Response to referee comments

(Referee comments are shown in black, our response is in blue and changes to the manuscript are shown in red. The revised manuscript is also included in this document.)

Page references are given to the updated manuscript as PXLY indicating that the manuscript has been updated on page X line Y.

Thank you again to both the reviewers and the editor for their helpful feedback. We appreciate the recognition of the significant improvement in quality of the manuscript since the original submission and are grateful to both reviewers and the editor for their role in this process.

Reviewer 1

The revisions that the authors have made on the manuscript have substantially improved its quality. I appreciate the responses to my major comments and the additional text that has been added. I have a number of remaining comments on this version of the manuscript. After these comments have been addressed, I believe the manuscript should be published and will make a meaningful contribution to the literature on modeling of floe size distribution in the Arctic.

P1L29-31: suggestion to reverse sentence so that the abstract ends on a positive note. "We note that although the WIPoFSD model is unable to represent potentially important features of annual FSD evolution seen with the prognostic model, it is less computationally expensive…possibly making this a stronger candidate for inclusion in climate models."

Manuscript modified as suggested.

P2L23: Are FSD models where shape is fixed always to a power law? If so, note

As far as we are aware, yes. The following clause has been added to the end of the relevant sentence: 'generally to a power law'.

P3L18-19: Note/make clear that the prognostic mixed-layer model is an ocean model

Corrected as suggested. 'prognostic mixed-layer model' has been updated to 'prognostic mixed-layer ocean model'.

P3L34-36: Suggest moving sentence beginning with "Full details…" to earlier in the paragraph, remove the sentence beginning with "Below we provide…"

Above changes made as suggested.

and simplifying "In section…" to "The adaption of lateral melt for FSD models will be introduced in section 2.1.4".

'In section 2.1.4 we will explain how this standard treatment is adapted for use with an FSD model.'

Now reads as:

'The adaptation of the standard CICE lateral melt treatment for use with FSD models is described in section 2.1.4.'

P4L39: subscript a missing from C^{f,floe}?

Corrected as suggested.

P6L5: appears that l_eff,n has . rather than ,

Corrected as suggested.

Section 2.2.2: It seems that I disagree with Reviewer 1, as I found it disorienting to read this material in the methods, which seems to me to belong in the introduction. In my read, I would suggest P6L8-38 should be in the introduction. One possible compromise would be to add sub-sections to the introduction (uncommon, but a feature I appreciate as a reader) such that the background on brittle fracture doesn't overwhelm the rest of the background presented. Ultimately, I suppose it should be up to the authors how to handle this organizational issue.

Given the disagreement between the reviewers on this point, our preference would be to leave the material referenced in section 2.2.2.

P7L23: missing "such" after sufficiently

Corrected as suggested.

P9~L18-25: Note explicitly that prog-best includes brittle fracture.

The following sentence:

'The prognostic FSD setup, prog-best, uses the standard 12 floe size categories outlined in Roach et al. (2018) and the 5 standard CICE thickness categories (Hunke et al., 2015).'

Now concludes with the following:

'and includes the brittle fracture scheme described in section 2.2.3.'

P9L25: Do you or could you briefly explore the sensitivity to these parameter choices? Or is there some references that could be included here? It would be good to clarify how sensitive the results are/are not to this fitting.

The following sentence has been added to highlight references that present results exploring the sensitivity to the parameter choices:

'Sensitivity studies to these parameter choices have previously been performed for the WIPoFSD model and the version of the prognostic FSD model considered here (i.e. including the brittle fracture scheme) in Bateson et al. (2020) and Bateson (2021a) respectively.'

Section 3.2: Apologies if this is explained elsewhere, but why is only the prognostic model compared to observations, and not WIPo-FSD? I suppose this is because the observational comparison is focused on showing that the brittle fracture improves comparisons, but it seems to me that comparing both would be useful, still. Explain reasoning here (or consider including in figures)

P11L23-24: Just to reiterate the comment above, I think this statement necessitates including the WIPo model results in Fig 4

The core assumption of the WIPoFSD model that is presented in this study is that the FSD can be approximated by a truncated power-law with a singular time-invariant exponent. The power-law distribution shown in all sub-plots uses the same exponent of -2.56 i.e. the same exponent used in *WIPo-best*. In Fig. 4 we only consider floes with a diameter of about 1700 m and smaller. The emergent FSD from the WIPoFSD model will only deviate significantly from the power-law distribution shown in Fig. 4 if l_{var} drops below 1700 m for the snapshots considered; model output shows this is not the case and generally l_{var} is around an order of magnitude larger for the case studies considered. As such, we do effectively compare WIPoFSD model output to observations.

In the first sentence of section 4.1, ‘power-law fit’ is replaced with, ‘power-law fit using the same exponent across all case studies’.

The following clarification has also been added to the first paragraph of section 4.1:

‘The power-law fit is used here to represent *WIPo-best*, since the WIPoFSD model is built on the assumption that the FSD can be approximated by a truncated power-law with a singular time-invariant exponent. In practice, the emergent FSD from *WIPo-best* will be identical to the power-law fit shown in Fig. 4 provided the floe size range included is consistently below l_{var}, which is the case for all the case studies considered.’

P12L5: I think this sentence perhaps overstates the impact still. Delete “significant”, and perhaps add that there is less change the the level of observational uncertainty.

Corrected as suggested. Following clause has been added to relevant sentence: ‘with the size of any changes well within observational uncertainty’.

P12L12: delete “generally”. Similar to comment above, make more clear that this is well within the observational uncertainty.

Corrected as suggested. Following clause has been added to relevant sentence: ‘with any differences between the simulations significantly smaller than the observational uncertainty’.

P12L18: Consider including the percent relative change of MIZ extent for the month with greatest difference?

The statement referred to here, ‘Overall, inclusion of FSD processes within CICE results in changes to extent metrics of order 1 x $10^5\,km^2$’, refers to the four different time series described in Fig. 6. The maximum percentage change will be different for the different timeseries. To address this, we have included a statement describing the general range that the percentage change spans across the four timeseries.

The following sentence has been added after the relevant statement:

‘This corresponds to a percentage change in extent varying between around 1 % to 10 % across the different months and regions considered.’

P12L19: Note that tense is inconsistent across sub-headings

Relevant sub-heading has been updated to ensure consistency in heading style.

P12L30: edit to specifically note that the ribbon shows 2 standard deviation, not just the range

The following statement in brackets has now been removed: ‘indicated by the width of the ribbon’.

Figure 7 is now introduced as follows, with the second sentence a new addition:

‘Figure 7 shows the percentage difference in the sea ice extent and volume for both *prog-best* and *WIPo-best* relative to *ref* averaged over 2000 to 2016, indicating the impact of each FSD scheme compared to assuming a constant floe size. The shaded region in Fig. 7 covers twice the standard deviation from the mean in each direction.’

P14L21: discriminates should be discriminate

Corrected as suggested.

P18L30: ice-thickness distribution has a hyphen here, but not elsewhere (ice thickness distribution). Suggest the author check throughout paper for consistency (i.e., remove all such hyphens)

Corrected as suggested.

Fig 1: Edits are generally helpful. However, the y-values in the lower subpanel are now inconsistent with the examples shown above. Please edit the bar sizes to agree with the furthest right example, and perhaps make it clear that this is where it's coming from (such as with corresponding letter or box)

Edited as suggested.

Fig 2: delete repeated sentence from caption

Repeated sentence has been removed.

Fig 4: As noted above, should add WIPo-best here, but also prog-best. Otherwise, the link between this figure and result (using prog-16) and subsequent results (with prog-best, 12 categories) is not explicit.

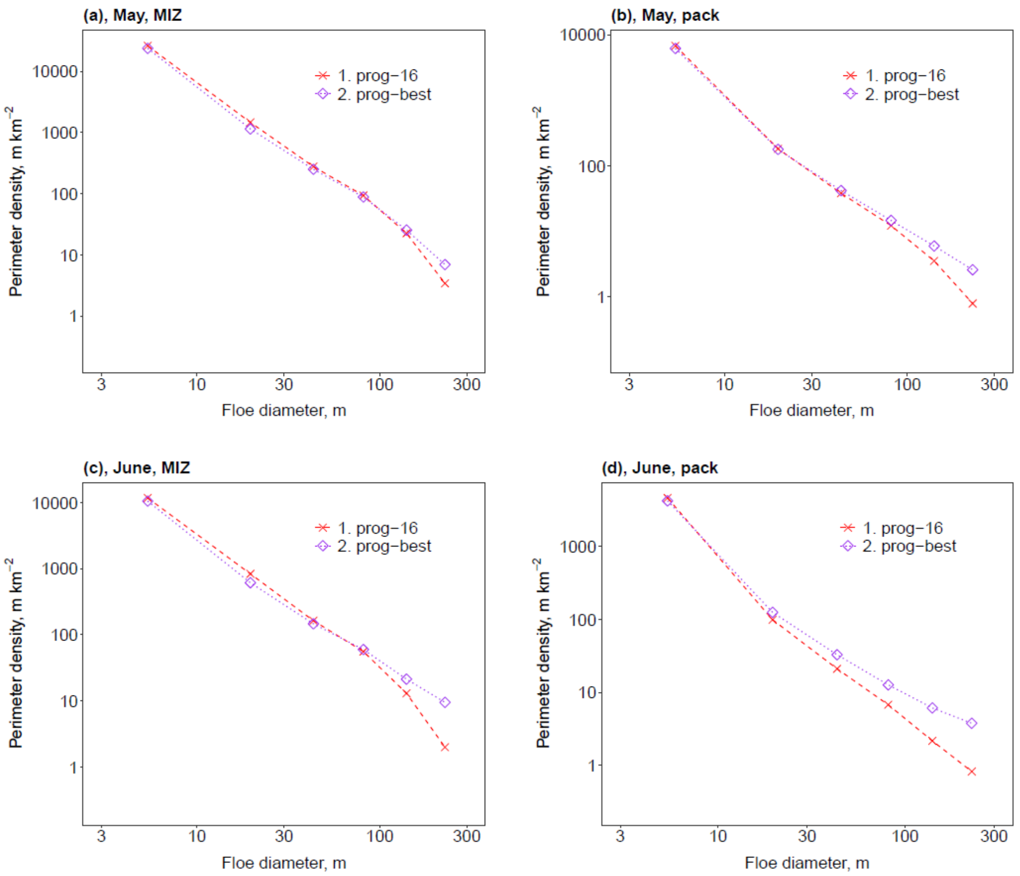

Figure A: The perimeter density distribution, $m\ km^{-2}$, of sea ice area as a function of floe size for April MIZ (top left), April pack ice (top right), August MIZ (bottom left), and August pack ice (bottom right). Distributions are shown for *prog-16* (red, cross, dashed) and *prog-best* (purple, diamond, dotted), all averaged over 2000 – 2016.

As discussed in section 3.1 in the manuscript, *prog-best* uses 12 floe size categories as this enables the range of floe sizes relevant to FSD impacts on the sea ice cover via lateral melting and form drag to be sufficiently resolved without excessive computational cost (an increase from 12 to 16 floe size categories increases CICE run time by about 60%). Given the aim here is to compare two alternative practical ways to model the FSD in sea ice and coupled climate models, it is important that neither method has a prohibitive computational cost. The reason we use output from *prog-16* rather than *prog*-best in the comparison to observations is due to the 'uptick' present in the emergent FSD from the prognostic model. By using 16 floe size categories rather than 12, the 'uptick' falls outside the range of floe sizes included in the comparison. Since the model output is renormalised according to the total sea ice area within the floe size range considered, the presence of the 'uptick' in *prog-best* FSD output within the range of floe sizes considered would preclude a useful comparison between *prog-16* and *prog-best*. Perimeter density from the smallest floe size categories is the most important metric in terms of FSD impact on the sea ice cover (due to the inverse relationship between lateral melt rate / floe edge contribution to form drag and floe size). Therefore, a useful way to address the link between *prog-16* and *prog-best* in the context of this study is to compare model output from both *prog-best* and *prog-16* from the smallest 6 floe size categories i.e. outside of the range of significant uptick influence for both setups. In Fig. A we consider model output averaged over 2000-2016 from both May and June i.e. the main months considered in Fig. 4. Here we show that the differences between the two models are negligible in the smallest few categories i.e. those that most strongly determine FSD impact on the FSD cover. We have added a comment to address this point in the manuscript.

The following comment has been added to section 4.1:

'Whilst in Fig. 4 we consider prog-16 with 16 floe size categories, for comparisons against *WIPo-best* on sea ice behaviour within CICE we consider *prog-best* with 12 floe-size categories since this represents a more practical setup of the prognostic FSD model for use in sea ice and climate simulations, as discussed in section 3.1. In a comparison of model output from *prog-16* and *prog-best* (not presented here) larger differences can be seen in the shape of the distributions in the larger floe size categories due to the presence of the 'uptick' but these differences tend towards negligible in the smallest categories i.e. those most relevant in determining FSD impact on the sea ice cover (e.g. Tsamados et al., 2015; Bateson et al., 2020).'

In addition, we have made minor adjustments to section 3.1 to ensure the reason why *prog-best* and *prog-16* use a different number of floe size categories is clearly explained.

Figures 7,8. Make colors consistent throughout, i.e., WIPo-best always blue, prog-best always red

As suggested, colours have been updated in Fig. 7 to ensure consistency with Fig. 8.

Figure 7 caption: replace "ribbon" with "shading"; remove "the region spanned by"

Corrected as suggested. Relevant sentence now reads as follows:

'The shading shows, in each case, plus or minus two times the standard deviation around the mean.'

Figure 10: I believe I commented on this in the previous round, but I still find it hard to believe that the thickness change is really on the order of meters. Please confirm that this should not be cm?

It is worth noting that a non-linear scale is used in this plot. Changes larger than 1 cm are shown on the plot, and thereafter categories are: 1 cm – 2 cm, 2 cm – 5 cm, 5 cm – 10 cm, 10 cm – 20 cm, 20 cm – 50 cm, 50 cm – 2 m. This quasi-logarithmic scale is used to identify locations where the largest changes are seen whilst also being able to identify broader areas of smaller changes in thickness. Fig. 10 shows that only a few grid cells have changes that exceed 50 cm (and of these even a smaller set will have changes that exceed 1 m). Whilst there is a more substantial area where changes exceed 10 cm, particularly for *prog-best* in September, changes in thickness mostly do not exceed 10 cm i.e. they are on the order of cm. Note these differences are comparable in order to previous studies using versions of these models e.g. Bateson et al. (2020), Roach et al. (2018).

Reviewer 2

This is the second review of the manuscript entitled Sea ice floe size: its impact on pan-Arctic and local ice mass, and required model complexity. I am overall satisfied with the answers to my previous review and the modifications to the manuscript. The work made by the authors has significantly improved the quality of the study in my opinion (congratulations to them), and I would be happy to recommend the paper for publication after some minor adjustments.

A general comment is that the paper is much clearer than it was, but also quite longer and with some repetitions between the results, the discussion and the conclusion. This is not a major problem, and the manuscript could be left as it is, but the authors should not hesitate to remove some comments that are repeated throughout the paper before publication (for instance the reasons behind the uptick).

P1L22: "We demonstrate that a parameterization of in-plane brittle fracture processes enables the prognostic model to achieve a reasonable match against the novel observations". I slightly disagree with this sentence. The study demonstrates that adding a term to the prognostic is needed to achieve a reasonable match against the novel observations, and that there are good reasons to believe that this term represents the effects of in-plane brittle fracture. Also, the manuscript spends some time discussing the effect of this parameterization and why it improves the results, so it might be worth summarizing these results in the abstract. I would therefore suggest writing something like: "We show that adding a term that relaxes the FSD towards a power-law enables the prognostic model to achieve a reasonable match against the novel observations in the summer. We suggest this term represents the effects of in-plane brittle fracture that break the larger floes (>2km) into mid-sized floes (100m-2km)."

We agree that it is important to be precise here on the methodology used, however, since we use a scheme that extends beyond a simple relaxation based on consideration of the relevant brittle fracture derived mechanisms, we are hesitant to describe it as a relaxation scheme in the abstract. We have updated the abstract to provide further details on what has been done and be more precise in describing the impact of the new parameterisation.

The following sentence in the abstract:

'We demonstrate that a parameterisation of in-plane brittle fracture processes enables the prognostic model to achieve a reasonable match against the novel observations.'

Has been updated to:

'We introduce a parameterisation motivated by idealised models of in-plane brittle fracture to the prognostic model and demonstrate that the inclusion of this scheme enables the prognostic model to achieve a reasonable match against the novel observations for mid-sized floes (100 m – 2 km).'

P6L24: "Clearly, brittle fracture events…size." I agree that this is very likely the case for the larger floes (>100m), but more uncertain for smaller floe size. Maybe add a comment on that (if you agree)?

Following sentence:

‘Clearly, brittle fracture events can have a direct impact on floe size.’

Has been updated to:

‘Clearly, brittle fracture events can have a direct impact on the size of larger floes and potentially also smaller floes.’

P7L18: “stronger regions of sea ice” I am not sure of what you mean by this expression. Regions where the ice is thicker/stronger?

Text has been modified to clarify that stronger in this context effectively refers to thicker ice.

P8L17: “it marks…larger floes”: I find the sentence a bit confusing.

Original clause has been clarified. Original clause:

‘it marks a transition from a regime where floes are being broken up to a regime where the number of floes is increasing due to the break-up of larger floes.’

Now reads as follows:

‘Wave break-up acts to reduce the number of larger floes and increase the number of smaller floes; l_{var} effectively marks the boundary between these two contrasting effects.’

P15L35: “This is an important…models.” Yes, but I suspect the quantities you are looking at remain quite constrained by the ocean and atmosphere forcing. I am therefore not certain the conclusion you draw from a stand-alone simulation can be extended to fully coupled climate models. I am aware this is discussed a bit further in the text, but I think it should be mentioned here too.

The following clarification has been added to the end of the relevant sentence: ‘though this conclusion needs to be confirmed using fully coupled climate simulations.’

P18L30: Same comment. Maybe replace “is” with “may”.

Corrected as suggested.