Review on "Automatic detection of Arctic polynyas using hybrid supervised-unsupervised deep learning" by Céline Heuzé and Carmen Hau Man Wong // submitted to The Cryosphere // July 2025

The authors present a study that uses a labelled dataset (daily polynya masks) to train a machine learning (ML) model to detect Arctic polynyas (areas with open water and/or thin sea ice) in sea ice concentration (SIC) maps (based on PMW remote sensing data). Presented results indicate that the model was able to identify polynyas with minimal false negatives, and most "false positives" were claimed by the authors to be correct detections of patterns with reduced SIC. The model was then also applied to the output of a particular climate model of the CMIP6 range, but the authors highlighted that there are several challenges attached to this kind of endeavour, i.e. identifying polynyas in more or less coarse resolution output fields with a (so far) varying level of detail in terms of sea ice information. Nevertheless, the presented approach and its first results make it a promising and seemingly adaptable tool for future studies that aim to continue the monitoring of Arctic polynyas in a rapidly changing environment.

The paper is very well written, reads fluently and the quality of the figures is overall good, with a few suggestions for further improvements given below as part of my "Specific comments". Otherwise, please also find my more general remarks and comments below. Overall, I consider the study to be highly interesting and hence very well publishable, but only after some more or less major revisions, additions and clarifications are addressed by the authors. These identified issues mainly revolve around some parts of the manuscript that would benefit from more care/detail in explaining the taken approach with its advantages/limitations, but also include the overly short discussions and comparisons on the thereby identified (interannual) polynya occurrences in the Arctic.

General comments

(1) Regarding the anti-correlation of false-positive polynya pixels (which the authors mostly consider as true) with SIE/SIA: the authors could comment on the circumstance, that the increasing number of polynya pixels (Fig.4) is likely related to another rather simple reason – the delayed fall freeze-up in many regions at or in proximity of the MIZ. In the present manuscript, I often had the feeling that it is not really separated between the regular freeze-up in late fall (~NOV/DEC), and "real" wind-driven polynya events in an otherwise more rigid ice regime. While this is certainly an issue for most studies on polynya detection in the Arctic during winter, I would at least appreciate a few additional sentences in the manuscript that properly addresses these methodical challenges.

(2) A bit related to my comment above: This issue also translates to the presented application to detect wintertime polynyas in MIROC6. There, the authors consider their model to perform well and note that "polynyas big and small are successfully detected". However, when taking a closer look at the presented SIC maps in Fig. 6 and 7, my impression is that most marked polynya areas are more likely part of remaining regions with MIZ characteristics (despite the GMM approach), while other areas of decreased SIC at well-known sites of polynya formation (such as the Laptev and Kara Seas, northern Baffin Bay, Beaufort Sea) are missed. I think it is great to see that the taken approach is somewhat flexible in terms of input data, but as the authors rightly acknowledge themselves – in this form it is purely a proof-of-concept that certainly requires some subsequent fine-tuning.

(3) In terms of presented results, I'm missing some sense of the long-term performance on annual / interannual timescales. A straight forward addition would be for instance maps of polynya occurrences (per winter or decadal averages) that would then help to identify if/how the here detected polynya sites around the Arctic compare to previous assessments in published literature. I noticed that the preprint by Wong et al. concentrates on that aspect (and more), so I guess such a comparison and related discussion should be rather easy to include here as well. It would also increase the impact of your study, beyond the current technical focus.

Specific comments

Abstract

L.1: "…areas with no- or thin-ice" → more commonly "open water and thin-ice"

L.11: "that the rigid traditional methods with fixed thresholds cannot identify" → not that I'm asking for a long explanation in the Abstract (please don't), but I'm asking myself – why exactly?

Introduction

L.15: "Polynyas, small openings in the pack ice, …" → in what sense small? Maybe worth to add some sort of reference for typical polynya sizes

L.17: Regarding the references for SIP in the Arctic – both of these cited papers had several follow-up studies over the years that introduced some changes/improvements to certain methodical aspects. Any particular reason why you picked these two?

L.19: "…and even impacts the large-scale atmospheric circulation" → not only large scale, but also regional effects on atmospheric boundary layer dynamics. Likely implied by "strong heat loss to the atmosphere" a few lines above, but not explicitly stated/referenced (e.g., Marcq and Weiss, 2012; Lüpkes et al., 2008).

L.24: I wouldn't use the terms "easy" and "way more tedious" here – that sounds like downplaying the referenced study. Maybe try to find a more "diplomatic" phrasing.

L.34: Why shouldn't they remain valid, and in what sense? Please explain your thoughts here.

L.38: "6 months" → six months

L.45: You could add a note here that this data set will be explained in a bit more detail in Section 2.1

L47: "Following Liu et al. (2025)'s work…

Data & Model architectures

L.56: In my opinion – this single sentence is not really necessary here as it just partly repeats the end of section one.

L.60-62: Why did you use this (now rather legacy) data set and not (for instance) the OSI-SAF equivalent? And: did you also try out alternative PMW data sets with higher spatial resolutions at some point? It would be interesting to see how your approach fares in regions where the grid resolution makes a real difference.

L.63: "no threshold for the area" → but I assume any pixel with a SIC > 0%, right? So not exactly *no* threshold.

L.67: "Briefly" → "In brief,"

L.84: "all NaNs" – I assume these contain the "pole hole" (basically filled up with 100% SIC as it seems?), land and missing pixels – something else? Details not given here.

L.86: "polynyas are rare events" → not true for every polynya, for instance the NOW polynya. If it's more or less open at more than 50% of the time during winter, I wouldn't call it rare. "Used to be rare" could be true however.

L.88: How do you define "scenes" here? 192 x 192 pixel grids or smaller subsets?

L.118: "the number of false negatives was not significantly affected" → any idea why?

L.136: "2 to 10 images" → i.e., 2 to 10 *daily* images, correct?

Figure 3: To be honest, I'm not sure if that large direct comparison of different class numbers is really adding much to the overall story. It could be more interesting to get an insight into seasonal differences in the GMM performance, and/or comparisons to other independent data sets.

Results and Discussion

Table 1 (caption): "269 million pixels in total" – what is total? All pixels or just those containing ocean?

L.167/171: "Fig.5" → Fig.4?

L.169: "the less sea ice, the more false positives" – any explanation for that (relates to my initial general comments)?

Figure 4: As they were mentioned in the text → one could add the corresponding regression lines and related statistics

L.187-189: "polynyas become harder to predict with traditional methods" → I can't defer that from Fig. 4 and Fig.5 alone, as you did not differentiate / illustrate the long-term evolution of the "traditional" threshold methods. And do you really mean "predict", and not rather "detect"?

L.192: "sea ice thickness data" → Please indicate which SIT you mean here, as other derived SIT estimates were obviously also utilized in earlier studies, incl. those prior to 2010

L.202: MIROC6 → reference missing;

L.202: “we don’t have any truthing data for it” → That admittedly sounds a bit handwavy, so can you at least try to list some model characteristics that could be behind this circumstance? In know that grid resolution is one prominent candidate here (as discussed towards the end of section 3.2), but maybe there’s more to it.

Figures 5, 6 and 7: Lat/lon values not indicated, should be added in at least one panel per figure. Fig. 6 and 7 are missing a clear depiction for the magenta colour, as in Fig. 5.

Conclusions

L.231: “It returns many false positives in the marginal ice zone, so we filter its results using an unsupervised Gaussian Mixture Model classifier that detects the open ocean / MIZ on each image.” → This effect could have been illustrated somewhere in Section 2.

L.234: “all polynyas are detected” → how can you be sure about that?

L.244: “that it indeed works” → “that it indeed works *to some extent*.”