Revision CryoSphere-tc-2020-361
Author: A. Melson
Title: Edge displacement scores

Overall general comments:

This study introduces a new distance-based approach for assessing the quality of the representation of the displacement over time of the sea-ice edge. The methodology introduced is promising, however the article is yet not mature and need some major revisions for being acceptable for publication in the CryoSphere. Major revisions aim to improve the structure and exposition of the article, as well as some aspects of the verification method.

Major revisions:

1. **Section 2.1 needs major rewriting**. Since the author aims to maximize generalization, the description is too generic and therefore it is not clear (until you read section 3). A not exhaustive list of suggestions (for improvement) are:
    - State more explicitly and since the beginning (e.g. at lines 33, in the paragraph at lines 39-45, and then at line 47) that you aim to verify the displacement over time of the sea-ice edge, and compare the forecast versus observed displacements. (e.g. at line 47 the phrasing "the difference in maximum edge displacement between two products" can be mis-interpreted as the "distance between observed and forecasted ice-edges").
    - Do not use "product". You can use "dataset", "ice-edge" or other terms. You can as well refer to model (or forecast) and observation, explicitly.
    - The titles of Sections 2.1 and 2.2 should include "displacement" somewhere, e.g. "2.1 verification measures of the displacement of a single ice-edge" and "2.2. comparison of the displacements of forecast and observed ice-edges"
    - Lines 58-59: "$L^{(1)}$,$L^{(2)}$ denote the sea ice edges for two representations … " is not clear at all: here **you need to state explicitly here that (1) indicate the time t_0 and (2) indicate the time $t_0+\Delta t$, and that you measure the ice-edge displacement that occurred between the times t_0 and $t_0+\Delta t$** (for either the forecast or the observation ice-edge). In fact, I suggest to change all the notation here (otherwise the reader will naturally associate 1 for forecast and 2 to observations), replacing (1) with (t_0) and (2) with ($t_0+\Delta t$), and then $d_n^{2:1}$ can become $d_n^{\Delta t}$.
    - The mathematical equations (3) and (4) need to be defined in a more mathematical rigorous way. Especially equation (4), the sign is not accounted for in the mathematical formula and there is a missing absolute value $|.|$.
    - Lines 62-63: the text is not precise; you need to state that in Eq. (3) you consider the minimum of all distances (between the single point in the edge at time $t_0+\Delta t$ and all points of the edge at time t_0). Similarly, for Eq(4) you need to state that you consider the max (over all n) of the $d_n^{\Delta t}$

- Lines 65: I believe this should state " … will return the largest *absolute* positive value … "
- Line 66: what happens if the $d_n^{2:1}$ are partially positive and partially negatives? Mention the canceling errors.

2. Section 2.1, Lines 70-75: **Table 1 (as well as Table 2** in Section 3) lists the frequencies of the histograms of the set of distances $d_n^{2:1}$ defned by Eq. (3), for the idealized model and observation ice-edges of figure 1. This is not clear from the text at line 75, nor from the caption (which I believe has a mistake, since it should refer to Eq. 3, and not Eq. 4) and the heading of the table is wrong (it should state "Frequencies" rather than "fraction of grid cells"). Here are my suggestions:
    - **Representing a distribution with a table is quite unconventional: I strongly recommend plotting the histograms of the distances $\mathbf{d_n^{2:1}}$ instead**. (You can choose if to leave the table or eliminate it, but please add the histogram). Please show the histograms also for Table 2.
    - Please rephrase the caption ("category distribution of displacement distances" is not clear; this can simply be "distribution of ice-edge displacements").
    - Please rephrase lines 71-74: you can shrink it all as a couple of sentences such "The maximum distance in Eq. (4) provides a single measure to examine the ice-edge displacement. However, it can be more informative to analyse the whole distribution of the displacements $d_n^{2:1}$ defined by Eq. (3), rather than their maximum only. This can be done by analysing the histogram of the displacements (Figure HIST) and its corresponding frequencies (Table 1); the distribution of the displacements $d_n^{2:1}$ can as well be represented by their cumulative probability distribution (Figure 2)."
    - Please rephrase line 75: "the distribution of selected distance categories" is not clear.
3. The **decorrelation length**, used to sub-sample the ice-edge, is lightly mentioned at lines 78-80. It is thereafter used (e.g. caption of Figure 2, and then more heavily at lines 110-115), however a more thorough definition and how the author calculate Δn is missing in the article. Please add some text about it (maybe this can be done in an appendix).
4. Section 2.3 is too long, needs rewording, and need to be accompanied by a visual example. I strongly suggest to:
    a) summarize it in few sentences, such as in Melsom 2019, page 617, left column third paragraph ("A variant … "). In the re-wording it is important that you still retain the explanation at lines 138-139, where you correctly state that you expect the distances to be smaller when adding ocean open boundaries and coastal lines (because when adding these artificial "fixed" edges you automatically include in your verification some perfectly matched edges, aka trivial skill).
    b) **Include subsection 2.3 in Section 3**, after the data description and prior the results (aka after line 173 and before line 174). You can actually split section 3 in three subsections: 3a Sea-ice data description, 3b open boundaries and coasts, 3c verification results. In this fashion the reader has an immediate visual

example (figure 3) on your need of adding these fictitious boundaries (especially when considering an Arctic sub-region).

c) The example presented at lines 180-187 and illustrated in figure 4 is excellent!

5. Give a new title to Section 3: Application of the new verification method on sea-ice forecast (this is not a "case study" … also at line 211).
6. Lines 115-119: **rather than considering the ranks**, and a variable numbers of bins, I suggest using the **quantiles**, which have a fix range in [0,1] (or which range between 0 and 100, if you consider centiles): in this fashion it is more natural to aggregate and compare (you avoid any issue related to the variable binning).
7. Lines 120-126: you need to state here that the larger is the quantile (or rank), the better the geographical correspondence between maximum observed displacement and max (or at least large) forecast displacement.
8. Lines 199-206: the expected histogram for a random process would be a flat histogram (each bin is equiprobable), and you could compare your histogram to a flat one as described in Wilks (2019). I do however suggest to simply describe visually the intuitive result: your histogram in figure 5 shows a mode for the highest rank, which shows some skill in detecting the location of the maximum displacement. Rephrase (and join) these paragraphs. The reference is:
    - Wilks, D, 2019: Indices of Rank Histogram Flatness and Their Sampling Properties. *Mon Wea Rev*, **147**: 763-769. https://doi.org/10.1175/MWR-D-18-0369.1

Minor (technical) revisions:

1. Lines 3 (and line 5): I suggest replacing "expansion" with "decline" and replacing "advancing" with "retreat". This is because in the first few sentences the authors relate with climate change, so it sounds a bit counter-intuitive to talk about sea-ice expansion (after reading the article it is clear that the methodology applies to both spring melting and autumn freeze-up, but in the abstract and introduction -if you maintain the climate flavour- you might prefer focussing on "decline")
2. Lines 28-32: citing literature from the verification community is welcome! You could add to Ebert and McBride (2000) also Davis et al (2006a,b), since MODE is now the most commonly used object-oriented verification method in the weather community. You could as well add that both these methods were designed for (several) precipitation-like (convex-shaped) features, and that there is no equivalent for a single linear feature such as ice-edge.
    a. Davis, C. A., B. G. Brown, and R. G. Bullock, 2006: Object-based verification of precipitation forecasts, Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.* **134**, 1772 - 1784, doi: 10.1175/MWR3145.1.
    b. Davis, C. A., B. G. Brown, and R. G. Bullock, 2006: Object-based verification of precipitation forecasts, Part II: Application to convective rain systems. *Mon. Wea. Rev.* **134**, 1785 - 1795, doi: 10.1175/MWR3146.1

3. End of line 33: please be specific in stating that "We begin this study by presenting a new algorithm for assessing the quality of representation of *the displacement over time* of the sea-ice edge … "
4. Line 34-36: please rephrase these two sentences to align with the suggestion of major revision 4b.
5. Line 39: replace "some idealized distributions are …" with "an idealized ice-edge is"
6. Line 49 (and 70): I am sure the Hausdorff distance was introduced earlier than Dukhovskoy et al (2015), can you please provide the original reference?
7. Lines 78-83: you can join these two sentences in a single paragraph. You should replace "when time serie results are examined" with "when results aggregated over multiple cases over an extended time period", or something similar (the point is that you aggregate multiple cases, and not that you consider a time serie)
8. Lines 85-87: rephrase these two sentences to twin them with the beginning of Section 2.1, something like "In the previous section we focussed on measures which describe the displacement of a single (forecast or observed) ice-edge. In this section we extend these to assess the differences in the displacements of the forecast versus observed ice-edge."
9. Rephrase lines 88-89.
10. In the caption of Figure 2, lines 3-4, replace "the mean separation distance difference" with "A measure of the overall displacement difference" (be careful, that grey-shaded integral <u>is not</u> the mean displacement difference). The idea of using the area between the two curves (the grey shaded area) is excellent!
11. Line 92: replace "property" with "attribute" (verification term).
12. Line 93: I suggest writing "a simple measure which provide this type of information is …"
13. Line 99: replace "site-specific" with "local".
14. Lines 99-100: write "… of the model ice edge in proximity of the maximum displacement found in the observations".
15. Line 110: "In order <u>*to*</u> examine … "
16. The original reference for the rank (Talagrand) histogram is Talagrand et al (1999) and Anderson (1996)
    a. Talagrand, O., R. Vautard, and B. Strauss, 1999: Evaluation of probabilistic prediction systems. *Proc. Workshop on Predict- ability,* Reading, United Kingdom, ECMWF, 1–25.
    b. Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate,* **9,** 1518–1530.
17. Line 171: change "integrated" to "interpolated" or "upscaled" (I imagine it is a mass-conservative up-scaling from the ice charts at 1km resolution to the SVIM 4km resolution domain). Is the smoothing (the second order checkboard suppression mentioned at line 161) performed on the observations too?
18. Line 172: what is meant here with "dry"?
19. Line 180: Eliminate "Category" and write "The distribution frequencies in Table 2 change only moderately when including open ocean boundaries and coastal lines."
20. Line 185: replace "unreasonable results" with "mis-matches ice-edge points".

21. Rephrase lines 188-189: I suppose you consider a fix number of nine ranks (and not nine randomly chosen points of the ice-edge).
22. Lines 193-198: please consider using quantiles, rather than ranks, as suggested in major comment 6.
23. The first paragraph of the conclusion need polishing / rephrasing.