Cryosphere

Response to reviewers

~~Defining Coastal Antarctic polynyas in satellite observations and climate model output to support ecological climate change research~~

Challenges in identifying Antarctic coastal polynyas in satellite observations and climate model output to support ecological climate change research

Laura Landrum[1], Alice K. DuVivier[1], Marika M. Holland[1], Kristen Krumhardt[1], Zephyr Sylvester[2]

[1] NSF National Center for Atmospheric Research, Boulder, CO, USA

[2] INSTAAR, University of Colorado, Boulder, CO, USA

Thank you to the reviewers for considerate reading and helpful reviews of our revised manuscript. We have gone over the reviews, and in response are submitting a major revision of the manuscript. The primary – and largest – changes came after reviewing the many comments that the previous version was too long, and could easily be shortened, clarified and possibly separated into 2-3 separate papers. We have shortened the paper considerably by focusing on polynyas identified in the two observational products and the hindcast model simulation and leave studies involving the fully coupled CESM2 to future work. This has enabled us on the current climatology, how the observations compare to each other and the model output. We have also changed the title such to better reflect the somewhat more narrow scope of this revised manuscript.

The second major change in this version is that we realized that we had not adequately defined polynyas or what we are identifying in the model using our polynya algorithm. Many (perhaps even most) coastal polynyas are of a size and shape that are not resolved by the 25 km X 25 km EASE grids of the NOAA and OSISAF satellite-based sea ice concentration (SIC) products, much less by the relatively coarser (typical) model grids of Earth System Models (ESMs). As reviewer #4 highlights, this leads to understandable confusion. Therefore we highlight how polynyas strictly defined by the manner of the World Meteorological Organization definitions of terms will not be resolved on these grids, but rather our polynya identification algorithm more accurately defines grid cells, or groupings of grid-cells, that have polynya-like features in that they are of lower SIC or thinner sea ice thicknesses (SIT) than surrounding cells. As a related result, ”polynya areas” are not truly areal sizes of individual polynyas (as sub-grid cell information regarding amount of open water in one continuous ”polynya” or multiple, smaller ”polynyas” or fraction of open water in leads vs polynyas, etc. is missing) but rather an areal size of the grid cells that have been labeled polynyas and will, by definition, be larger than the actual areas of strictly-defined polynyas. This difference in definition is stated in the introduction and helps to frame the work.

A third major change is that we followed the suggestion of reviewer #2 and changed our analysis from monthly to seasonal, with a focus on two seasons – namely winter (June-July-August; JJA) and spring (September-October-November; SON) due to the importance of these seasons on coastal polynya formation and subsequent impacts on Antarctic marine NPP in the sea ice zone.

Below we highlight further revisions based on reviewer feedback showing *reviewer comments* and our response.

Referee #2: Tarkan Bilge, tarlge@bas.ac.uk

Suggestions for revision or reasons for rejection

I would like to thank the authors for their detailed response to my previous review. Reiterating my comments from the beginning of my last review, I see this manuscript as providing an original and important contribution to the field in the form of a detailed analysis of polynya identification methodology.

Thank you again for your overall support and helpful review of this work.

Below I have included a couple of issues to bring to the authors' attention, given the revised manuscript.

In comparing different grids, temporal resolutions, products and thresholds across different regions and months, the authors have presented a very thorough analysis of polynya identification. In my opinion, given that the manuscript has expanded to contain another observational product, it would more concise and impactful if the results were presented seasonally rather than monthly where possible, and results from different regions were only displayed where the authors are making a point which connects a difference in two regions to a physical mechanism. For the remainder of cases where the reasons for specific regional discrepancies are not discussed, I think it is sufficient to have these regional figures in the Supplemental. Ultimately, the results for individual months and regions do not appear in the manuscript conclusions beyond the broad finding that there is a regional and seasonal sensitivity to polynya identification (which is an important finding but does not require detailed monthly plots in the main manuscript to substantiate).

Thank you for this suggestion. We have revised the manuscript and focus on seasonal analysis rather than monthly, as well as removed several of the regional figures (now also new) to the supplemental information.

Making these changes would have the following benefits: 1) The figures would display less data, making them easier to view, and still providing seasonal information (e.g. SON instead of S, O, N). 2) The manuscript text would be more concise, and avoid important results being obscured by months in which there are exceptions to those results, e.g. L408-419. 3) The authors would have to perform fewer statistical tests. The way the significance tests are currently set up, having monthly resolution for each individual region, means there are so many individual significance tests that we would expect to see quite a few false positives. For example, in Figure 5 alone there are 240 statistical tests at the 95% level, one might expect 12 false positives to appear in this plot if no multiple-hypothesis significance methods are employed. Particularly if the authors are then picking out individual test results for discussion, I am not confident that this is a robust way of performing this analysis.

Agreed! This has been very helpful for reducing the number of statistical tests, and presenting results.

On this first point, I do recognise the importance of monthly resolution on e.g. Figures 9 and 11 for discussion surrounding the shift in NPP seasonality as an exception.

Secondly, from my understanding, there is very little correlation between satellite and model polynya variability in the temporal dimension (e.g. Figure 5e), and the large ensemble seems to even disagree on the sign of the trend during the historical period (e.g. Figure 10a,b). I think this uncertainty should be reflected in Section 5.3. Perhaps this caveat should be mentioned near L745 or L820. I think it is difficult to talk about the changing role of polynyas on NPP when it seems to me that there is not much confidence on whether polynya area will increase or decrease in the future from model runs.

This is a very good point. We have removed the discussion of the CESM2 fully-coupled results, both historical and future. We have also clearly stated that the lack of correspondence in trends and temporal variability in identified polynya-like features between not only the model and the observations, but also between the two observational products are remarkable. In particular, the magnitude, significance and even sign of the trends (in the integrated time series) can differ in the observation products by retrieval algorithm (NOAA vs OSISAF), grid (EASE vs CESM; daily vs monthly) and SIC threshold value!

In connection to this above point, (but not a revision suggestion as such) have the authors considered splitting this work into two separate manuscripts? It seems to me that there is great value in a study on polynya identification across different products, grids, variables and thresholds even without the ecological application suggested in the manuscript. In my opinion the authors have made a valuable contribution in the form of the polynya identification analysis, but the narrative surrounding the NPP in relation to this does not seem as strong. If in a separate manuscript, the authors considered another future emissions scenario (or set of large ensemble runs), I think they could create a stronger narrative about what will drive NPP changes in the future, and I think that would merit its own dedicated study.

Yes, agreed. We have split the work and removed future scenario changes (and indeed fully coupled CESM2 analysis) and left that for future work. We have kept some of the NPP analysis in this current revision however as we feel it is an important indicator that the ”polynya-like features” identified in the model are functioning, in the marine ecosystemin a manner similar to what we would expect from observational studies – namely as ”hotspots” of marine NPP compared to the rest of the sea ice zone.

Many thanks to the authors for their considerable time spent revising the manuscript since the last round of reviews, ultimately I find this to be a very worthwhile study.

Report #2

Anonymous referee #1

Suggestions for revision or reasons for rejection

My main concerns with the first version of the draft were that the authors were doing too much, trying too hard to make the model and observations match, and the objectives were unclear. The objectives have been clarified, and the structure has somewhat improved, but the manuscript is still very unequal, especially now that in response to a comment from the other reviewer the authors chose to do even more. My comments here mainly focus on the remaining issue that JRA-CESM is biased – but it is ok, all models are.

The majority of the result section tries to convince us that JRA-CESM has an accurate representation of coastal polynyas, in order to then briefly use it to investigate present and future NPP in polynyas. To reach this conclusion, the authors have to use an unrealistically high threshold for polynya detections in observations.

Comment 1: You argue that you chose this threshold so that your 1979-2020 climatology reaches the same values as Arrigo and van Dijken (2003), Tamura et al. (2008), and Nihashi and Oshima (2015). But these three references computed polynya areas over very different time periods from each other and than you did:

- Arrigo and van Dijken was 1997 – 2002

- Tamura et al., 1992 – 2005

- Nihashi and Ohshima, 2003 – 2011

So, what exactly did you compare?

The choice of threshold to use is somewhat arbitrary, as there is no clear, ideal choice. Larger thresholds will result in larger polynya areas, and usually larger number of polynyas, and vice versa. We've added text to help clarify the ambiguous nature of threshold choice, and we picked the ones we did (line 407-413):

> The thresholds chosen are based on results from the range of possible thresholds and represent a compromise with regards to number of polynyas and total integrated southern hemisphere (SH) polynya area. We aim for a total number of polynyas within the observational range, winter-spring, for sea ice production (13; Nihashi & Ohshima, 2015; Tamura et al., 2008) and high marine productivity (37-46; Arrigo & van Dijken, 2004; Arrigo et al., 2015). The total area of polynyas in our method represents the area of grid cells labelled as containing polynya-like features, and therefore we pick thresholds that lead to polynya-gridcell-areas at least as large if not larger than the lower ranges of published observational estimates for these seasons (~$1x10^5 km^2$; e.g. Arrigo & van Dijken, 2004; Arrigo & van Dijken, 2004; Tamura et al., 2008; Nihashi & Ohshima, 2015).

We also present many of our results across a range of thresholds, with conclusions that do not depend on actual threshold values. The thresholds we pick might indeed be unrealistic if we were able to resolve individual polynyas – which we are not and cannot on grids that are often larger than the dimensions of polynyas themselves. In this sense, a grid cell labeled a "polynya" is more accurately described as containing enough open water (or thin sea ice) compared to surrounding cells that it contains a polynya-like feature (e.g. see Appendix B). This is one of the many complications of identifying and comparing "polynyas" from gridded products that may not truly resolve individual polynyas per se but rather grid cells that may contain polynyas based on the threshold used.

Comment 2: Arrigo and van Dijken (2003) came to their value using at 10% sea ice concentration and Nihashi and Ohshima (2015) a 20 cm sea ice thickness thresholds – Tamura et al. (2008) used heat fluxes and can be ignored in this comment. Yet you have to use 85% SIC and 40 cm SIT to reproduce their results. How do you explain this discrepancy?

Thank you for pointing out that we did not communicate clearly about metric choices, and the somewhat arbitrary nature of them, as well as how grid resolutions will impact the size of possible polynyas identified.

These different thresholds are for different grid resolutions. The 10% threshold that Arrigo and van Dijken used was for 6.25 km subpixels which were labeled "ice" or "water" based on the 10% SIC threshold. If, for example, one wants to allow a body of water of 35 km^2 in size to be considered a "polynya" (0.9*6.25*6.25 km^2) this would require a 91% SIC threshold on a grid cell of 3.2 x103 km^2 (roughly the size of many coastal grid cells in the CESM). Thus, from a size of an individual polynya, a threshold of 85% is actually somewhat lower than a 10% threshold on the grid size used by Arrigo & van Dijken. Likewise, the work of Nihashi & Ohshima used the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) data – which is a 6.25 km grid. We used grid cell averaged sea ice thickness – thus our threshold of 0.4 cm thick is for an averaged thickness of 0.4 m over grid cells of ~$3.2 \times 10^3 km^2$ – a much larger area. Our threshold is indeed different, but it is not necessary contradictory. This is one of the complications we hope comes across in identifying polynyas on grids, particularly coastal polynyas that are frequently smaller and thus not resolvable on 25x25 km EASE grids much less nominal 1° model grids.

Comment 3: My comment about the flow of the manuscript remains for the beginning of the results section. You use these fixed sea ice thresholds on Figure 2, but only explain where all these thresholds come from, especially so for the modelled data, on Figure 3. At least include some cross-reference, or change the order of the subsections.

Comment 4: What figure 3 really shows about JRA-CESM is
- a large discrepancy between the SIC and SIT estimates;
- using SIC, a dramatic underestimation of the polynyas in winter and overestimation in spring, regardless of the thresholds;
- using SIT, a large overestimation in both seasons.
You find a work around by changing the thresholds for the model so that the total SH-areas

match, but then figure 6 further shows that the outcome is inconsistent biases in the different subregions.
As per my previous review, my main issue with your manuscript is the framing of these findings. It is ok to use the model you have access to, regardless of its imperfections, as long as the biases are acknowledged – which they are here. So I would still recommend shortening the first 28 pages of the manuscript and spend less time trying to convince us that the discrepancies solely come from bad diagnostic choices. In particular, I would remove the temporal correlation analyses.

To summarise: better justify your very high thresholds, and consider shortening / reframing up till section 4.4

Thank you for pointing out how awkward our previous submission was. We have reframed the manuscript as well as narrowed the scope – and as a result it is significantly shorter, with fewer figures, and hopefully clearer text.

Report #3

Submitted on 30 Jul 2025
Anonymous referee #4

Suggestions for revision or reasons for rejection

Review of

Summary: This is a complex paper with the aim to cover quite some ground in our understanding and our capabilities to detect and map Antarctic polynyas - both in observations and in results from numerical modeling - with the overarching goal to arrive at an improved quantitative knowledge about how marine net primary production (NPP) is connected to the circum-Antarctic polynya area now and in projections of the future climate conditions. Besides the many objectives of this contribution I distilled that the authors want to come up with recommendations about how to reliably detect polynyas as a sub-grid scale phenomenon in coarse resolution climate models. The authors chose a simple sea-ice concentration threshold based approach to delineate polynyas both in observational and modeling data. They investigate the influence of different grid resolution and of different temporal sampling and come up with a rich set of results from correlation and trend analyses, partly only based on observational data, partly mixing both worlds.

General Comments:
GC0: I have one (first) general comment on this manuscript. I am hesitant to think that the way the study has conceptualized is delivering the results the authors may have wanted to achieve. I understand that the authors wanted to use satellite observations to test how well ESM models delineate polynyas and how accurate the (integrated) polynya area is compared to observations. One focus of the investigation was on the metrics of how to delineate polynyas in the model grid cells. With the then selected and/or optimized metric you want to get reliable polynya area estimates from the ESM data and transfer that information to estimates of the NPP - now and in the future. The conceptual drawback I see is that the paper does not convince with the generation of a well-analysed, well-evaluated polynya location, occurrence and area data set based on observations. The threshold based method used is a poor-man's approach to identify polynyas. While I understand that for ESM analysis it possibly has to be that approach, this is not the state-of-the-art using satellite observations. Hence, the estimates that are based on observations and are presented here are not fit-for-purpose. The "processing chain" could have been (if you need to stick to the thresholds approach): Select a number of key polynyas around Antarctica. Get high-resolution estimates of their area for a reasonable time period. Apply different SIC thresholds to ONE SIC product (You could use OSI SAF because it is a state-of-the-art algorithm / product with physically based uncertainty estimates and the possibility to go down to SIC = 0% if need be; but it might also be a good solution to seek for SIC products at considerably finer grid resolution, e.g. ASI-AMSR2, offering 3.125 km grid resolution.) and test with which threshold you achieve the best result in comparison to the independent data. Generate data sets (maps + time series) of the polynya location, occurrence and area based on SIC observations and the selected threshold. Get an estimate of the uncertainty by playing around with the SIC threshold (+/-10% and +/- 20% for

instance). Upscale the maps obtained to the ESM grid resolution and use these as the "truth". Use daily ESM SIC data and apply different SIC (and SIT) thresholds to develop an approach of how to best identify polynyas in the ESMs. Again develop a feeling how sensitive the obtained polynya area is to changes in the applied SIC threshold. Once you have that you could - if need be - generate monthly data sets of polynya location / area based on the obervations and based on the ESM data. But, if I am honest, if the model allows daily output I would also use it - last but not least because of your overarching goal to look into NPP dynamics which might be dominated a lot by shortlived (a few days) blooming events which you would lose in the monthly context.

Thank you for your general comments which have helped us recognize how we needed to reframe the text, define polynyas or rather “polynya-like features” and better communicate our results. We have significantly reduced the manuscript so that we can better explore our primary goal of highlighting challenges in identifying and comparing climatologically and biologically relevant Antarctic coastal polynya-like features in gridded data, both satellite and model. While a threshold-based method is indeed a “poor-man’s approach” to identifying polynyas, it is a method than can be applied to gridded data that neither resolves polynyas (due to size) nor contains sub-grid scale information necessary for more complex methods. We could indeed use more sophisticated methods and sub-pixel satellite information to better identify strictly defined (per the WMO definition) individual polynyas and use these individual polynyas to create maps to use as a “ground truth”.

We chose not to use this approach primarily because it further complicates comparisons between observations and models in that polynyas will be identified using different methods, and we felt would be better addressed in a separate study that would take this approach with the observational data itself – in other words, to identify strict polynyas from high resolution data products, create maps of these, and compare these maps to maps of polynya-like-features identified in the observational products on the original EASE grid as well as regridded to coarser grids, and thus quantify how polynya information (area, location, variability, number, etc.) may change given different methods applied to different observational products. We also feel this is beyond the scope of this work.

Earth system models are being used with increasing frequency to study climatological relevant mechanisms involved in deep water formation, bottom water circulation, and marine ecosystems in ice dominated regions (e.g. Duffy et al., 2024; DuVivier et al., 2023; Mohrman et al., 2021; Jeong et al., 2023) – which in Antarctica involves studying polynyas. Published studies use different thresholds for identifying coastal and open water polynyas from model output, and for comparing this output to polynyas identified using the same method from observational products, without fully exploring the many complications involved including grid resolutions, differing definitions of sea ice concentrations and thickness and methods for calculating them, threshold choices, season and region of interest. We anticipate that more studies will be forthcoming exploring polynyas in the climate system, and thus we believe it is important to demonstrate many of the challenges involved as a baseline for future work.

Both the NOAA and OSISAF products are used extensively by multiple climate modeling centers for model validation, etc., and in all likelihood these products will continue to serve in this regard. Furthermore, most research that we are aware of do not use both of the products but rather choose one. One of our main conclusions is that the choice of observational product can influence the result, and feel one of the strengths of our submission is that it includes two broadly recognized, accessible and utilized observational products.

GC1: What are the different methods to delineate polynyas in satellite remote sensing data? What do they have in common? Where are these different from each other? How have there results been evaluated? What are the different SIC thresholds that have been used during the past decades to define polynyas? What are the conclusions of all this for your work? What is your definition of a polynya (used in this manuscript) and how (if so) does it differ from the definition given in the WMO sea ice nomenclature? To give an answer to these question seem a pre-requisite to me to use the output of polynya "retrieval" algorithms applied to satellite remote sensing data to kind of guide which method to use to define polynyas in ESM data.

These general comment was very helpful in pointing out omissions we made, particularly in setting the stage for this work, including defining polynyas. We have greatly revised the introduction and added definitions for polynyas, as well as referenced previously published work with remote sensing data to identify Antarctic polynyas to add context to this submission.

GC2: Please motivate better your choice of the SIC products used - especially with regard to spatial resolution and with regard to the need to operate with a long time series. Please also take into account that the two SIC products you used are provided on different grids. I did not find any notion about the NSIDC polarstereographic grid of the NOAA/NSIDC SIC CDR and how your treated that in comparison to the EASE grid. Also into this topic falls my impression that your manuscript would benefit from a more careful description of the skills and limitations of passive microwave SIC retrieval algorithms to derive SIC values for low sea ice concentration and thin sea ice. See my specific comments in this regard.

The NOAA/NSIDC CDR data that we downloaded for the southern hemisphere was on a 25x25 km EASE grid, it was not on a polar stereographic grid – which is why the polar stereographic grid was not mentioned.

GC3: One aspect which I thought would be high up on the agenda of this contribution but which I could not find treated in a convincing manner is the optimization of how to delinear polynyas in global climate models. I may repeat a bit of what I stated in GC0 here ... but what I am missing is the description of why the authors ended up at the SIC and SIT values chosen; while I acknowledge that there is an attempt to fit thresholds used in the modeling world to those used in the observations I have problems to find clearly stated information about the sensitivity of the approach, how accurate polynya area derived the method chosen is, and how and with which independent observations you carried out your evaluation. I am aware that a number of your figures point into that direction but I have problems to distill quantitative information out of these and I would not be able to repeat your approach to check its validity or to ask one of my students to repeat the study using a different observational SIC product and a different GCM.

We have added text and panels to figures to highlight a range of results in not only polynya areas, but also numbers and trends.

We have responded to many of the specific comments, however the significant re-framing, restructuring and major modifications of this revision make it substantially different. As such, we are not entirely sure when some comments are still relevant considering the modifications we have made, and apologize in advance if by chance we have not responded sufficiently to any comments that you still consider important to address.

Specific comments:
L13: "no significant impact ..." --> This is an interesting result, suggesting that at daily scale, variations in the overall area of coastal polynyas cancel out? I am not so sure I believe this is true.

This has been removed from the abstract, which is also substantially different in this revision.

L31-33: "Open water ... more sea ice." --> Three comments here. You could already now mention on which type of polynya you are focussing in your contribution. You could mention whether some of the polynyas are perhaps of mixed type where both upwelling and katabatic winds play a role. You could come up with 1-2 more explaining sentences about persistence of coastal polynyas and their role in deep water modification (in more detail and more concrete then you have done so far), and also about the fact that especially in the Eastern Antarctic there is a very close interplay between land-fast sea ice and icebergs on the one hand and the generation of coastal polynyas on the other hand. There exist quite some publications into this direction which I invite you to find and cite.

Thanks for point this out. We have added a couple sentences to this section, and also a sentence in the discussion landfast ice will not influence locations of coastal polynyas in the CESM, as the model (and most ESMs) does not simulate land fast ice.

L45-47: "Defining polynyas ... functioning." --> I agree on the polynya functioning part here but with respect to the "general" definition, I invite you to take a look at the WMO sea ice nomenclature where you can find a definition of a polynya. I also recommend to cite this nomenclature here. I note however now, after having read you paper completely, that I am not sure that you developed your own definition of what a polynya is.

We have cited the WMO definition of polynya and added text describing our definition of polynyas and how that is different than the WMO definition

.
L65-71: "Furthermore ... 2015)" --> These sentences about sea ice concentration retrieval algorithms and products based on satellite observations are not sufficiently well put into context with polynyas. There are 2 - 3 methods to define polynyas / infer the polynya area from satellite remote sensing data and I would have thought it is more straightforward to first introduce those before eventually moving on to sea ice concentration products.

This revision has been reframed and we have added text explaining why we use the threshold method for this work.

L71-82: Also the remainder of this paragraph is still detached from the main topic of this contribution: polynyas. In this part there are several statements with respect to satellite remote sensing (products) for sea ice concentration and sea ice thickness that do not necessarily hold the way written. I don't "buy" the general statement that algorithms using passive microwave data generally underestimate sea ice concentrations at low concentration levels. What is correct is that the uncertainty of the retrieval is considerably larger due to the reasons laid out by you already and thanks to an often elevated fraction of thin ice and/or quite diverse ice types which both complicate the retrieval.
"to very thin thicknesses" needs to be quantified. I am also not sure whether this general statement holds because I would think ESMs use sea ice thickness categories and might also provide a grid cell mean sea ice thickness rather than the thickness of the ice that is actually present with that grid cell.

Yes, you are correct. We have removed *"to very thin thicknesses"* and have added text explaining more carefully SIT from ESMs.

Your references for SIT retrieval focus on the Arctic. It might be good to learn something specific for the Antarctic in this regard.

We have added Antarctic specific references.

The statement that SIT from satellite products has limited capacity for identifying thin sea ice is not correct. First of all there is the SMOS sensor which has been used successfully in the Arctic but recently also in the Antarctic to derive SIT of thin sea ice. Secondly, there are at least 2 approaches to use either thermal infrared or passive microwave satellite remote sensing to compute the sea ice thickness - especially in polynyas.

Satellite products of the SIC are typically daily or twice daily, that is true, but again, when it comes to polynyas there are a few studies that use swath data, so data from single overpasses, to delineate polynyas and hence get an improved and more accurate estimate of the actual polynya area change during the day. Polynyas are quite dynamic and can open and close during substantially shorter time spans than a day.

Yes, there are alternative methods of retrieving SIT and also delineating polynyas from swath data. We focus on comparisons with gridded SIC products so that we can apply the same method to both observational and model data, have a long time record, and give context for making observational and model comparisons. We have re-framed the paper to [hopefully] make this more clear.

L84: The information given in this paragraph needs to come before you go into details about sea ice concentration and sea ice thickness products so the reader understands that "Ahaaaa ... people have used sea ice concentration data or thickness data to detect polynyas and to compute the polynya area." I invite you to search for contributions mentioning the Polynya

Signature Simulation Method (PSSM), for contributions where polynyas are identified using SAR (e.g. COSMO-Skymed), and for contributions by Nihashi and Oshima // Fraser et al. // Nakata et al. to learn more possibilities to infer polynya area and polynya dynamics from satellite observations.

We have referenced previous work using observational products and PSSM and related work to infer polynyas.

L109: My immediate take on these five "primary" goals which are followed by a "we also seek to ..." is that this is far too much to be dealt with within one contribution. The sole intercomparison of the various existing polynya identification and area retrieval methods based on satellite observations is a topic which would deserve ONE paper on its own.
I cannot speak about the ESM world but I would assume that the diversity of approaches to define a polynya in ESMs is much smaller and is mainly a function of model resolution and time step as well as whether and how polynyas are parameterized. The spatial and temporal resolution of the wind forcing - driven by the resolution of the topography - plays a role, the representation of land-fast sea ice and, of course, how realistically the ESMs represent the oceanic heat flux and/or ice shelf cavity processes that modify water masses near ice shelf fronts. Putting all this on the table I am not convinced that this contribution can come up with a credible "optimal polynya metrics".

The 2nd primary goal also reads like it would deserve an own paper.

The 3rd primary goal reads like an analysis of ESM data ... hence it is not entirely clear why you need all the exercises related to the satellite observations and the metrics (i.e. goals 1 and 2) to come up with this part of your contribution.
Finally, the 4th and 5th goal seem to be linked more closely to goal 2 and seem not to be related to either observations or biogeochemistry.

So, in short, looking at the goals I see 3 papers here. One targets goal 1 (but possibly without the model component); one targets goal 3; and the final one targets goals 2, 4 and 5 and the part how do I define polynyas in ESMs.

We have reduced the paper significantly, and now have four primary questions that we address:

1. Are coupled earth system models (ESMs) capable of capturing climatologically and biologically relevant coastal polynya-like features?
2. How to best use observational products to validate coupled earth system model simulations of polynya-like features?
3. Do these polynya-like features in ESMs occur in regions/areas where polynyas and polynya-like features are identified in observations?

4. Do coastal polynya features in ESMs function in marine ecological processes as we believe they do in the real world?

L110: Given the multiple possibilities that exist to delineate polynyas in satellite remote sensing data that - mostly if not all - work without "time series" I am not sure I understand the incentive to look into time series data. This needs to be motivated better.

L135 / section 2.1: Okay, so you only focus on using a sea ice concentration threshold as the metrics to "define" what a polynya is. This does not result sufficiently well from your introduction / motivation.

I was wondering whether there are more recent studies dealing with an intercomparison between the NOAA/NSIDC sea ice concentration CDR and the one from OSI SAF. I am sure there are and I am sure in those you might find useful information about the limitations of the two products you used.

I invite you to clearly spell out which product version you are using. I assume OSI SAF is OSI-450a / OSI-430a, i.e. the CDR and the iCDR, but I am not sure whether you used v3 or v4 of the NOAA/NSIDC CDR.

Both versions are are now spelled out.

I invite you to also clearly motivate why you are using SIC products that come at a comparably coarse grid resolution. As far as I have understood you (from your introduction), the main investigation about long-term trends will be based on model data and you will only use the observations to define an improved metrics to define polynyas across ESMs. Therefore, I was wondering whether it would not be considerably better to look into sea ice concentration products that come at a finer grid resolution, i.e. at 12.5 km grid resolution based on AMSR-E / AMSR2 or at 3.125 km grid resolution based on AMSR2. This is not made sufficiently clear.

L152/153: The OSI SAF SIC product is indeed provided on an EASE 2.0 grid. The NOAA/NSIDC CDR is not - neither v3 nor v4 nor v5; these are all on a polarstereographic grid so that whenever you compute an area in squarekilometers you need to take into account the actual size of the grid cells which changes with latitude.

L183-184: "compare well" --> I was wondering whether you would like to specify this a bit better. I was also wondering whether correct representation of the SIE is really such a good measure when it comes to assessing the skills of a model which sea ice concentration is subsequently used together with a threshold approach to delineate polynyas. What are the differences in SIC for the different SIC categories - especially around the threshold that I assume will be used and introduced later in this paper?

L243/244: "Passive microwave ... 10%" --> This statement is not true as it stands. What is

true is that at lower sea ice concentrations the uncertainty budget increases but nevertheless the emissivity difference between open water and sea ice still allows adequate SIC retrieval. There have been publications about this by Soren Andersen et al. in 2006 and/or 2007, I guess. So, this cutoff that is applied by NOAA/NSIDC is a simple approach to discard those SIC values that might have a larger error budget. And one can argue whether one should use 10%, 15%, 30% or whatsoever ... Also, most algorithms apply special filters to filter out grid cells affected by weather - based on the gradient ratio of certain brightness temperatures; these filters cause low sea ice concentration values to be filtered out as well.
If you want to better understand this issue then I recommend you to read the Lavergne et al. (2019) paper in a bit more detail. There you will also find the information that the OSI SAF product actually contains the raw, unfiltered SIC estimates which allow a user to utilize SIC values down to 0%. You will also find information about physically-based (by the retrieval settings) and mathematical (by the gridding process) uncertainties of the OSI SAF SIC product you are using.

L274: How about sea ice that is even thinner than 5cm?

L251-253: I can follow your discussion about that SIT might be a better measure to "define" a polynya. But to do that and to further argue about the pros and cons you first need to come up with a clear definition of the polynya area. Is it i) only the area of open water? Well, then under very cold conditions the polynya is very small - because of the freezing. Or is it ii) the area of open water and frazil / grease ice and the thin ice downstream of the polynya up to a thickness of x cm? In that case you need to define x and depending on that definition a polynya might be of substantial size even under very cold conditions (at least for a day). Or is it iii) somewhere between i) and ii), following the polynya model of Pease et al., defining the polynya area via the so-called frazil ice collection depth? To my opinion, the action item for you that is coming out of these lines is to answer the question: What is your definition of a polynya? And why?

L258-266: I don't understand this step and I suggest not to do it. I would not modify any of the input data sets for your inter-comparison beyond regridding or the like. But applying the thresholds and filters that are used on the observational side to the model side does not sound overly credible to me - for the reasons I laid out earlier in my comments.

Also, you might want to bear in mind that the results shown in Ivanova et al. 2015 with respect to thin ice are based on data of the SMOS sensor; the thin ice values used in that paper were computed from SMOS data. This computation has its limitations in case the sea ice concentration is not near 100%. Also, the sea ice conditions that are covered by that investigation do not reflect the conditions in a polynya. There one has frazil and grease ice, small pancakes and then all sorts of young and new ice often co-existing with open water whereas the investigation shown by Ivanova et al. 2015 is mainly based on thin ice regions of larger extent (i.e. sheets of nilas / grey and grey-white ice) as they occur in the Arctic, e.g. towards Siberia, during fall freeze-up. Hence the applicability of that "filter" to your model data does not appear credible to me.

Finally, using just 50% of original modeled SIC in case the SIT is between 5cm and 20cm *appears a far too large modification (I deliberately do not say correction). If you look at the Ivanova et al. 2015 paper, figure 4, you can see that in the two reasonably populated thin thickness bins (0.1m-0.15m and 0.15m-0.2m) OSI SAF takes values of 80% and 87% and Cal/Val, which you could see as a surrogate for the Comiso Bootstrap, takes values of 82% and 90%.*

Overall, I would say this comparison of "degraded" model SIC values is of limited value and I propose to discard it.

L268 / Section 3.4 / the entire paper: I don't understand / did not find it laid out in a very transparарent way, how the thresholds that are used henceforth were selected. I could not find an investigation that would tell me exactly why, e.g., for OSI SAF 75% SIC are used (for all cases, i.e. daily, monthly, re-gridded) while for NOAA/NSIDC CDR the thresholds are 1st different to OSI SAF and 2nd also different for the cases used. Possibly I overlooked this very important part in your paper, I am sorry. In some sense this resonates with my general comment GC0.

We have rewritten much of the manuscript and added text re. our threshold selection (see also our response above to GC0).

L306: What made you to chose NOAA/NSIDC CDR? Isn't the choice made in the NOAA/NSIDC SIC CDR to use the maximum value of two SIC products inferior to the well-thought through concept used in OSI SAF to derive a SIC CDR in a physical consistent way, with physically based uncertainty estimates? OSI SAF offers more information to be used than the NOAA/NSIDC CDR.

We feel it is insightful to use both products, as they are both so widely used in the community, and the comparison between the two of them provides further insight into model-observational comparisons particularly w.r.t. polynyas.

Figure 1: Just for clarification: The shaded area around the NOAA/NSIDC CDR curve represent +/- one standard deviation of the mean over the 42-years period at the respective longitude. Okay. But the results from the model represent both: a mean over the 50 ensembles PLUS the mean over the 42 years? Did you first compute the ensemble mean for every year and then average over the years? In short: My impression is that the shaded areas around the 3 curves are not directly comparable because they are based on different statistics.
I was wondering in this context, whether - in general - it might be more informative to use the median instead of the mean.

We have removed the CESM2-LE.

L337+ I am very sorry to have missed that but why are you also separately showing results derived from the NASA-Team SIC algorithm? As you indicate, it is very often biased low (often by 10-20%) compared to the Comiso Bootstrap algorithm SIC or independent estimates of the SIC and it is hence not overly surprizing that SIC fall below 85% and with that trigger the count of a polynya - either way, a coastal or an open water one. In order to keep the results *and the text refering to the NASA Team SIC results, you need to motivate why you include this product in addition much better. The simplest solution (and more credible as well) would be to simply discard all results based on the NASA-Team algorithm.*

Agreed, it does not add to the manuscript to include NASA team and we have removed it from our examples.

.

L347/348: I am happy with your decision to focus on coastal polynyas because your results for what you call "open water polynyas" are not always convincing and are in particular not matching with the general idea that exists about what an open water polynya is (the one in the Weddell Sea near Maud Rise for instance or the Cosmonaut Sea polynya).

L356-358: I can understand that you need to use different thresholds when defining polynyas in the ESM models and in the observational data. But I don't understand the necessity and the scientific reasoning to use different thresholds for different observational data sets if - as is the case - these are based on the same satellite data with the same native and the same grid resolution. Therefore, for OSI SAF and NOAA/NSIDC the same SIC thresholds should be used. See also my comment made in general to section 3.4 further above.

L377/378: See my comment to Figure 3.

L384/385: "underscoring ..." --> Well, yes, also this is not an overly surprizing result. Especially when using the SIC and the SIT as a metric to define a polynya this will apply. While satellite remote sensing offers other possibilities to identify polynyas, this seems to be more limited in ESMs - especially because in ESMs polynyas are predominantly sub-grid scale phenomena while in observational data polynyas can be resolved when using the correct observations. I am not sure where you are heading for with these illustrations.

L385-388: "Additionally ..." --> Okay, this is all interesting ... but nobody (?) would regrid a 25 km (or even finer resolved observational product) to 1 degree grid resolution with the incentive to derive a trend. One would use the finest resolution possible to get the most accurate information. I guess what I want to state is that it is already clear from what you stated above that there are serious differences in the polynya statistics when deriving them on the native grid to deriving them after re-gridding and there is no need to expand on this also towards the trend analysis.

L403-419: I am sorry, but I don't see a convincing scientific rationale for this inter-comparison. First of all: defining polynyas in satellite observations by means of a SIC threshold is a simple approach with considerable limitations and not the state of the art. There are other means to do so and, an aspect that has not been mentioned in your paper so far sufficiently well, the sensitivity of the used SIC algorithms and products to the diversity of the sea ice types that are encountered in a polynya render usage of a SIC threshold to define a polynya a rather arbitrary approach.

Secondly, as stated already, nobody in the observational community would regrid such data to a coarser resolution to then do a trend analysis. Polynyas are so dynamic that the idea would

rather be to go down to even finer grid resolutions and to swath data than to daily gridded data.
We are not recommending regridding to a coarser resolution for the identification of polynyas, however we are recommending that when comparing observations and model output it is important to consider them on the same grid. This is a common step when climate modelers compare climate model output to observational data on climatological scales.

I propose to reduce this part of the paper considerably.

L500: I might have overlooked that but since you will be using the 1 degree outcome of the exercise presented in this chapter 4.2 to state something about the reliability and accuracy of the ESM output: Did you do any evaluation against independent, high-resolution estimates of the polynya location and polynya area? I mean, after all, you are applying a SIC threshold algorithm which in that extent has not yet been applied that extensively to delineate and map polynyas. Hence the credibility of the results need to be illustrated. Otherwise this data cannot serve as "a truth" for the ESM model output. This resonates with general comment GC0.

L572-580: I think talking about these complexities would be more credible if you would have i) evaluated your satellite-based products against independent data and ii) if you would focus on those regions of the Southern Ocean where the ESM is really capable to simulate the seasonal cycle of the polynya area reliably - which I assume is basically the Weddell Sea, the Ross Sea and the Amundsen Sea - whereas the number of sea-ice covered model grid cells along the East Antarctic coast is often so small that there is no proper delineation of polynyas possible.

L591/592: Isn't the argument of rapid freeze-over also applying to the Weddell Sea where most of the polynyas are situated quite far South and hence are experiencing similar if not even colder conditions than the polynyas along the East Antarctic coastline?

L594-606: I have difficulties to understand the scientific rationale behind showing these regional trends - especially when looking at data with 1 degree grid resolution and monthly temporal resolution. Also, I thought that this intercomparison between CESM-JRA and the observations was done to check the reliability of the modeled against the observed polynya locations and areas. I see limited value to look into integrated polynya area. It would make, in my eyes, more sense to look into specific key polynyas: Ronne-Filchner Ice Shelf polynya, Cape Darnley Polynya, Mertz Glacier Polynya, Terra-Nova Bay Polynya and Ross Ice Shelf Polynya. Finding a good match in occurrence and size of these polynyas over time would help us much more to understand what the capabilities and difficulties are with the CESM-JRA model to come up with a reasonable estimate of polynya location and area. I am not convinced we learn a lot from this regional trend analysis which, to my opinion, is too much influenced by artefacts in the polynya identification method, both in the observational and the model data.

It might, in this context, also make sense to keep in mind that the grid cell area of a ESM grid cell is about 500 sqkm, hence when you talk about a trend in polynya area of 1000 sqkm / decade you are talking about 2 grid cells. What is your uncertainty estimate of how well you *can delineate a polynya in the model?*

L611: "Satellite imagery ..." --> This statement is not true as it stands because SIC retrieval algorithm can reliably provide SIC values of less than a few percent and they detect sea ice even when it is very thin (1-2 cm); The SIC estimates are biased low, yes, but the detection works. Whether and to which degree a SIC retrieval algorithm is biased low in the presence of thin sea ice depnds on the tie points chosen.

L643 "significantly more" --> can you elaborate and provide a number here? As a rule of thumb, from observations it is said that about 1% of the area bound by the ice edge is covered by polynyas (I guess this has been written down in one of the papers by Tamura et al.).

L651/652: "These results ..." I was wondering whether you can comment on whether CESM has a snow cover on sea ice. Snow on sea ice limits light availability considerably and I am not sure this is taken into account in your investigations sufficiently well.

Yes – snow on sea ice is critically important for albedo, conductive heat flux, light limitations. The CESM does have snow on sea ice.

L725: "differ only by retrieval algorithm" --> But the differences in the retieval algorithms are substantial - especially in light of that the NOAA/NSIDC CDR is in fact the maximum SIC value of two SIC products which often differ by up to 20% in absolute terms.

L752/753 "thus monthly data ..." --> When I look at Figure 7, i.e. the correlation between NOAA/NSIDC and OSI SAF CDR based polynya identification it is obvious that the correlation is highest for the daily data. For me this is an indication that an adequate coincidence in the derived polynya area requires daily temporal resolution.

When I look at Figure 6 and compare the polynya areas derived with the three different settings I can also see considerable differences between daily, daily regridded and monthly regridded data. From those I would not conclude that using monthly data may be "sufficient".

And: Phytoplankton blooms may occur only for a number of consecutive days and certainly there are peaks of activities that are linked to light availability not only because of the sea ice conditions but also because of weather and hence shortwave radiation conditions. Given that the polynya area quickly (aka within hours) responds to temperature and wind speed and direction changes and hence changes on daily if not even sub-daily scale, I am not convinced that the statement given here, that monthly data may be sufficient, fits well with your overarching aim to also say something about the NPP and its dynamics.

L810-812: I don't think that further investigations of "these contradictionss" as you state them are beyond the scope of your work. Rather on the contrary. To my opinion, your investigation of the different methods to delineate polynya area from observational data is not yet complete. See my general comment GC0.

Figure A1: I don't understand why you use two different SIC thresholds in your schematic *illustration. It implies the question: What if the SIC is between 50% and 90%?*

L907: Better: "and two polynya grid cells in (b)"
Figure B3 clearly advocates to use an as small grid cell dimension as possible because case b) results in an identified total polynya area that is closer to what you call "integrated polynya area" than case a)

Editoral comments / Typos:
L58/59: "will consequently ... form." --> This statement does only apply to open water polynyas, does it?

L64: What are "polynya-like" features? Are you also refering to leads?

L223: "25km^2" needs to be "25 km x 25 km"

L226: "25/45 km" --> 45 km reads strange. Are you sure it is not 50km?

L248: Please specify what you mean by "very small fractions"

L264/265: The daily sea ice concentration data you are using are either based on daily average gridded brightness temperature maps (NOAA/NSIDC) or were computed at swath level and subsequently gridded (and averaged) to form a daily average gridded product (OSI SAF). With that there is no "instantaneous snapshot" - at least not more than what you have in a model.

L278: Is this a typo? "0.4 cm" aka 4 millimetres ...

L326: I guess it needs to be "Figs. 3-4" here?

L330: The text says "July 15".

L346/347 "This is likely ..." --> True, but July 15 is not the melt season.

L369: "are much quite small" ?

L761: "higher / lower resolution" --> I suggest to use "finer" (i.e. smaller grid cell size) and "coarser" (i.e. larger grid cell size) because, applying a fixed SIC threshold of, say 75%, will lead to a larger polynya area in the coarse resolution than the fine resolution grid because of the smearing effect.

Thank you, we have changed our wording.

L767-769: You are using "very high" two times here. Would you mind to be a bit more specific here? Maybe a sentence that details "typical" values used to identify polynyas in SIC data would put the first "very high" into context when you compare it with the values you used *here? And the second "very high" could potentially be solved by writing "near 100%"?*

L784-787: Also here "very high" and "very thin" could / should be replaced by more specific values. You also might want to take into account the comments I made earlier about your statement of the capabilities of the satellite SIC retrieval.

L803: Another aspect that you could mention in this paragraph is the dependency of polynya formation in the existence of fast ice / icebergs - as demonstrated through various publications. How realistic is the fast ice formed in CESM?

Also, the oceanographic setting in the model might play a role - depending on how realistically the model simulates processes related to ice shelf cavities and nearby polynyas.

Finally, many Antarctic polynyas are bound to ice shelves. These calve and consequently the boundary for the polynyas change over time - but at the same time very radpidly from one year to the next. How does your investigation takes this into account?

Comments to the supplements:
Figure S3 caption line 4: "The horizontal dashed lines ..." because there are also vertical dashed lines.

Figure S4: Do I need to understand why there are no blue dashed lines?

Fig. 7: I know, this is just a supplemental figure but it takes quite some time to digest what is in the panels. I was wondering whether it would make sense to more clearly explain which data sets are correlated in the panels on the left hand side. I have difficulties to understand the term "EASE2" since this does not apply to an algorithm. I also suggest to be more consistent and, e.g., remove the "CDR" so that the names of the products are the same throughout the figure. I suggest to use "Month" instead of "Month/Season" as the x-axis annotation because there are no seasonal data shown. I recommend to refer to the panels only by a) to f) without the name that is put behind currently. I recommend to find a different solution to express which products are correlated. The usage of ":" is not very intuitive in that regard. You could perhaps write at the Y-axis "Correlation OBS vs. JRA" in panel e) and fine similar solutions for the other occasions. I also suggest to try to avoid that data points overlap with the symbols used as a legend. The caption contains a typo in the name of the Bellinghausen Sea.

Most of the above-given comments also apply to other figures of the same type.

We have removed many of the supplemental figures, added a couple new ones.