Reviewer 2

Moreno-Chamarro et al.: Impact of the ice thickness distribution discretization on the sea ice concentration variability in the NEMO3.6-LIM3 global ocean–sea ice model. The authors investigate ice thickness distribution (ITD) categories in NEMO-LIM and how they impact sea ice concentration variability. They use k-means clustering as a technique in tandem with three observational based SIC datasets. The authors do not find an optimal configuration as results in the Arctic and Antarctic have opposite responses to ITD changes, so no clear benefit to NEMO-LIM is determined from changing ITD.

Overall, I believe this will be suitable for publication with a few major/moderate changes. I felt that the scientific significance and quality were good to fair, but could be improved with some expansion in the text. The Scientific Reproducibility is also fair, which again could be improved with further clarification in the text. The Presentation quality was excellent.
We thank the Reviewer for the appreciation and the thoughtful comments. In the following we answer each specific point (in blue).

Specific Comments:

- One of the biggest concerns I have about this paper is that it doesn't generalize to modeling in general beyond NEMO-LIM to provide insight about modeling in general. I realize that this is for the NEMO special issue, however, it currently feels a bit like a sensitivity experiment to determine optimal model configuration but not otherwise generally of interest to the community of sea ice modelers who may be setting up their own models using LIM or other sea ice models.

  This begins in the introduction where there should be a brief discussion of previous work about why 5 ITD categories have been chosen in the past due to volume studies (Lipscomb 2001, Remapping the thickness distribution in sea ice models, doi: 10.1029/2000JC000518; Bitz et al. 2001, Simulating the ice-thickness distribution in a coupled climate model, doi: 10.1029/1999JC000113). In fact, in Bitz 2001 one of the conclusions is "…the concentration of open water and thin ice, which is relatively insensitive to the number of categories beyond M=5," which is directly relevant for this paper. Why weren't these cited? If anything, studies using CICE that agree with these results should strengthen your results because they become more robust across models.

We thank the reviewer for the references. We opted for an Introduction briefly reviewing previous research on the impact of the ITD in climate models since a longer, more detailed one is provided in the companion paper Massonnet et al. (2019). This was also done because nearly all of the previous studies have focused on the mean climatological state of the sea ice (the focus in Massonnet et al., 2019) and not on its variability (our focus). Thus, whereas Bitz et al. (2001) was indeed cited in the Introduction as an example, Lipscomb (2001) was not. Both works are now cited in the revised manuscript. The Introduction has further been clarified on this point and extended following the Reviewer's suggestion (Lines 48–55).

In the discussion and conclusions section you should add more information about how these results might be directly relevant in coupled models. This is brought up briefly but could be fleshed out and suggestions for how to test this would be useful. Additionally, you mention that parameterizations and parameter values are tuned for 5 categories (line 359). Can you specify which of these might be directly affected or changed? Are similar parameterizations present in other sea ice models? How can this be generalized for the community?

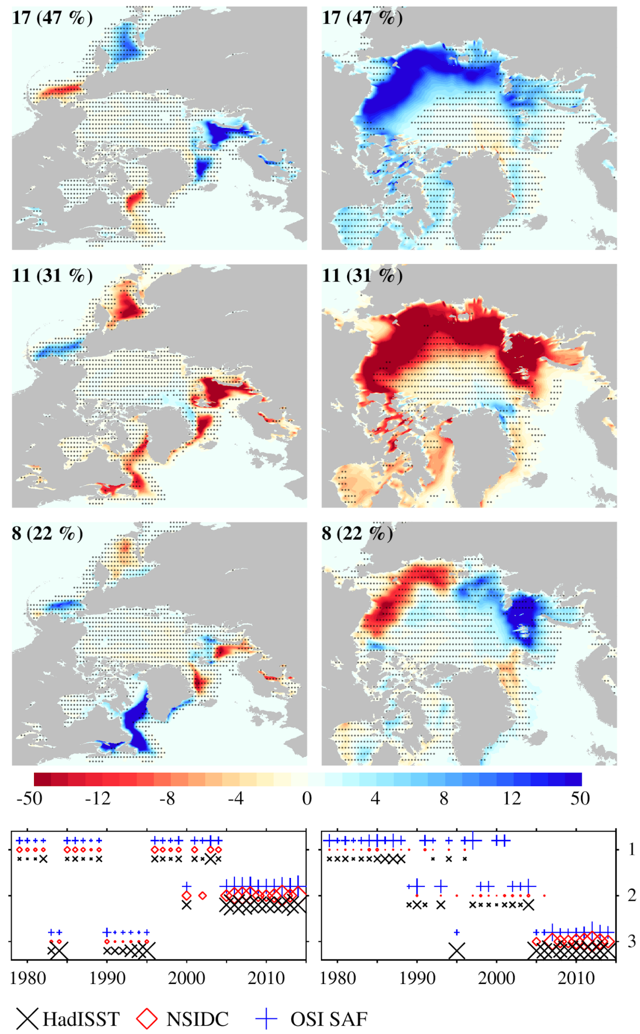
The Discussion section has been rewritten in full to accommodate these suggestions. Now dedicated parts discuss the tuning parameters that might need adjustment in the model (Lines 383–389), the potential relevance for other sea ice models beyond NEMO3.6-LIM3 (Lines 390–406) and for fully-coupled modes (Lines 407–415).

- The methods need clarification, particularly for replicability purposes. In particular, I found these sections to need to be expanded.

  1) At line 145/Figure 2 the Arctic "winter" cluster was defined but didn't include April. What threshold values for these groups were used? Are your results sensitive to including different months? These things should be tested.

  We opted for the standard definition of a season of three months. To define the winter (summer) season, we search for the largest correlation coefficient between the monthly clusters and the adjacent second largest value in the winter (summer) half year (correlations coefficients are plotted in Fig. 2). The two seasons must be and are consistent across the three observational datasets included in the analysis. This method renders two seasons in which monthly cluster agreement is consistently high: JFM and ASO, on which we then base the whole paper analysis. Although monthly clusters in April and in previous months show good agreement, agreement is smaller than across JFM. This method hence leaves April outside the winter season. We note our winter season agrees with the analysis in Close et al., 2017, where monthly Principal Components of the sea ice concentration show that January, February, and March have similar modes, different from those in November, December, or April. For these reasons, we have decided to keep our definition of a winter season without April. This point has nonetheless been clarified in the revised manuscript (Lines 156–160).

In addition, to show that including April in winter would actually have had little impact on the analysis, Response Figure 1 (below) shows the test where the clusters are extracted in the 4-month seasons January–April (as winter) and July–October (as summer) in the Arctic. The main difference with respect to the clusters in JFM and ASO (Fig. 5) is that the second and third clusters have switched positions. Cluster patterns and years of occurrence are however virtually identical as those in JFM and ASO (c.f. Fig. 5). The Response Figure 1 is not included in the received manuscript, but we do mention in the text that results are not sensitive to the specific season definitions.



Response Figure 1: As in Fig. 5 in the main text, but in JFMA (left) and JASO (right)

2) The % values in Fig.5/6 refer to occurrence, can you translate these values to number of months or something to better indicate what this means?

The value is the percent of years in the period 1979–2014 whose anomaly pattern is the closest to a particular cluster. The number of months is now shown together with the percent value in Figs. 5 and 6, and Supp. Figs. 2, 3, and 5.

3) Section 3.3.1 – are these correlation differences statistically significant from one another? Can you clarify what you mean by these are significant?

In the updated manuscript, statistical significance of the difference between correlation coefficients is tested using Fisher's z-transform assuming a two-tailed significance level of 0.05. Given the large number of coefficient pairs for which differences might be tested, we simplify this by comparing only the median values between an ITD discretization and the one immediately below within the same discretization type: for example, S1.50 is compared with S1.30, the latter with S1.10, and so on; S2.15 is compared with S2.11, and so on; S1.03, S2.03, and S3.05 are all compared with S1.01. The results of these significance tests are now included and discussed in the revised manuscript (see the new Figs. 7, 8 and 10, and Supp. Fig. 6).

4) Line 263 – how was the polynomial determined? Can you provide information about this?
Detrending is done by removing a spatially varying 2nd degree polynomial fit with respect to time using the 'Trend' function in the s2dverification R package [Manubens et al., 2018]. This is now indicated in the revised manuscript (Lines 145–146).

- If there is not a lot of information gleaned from the de-trended Arctic analysis, then why is it presented? Can this be condensed somehow since the variability analysis primarily shows the forced trend without being de-trended?
We would like to keep this section in the main text. The analysis of detrended data is actually critical to characterizing interannual variability in summer, a season which is dominated by the long-term melting trend in the Arctic. Without detrending, Arctic clusters mostly capture this trend (compare Figs. 5 and 9 in the main text). The analysis of detrended data further shows that ice thickness distributions with more than 30 categories can help improve model–data agreement in the Arctic, at a cost of making the simulations computationally more expensive.

- I think that if possible you should consider including Supplemental Figures 4 and 7 as regular figures since they are referred to in detail.
We would like to keep them in the Supplement. Both Figures are only briefly discussed in the manuscript and add little extra information to the discussion of the results. And although the number of figures is not a constraint for publication in GMD, we think 11 main Figures is already a high enough number.

Technical corrections:

Line 81: misspelled “concentration”

Corrected

Lines 269-275: It looks to me like patterns 2 and 3 are both dipoles but opposite patterns. Can you clarify where the quadrupole is?

We call quadrupole a cluster that shows four dominant poles in Arctic sea ice concentration, regardless of the sign. In the winter Arctic this usually means a pole in the Labrador, Barents-Greenland, Okhotsk, and Bering seas respectively. This follows the definition of the quadrupole in Close et al., 2017. Both clusters 2 and 3 in Fig. 9 would therefore fall in this definition. We note, however, small differences between the two. In cluster 2 the pole in the Labrador Sea dominates and dominates in years of strong positive winter NAO. We interpret this as the wind-driven signature of the NAO on the ice concentration [Bader et al., 2011]. Cluster 3 is instead closer to cluster 1 in not detrended data and dominates in similar years. They both further resemble the quadrupole pattern analyzed in Close et al., 2017. This point has been clarified in the revised manuscript.

Lines 296-298: sentence is confusing. “…suggesting that this configuration poorly captures the forced variability but does capture interannual variability as well as any other configuration.”?

This has been clarified.

The stippling markers are used to indicate significance in Fig. 11 but insignificance in Fig.5. It would be nice if they were used consistently.

Both Fig. 11 and Supp. Fig. 7 have been modified as suggested.

The first two paragraphs of the discussion were clear and concise. The next three are a bit confusing and all over the place. I’d suggest you rearrange in the following order: 1. One category has worst results necessitating multi-category sea ice models like LIM3 or CICE; 2. The standard configuration is 5 ITD levels; 3. Adding more thin categories decreases agreement; 4. Having 30+ categories can improve some but is significantly more expensive at double the cost, which is clearly significant for coupled models

Both the Discussion and Abstract have been rewritten following the Reviewer's recommendation.