Review on revision of “Causes and Evolution of Winter Polynyas over North of Greenland” by Younjoo J. Lee, Wieslaw Maslowski, John J. Cassano, Jaclyn Clement Kinney, Anthony P. Craig, Samy Kamal, Robert Osinski, Mark W. Seefeldt, Julienne Stroeve, Hailong Wang.

The paper describes the performance of the fully-coupled Regional Arctic System Model (RASM) with respect to the simulation of polynya events north of Greenland. A 42-year long simulation (1979-2020) is analysed in combination with satellite products and weather station data. Additionally, two ensembles are generated by forcing RASM with output from the Community Earth System Model (CESM) Decadal Prediction Large Ensemble (DPLE) simulations. The two ensembles, initialized in December 1985 and December 2015, are investigated with respect to precondition of winter polynya events.

Although the polynya in 1986 is included now in the revision the main part of the paper has not changed much. I am still not satisfied with the revision for several reasons and still think that the paper needs major revisions.

Main points of criticism:

(1) It is still not clear to me what the scientific added value of this paper in comparison to Moore et al. (2018) and Ludwig et al. (2019) is. This should already be clearly stated in the abstract. That the polynya in 2018 is caused by mechanical redistribution is already known from Moore et al. (2018) and confirmed by Ludwig et al. (2019).

(2) The two ensembles generated by RASM are not analyzed in depth. For instance, it is mainly only mentioned that “The frequency of polynya occurrence had no apparent sensitivity to the initial sea ice thickness in the study area pointing to internal variability of atmospheric forcing as a dominant cause of winter polynyas north of Greenland.” This is much too general. I have my doubt, that there is any change at all in the statistics of the upper atmosphere (see below) but ANOVA is the technique to use for exactly this kind of problems (https://en.wikipedia.org/wiki/Analysis_of_variance).

(3) There are some typos and grammatically curious formulations in the revision that makes me wonder if the revision was done with the necessary care and if one of the native (American-) English speaking people has seen the revision.

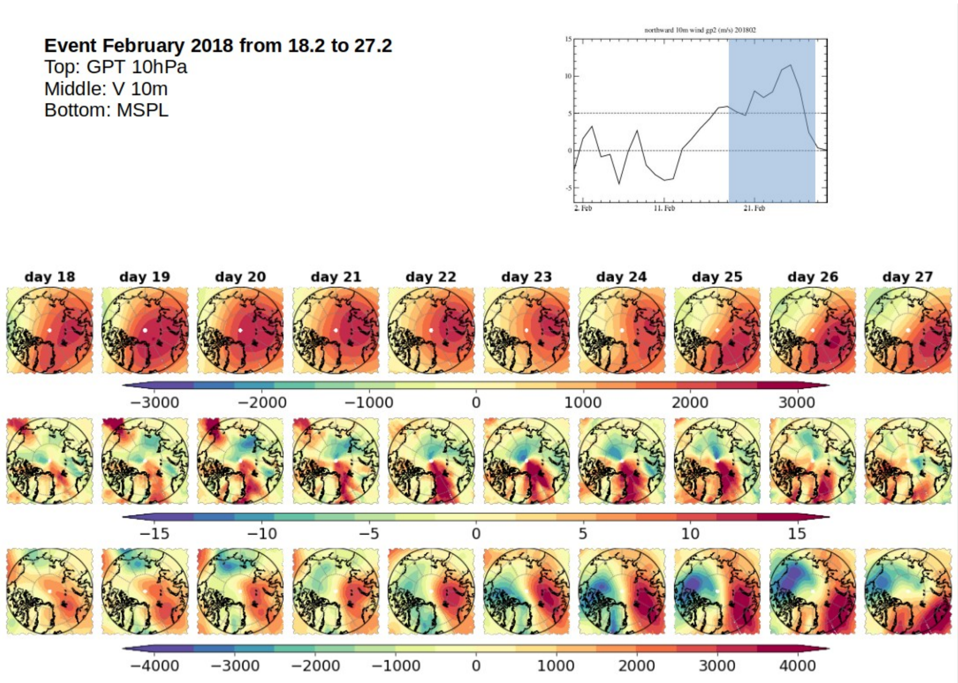
(4) Connected to the second point. I am suspicious that the reduced number of polynyas in the second ensemble (16 versus 25) is just an artifact of the metric used (more then 10km3/day outflow of ice volume for at least three days). The mean thickness in the region of the 1985 ensemble is 3.7m and of the 2015 ensemble 2.8m. Because the metric is based on volume outflow one would expect even without any change in the wind statistics in the two ensembles a reduced number of occurrences in the 2015 ensemble, namely 25 * 2.8m/3.7m ~ 19. As pointed out under (2) it has do be done a fair statistical analysis (ANOVA) to check if there is a significant difference at all between the two examples regarding the winds that cause the polynyas (there will be near surfaces differences, of course, with respect to the energy balance because the mean SIT is different).

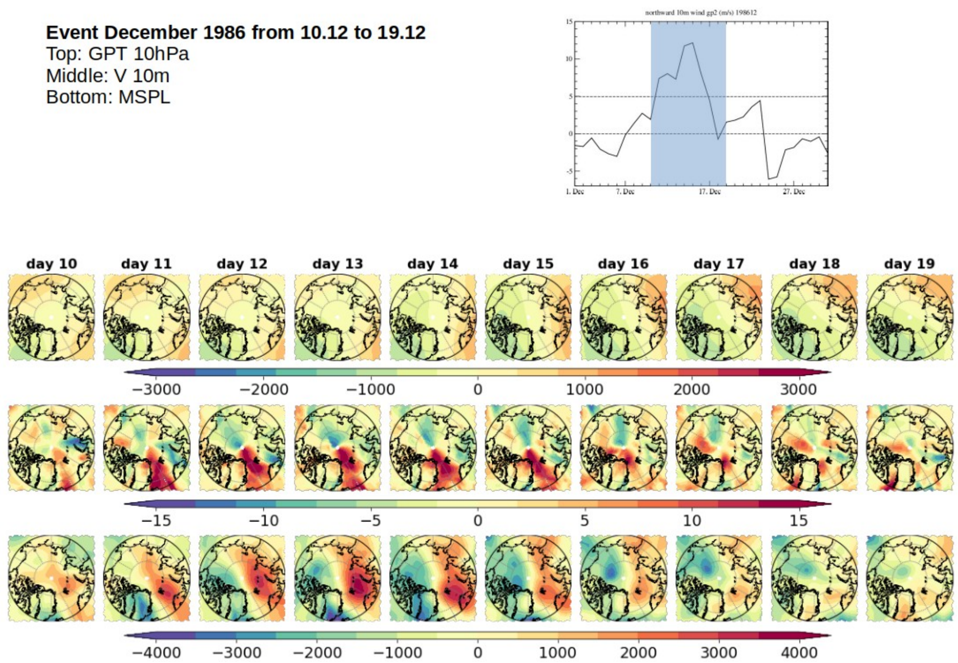
(5) I mentioned already in the first revision that I find the performance of RASM in simulated the 2018 polynya disappointing. Downscaling is expected to add details to

coarse resolution model results but in this case CFSv2 exists in higher resolution (about 0.2 x 0.2 degree) then WRF (about 50km). This should be clearly stated in the manuscript and that 'downscaling' is only done, if at all, with respect to the sea ice and ocean (see e.g. line 162), i.e with respect to the hindcast run the atmospheric variables are rather upscaled then downscaled. I suspect that the relatively coarse WRF resolution is responsible for the location mismatch of the observed and simulated 2018 polynya (see Fig. 2). It would be nice to give some information about the resolution of CESM-DPLE as well.

(6) There is almost no information given about the coupling of CFSR/v2 or CESM-DPLE. At least some information on the technique, the variables used and the height levels (that information is given in the discussion section but it should be given in the methods section together with the other information).

(7) Use of the AO: "The daily AO index is constructed by projecting the daily (00Z) 1000mb height anomalies poleward of 20°N onto the loading pattern of the AO" (https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.shtml) while the loading patterns are derived from monthly mean 1000mb height anomalies. Said that, it is clear that the daily AO is, as well as the monthly AO, a statistical mode that characterizes the Northern hemispheric state of the atmosphere but is not well suited to characterize local atmospheric states (as e.g. in the polynya region). This is reflected by the low correlation coefficient (e.g. 0.39 and 0.45 in line 313) which means that only about 20% of the variance of the near surface temperature can be 'explained' by (a lagged) AO response (neither the time period used for the calculation of the correlation is given nor it is explained who is leading whom). In other words: About 80% of the variance of the near surface temperature are not correlated to the AO, i.e. independent from it which raises the question why the daily AO is considered at all. The confidence level that is given additionally is not of much help and I wonder if the auto-correlation of the time series is considered or if the daily values are considered as being independent (the auto-correlation might reduce the number of degrees of freedom in the statistical test dramatically). In my own analysis of the polynya events 1986 and 2018 I found the plots attached below much more helpful as any correlation coefficients. They show nicely that southerly winds are caused by high pressure systems over the Barents Sea in both events, but that 2018 is connected to a SSW and 1986 not (if the SSW is causing the winds in 2018 or if that is just a coincidence cannot be answered in my opinion but obviously similar strong winds can occur without an SSW event (1986)).





Minor point:

Section 3.2: Ad hoc it is not clear to me why SOM should be used instead of conventional Empirical Orthogonal Function analysis. A short explanation might be helpful.

Line 235 – 240: Likely the location misfit of the polynya (Fig.3) can be attributed as well to the relatively low resolution of WRF.

Section 4.1.2: I wonder why 1986 is not mentioned in this section. Are no station data available? If yes, that should be mentioned. If data are available they should be discussed as well for 1986.

Line 250: I find the sentence “Figure 5 shows the RASM thermal sea ice surface, lateral and bottom melting terms were all negligible (< 1 cm) over the study region when integrated for the whole month of February 2018.” hard to comprehend. May be better to write “ Figure 5 shows that the thermal **ice melting terms of RASM (at the surface, lateral, and at the bottom)** are all negligible (< 1 cm) over the study region when integrated **over February 2018**.”

Line 270-271: “… due to the rapid ice growth during the polynya opening …”. The ice is mainly growing after the opening. I suggest: “… due to the rapid ice growth **following** the polynya opening …”

Line 275-276: “… and have found that the polynya development was associated with strong and persistent winds from the south-southeast.” This is no new finding. I suggest: “… and have found that the polynya development was associated with strong and persistent winds from the south-southeast **in agreement with Moore et al. (2018) and Ludwig et al. (2019)**.”

Line 314: See main point of criticism (7). If you want to stick to the comparison with the AO: “… lagged by approximately two weeks.”: I suggest to make clear that the near surface temperature is leading the AO (which means that the AO can not be causing the temperature anomalies), ie. I suggest “**… while the AO is leading by** approximately two weeks.”.

Section 4.3: Please check for grammatical correctness. Especially the last sentences “However, the mean turbulent heat flux was much less in December 1986 than in February 2011 even though it was a larger event in terms of polynya size and wind intensity. This is possibly due to the fact that sea ice was thicker in 1986; for example, the mean SIT was 4.4 m for 5 days before the polynya (Table 1). Due to large open water areas in December 1986, the integrated turbulent heat loss was much larger compared to the polynyas in February 2011 and 2017.” are **not understandable for me**.

Line 441-455: See major point (4).

Line 456: ‘longest’ → ‘largest’ ??? or ‘longest lasting’ ???

Line 493-497: “Overall, the more frequent winter polynyas, produced in a thicker sea ice regime between the two 30-year apart ensembles, implies that changes in SIT are not significant contributors (at least up to now) to the generation of such events for this region during wintertime. Therefore, the findings support that polynyas becomes prevalent when southerly winds are more persistent and stronger in northern Greenland.” See my concerns outlined in major point (4). The last sentence I do not understand.

Line 500: “… which means that the model is prescribed with reanalysis or gridded products on every grid cell.”. That sub sentence is simply untrue. Correct is that forced sea ice-ocean models are one-way coupled. But the reanalysis is not prescribed as is stated by the authors. 10m wind is acting via drag formulations on the ice, 2m temperature and humidity act via sensible and latent heat fluxes calculated at the surface by the model and normally downward long- and short-wave fluxes act at the surface but the net fluxes are calculated as well by the model. This sentence should be revised.

Line 561: The sentence “By taking advantage of an ensemble approach, the internal variability is better assessed with respect to the occurrence of such coastal polynyas during **extereme** events.” needs justification. From which analysis presented I can deduce that with the ensemble approach the internal variability is better assessed? What is meant by ‘better’? Better compared to forced sea ice-ocean models? That would be a trivial sentence, of course. Note the typo in ‘**extereme’.**

Line 637: I suggest to change the sentence to: “However, the mean turbulent heat loss in the study region **during the polynya in 2018** was about 61 W/m 2 (with a maximum of 124 W/m 2 **at day XX**), which is in good agreement with the results of Ludwig et al. (2019) based on the forced sea ice-ocean model NAOSIM (mean/maximum 40 and 124 W/m 2 , respectively).

Line 656: ‘that that’ -.> ‘that’