	$\widehat{D^{IE}_{H}}$	Δ^{IE}	$\widehat{\Delta^{IE}}$
Forecast 4-3	35	47	150	31	43	150	-14	-15
Forecast 5-29	98	230	1560	31	39	130	-87	-23
All 5-day forecasts	69	116	720	37	48	175	-55	-27
Bootstrap fraction	0.25	0.27	0.28	0.17	0.15	0.15	0.39	0.35

	FSS		IIEE displacement metrics				
	D^{FSS}	D^{IIEE}_{AVG}	$\widehat{D^{IIEE}_{AVG}}$	$\widehat{D^{IIEE}_{RMS}}$	$\widehat{D^{IIEE}_{MAX}}$	Δ^{IIEE}	$\widehat{\Delta^{IIEE}}$
Forecast 4-3	45	29	61	69	100	-14	-40
Forecast 5-29	48	34	57	61	91	-27	-48
All 5-day forecasts	61	39	79	86	119	-29	-64
Bootstrap fraction	0.28	0.18	0.18	0.18	0.15	0.33	0.30

Table 3. Results for the various sea ice edge displacement metrics. *Forecast 4-3* and *Forecast 5-29* results are metrics for the forecast for 2017-04-03 issued on 2017-03-30, and for the forecast for 2017-05-29 issued on 2017-05-25, respectively. *All 5-day forecasts* results are averages for all weekly 2017 forecast bulletins with a 5 day lead time. Bootstrap fraction is the difference between the 95 percentile and 5 percentile values from a bootstrap analysis, divided by the corresponding average value. All values are in km. See the text for details.

	IIEE area metrics		Fractions skill score		
	A^{IIEE}	α^{IIEE}	$n = 3$	$n = 7$	$n = 11$
Forecast 3-4	220	-110	0.35	0.63	0.75
Forecast 5-29	210	-167	0.30	0.54	0.68
All 5-day forecasts	260	-186	0.30	0.49	0.59
Bootstrap fraction	0.21	0.36	0.20	0.17	0.15

Table 4. Supplementary metric scores for the forecasts displayed in Fig. 5 and the corresponding 2017 average values. IIEE area scores are given in units of 1000 km^2.

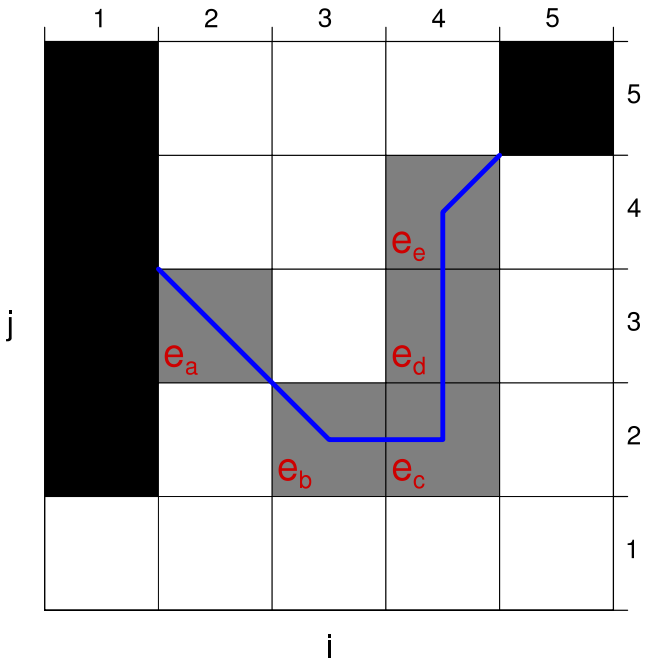

Figure 1. Schematic illustration for computation of ice edge length. The ice edge is displayed by the labeled cells that are filled grey. Black cells correspond to land. The algorithm we present here for calculation of the ice edge length yields a value that corresponds to the length of the blue line, see the text for details.

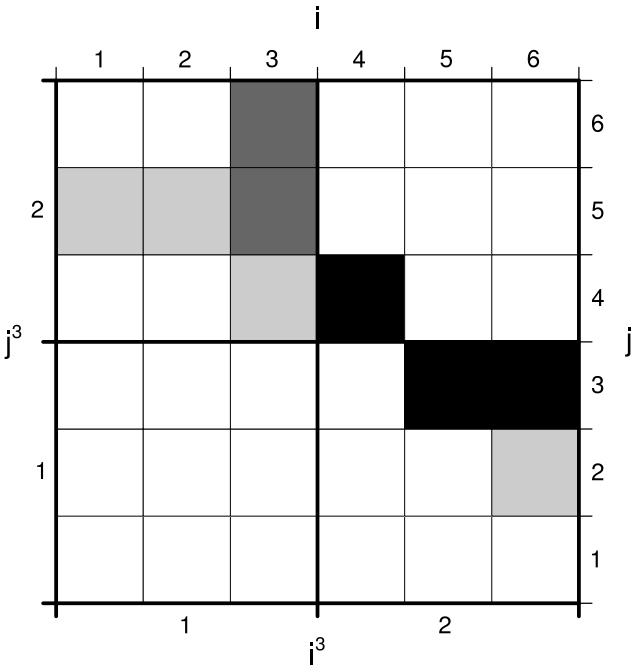

Figure 2. Schematic illustration for computation of fractions skill score for gridded contour lines. Gridded lines representing the ice edge of the model product and the observational product are shown as light gray boxes and dark gray boxes, respectively. Grid cells where the two lines overlap are black. The original grid is displayed by thin grid lines with x-axis indices at the top and y-axis indices to the right. Thick grid lines correspond to the grid of a neighbourhood with an extent of 3 grid cells ($n = 3$), with x- and y-axis indices at the bottom and to the left, respectively. See the text for details.

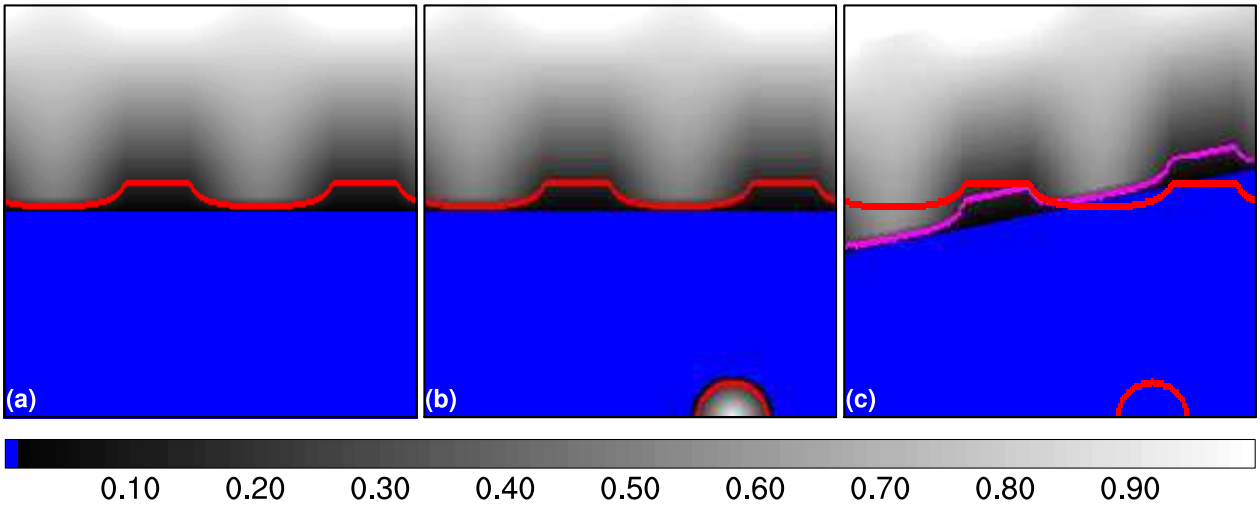

Figure 3. Sea ice concentrations representing (a) reference observations, (b) modified observations and (c) model results. The ice edges in the observational and model products are drawn as red and magenta lines, respectively. (These lines are drawn with three times their actual thickness in order to accentuate the edges graphically.) Note that the ice edge from the modified observations has been added in (c). Blue color represents ice free conditions, and the gray scale used for sea ice concentration is displayed by the label bar at the bottom.

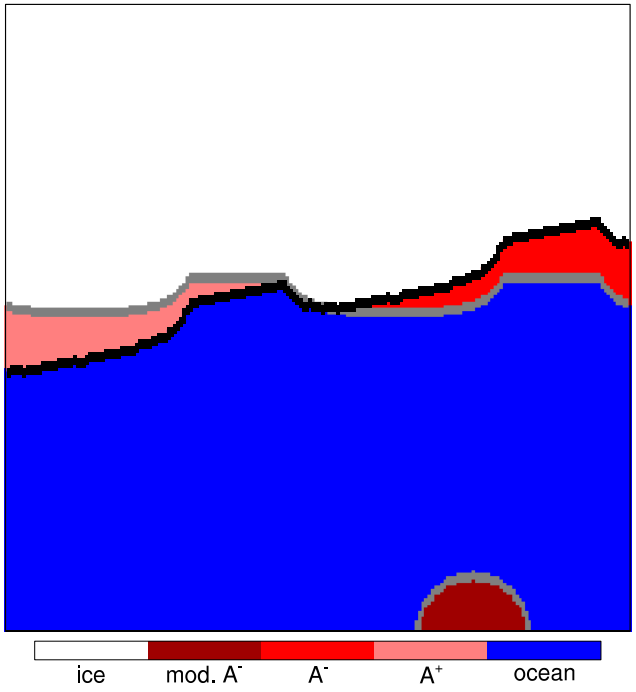

Figure 4. Depiction of areas used for computing the IIEE metrics. The pink region corresponds to the A^+ area given by Eq. 9, whereas the A^- area given by Eq. 10 is in red. The additional A^- area in the *Modified case* is in dark red. Ice edges are displayed as gray lines (observations) and black lines (model results). (These lines are drawn with three times their actual thickness in order to accentuate the edges graphically.) Regions where all products are on the open ocean side of the ice edges are blue, while regions which are inside the ice edges in all products are white.

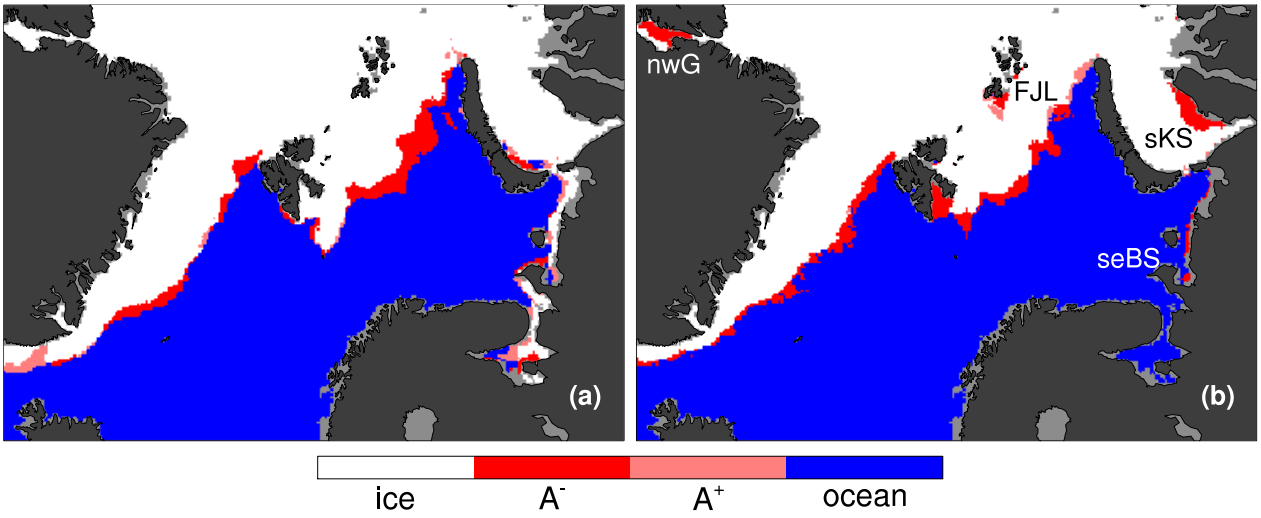

Figure 5. Map displaying the IIEE regions for two forecasts. Panels a and b display the results for the forecast for 2017-04-03 issued on 2017-03-30, and for the forecast for 2017-05-29 issued on 2017-05-25, respectively. Areas displayed in gray are not included in one or both products, and are excluded in the present analysis. The following regions with ice edge discrepancies are labeled in panel b: near Franz Josef Land (FJL), southern Kara Sea (sKS), northwest of Greenland (nwG), and southeastern Barents Sea (seBS). The displayed region is nearly the same as the region with ice chart data (a slight zooming was applied in order to highlight features of interest, so narrow bands of ~~nodes~~ grid cells from the ice chart data to the right and to the bottom are not shown). The color codes for the various IIEE regions are the same as in Fig. 4.

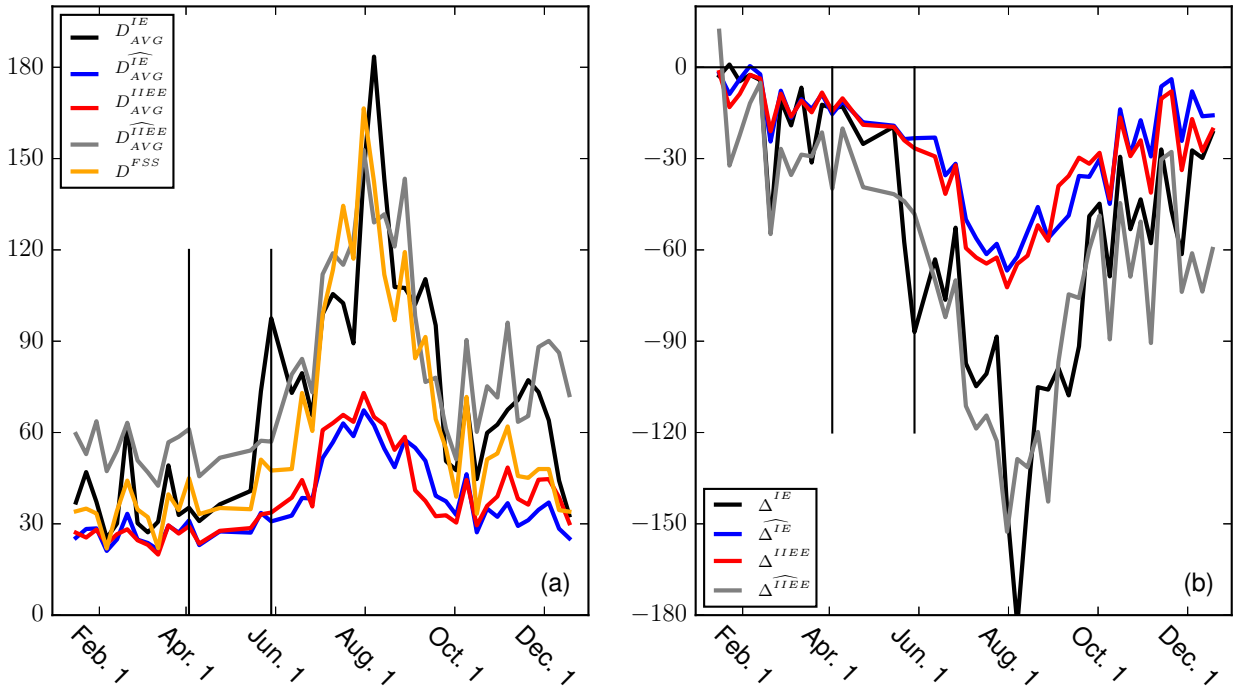

Figure 6. Time series for (a) mean displacement and (b) bias metrics as defined in Sect. 2. All results are for the 5-day forecasts. Vertical lines correspond to the two forecasts that were analyzed in Sect. 4. Values along the vertical axes are in units of km.

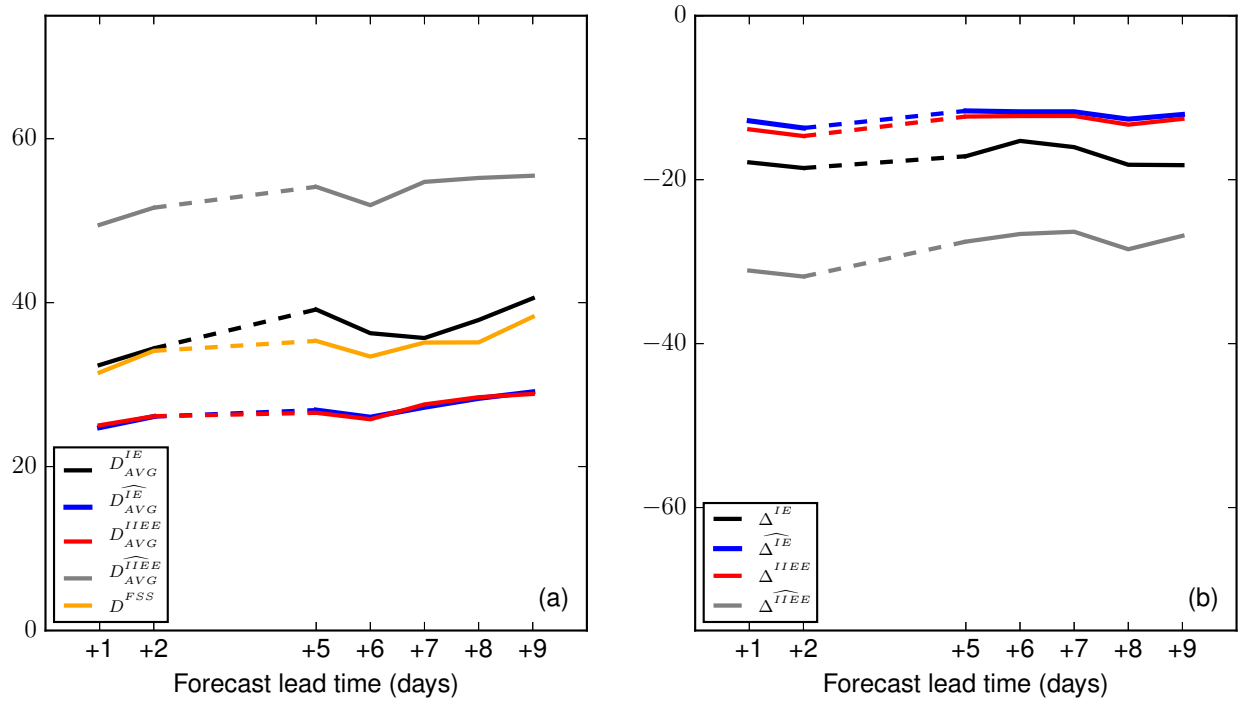

Figure 7. Metrics for (a) mean displacement and (b) bias, as functions of forecast lead time, in days. These results are based on forecast bulletins from January 2017 to mid-May 2017. Note that lines for $\widehat{D^{IE}_{AVG}}$ and D^{IIEE}_{AVG} in (a) nearly overlap, as do lines for $\widehat{\Delta^{IE}}$ and Δ^{IIEE} in (b). Values along the vertical axes are in units of km. Ice charts are not produced on Saturdays and Sundays, which correspond to forecast lead times of +3 days and +4 days, respectively. Dashed lines are thus used to indicate the lack of analysis for these two days.

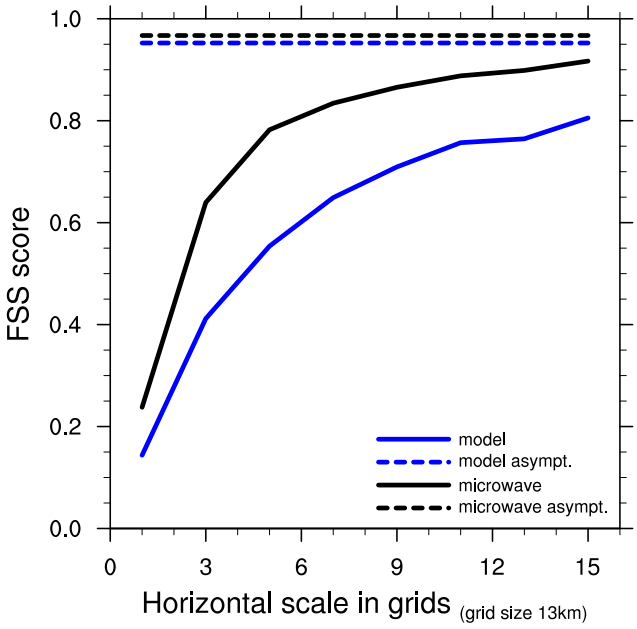

Figure 8. ~~(~~Fractions skill score as a ~~) Time series~~ function of ~~two metrics ratios~~ resolution, for ~~forecasts with a~~ five-day lead time ~~of 5 days. Vertical lines correspond to cases for which results are discussed in detail. The left and center vertical lines correspond to the two~~ model forecasts ~~that were analyzed in Sect. 4, whereas the~~ *vs.* ice chart data (blue line~~to the right is for the situation displaced in the right panel.~~) and microwave data *vs.* ice chart data (~~b~~black line)~~Detail of IIEE in~~ . Dashed lines show the ~~Fram strait~~ asymptotic FSS values as defined by Roberts and Lean (2008) (~~the region between Greenland and the Svalbard archipelago~~their Eq. 8)~~on 2017-09-11~~. These results are based on forecast bulletins and microwave data from January 2017 to mid-May 2017.

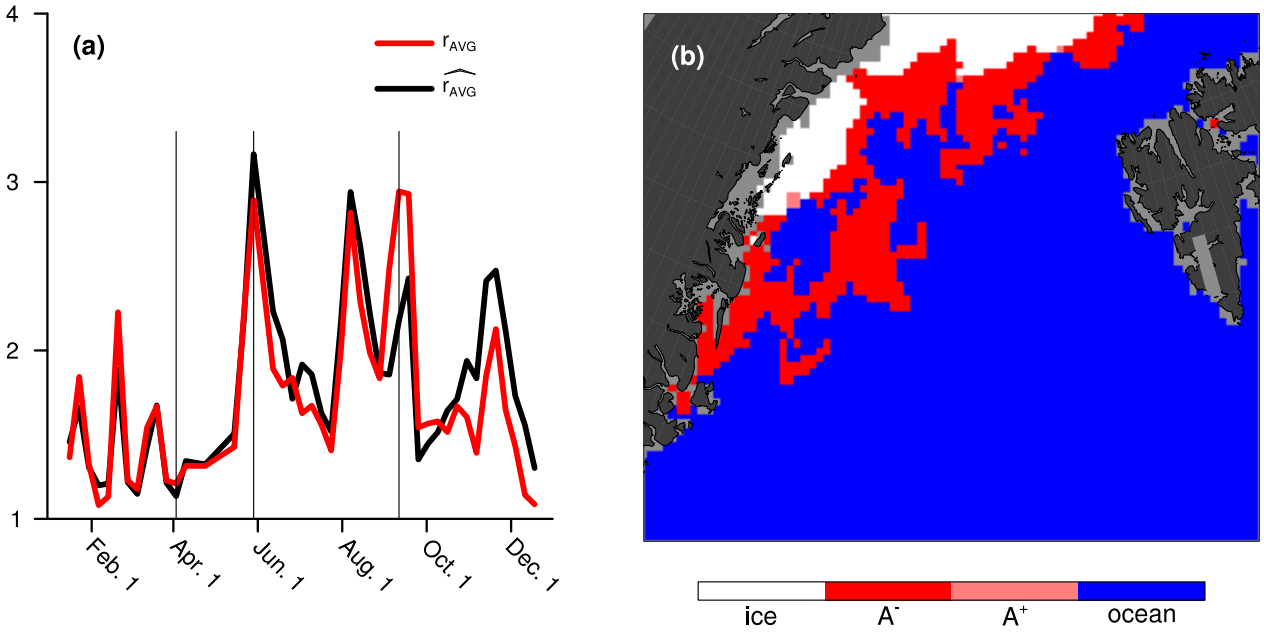

Figure 9. (a) Time series of two metrics ratios for forecasts with a lead time of 5 days. Vertical lines correspond to cases for which results are discussed in detail. The left and center vertical lines correspond to the two forecasts that were analyzed in Sect. 4, whereas the line to the right is for the situation displayed in the right panel. (b) Detail of IIEE in the Fram strait (the region between Greenland and the Svalbard archipelago) on 2017-09-11.

Supplementary Information

Arne Melsom[1], Cyril Palerme[1], and Malte Müller[1]

[1]Norwegian Meteorological Institute

Correspondence: Arne Melsom (arne.melsom@met.no)

S1 ~~Definitions and illustrations for selected ice edge~~ Separation based IIEE displacement metrics

Most definitions were provided in the main text. Here we present some supplementary information which ~~mainly~~ is concerned with metrics $\widehat{D^{IIEE}}$ that were not included in the recommended set in Sect. 6.3 in the main text. ~~Also, two figures are included to provide readers with illustrative information regarding the definition and interpretation of some metrics.~~

S1.1 ~~Separation based IIEE displacement metrics~~

Provided that the model initialization of the sea ice fraction is close to the observed ice edge fraction at that time, IIEE areas can be expected to emerge as the model ice edge drifts away from the observed edge with an increasing forecast lead time. This evolution is expected to frequently give rise to elongated IIEE areas, and we here adopt the maximum distance inside an IIEE area as the scaling length.

An illustrative example for IIEE and derived metrics is provided in Fig. S1. Here, gray shaded ~~grids represents grids~~ cells represents grid cells in IIEE area ia, while white ~~grids~~ cells are outside of the IIEE domain. The scaling length l^{ia}_{max} is indicated by the dashed line. Note that when computing the scaling length we have chosen not to include IIEE ~~grids~~ grid cells with only a single IIEE grid cell neighbour (given by light gray shading in the figure).

Since the definitions of a^{ia} and l^{ia}_{max} take adjacent dry nodes into account, we adopt the hatted notation as introduced in Sect. 2.1 in the main text. The resulting displacement for this area is given as

$$\hat{d}^{ia}_{IIEE} = a^{ia} / l^{ia}_{max} \tag{S1}$$

Note that in theory, a node may be adjacent to two IIEE areas. In such cases, we divide the node's area equally between the two relevant IIEE areas.

A solitary IIEE node is formally treated as a separate IIEE area, with scaling length set to the (average) resolution. Furthermore, let A_0 be the total area of the ~~grids~~ grid cells where the two ice edges overlap. Letting N_A be the number of IIEE areas, we introduce a set of four corresponding displacement metrics here.

1. The root-mean-squared displacement:

$$\widehat{D^{IIEE}_{RMS}} = \left[\frac{\sum_{ia=1}^{N_A} a^{ia} \left(\hat{d}^{ia}_{IIEE} \right)^2}{A_0 + \sum_{ia=1}^{N_A} a^{ia}} \right]^{1/2} \tag{S2}$$

2. The average displacement:

$$\widehat{D^{IIEE}_{AVG}} = \frac{\sum_{ia=1}^{N_A} a^{ia}\hat{d}^{ia}_{IIEE}}{A_0 + \sum_{ia=1}^{N_A} a^{ia}} \tag{S3}$$

3. The displacement bias:

$$\widehat{\Delta^{IIEE}} = \frac{\sum_{ia^+=1}^{N_A^+} a^{ia^+}\hat{d}^{ia^+}_{IIEE}}{A_0/2 + \sum_{ia^+=1}^{N_A^+} a^{ia^+}} - \frac{\sum_{ia^-=1}^{N_A^-} a^{ia^-}\hat{d}^{ia^-}_{IIEE}}{A_0/2 + \sum_{ia^-=1}^{N_A^-} a^{ia^-}} \tag{S4}$$

4. The maximum displacement:

$$\widehat{D^{IIEE}_{MAX}} = \max(\hat{d}^{ia}) \tag{S5}$$

In order to shed some light on the relation between the D^{IIEE} metric and $\widehat{D^{IIEE}}$ we consider an idealized case where two products' ice edges are y symmetric to each other, and form IIEE in the shape of two rectangles, connected by a line where the edges overlap. A sample configuration of such an idealized case is displayed in Fig. S2. Now, take the width (in the x-direction) of the rectangles to be w_1 and w_2 ~~grids~~ grid cells respectively, while the length of the mutual edge in between is w_0 ~~grids~~grid cells. The height of the two rectangles are set to h_1 and h_2 ~~grids~~grid cells, respectively.

Then, for D^{IIEE}_{AVG} we have

$$\begin{aligned} A^{IIEE} &= w_1 \cdot h_1 + w_2 \cdot h_2, \\ L &= h_1 + w_1 + w_o + h_2 + w_2 \end{aligned} \tag{S6}$$

where L is the ice edge length for both products. Consequently,

$$D^{IIEE}_{AVG} = \frac{w_1 \cdot h_1 + w_2 \cdot h_2}{h_1 + w_1 + w_0 + h_2 + w_2} \tag{S7}$$

To determine $\widehat{D^{IIEE}}$ we first find that

$$\begin{aligned} \hat{d}^{(1,2)}_{IIEE} &= w_{(1,2)} \cdot h_{(1,2)}/l^{(1,2)}_{max}, \\ l^{(1,2)}_{max} &= (w^2_{(1,2)} + h^2_{(1,2)})^{0.5} \end{aligned} \tag{S8}$$

Furthermore, $A_0 = w_0 \cdot 1$, and introducing these quantities into Eq. S3 we find

$$\widehat{D^{IIEE}_{AVG}} = \frac{w_1^2 \cdot h_1^2/(w_1^2 + h_1^2)^{0.5} + w_2^2 \cdot h_2^2/(w_2^2 + h_2^2)^{0.5}}{w_0 \cdot 1 + w_1 \cdot h_1 + w_2 \cdot h_2} \tag{S9}$$

Now consider some selected cases:

Case 1 Identical squares, *i.e.*, $w_1 = w_2 = h_1 = h_2 = w$; $w_0 = \nu w$. Then,

$$\frac{\widehat{D^{IIEE}_{AVG}}}{D^{IIEE}_{AVG}} = \frac{1 + \nu/4}{1 + \nu/(2w)}\sqrt{2} \geq \sqrt{2} \tag{S10}$$

The latter inequality follows since $w \geq 2$. To take an example, assume that the squares have sides with 20 ~~grids~~grid cells. Then, if $\nu = 1/4$ (the squares are 5 ~~grids~~ grid cells apart) the fraction in Eq. S10 is approximately 1.5. If $\nu = 4$ (a separation of 80 ~~grids~~grid cells) the fraction has a value of about 3.

Case 2 Different sized squares, *i.e.*, $w_1 = h_1 = w;\ \ w_2 = h_2 = \alpha w;\ \ w_0 = \nu w$. Then,

$$\frac{\widehat{D^{IIEE}_{AVG}}}{D^{IIEE}_{AVG}} = \frac{1+\alpha^3}{1+\alpha^2}\ \frac{1+\alpha+\nu/2}{1+\alpha^2+\nu/w}\sqrt{2} \tag{S11}$$

Consider the case $\alpha = 1/4$, and set $w = 20$ ~~grids~~grid cells. Then, the fraction in Eq. S11 becomes about 1.7 and 2.5 when we set $\nu = 1/4$ and $\nu = 4$, respectively.

Case 3 Identical rectangles, *i.e.*, $w_1 = w_2 = w;\ \ h_1 = h_2 = \delta w;\ \ w_0 = \nu w$. Then,

$$\frac{\widehat{D^{IIEE}_{AVG}}}{D^{IIEE}_{AVG}} = \frac{1}{\sqrt{1+\delta^{-2}}}\ \frac{1+\delta^{-1}(1+\nu/2)}{1+\delta^{-1}\nu/(2w)} \tag{S12}$$

In the model results, the IIEE areas are usually elongated in the direction ~~parallell~~ parallel to the main direction of the ice edge, *i.e.*, $\delta < 1$. When we investigate the case $\delta = 1/4$ and again set $w = 20$ ~~grids~~grid cells, the fraction in Eq. S12 becomes approximately 1.35 and 2.3 for $\nu = 1/4$ and $\nu = 4$, respectively.

Based on these idealized examples, we will expect that the definition of $\widehat{D^{IIEE}_{AVG}}$ leads to values that are larger than the corresponding values for D^{IIEE}_{AVG}. If the results from the idealized examples are representative in operational applications, the ratio of these quantities will be in the approximate range of 1.5-3.

S1.1 ~~Fractions skill score~~

S2 Comparison of results from ice charts and microwave data

~~An idealized example provided to shed light on FSS metrics is given in Fig. ??. Here, two gridded contour lines are displayed by filled boxes. On the original grid the two lines extend over 9 and 12 grid cells, respectively, including four cells where they overlap. Let us associate the gridded line shown by light gray and black boxes with observations of the sea ice edge, and take the dark gray and black boxes to represent a model result. Then, the ice edge fractions for a neighbourhood size $n = 3$ becomes~~

$$I_o^{n=3} = \frac{1}{9}\begin{pmatrix} 0 & 0 & 0 \\ 3 & 1 & 3 \\ 0 & 2 & 0 \end{pmatrix} \quad ; \quad I_M^{n=3} = \frac{1}{9}\begin{pmatrix} 2 & 0 & 2 \\ 2 & 1 & 3 \\ 0 & 2 & 0 \end{pmatrix}$$

~~and we find that $\mathrm{MSE}^{n=3} = 9/9^3$, $\mathrm{MSE}^{n=3}_{\mathrm{ref}} = 49/9^3$ and the fractions skill score for a neighbourhood size $n = 3$ is $\mathrm{FSS}^{n=3} = 40/49 \approx 0.82$~~
~~For the skill score with the original 9×9 grid we have $\mathrm{MSE}^{n=1} = 13/81$, $\mathrm{MSE}^{n=3}_{\mathrm{ref}} = 21/81$, and we find that $\mathrm{FSS}^{n=1} = 8/21 \approx 0.38$~~We present results from a comparison of the microwave data for sea ice concentration and corresponding results from ice charts, *i.e.*, a comparison of two observational products. The microwave data have been assimilated by the Arctic Ocean Physics Analysis And Forecast product. Differences between assimilated data and the product used for subsequent validation of model results can potentially significantly affect the validation results. Thus, the purpose of this supplementary analysis is to provide information that shed light on the effect of using an independent data set (the ice charts) has on the validation results.

Moreover, we note from Eq. s 18 and 19 in the main text that the FSS score will not change if we introduce a set of additional grids where neither product has an ice edge, provided that non-events dominate events (*i.e.*, the first term in Eq. We repeat the analysis in Sect. 5 after having replaced model results with microwave data. The tabulated metric values are provided in Tables S1, S2, and the temporal evolution of average displacement metrics and the displacement biases are displayed in Fig. S3. The corresponding results for the model product are given by Tables 3, 19 is used, here, that the number of nodes without an ice edge is larger than the number of edge nodes) . This observation has consequences for two different aspects in the present study.

First, when modeling the ocean, dry nodes are usually not considered to be part of the computational domain, and are assigned a special value in numerical results . When integrating over a neighbourhood as in Eq. 22 in the main text one option would be to discard the grids that are dry in the original representation.We will then be left with a result which has a non-constant neighbourhood size of n^2 where dry nodes are not present,and $< n^2$ for neighbourhoods where dry nodes are present.Here, we choose to avoid the problem of non-constant neighbourhood sizes by adopting $I_o = I_m = 0$ for dry grids.

Second, 4 and Fig. 6. We note the the metrics for average displacements for the microwave product (Fig. S3a) are generally about half of the values when compared to the grid for n=3 indicated by thick lines in metrics computed by the model product (Fig. 6a). The exception is the period leading up to the sea ice minimum, when the discrepancy between the two observational products is about the same as revealed by Fig. **??** is only one of nine possible configurations. Since the FSS results are not affected by additional grids where neither product has an ice edge, we can expand the original domain by adding a padding region of n-1 grids. In the case of n=3 all configurations are attained by shifting the neighbourhood by 6a, and even higher episodically.

Regarding the results for the bias, we note that the ice edge position in the model product is biased negative throughout the year (corresponding to a larger sea ice extent in the ice chart data). When we compare the microwave data with the ice chart, the bias is generally approximately 0, 1 and 2 original grids in both directions. The average FSS score from all of the configurations will be used, since the alternative is a set of results that will depended on an arbitrary configuration subset choicebut again large discrepancies between the two operational products are seen in the bias metrics values during the period that leads up to the sea ice minimum.

S3 Map of GODAE regions

The map of GODAE regions in the Arctic Ocean and adjacent seas, which was referred to near the end of the main text, is available as Fig. S4.

	Ice edge displacement metrics							
	D^{IE}_{AVG}	D^{IE}_{RMS}	D^{IE}_{H}	$\widehat{D^{IE}_{AVG}}$	$\widehat{D^{IE}_{RMS}}$	$\widehat{D^{IE}_{H}}$	Δ^{IE}	$\widehat{\Delta^{IE}}$
Microwave 4-3	14	19	170	13	16	87	1	1
Microwave 5-29	32	68	490	16	20	66	21	4
All microwave data	38	59	350	23	30	110	-26	-13

	FSS	IIEE displacement metrics					
	D^{FSS}	D^{IIEE}_{AVG}	$\widehat{D^{IIEE}_{AVG}}$	$\widehat{D^{IIEE}_{RMS}}$	$\widehat{D^{IIEE}_{MAX}}$	Δ^{IIEE}	$\widehat{\Delta^{IIEE}}$
Microwave 4-3	18	12	28	34	65	0	1
Microwave 5-29	21	15	31	36	74	4	11
All microwave data	40	23	50	56	82	-13	-32

Table S1. Results for the various sea ice edge displacement metrics, when microwave data are compared to ice chart data. *Microwave 4-3* and *Microwave 5-29* results are metrics valid for 2017-04-03 and 2017-05-29, respectively. *All microwave data* are averages for all weekly 2017 data, on dates for which results are examined Sect. 5.

	IIEE area metrics		Fractions skill score		
	A^{IIEE}	α^{IIEE}	$n = 3$	$n = 7$	$n = 11$
Microwave 3-4	81	-48	0.68	0.85	0.90
Microwave 5-29	96	1	0.64	0.80	0.86
All microwave data	92	26	0.51	0.68	0.74

Table S2. Supplementary metric scores for microwave data *vs.* ice charts. IIEE area scores are given in units of 1000 km^2.

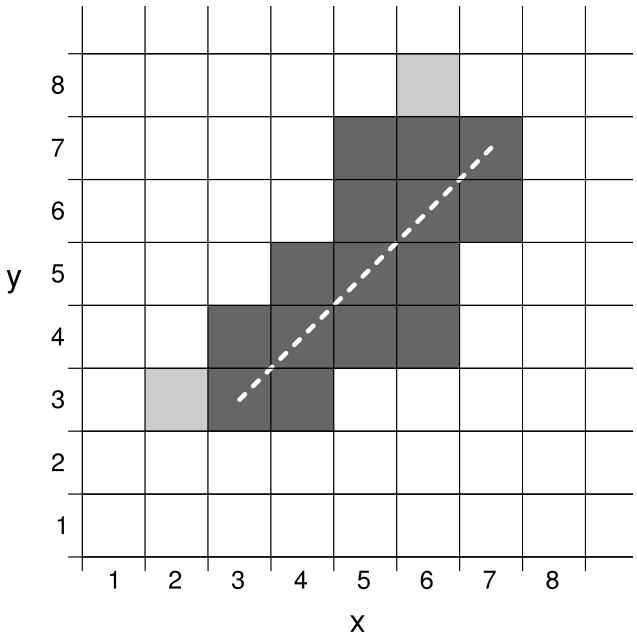

Figure S1. Illustration for scaling length of continuous IIEE areas. Here, the IIEE area is shown as gray shaded ~~grids~~grid cells, which in this example is a 17 grid cell area. When determining the scaling length, IIEE area ~~grids~~ grid cells with only one IIEE area grid cell neighbour are disregarded (light gray shading). The scaling length is then set to the largest distance between the centers of the remaining IIEE area ~~grids~~grid cells. This distance is indicated by the white dashed line. The displacement given by Eq. S1 in the metrics defined in Sect. S1 of this continuous IIEE area is then the area (17 ~~grids~~grid cells) divided by its scaling length.

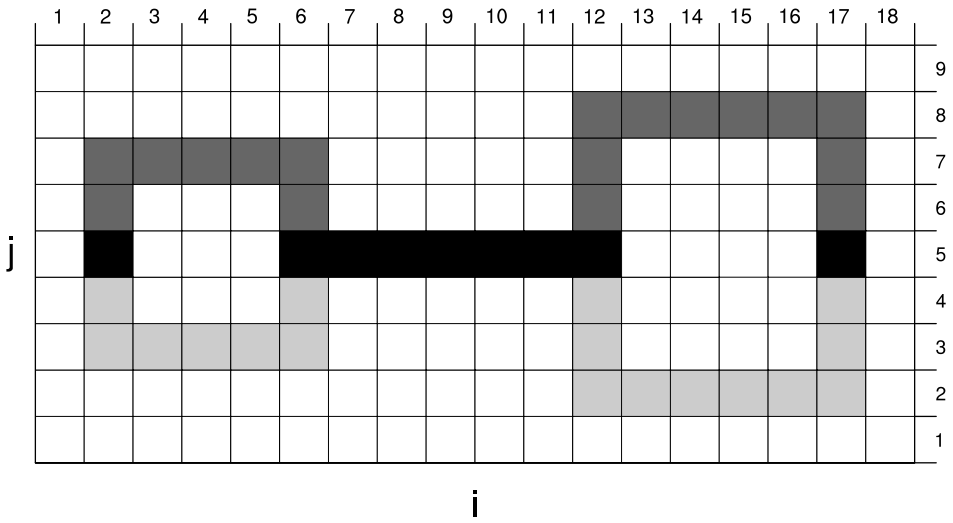

Figure S2. ~~Illustration for computation of fractions skill score for gridded contour lines. One of~~ Diagram displaying a sample idealized situation with IIEE areas have taken on the ~~gridded lines is shown as light gray boxes~~shapes of two rectangles, ~~whereas the other is shown as dark gray. Grids~~ connected by a straight line where the ice edges in the two ~~lines~~ products overlap~~are black~~. ~~The original grid is displayed by thin grid lines with x-axis indices at the top and y-axis indices to the right. Thick grid lines correspond to the grid with $n = 3$~~Here, ~~with x- and y-axis indices at the bottom~~ $w_1 = h_1 = 4$; $w_0 = 6$; $w_2 = 5, h_2 = 6$ where subscripts 1 and 2 correspond to the left and right rectangles, respectively. See the text for details.

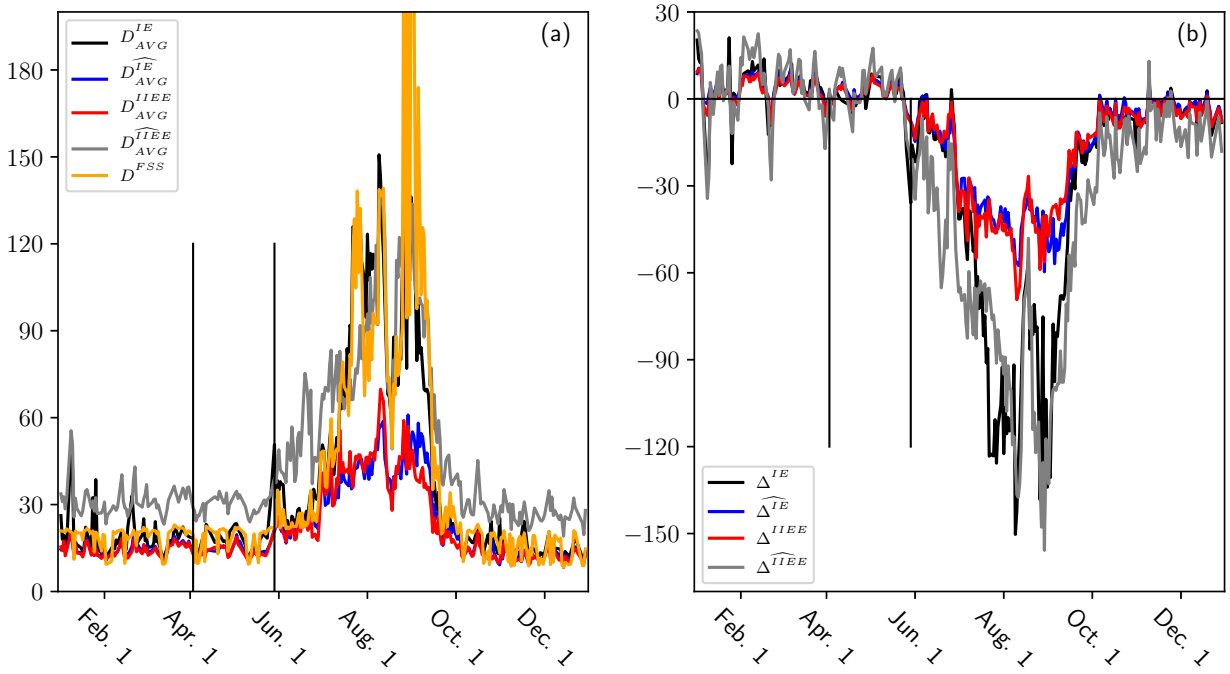

Figure S3. Time series for (a) mean displacement and (b) bias metrics as defined in Sect. 2, for the microwave product *vs.* ice chart data. Vertical lines correspond to the valid time of the two forecasts that were analyzed in Sect. 4. Values along the vertical axes are in units of km.

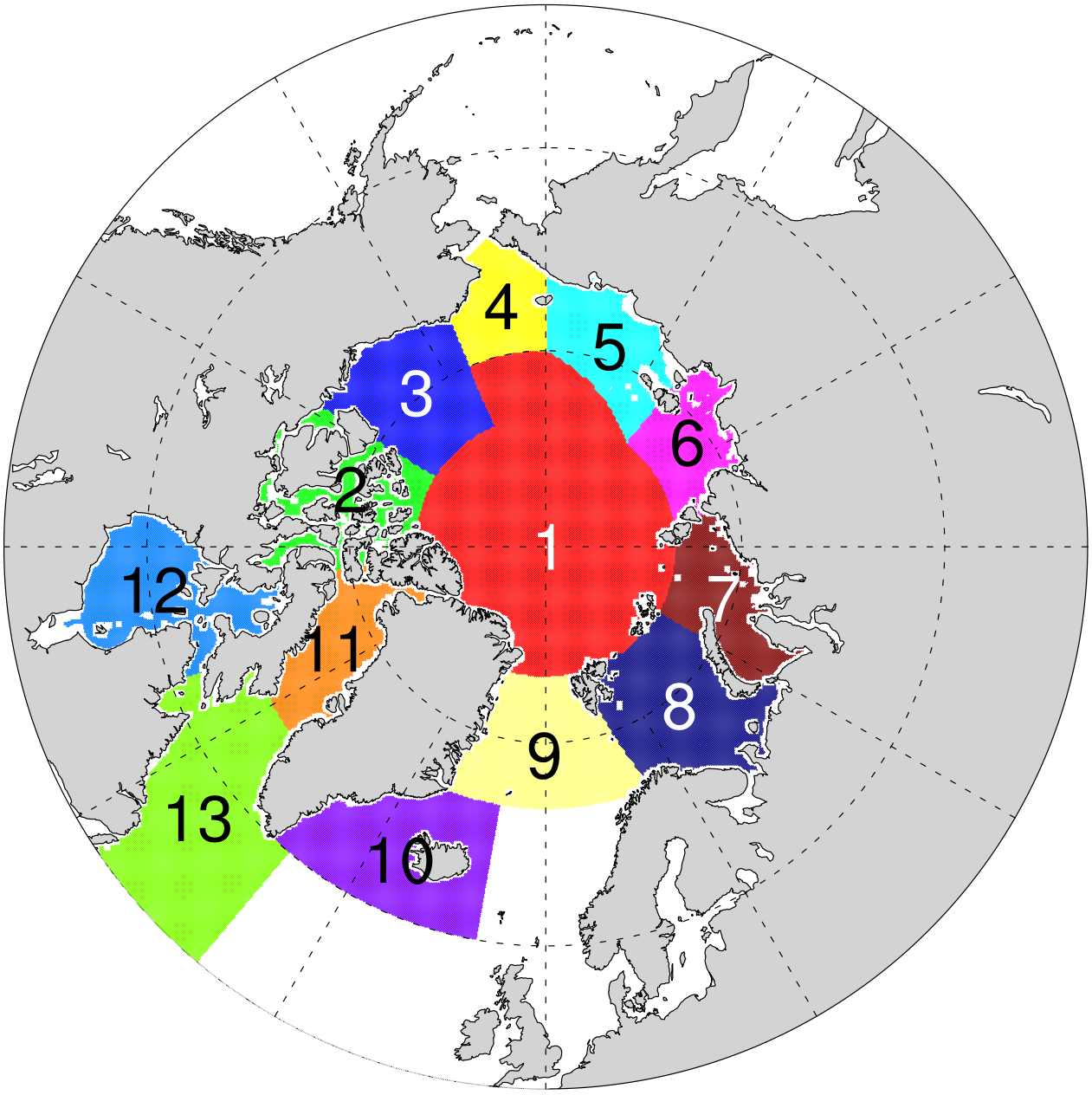

Figure S4. Arctic sub-regions as defined in GODAE OceanView. The numbered regions are (1) Arctic Deep Basin, (2) Queen Elisabeth Islands, (3) Beaufort Sea, (4) Chuckchi Sea, (5) Siberian Sea, (6) Laptev Sea, (7) Kara Sea, (8) Barents Sea, (9) Greenland Basin, (10) Southeast Greenland, (11) Baffin Bay, (12) Hudson Bay, and (13) Labrador Sea.