In this document we have compiled our response to both referee and editor comments for the manuscript entitled, 'Sea ice floe size: its impact on pan-Arctic and local ice mass, and required model complexity'.

This document includes the following:

In this document referee / editor comments are shown in black, our response is in blue, and descriptions of changes to the manuscript are shown in red. Red is also used to indicate where the updated manuscript has been modified from the original submitted manuscript.

Response to Anonymous Referee #1

Firstly, we would like to thank the reviewer for providing such thorough and thoughtful comments on our manuscript. They have been invaluable in improving the manuscript.

This is a review of the manuscript entitled Sea ice floe size: its impact on pan-Arctic and local ice mass, and required model complexity. In this manuscript, the authors investigate the differences between two ways of representing the floe size distribution (FSD) in sea ice models. The first approach uses a prognostic floe size model, in which floe size evolves freely depending on some physical processes (breakup, lateral melting, welding…). The second approach is simpler, as it constrains the floe size distribution to always obey a truncated power-law. Only the upper-limit of this power-law varies with processes affecting the floe size. After having described the two models and their implementation in a stand-alone version of the sea ice model CICE, the authors evaluate the simulated floe size distribution against available observations in the summer. They show that the original prognostic model leads to unrealistic results in the absence of a process able to break the largest floes considered in their FSD. They suggest this process corresponds to brittle fracture of the ice, and that it can be represented by relaxing the FSD in the prognostic model towards a powerlaw. They further investigate the impact of the two FSD models on sea ice extent and volume in Pan-Arctic simulations. They find little evidence of any significant improvement of model results related to the addition of FSDs in the sea ice model. They discuss the differences between the two FSD models, as well as their advantages and drawbacks.
The manuscript is overall easy to read, and the method followed by the authors is rigorous and well explained. I acknowledge the quality of the work that has been done, but I think there are a few problems to fix before I can support the publication of the manuscript.

General Comments:
The manuscript is in general well-written and not very long, however I find that some topics are repeated and too much time is spent describing results that are, in my opinion, not key to the study. I find it particularly detrimental to the potential impact of the study, as it makes the paper confusing in places, and the interesting findings and discussions are a bit lost among things that have already been long discussed in previous studies. I think there is potential for this manuscript to address the whole sea ice community, but in its current state I don't believe anyone not familiar with FSD modelling would get what the key findings are. I will try to highlight these problems and suggest ways of improving the manuscript in my specific comments.
My second main comment concerns the brittle fracture mechanism. This process occupies a large amount of the manuscript, however I am not fully satisfied with the way it is presented and discussed. I have the feeling (but I might be wrong, in which case I am sorry) that the authors found out during their evaluation against observations that a mechanism was missing in the prognostic model to break the largest floes into floes of "mid-range" sizes. They realized that observations were showing a power-law behaviour, and therefore improved the prognostic model by adding a relaxation towards this power-law. To explain this behavior, they suggest that brittle-facture is a good candidate, even though it has not really been demonstrated before for this spatial scale. However, the way it is presented in the paper is confusing, mixing LKFs in pack ice, fragmentation by waves and scientific intuition. The study does not do enough to justify the use of this relaxation in the winter in my opinion. I also find the discussion about this process a bit shallow, particularly as it is a major change compared to the prognostic model used in Roach et al. (2018). My recommendations would be a) to present this new process more carefully, i.e. introduce it as a relaxation of the prognostic FSD

towards a power-law, b) motivate this introduction by the fact that the prognostic model fails to reproduce observed FSDs in the summer without this relaxation, c) discuss what this relaxation represents (and I agree with the authors that brittle-failure is a good candidate), and when and where it should be applied. I will also explain these recommendations and my criticism in more detail in the specific comments.
Thank you for these overall comments regarding the manuscript. To address the concerns regarding the potential impact of the study, we have made significant edits to both the discussion and conclusions to highlight the novel aspects of this study and to clarify why the results are relevant to the broader sea ice community. We have also moved the motivation for the brittle fracture scheme out of the introduction and into section 2.2.2 as suggested in the specific comments. To address concerns regarding the brittle fracture scheme, we have made significant edits to section 2.2.2, including several new references, to more clearly explain and motivate why we introduced this scheme and the model choices made. This includes a more careful explanation of the physical interpretation of the restoring timescale used in the scheme and the limitations of this scheme. We have also reviewed how brittle fracture and the brittle fracture scheme are presented in the results, discussion, and conclusions in respect of the modifications made to section 2.2.2. Full details on the modifications to the manuscript described above can be found in the response to the specific comments below.

Specific comments (major):
P2L28 to 44. This is the first introduction to the brittle-fracture process. It is quite long, and for a reader that has not read the full paper yet, I believe the relationship with the rest of the introduction is very obscure. I also believe this level of detail would fit better in section 2.2.2. The paragraph first explains in detail the mechanism of LKF formation in pack ice, then switches to Perovich et al. (2001) that relate floe breakup to melt as thin ice is very weak, and finally refers to Kohout et al. (2016) who report flexural failure (and not in-plane failure) of ice under wave action in places where ice strength is minimal. I think I get that the authors want to say that processes responsible for LKFs at scales >1km might affect floe size at scales <1km, and that heterogeneity in the ice strength/thickness exists at these scales that would ease brittle fracture, but this does not appear explicitly in the text. I also think the mechanical behaviour of sea ice depending on the spatial scale of interest is an open question, which should appear more clearly in the text. Details about the spatial scales discussed in each reference and the one of interest for the study are missing.
As suggested, the full introduction to the brittle fracture process has been moved to section 2.2.2 (see the first two paragraphs in section 2.2.2 within the updated manuscript). The text has also been modified to address the comments above including providing details of the spatial scales considered in the references, a discussion of the scale variability of brittle fracture processes in sea ice, and that gaps in understanding remain within this topic.

I also find the conclusion ("These observations and model studies collectively suggest that brittle fracture processes impact floe size in winter and that the resulting pattern of linear features resulting from brittle fracture may also influence floe breakup in the subsequent melt season.") quite far-fetched and not well motivated with these references. Unless the manuscript addresses sea ice rheology at floe scale (which is not the case to me), I find this statement too strong to belong to the introduction.
In this introduction, I would recommend only mentioning the fact the prognostic model used by Roach and others has not been thoroughly evaluated against FSD observations, and that some processes might be missing. For instance, brittle fracture of ice might occur as the ice is thinning in the summer (Perovich et al., 2001).
We have moved the paragraph that discusses brittle fracture to section 2.2.2. Instead, as suggested, we add the following to the existing paragraph discussing the state of FSD modelling:

‘The limited spatial and temporal coverage of floe size observations has prohibited effective evaluation of these models, though there have been recent efforts to develop satellite derived FSD product to enable such evaluations (Horvat et al., 2019). It is nevertheless anticipated that important processes are not yet represented in these models. For example, thermodynamically-driven break-up of floes along existing cracks and refrozen leads in the sea ice cover (Perovich et al., 2001).’

P4L4->L10: The lateral heat flux…
I find this paragraph a bit hard to follow:
a) I don't understand why it is important to explain how heat fluxes are dealt with in CICE here. Please inform the reader of their use in this study. Also, it would be nice to highlight if this is a change from the “standard” in CICE, or if these are all default settings (as is done well further in the text for other code changes).
One of the two ways in which CICE has been adapted to use an FSD model is by modifying how lateral melt is calculated (the other is the calculation of momentum exchange between sea ice, ocean, and atmosphere via the form drag scheme). As such, we believe it is important to provide complete details of how lateral melt is treated in standard CICE, to make it clear how this treatment is different in the modified setup.
The final sentence of the first paragraph in section 2.1.1, ‘An overview of lateral melt treatment within CICE is presented here.’ is replaced with, ‘Below we provide an overview of features of standard CICE that are pertinent to the evaluation of the lateral melt volume. In section 2.1.4 we will explain how this standard treatment is adapted for use with an FSD model.’
Additional sentences have also been added to the introduction and the opening of section 2 to add clarification regarding the importance of the lateral melt treatment in this study.

b) F_frzmlt is computed as…The authors might want to write the equations instead of describing them in the text, that would likely improve the readability.
Manuscript updated as suggested.

c) Why is F_frzmlt capped, and why is it important?
This is a component of the standard CICE formulation for calculating the lateral and basal met volume. We have mentioned this as it is a component to how lateral melt volume is calculated in the standard CICE model. This cap is sufficiently large to be effectively immaterial to any calculations.
The modifications made to point (a) above also apply here (i.e. clarifying the importance of explaining the treatment of lateral melt in standard CICE).

P5L12: Please clarify the definition of l_eff. What do you call “the perimeter of a FSD”?
Original definition, ‘l_eff is the floe diameter that has the same perimeter per unit sea ice area as a given FSD‘, has been replaced with, ‘l_eff is the diameter of the set of identical floes that has the same total perimeter as a set of floes of variable size with the same total ice area’.

P5L16: You might want to explain briefly why l_eff is better than the average floe size (I believe this is mentioned somewhere else in the text, but it would fit nicely here).
The following clarification is added: ‘l_eff is applicable here because the lateral melt volume is proportional to the total floe perimeter.’

P5L28: Note that the in-ice… I don't understand this sentence. Could the authors clarify the difference between their parameterization and the one by Roach et al. (2019)?
The following section:
‘Note that the in-ice wave scheme used by Roach et al. (2018) has been adapted here to calculate H_m0, the spectral height parameter, and λ_p, the wavelength corresponding to the peak wave

energy, within the sea ice-covered grid cells for use with the wave-dependent floe formation parameterisation.'
Has been replaced with:
'Unlike Roach et al. (2019), we do not use a separate wave model coupled to CICE to calculate the necessary wave properties within the sea ice-covered grid cells for use with the wave-dependent floe formation parameterisation. Instead, we adapt the scheme used in Roach et al. (2018), which calculates in-ice wave properties using an extrapolation from forcing external to the sea ice cover, to calculate the necessary in-ice wave properties.'

Section 2.2.2:
For a reader that has not been through the Results section, the motivations behind the addition of this process remain very obscure. I think part of the motivations currently in the introduction (e.g Perovitch et al., 2001) would fit nicely here. Hinting at the results a bit would also help. Stating that the prognostic model was found to fit poorly with observations without this model would really help to understand the addition of this process. The authors should at least reassure the reader by saying that they are going to investigate the effect of this addition by running two experiments, one with this modified prognostic model, and one without it.
We have moved an updated version of the motivation of the brittle fracture scheme from the introduction to section 2.2.2 (first two paragraphs). We have also included the following sentences at the start of section 2.2.2 to explain the need to introduce new model physics to the prognostic model:
'It will be shown in section 4.1 that the prognostic model struggles to capture the shape of the observed FSD for mid-sized floes. Sensitivity studies show that it not possible to modify existing parameterisations in the prognostic FSD model to substantially improve model performance against observations (Bateson, 2021). This suggests there are important processes currently not represented within the prognostic model.'
The section also now concludes with the following line:
'Results will be presented in section 4.1 to demonstrate that the inclusion of this new brittle fracture scheme significantly improves prognostic FSD model performance against observations in simulating FSD shape for mid-sized floes.'

This condition means… It could be worth giving the physical interpretation of this sentence, as not all readers are familiar with FSDs.
The following clarification has been added: 'i.e. only when the ratio of larger floes to smaller floes exceeds a given value'

Fracture events occur regularly through autumn, winter and spring within the pack ice to form linear features like leads, which subsequently freeze up again…
I find the current discussion quite vague, and not very well linked to other references in the literature. The way I understand it, the authors assume that at subgrid scale, leads create a network of cracks that define floes in a kind of mechanical strength sense, even though they are separated by thinner ice and not open water. Distribution of these floes is assumed to follow a power-law of exponent -2 as it results from successive fracture events. Between these floes there are therefore weak joints made of thinner ice that will melt faster in the summer than the surrounding ice, to the point where their strength will become very weak (Perovitch et al., 2001). The idea here is therefore that sea ice has a memory of fracturing events, is this correct? If so, this is not so different from the definition of damage in brittle rheologies (cf. the work of Weiss that is already cited, and the models of Dansereau or Rampal, see references at the end of the review), and it could be worth commenting on that, if not here, maybe in the discussion. If not, the text needs to be clarified to make clearer what

the actual point being discussed is.
The summary above is indeed consistent with our thinking of the purpose of the brittle fracture scheme. We have made updates to the relevant text to make these points clearer.
The relevant paragraph in section 2.2.2 now reads as follows:
'A value for the restoring timescale, τ, needs to be determined. Both direct and indirect mechanisms have been discussed above describing how brittle fracture can impact the sea ice cover. Fracture events occur regularly through autumn, winter and spring within the pack ice to break up floes and form features such as leads, though these generally freeze up again. The result of these fracture events is to create a network of linear features that define weaker regions of ice interspersing stronger ice. Idealised models of brittle fracture suggest that the size distribution of the stronger regions of ice follow a power law with an exponent of -2. The linear features are then vulnerable to increased thinning and melting, increasing the likelihood of break-up along these features during late spring and summer as the sea ice retreats. This effectively 'releases' the floe size distribution defined during brittle fracture events outside of the melt season. It is this second mechanism that is of more relevance when considering the impacts of the FSD on the seasonal retreat of the Arctic sea ice.'
We thank the reviewer for highlighting the similarities between the brittle fracture mechanism suggested and the concept of damage in brittle rheology. This is an interesting comparison and worth noting in the manuscript.
We have added the following discussion to the end of section 5.1:
'An interesting comparison can be made between the treatment of brittle fracture within the prognostic model presented here and recent developments introducing the concept of 'damage' to the treatment of rheology within sea ice models (Dansereau et al., 2016). One such sea ice rheology, named the Maxwell-elasto-brittle (Maxwell-EB) rheology, has been applied within the continuous and fully Lagrangian sea ice model neXtSIM (Rampal et al., 2019). This new rheology retains a 'memory' of any fracture events, effectively tracking how 'damaged' the sea ice cover is, and modifies the sea ice properties accordingly. This concept of 'damage' has clear parallels with the mechanism discussed above of how winter in-plane brittle fracture events can determine how the sea ice breaks-up in summer and may therefore present a useful basis for the development of a full parametrisation of brittle fracture processes for use in FSD models. Boutin et al. (2021) also demonstrated that the Maxwell-EB rheology can be combined with an FSD model in order to explore how wave break-up of floes can impact sea ice dynamics, highlighting an application of floe size modelling not considered in this study.'

Freezing/floe welding/convergence will act as mechanical healing that will erase any memory of fracturing events, in the winter at least (e.g Rampal et al., 2016). How would it compete with brittle failure in your model? This point is quite important to me, as Roach et al. (2018) present how FSD processes are balanced in the prognostic model over the year, and here a modification is introduced that likely breaks this balance. In a sense, it also relaxes the prognostic model towards a truncated power law with upper limit the maximum floe size in the model, does it not? This goes against what I think is the original philosophy of the freely evolving FSD of the prognostic model but this fact is not really discussed here.
We assume when you refer to the 'balance' of the model you mean that the average ratio of small floes to large floes stays roughly constant over large timescales and there is not an imbalance in processes such that the model tends excessively towards floes that are too small or too large. Roach et al. (2018) demonstrated this by including some interesting and important figures demonstrating the tendency (both scale and sign) of individual processes in changing the total sea ice area within each floe size category. Obviously the introduction of a new process will perturb the balance and can influence the magnitude of other processes that impact the FSD such as the welding of floes, but in this case we can be confident that a reasonable 'balance' still exists since we compared model output to observations and the effective floe size metric shows significant spatial and temporal

variability (which would not be the case if very small or very large floes dominated the distribution). Regarding the second point, the FSD in the prognostic model including brittle fracture is still able to freely evolve since the brittle fracture scheme is applied at the scale of floe size categories rather than over the whole distribution. Whilst the brittle fracture scheme does, of course, influence the emergent FSD shape, sensitivity studies (not presented here but can be found in chapter 7 of Bateson, 2021) performed using the prognostic FSD model with brittle fracture show other processes continue to have a significant and comparable influence on the emergent FSD shape.

However, the brittle fracture-derived mechanisms operate over different timescales and scale with different properties. à This is a very vague sentence.
This point has been significantly expanded upon in section 2.2.2 to clarify the point being made and also to reflect some of the other edits made to this section (this should also address the point below about the value of tau in summer vs winter):
'The use of a fixed timescale makes it difficult to capture both the direct mechanism of brittle fracture impact on floe size, which dominates outside of the melt season, and the indirect mechanism via thermodynamic weakening, which is more important within the melt season. The latter mechanism has been prioritised in this case in determining the timescale given it is the FSD state in the melt season that is of primary importance for understanding FSD impacts on the Arctic sea ice (Bateson et al., 2020). Just considering the thermodynamic weakening mechanism, the use of a fixed timescale is still a simplification given the significant spatial and temporal variability of relevant factors to this mechanism such as melt rates, ice strength, and dynamic forcing.'

The timescale for a crack or linear feature in the sea ice to fully melt through is taken to be of the order of 1 month. For simplicity, τ is here set to 30 days.
I find it very confusing to relate the time scale of brittle fracture, which is a dynamical process occurring in very little time, to the time scale of sea ice melt. To me 30 days is not related to the fracture itself, but to the reduction of ice strength that tends towards 0 as ice melts, making it very sensitive to brittle failure. It would also be nice to a provide more quantitative details to the reader: what thickness/melt rate are you considering here to end up with tau=30days?
The discussion of τ in section 2.2.2 has been modified to clarify what it is supposed to represent and to provide qualitative details, as suggested:
'In this context, τ, the restoring timescale, refers to the timescale for the sea ice to thin sufficiently that the sea ice is vulnerable to in-plane fracture events along existing weaknesses. Sea ice thickness away from the ice edge at the start of the melt season is generally in the range of 1 – 3 m. Vertical melt rates are of the order of 5 – 15 cm day^{-1}. Therefore, significant thinning can generally occur over timescales as short as a week up to a couple of months. For simplicity, τ is here set to 30 days.'

Also, to me, this justification does not motivate the addition of the brittle-fracture process in the freezing season. With this motivation, tau should tend towards infinity in the winter. The lack of discussion about this process in winter is particularly detrimental to the study as modelled FSDs are only evaluated in the summer (or at least melting season) in section 4.1. The seasonal impact of this relaxation process should be discussed: does it overwhelm the floe size growth process in freezing conditions, or is it negligible?
As mentioned in the manuscript, this treatment of brittle fracture is supposed to be a simple approximation that has been included since, without this process, the prognostic model struggles to capture the shape of the distribution for mid-sized floes. As such, there are significant limitations with the treatment, including the use of a fixed tau throughout the year. We chose not to include these results here since the focus of the paper is intended to be the comparison between the two FSD models, but in Chapter 7 in the thesis from Bateson (2021), a series of sensitivity studies are presented with the version of the prognostic model used here including brittle fracture,

demonstrating that the new brittle fracture scheme does not dominate the shape of the distribution and other processes continue to influence FSD shape, including winter floe growth processes. This is an important point to at least mention within the manuscript, however.
The following comment has been included in the relevant discussion section i.e. 5.1:
'However, sensitivity studies show that the brittle fracture scheme does not dominate the shape of the emergent FSD and other processes continue to be important in the evolution of the FSD, particularly winter growth processes such as floe formation and welding (Bateson, 2021).'

Section 4.1

P9L39: Overall, the inclusion of the quasi-restoring brittle fracture scheme represents a significant improvement in the ability of the prognostic model to capture the shape of the FSD for mid-sized floes.
I agree, but you have only shown it in summer.
'Over the period May – July' has been added to the end of this sentence.

P9L42: […] does not include floes smaller than 100 m in the comparison, which are particularly important for determining the impact of the FSD on the sea ice mass balance.
These floes are important for lateral melting, but is lateral melting important for the mass balance? This is not what I retain from this manuscript, or from studies such as Bateson et al. (2020) and others, at least not for the Pan-Arctic mass balance.
Several studies, including Bateson et al. (2020, see case B in table 3) and Smith et al. (2022) have shown that where an FSD is dominated by smaller floes or where the fixed floe size is of the order of metres, the change in mass balance can become significant (orders of 10% change or larger in summer). However, we recognise that in the context of this study, it would be more appropriate to highlight the sensitivity of local sea ice properties rather than pan-Arctic properties, since we do find larger impacts on these scales.
'on the sea ice mass balance' is replaced with 'on sea ice concentration and thickness'.

P10L5: Whilst a reduction in ice area fraction in the largest category and an increase in the smallest category can be expected, the change in ice area fraction in the remaining categories depends precisely on the balance between ice area fraction lost from that category and ice area fraction gained from the adjacent larger category.
I find this sentence very unclear. Do you mean that, in the absence of brittle fracture, the very large floes (that are not used in the comparison with observations) occupy a significant fraction of the ice covered area in the model, and that prog-16-nobf demonstrates that a breakup mechanism of these large floes is missing in the original prognostic model ?
Intuitively, it would be expected that more breakup -> less larger floes and more smaller floes -> steeper negative gradient, but the results here show the reverse. The point being made here is that this intuitive response does not apply to a distribution since, non-withstanding the smallest and largest categories, all remaining categories have both a source and sink of floe area.
The following clarification has been added at the end of this section: 'In this case, the presence of the 'uptick' shown in Fig. 2 for the prognostic model without brittle fracture results in the source of floe area being larger than the sink for most floe size categories and a net reduction in gradient overall from including brittle fracture.'

Section 4.2:

To me, this section could be quite a bit shorter, given the few changes introduced by the FSD and shown in Figures 5,6,7. For instance, I am not sure that Figure 5 is really needed and even Figure 6 could be simply summarized in the text. As it is, I felt like I was reading details about

why the addition of FSDs in models is relatively useless, which does not really help to convince me of the interest of this paper. In my view, there are some topics addressed in the Discussion section that would deserve more highlights than the Pan-Arctic impact of FSDs, which is less than a small change in the ice albedo for instance.
P11L34 Previous studies e.g. Bateson et al. (2020) and Roach et al. (2018), have shown large FSD model impacts locally even where Pan-Arctic impacts are small.
Exactly! So it might not be worth the price of 3 figures.
The purpose of this manuscript is to provide a full comparison and assessment of the impact of the two approaches to modelling the FSD on the sea ice cover, and we believe that these figures provide important context for the comparison, even if what they show is a 'null result' in terms of the impact of either FSD model being significant for the considered metric. However, given the present length and large number of figures within the paper, we recognise the need to remove at least one figure. As such, we have decided to removed Fig. 6 from the manuscript, since we believe that this figure offers the least additional insight of Figs 5-7.
Figure 6 and any references to this figure have been removed from the manuscript. Figure numbering has been updated accordingly.

P11L19 Bateson et al. (2020) demonstrated...
They did, but so did and others before (It is reported in Tsamados et al., 2015, which already involve some of the co-authors of this manuscript).
Both Roach et al. (2018) and Tsamados et al. (2015) identified that the increase in lateral melt was compensated by a reduction in basal melt, but only Bateson et al. (2020) went on to show that this effect could be primarily attributed to the physical reduction of sea ice area in locations of high basal melt.
The relevant two sentences have been updated with additional references:
'Several previous studies, including Tsamados et al. (2015) and Roach et al. (2018), found that increases in the lateral melt volume resulting from higher floe perimeter were compensated by a reduction in the basal melt. Bateson et al. (2020) demonstrated that this compensation effect was shown to primarily be a result of the physical reduction of sea ice area in locations of high basal melt.'

P11L28: The similar magnitude of change in the total melt also means that the results shown in Figs 8 and 9, where the sea ice volume is lower in both September and March for prog-best compared to WIPo-best, are unlikely to be driven by an increase in the total melt.
This is a nice teasing of the discussion, but you should either discuss it here, or refer to the section where it is discussed, otherwise it is quite upsetting for the reader.
The following additional comment has been added: 'This point will be discussed further in section 5.2.'

Section 4.3.3 is interesting. It could be worthy of a bit more in-depth analysis (particularly if section 4.2 is shortened). For instance, why do you see a relatively low l_eff in the Chukchi/Siberian area for the prognostic model in March? Could you suggest what is driving this drop? L_eff in March with the WIPoFSD model is more like one would expect, with lower floe size found around the ice edge, where wave activity is strong. If the authors wanted to define a MIZ based on l_eff, they would likely find a pretty good agreement between this MIZ and the one based on wave-activity suggested in Horvat et al. (2020). That could be worth a mention, given the authors already refer to this study.
Section 4.3.3 has been expended to include the following comments, as suggested above:
'A good case study is the relatively low l_eff seen in the Chukchi Sea during March and June. The floe formation mechanism is important to FSD evolution in this region of the Arctic since it experiences ice-free conditions for at least part of the year. Higher wave activity is also expected in this region

due to an increased fetch via the Bering Strait, and this will increase the proportion of floes that form in smaller floe size categories. Other regions that experience ice-free conditions are generally more sheltered from wave activity due to adjacency to continental land mass. The only comparable regions in terms of wave exposure are the Greenland Sea and Barents Sea, where lower values of l_eff can also be seen.'
And:
'One point of interest here is the regions of reduced l_eff shown for the WIPoFSD model appear to correspond well with the MIZ defined using wave activity presented in Horvat et al. (2020), which suggests that a possible application of FSD models would be an alternative way of defining the MIZ compared to the sea ice concentration-derived definition.'

Results for the prognostic model also differ sensibly from the one shown in Figure 4 and 5 of Roach et al. (2018) manuscript. Could the authors suggest why?
Regarding the differences with Fig. 4 and 5 in Roach et al. (2018), the leading order difference appears to be that our setup shows a reduction in sea ice thickness across the sea ice cover rather than some regions of decrease and some regions of increase as shown in Roach et al. (2018). This makes sense given we have introduced the brittle fracture scheme, which will have a net effect of reducing the effective floe size and therefore increasing the lateral melt rate. It is difficult to accurately compare the magnitude of the changes due to the very different scales used by the two sets of figures, but broadly they are of comparable order.

Figure 11 shows much higher spatial variability in l_eff for prog-best compared to WIPo-best. Further analysis (not presented here) indicates the high spatial variability in $leff$ for the prognostic model cannot easily be attributed to a single process but is particularly sensitive to the floe formation mechanism, brittle fracture scheme, and welding, all processes not explicitly represented in the WIPoFSD model. Processes included in the WIPoFSD model, such as wave break up of floes and lateral melt, are not found to have a large impact on the spatial distribution of l_eff within the prognostic model.
I think this paragraph briefly addresses what is missing in the manuscript that, in my opinion, would really increase its significance. I don't know how far the authors can go in their analysis with the simulations they already have, but it would really improve the paper to discuss the contribution of the different processes a bit more. This is particularly true for the brittle fracture process. The manuscript in its current state sometimes sounds like a criticism of the prognostic model as used by Roach et al. (2018) but does not really show the impact of the changes they made on the model (except for the FSD in the summer).
The referenced paragraph has been edited in response to a previous comment to explain the region of reduced effective floe size in the Chukchi Sea in March and June. Whilst this is a single case study it does at least provide further details of how individual processes (in this case the floe formation mechanism) can influence the spatial distribution of the effective floe size. We appreciate the point that it would be interesting to explicitly evaluate the contribution of individual processes, but this has been a focus of previous papers (Roach et al., 2018, in the case of the prognostic model, and Bateson et al., 2020, in the case of the WIPoFSD model). Our focus in this study is to compare and discuss two alternative methods of modelling the FSD and given the significant differences in model structure and how individual processes are represented within the model we wanted to keep the focus on overall model performance and impacts. It is worth noting that chapter 7 in Bateson (2021) presents a series of sensitivity studies using the prognostic model (including the brittle fracture scheme) that explores the role of individual processes extensively but, as the length of this chapter should indicate, a complete presentation and discussion of these studies would be a paper in its own right. To be clear, the intention of this paper was never to be critical of the prognostic model as presented in Roach et al. (2018). We have added a comment to section 4.1 that we hope should make clear the purpose of evaluating FSD model performance against observations of floe size.

The following comment has been added to section 4.1:
'The purpose of this comparison against floe size observations is to ensure that the WIPoFSD and prognostic model setups used in this study perform comparably well to the same dataset. There are limitations to this evaluation of model performance, however. In particular, floes smaller than 100 m or larger than 2 km are not considered for the reasons outlined in section 3.2, and the former are especially significant for determining the impact of a given FSD on sea ice concentration and thickness.'

P12L30 Section 5.1
I feel like a lot of things in this section are repeated but not necessarily developed. I have already given several comments about the brittle-fracture mechanism. I think a lot of the problems I have with this addition could be solved with more clarity in its presentation.
Significant modifications have been made to this section to reflect the modifications made to section 2.2.2 in response to earlier comments. The points made in this section have also been developed to better explain their relevance.

Section 5.3
P14L26: I think the performance aspect, even though it is quite short, is very important for people that would like to use FSDs in the future, but that are not necessarily experts. It could be nice to highlight this a bit more, maybe by starting this section with this topic. I think it would also be fair to refer to the preprint of Horvat and Roach (2021), as it tries to address some of the shortcomings of the prognostic model.
The relevant paragraph has been moved to the start of the section as suggested. This section also now references the suggested preprint:
'It is worth noting that future advancements in modelling techniques may reduce or mitigate the computational expense or complexity of either model e.g. Horvat and Roach (2022) presented a machine-learning-based parameterisation to simulate wave break-up of sea ice floes that can replace the existing treatment of wave break-up in the prognostic model. The study found that CICE simulations including the prognostic model with this new parameterisation have an approximately 40% longer run time than CICE simulations without the prognostic model i.e. a comparable cost to the WIPoFSD model.'

P15L1 Section 5.4:
Lots of points are discussed, but they are a bit all mixed. Maybe cut this sub-section into paragraphs to clearly show the structure of the argument.
As suggested, this section has been separated into paragraphs and slightly edited to make the individual points being made clearer.

P15L19 Conclusion
The conclusion is a bit long. I think the significance of the paper would appear more clearly with a better hierarchy in the importance of the findings developed in this manuscript. To me, it is not clear what are the most important results according to the authors.
The level of detail in this conclusion is too high, a lot of things are repeated (motivation for the inclusion of the brittle-failure, no improvement of the models at Pan-Arctic scale, utility of l_eff…) that could be removed in my opinion.
Significant edits have been made to the conclusion to both shorten it, remove unnecessary repetition, and increase the prominence of the key findings in the study.

Future work should focus on the development of a full physical treatment of the impact of brittle fracture on the FSD.
Again, would it not be interesting to relate this with the work carried out on emerging brittle

rheology models (e.g. Dansereau et al., 2016)? Or to the work of Rynders that is already mentioned in the introduction? As it is, the paper does not really demonstrate the interest of using FSDs, which reduces its potential impact, at least in my opinion. Giving more context would highlight how the comparison made in this manuscript can contribute to the future of sea ice modelling.
As suggested, we have related this work to the work of Dansereau et al., (2016):

'Whilst the quasi-restoring scheme presented here is a useful tool to improve prognostic model performance and based on idealised models of brittle fracture, its current formulation relies on significant approximations. The concept of defining the 'damage' of a given area of sea ice such as used within the Maxwell-EB rheology (Dansereau et al., 2016) presents a promising basis for future developments of the brittle fracture scheme.'

Minor comments:

General:

I believe this is the first manuscript I have read that does not use a chronological order when citing 2+ references. This is not a big deal, but you might want to change that.

When citing multiple reference, we applied the same principles as used to determine the order of the reference list. Having reviewed the Cryosphere style guide, no clear guidance is given on this, and in fact the example given for multiple references is chronological rather than alphabetical.

As suggested, citations have been updated to be chronological rather than alphabetical.

The image resolution of the figures seems in general quite low to me. This is purely aesthetics, but it gives a "draft" impression of the Figures.

We believe this to be an artifact of importing the figures into a word document and therefore should not be an issue for the publication of any final paper.

Readability of graphs would also be improved with more ticks (Figure 4) or maybe a grid in the background (Figure 2,5,6,7,9).

Our personal preference is to not include grids within graphs since this can make figures appear cluttered and distort how they are interpreted.

Ticks have been added to Fig. 4 as suggested.

Aesthetics again, but the authors should consider using roman (normal) text to subscripts in equations/variable names when they have more than 1 letter. It improves the readability.

The Cryosphere journal style guide suggests that Microsoft Equation Editor should be used for equations and variables when compiling a manuscript with Microsoft Word. The font used for equations is the standard when using the Microsoft Equation Editor and cannot easily be changed. The font of equations / variables will also be updated during typesetting prior to final publication (if this paper is accepted).

I find the name "WIPoFSD" a bit hard to read (too long for an acronym, and not straightforward to pronounce). The authors might consider using a shorter name, or a name that would be related to a key property of this model (fixed-shape FSD model?).

The name WIPoFSD was defined in a previous paper (Bateson et al., 2020) so renaming it here would create an inconsistency in the literature.

Abstract:

P1L13à17: The beginning of the abstract would gain from being a bit more synthetic/sharper (until […] “In this study”.

The first few sentences in the abstract have been modified to improve flow and clarity:

‘Sea ice is composed of discrete units called floes. Observations show that these floes can adopt a range of sizes spanning orders of magnitude, from metres to tens of kilometres. Floe size impacts the nature and magnitude of interactions between the sea ice, ocean, and atmosphere including lateral melt rate and momentum and heat exchange. However, large-scale geophysical sea ice models employ a continuum approach and traditionally either assume floes adopt a constant size or do not include an explicit treatment of floe size.’

Introduction:

P1L36: The number of references for the mechanical response of sea ice to stress is quite high compared to the rest, given this is not the main topic of the paper. I acknowledge these references are relevant to this topic. My main concern is that as all these references are linked to only one team working on this topic, it gives a misleading picture of the field to the reader (see for instance the studies by Shen et al., 1986 a,b; Williams et al., 2017; Boutin et al., 2021…). As mentioned earlier, the link between this manuscript and these references could fit well in the conclusion, so maybe the authors should move some of these references there.

Original reference list:

Feltham, 2005; Rynders, 2017; Rynders et al., 2020; Wilchinsky and Feltham, 2006

Has been updated with:

e.g. Shen et al., 1986; Wilchinsky and Feltham, 2006; Rynders et al., 2020

P1L42: I was a bit confused by the word “province”. Whether it is correct or not, I would recommend using the word “region” that is clearer for international English speakers/reader.

‘Province’ changed to ‘region’ as suggested.

P2L2: “here” is a bit unexpected given the introduction has just started.

The ‘here’ is included since the MIZ does not have a singular, unique definition.

‘Here’ has been changed to ‘in this study’ to avoid confusion,

P2L9: “Note that all…” This sentence breaks the flow of the introduction a bit. The authors might want to move it a bit earlier or find a smoother way of stating this fact (just a suggestion obviously).

This line has been removed and FSD in the previous sentence replaced with ‘non-cumulative floe number density’ to improve flow without loss of information.

P3L4: “…FSD in the model is actively constrained according to observations, in this case by approximating the FSD as a power law.” All observations do not conclude that the FSD follows a power-law. Horvat et al., (2019) does not for instance.

‘According to observations’ has been removed to avoid incorrect inferences.

P3L6: "though with some dependency on model structure such as how the FSD is discretised over floe size categories." This has been addressed in previous studies I believe (Horvat and Tziperman, 2015?), it could be nice to refer to them.

Horvat and Tziperman (2015) do not appear to test sensitivity to model structure. Zhang et al. (2015) test model sensitivity to spacing of floe size categories for a different prognostic FSD modelling approach and do not find a significant effect (though this result may not hold for the prognostic model being used here). Given the significant uncertainty regarding this statement, I have removed it from the text.

P4L12: I find the beginning of section 2.1.2 a bit confusing. I suggest starting with one sentence to summarize why the MLD matters in your CICE setup. The second sentence of this paragraph would make a better start for instance. The way section 2.1.3 is introduced is much clearer for instance.

The opening three sentences to section 2.1.2 have been modified to the following:

'Ocean mixed-layer properties are important in determining lateral and basal melt rates, which are both relevant for evaluating the impact of floe size on the sea ice cover (e.g. Bateson et al., 2020). Here, a modified version of the prognostic bulk mixed-layer model of Petty et al. (2014) is used rather than a constant prescribed mixed-layer depth, to better represent sea ice-mixed layer interactions and feedbacks without the complexity and computational expense of a full ocean model.'

P7L1: Could you remind the reader of what l_var is (physically)? There are a lot of floe size names in this paragraph, it is quite easy to lose the reader.

The following line has been added to the relevant section: 'Bateson et al. (2020) suggested that l_var can be taken as representing the history of a given area of sea ice in terms of physical processes that affect the FSD.'

P7L9: "The broader impacts of a power-law distribution on the sea ice cover can be explored whilst also including spatial and temporal variability of the FSD within the model. For mechanical processes such as wave break-up, the use of $lvar$ is particularly suitable" I find these sentences a bit vague. The link with the rest of the paragraph could be more explicit.

The referred to sentences (in addition to both the prior and subsequent sentences) have been updated as follows:

'The appeal of this approach is that it is both simple and enables an exploration of the broader impacts of a power-law distribution on the sea ice cover whilst retaining spatial and temporal variability in l_eff. For mechanical processes such as wave break-up, the use of l_var is particularly suitable; 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.'

P7L13: "For thermodynamic processes it makes less intuitive sense. It is not possible to define two clear regimes; instead, floes across the distribution reduce in diameter by the same magnitude in response to a lateral melting event. Here, we have modified the lateral melting scheme to calculate the change in $leff$ rather than $lvar$, since it is possible to calculate exactly how $leff$ would change in response to a given perturbation of the FSD." Same comment here, I am afraid that a reader that is not familiar with FSD modelling would get quite confused. A few more details about the physical reasons behind these statements could improve the readability.

This section has been modified to add clarity:

‘However, it is not possible to define two clear regimes of how floe size would change in response to lateral melting; instead, floes across the distribution reduce in diameter by the same magnitude in response to a lateral melting event. Here, we have modified the lateral melting scheme to calculate the change in l_eff rather than l_var, since it is possible to calculate exactly how much the total floe perimeter, and therefore l_eff, would increase or decrease in response to any change in the FSD.’

P7L28: The reference for the CPOM CICE could be given here instead of further in the text.

Schröder et al. (2019) tested a series of different parametrisations and parameter choices within CPOM CICE, not just those adopted here.

P7L34: It is likely a very naive question, but why do the authors use a winter climatology?

Note the winter climatology is applied for deep ocean properties only. This is an approximation, but it is a reasonable approximation for two reasons. Firstly, changes in deep ocean properties occur over much longer timescales compared to the surface ocean. Secondly, mixed-layer deepening tends to happen during periods of freeze-up.

P9L21: Please give the spatial and temporal resolution of these datasets.

The following line has been added: ‘Both datasets have a spatial resolution of 25 km x 25 km and a temporal resolution of 1 day.’

P9L25: Another (important) reason PIOMAS is used as a reference it that it has been carefully evaluated against available sea ice thickness observations. See for instance:

Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok, 2011: Uncertainty in modeled Arctic sea ice volume. J. Geophys. Res., 116, C00D06, https://doi.org/10.1029/2011JC007084.

Thank you for pointing this out. The following clause has been added to the relevant sentence: ‘, it has been evaluated using available observations of sea ice thickness (e.g. Schweiger et al., 2011).’

P9L36: It would help to add the panels of interest in the references to Figure 4.

Where relevant to specific panels and not the full figure, the text in the manuscript has been updated to clarify the specific panels being referred to in Fig. 4.

P9L35: “In particular, the slope of the distribution is much steeper (more negative) for the model output than observations.” It would help to give a physical interpretation of this statement.

The following clarification has been added to the relevant sentence: ‘i.e. the model predicts smaller floes within the range 104.8 m - 1892 m take up a much larger proportion of the total sea ice area than is suggested by observations.’

P9L40: […] a significant improvement in the ability of the prognostic model to capture the shape of the FSD for mid-sized floes. In summer.

Following clarification added: ‘over the period May – July.’

P10L25: I am a bit confused by this “However,”.

The relevant comment has been removed since it was part of the discussion of Fig. 6, which has been removed from the manuscript.

P14L5àP14L13 Therefore… This is interesting, it would gain from being a bit clearer.

Section has been edited to improve clarity:

'This means that increasing the lateral melt contribution to the total melt increases the loss of thick ice in a given melt season. Vertical sea ice growth rates are inversely proportional to the sea ice thickness. Therefore, whilst the moderate reductions in thickness across large areas of sea ice from basal melt can be recovered within a single freeze-up season, the recovery of thick ice that has completely melted out from lateral melt will take several seasons of freeze-up to recover despite being over a smaller area. The reason for larger reductions in sea ice volume for prog-best compared to WIPo-best may therefore be a result of a changes to the ice thickness distribution that emerge due to the higher lateral to basal melt ratio for prog-best compared to WIPo-best.'

P15L15: "reduce the sea ice mass balance" This expression is confusing. The comparison with the results by Roach et al. (2019) in the next sentence is also a bit unclear to me.

'reduce the sea ice mass balance' has been replaced with 'result in a significant increase in total melt and a large corresponding reduction in sea ice volume.'

The following additional sentence has been added prior to the final sentence to clarify the comparison to the results in Roach et al. (2019):

'Conclusions regarding the role of the FSD in Arctic sea ice evolution also do not necessarily extend to the Antarctic.'

Caption of Figure 4: month(s)

Why is the "s" between brackets?

Some panels include data from a single month and others include data from several months.

"prog-16 performs particularly well in the Fram Strait and East Siberian Sea but less well for the Chukchi Sea. It represents a significant improvement to prog-16-nobf in all three locations."

I do not think this comment should be part of the caption.

Comment removed as suggested.

Caption of Figure 7:

All three simulations generally lie within the range spanned by the observational products except for pack ice extent in March after 2010.

I do not think this comment should be part of the caption.

Comment removed as suggested.

Figure 9:

It would be better to use the same vertical scale, at least for the extent (a,c) and volume (b,d).

As it is, it looks like differences between models are larger in March than in September.

Whilst we appreciate the point being made here, our intention with these plots is to present a comparison between the two simulations rather than the differences at different times of year and we have used the vertical scale for each subplot that we believe best serves this purpose. The range of the plots in September are about 4-5 times that in March. Therefore, using the same scales for each would result in the data points in March covering about a 20 % fraction of the figure, reducing the ability to clearly identify the differences between the two simulations.

Figure 12:

I was a bit confused by the use of the blue and red colormaps in section A, as these colors are later used to represent a positive/negative difference in sections B and C. I would recommend using the same colormap for all quantities that are not a difference, for instance the pink/purple colormap used for panels B(e,f) and C(e,f) could be used for all panels in section A.

We agree that the use of blue and red may lead to confusion with the difference plots. We have decided to use different colours to the pink/purple scheme used in B/C(e,f) to make it clear that different sea ice metrics are being presented in these different plots.

The red and blue colour schemes in Fig. 12A have been replaced with orange and green colour schemes respectively.

References:

Boutin, G., Williams, T., Rampal, P., Olason, E., and Lique, C.: Wave–sea-ice interactions in a brittle rheological framework, The Cryosphere, 15, 431–457, https://doi.org/10.5194/tc-15- 431-2021, 2021.

Dansereau, V., Weiss, J., Saramito, P., and Lattes, P.: A Maxwell elasto-brittle rheology for sea ice modelling, The Cryosphere, 10, 1339–1359, https://doi.org/10.5194/tc-10-1339-2016, 2016.

Horvat, C. and Roach, L. A.: WIFF1.0: A hybrid machine-learning-based parameterization of Wave-Induced sea-ice Floe Fracture, Geosci. Model Dev. Discuss. [preprint], https://doi.org/10.5194/gmd-2021-281, in review, 2021.

Rampal, P., Bouillon, S., Ólason, E., and Morlighem, M.: neXtSIM: a new Lagrangian sea ice model, The Cryosphere, 10, 1055–1073, https://doi.org/10.5194/tc-10-1055-2016, 2016.

Tsamados, M., Feltham, D., Petty, A., Schroeder, D., and Flocco, D.: Processes controlling surface, bottom and lateral melt of Arctic sea ice in a state of the art sea ice model, Philos. T. R. Soc. Lond, 373, 20140167, https://doi.org/10.1098/rsta.2014.0167, 2015.

Shen, H. H., Hibler, W. D., and Leppäranta, M.: On Applying Granular Flow Theory to a Deforming Broken Ice Field, Acta Mechanica, 63, 143–160, 1986. a, b

Williams, T. D., Rampal, P., and Bouillon, S.: Wave–ice interactions in the neXtSIM sea-ice model, The Cryosphere, 11, 2117–2135, https://doi.org/10.5194/tc-11-2117-2017, 2017.

Response to Anonymous Referee #2

Firstly, we would like to thank the reviewer for providing such thorough and thoughtful comments on our manuscript. They have been invaluable in improving the manuscript.

The manuscript by Bateson et al., "Sea ice floe size: its impact on pan-Arctic and local ice mass, and required model complexity" compares two of the main approaches for incorporating floe size distribution into a sea ice model (both using CICE) with observations in two ways. The first compares the floe size distribution with new FSD estimates from satellite imagery; the second is an evaluation of Arctic sea ice mean state over approximately the past 3 decades. With the newness of these models and the community focus on their implementation, this work is well justified. The manuscript is generally easy to read and complete. However, I have a number of concerns about how the comparison has been completed, and the presentation of the results.

Major comments:

Journal fit. This fundamentally is presented as a model evaluation study. Is The Cryosphere the appropriate venue for this? Would GMD perhaps be a better fit? If published in TC, the authors should more clearly address and center what new science is presented.
The Cryosphere has previously published numerous sea ice modelling-focused studies e.g. Bennetts et al., 2017; Bateson et al., 2020; Smith et al., 2022. We therefore strongly believe that this study focused on the numerical modelling of sea ice is also within the scope of The Cryosphere. In support of this we would also like to highlight that: (a) this is first study to motivate and incorporate an explicit treatment of brittle fracture into FSD models; (b) this is the first study that has compared different approaches to modelling the FSD in simulating the observed FSD shape; (c) this is the first study to present a direct comparison of the impacts on the sea ice cover of two alternative paradigms to modelling the FSD. Points (a) and (c) in particular highlight how The Cryosphere is a more appropriate fit for this research over GMD. In our revised version we highlight this novel science more clearly.

Comparison with FSD observations: I'm fundamentally a bit hesitant that floe size distribution models in global climate models are yet at a point where we expect them to match with observations (from specific location) as many other, more rigorously tested and developed features of models, can still not do so. To phrase this as a question: Why do you expect the models to represent realistic floe size distributions at a given point? Do you think there are other model factors that this representation is sensitive to, such as thickness distribution? Please discuss how other factors might impact this comparison
The points you raise are important and something that we have considered. You are entirely correct that there are several good reasons why FSD models may perform poorly in capturing the observed distribution at a given location that could result from entirely different aspects of the model (e.g. a poor representation of the ice thickness distribution, or in-ice wave fields). However, the reason we felt able to perform this comparison is because we found a remarkable consistency in the FSD plotted across the different locations and time periods considered. Figure A below shows all the observations included in our analysis included within the same figure. Whilst there is clearly variability between these different sites, the variability is still much smaller than the differences between the prognostic model without brittle fracture and the observations across all the case studies considered. We cannot expect any FSD model to precisely replicate an observed FSD, but we can expect that a simulated FSD should be within the variability in FSD shown by observations if an FSD model is accurately capturing the relevant processes. Figure 4 shows this not to be the case for the prognostic model without brittle fracture across all the case studies considered.

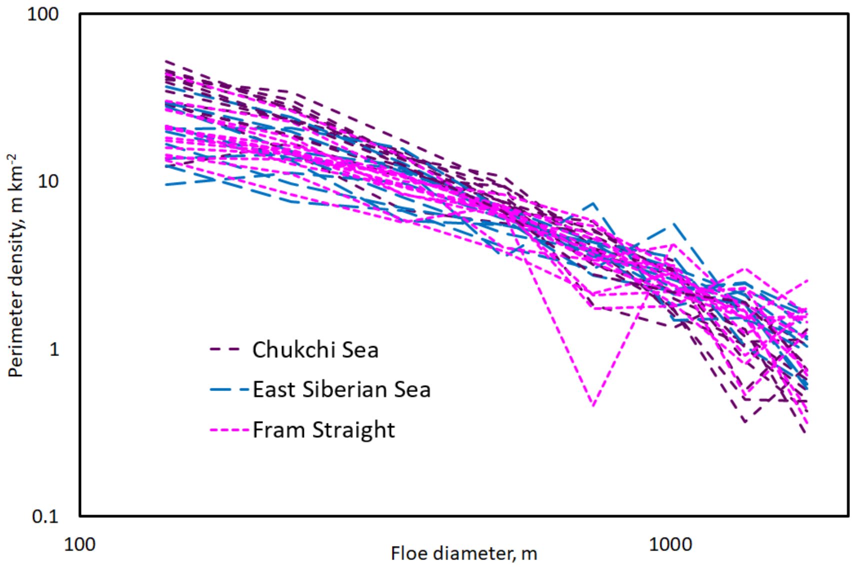

Figure A: A summary of the perimeter density distribution reported from 37 different LIDP satellite images in three different locations: Chukchi sea (plum, medium dash), Fram straight (pink, short dash), and East Siberian sea (blue, long dash).

The following comment has been added to section 4.1:
'When comparing observations across the sites considered in Fig. 4 there is clear variability between the different case studies, but this variability is substantially smaller than the differences between the prognostic model without brittle fracture and the observations across all the case studies considered. It cannot be expected that an FSD model can precisely replicate an observed FSD given other differences will exist between the model and the observed sea ice state such as ice thickness and concentration, but it can be expected that a simulated FSD should be within the variability in FSD shown by observations if an FSD model is accurately capturing the relevant processes.'

Comparison of sea ice mean state. It is worth noting that CICE, as most models, has had parameters largely tuned to best represent current state. As a result, the comparison of model with no additional tuning to observations (of sea ice extent and thickness) seems poorly motivated. Would we expect it to improve representation of mean state without tuning of other variables?
The general setup of CICE used in this study (when using a fixed floe size of 300 m) has been adopted from Bateson et al. (2020), where the reference setup was shown to perform well against observations of sea ice extent. Similarly, Figs 5-6 in this study show the reference state performs sufficiently against observations for our requirements. If the changes resulting from the inclusion of FSD processes are not overly large, we will remain within a realistic sea ice state and would expect to see a similar impact on sea ice state from the inclusion of FSD processes relative to the reference case after retuning the model. Where significant changes to the sea ice state can be seen, we can then refer to known biases in simulating Arctic sea ice (e.g. Ivanova et al., 2016; Notz and the SIMIP Community, 2020) and consider whether the changes to the mean sea ice state would counter these biases.

Notz, D., & SIMIP Community (2020). Arctic sea ice in CMIP6. Geophysical Research Letters, 47, e2019GL086749. https://doi.org/10.1029/2019GL086749.

Additionally, the implementation in a forced, standalone setup is likely to see less response in sea ice state, without the possibility of atmosphere and ocean feedbacks. What I think is more interesting is changes to sea ice mean state between models, which can suggest something about how different physics and processes relating to floe size feedback onto other sea ice characteristics. However, this is hard to see in a forced (rather than coupled) model, where feedbacks are limited. I think it is worth focusing on differences in the seasonal cycle and maps between the model which may impact these feedbacks, in the absence of coupled runs. Do these suggest improvement in how ice evolves? In short, I suggest that in the absence of additional, fully-coupled runs, the conclusions should be re-framed.

What we achieve in this paper is an understanding of direct FSD impacts on the sea ice state (i.e. via increases to lateral melt rate, form drag). A significant advantage of running standalone sea ice models for such studies is we can characterise these direct impacts on sea ice state and more easily identity the mechanisms responsible for these changes. Internal variability in a climate model e.g. of atmospheric heat content, would make distinguishing the impacts of FSD processes extremely challenging. Of course, the inclusion of feedbacks with the atmosphere and ocean may increase the impact of FSD models on sea ice state (or indeed reduce it) and we do acknowledge the absence of such feedbacks as a limitation of this study within the manuscript. The challenge we would envision with refocusing the paper onto how changes in the sea ice state might influence any feedbacks with the ocean and atmosphere is we would not be able to conclude on the magnitude of any changes to feedbacks, so the conclusions reached would be very speculative.

I was left confused how what appears to be substantial, meaningful changes in sea ice thickness in Fig. 11 (5-50 cm across much of the Arctic) corresponds to almost no change in volume in Fig. 5. Perhaps it is just an interpretation error on my part, but the presentation needs to be clarified to illuminate whether there are meaningful changes in ice state, or not.

It is useful to also consider Fig. 8 here e.g. Fig. 8 shows an average reduction in volume of about 4 % for *prog-best* compared to *ref* over 2000 – 2016. Fig. 5 shows the average September volume over this period was about 7.5 x 10^3 km^2, so 4 % of this would be about 0.35 x 10^3 km^2. Given the scale for the relevant panel spans about 3 – 15 x 10^3 km^2, the spacing between the three simulations can be expected to be very small. Figure 5 is not a useful figure for interpreting the differences between the simulations; it exists to compare all three simulations to observations. A 4 % reduction in volume would correspond to a 4 cm average decrease in sea ice thickness for a sea ice cover of a fixed 1 m thickness, and whilst these numbers represent very rough averages, they do indicate that the scale of change shown in Fig. 11 is consistent with the results in Fig. 8.

Minor/specific comments:

P1, L1: It would be helpful to provide a brief introduction to the range of floe ice sizes and key processes (and why it is useful to capture it with a distribution, as is often done for thickness)

As noted below, these points are addressed later on in the introduction.

P1, L16: I would suggest that the sentence beginning with "Observations show…" should go before the sentence beginning with "Large-scale…" on L15.

Corrected as suggested.

P1, L36: Is cluster of sea ice into larger floes really a process impacted by floe size, or is it primarily a process in determining floe size?

A quote from the discussion / conclusions of Herman (2012): 'As shown here, clusters of ice floes have a power-law size distribution with an exponent α dependent on (and generally different from) the exponent α_r of the FSD. Assuming the clustering-freezing scenario as one of the floe-generation mechanisms, this should lead to changes in the FSD exponent.' i.e. clustering of sea ice is both impacted by and impacts floe size.

P1, L37-38: May be worth considering additionally/alternatively citing Keen et al., 2021, which summarizes CMIP6 sea ice models and show that most use some derivation of CICE or LIM, which all have the same lateral melting parameterization
Additional reference added as suggested.

P1, L39: Floe size is also not considered in dynamics.
Following is added to sentence: 'or dynamics (Tsamados et al., 2014)'.

P2, L5: Why? Need to briefly describe the floe size distribution to justify why a power law is used (i.e., that typically more small floes). Replace "…generally fitted to a…" with "…summarized by fitting to a…"
P2, L17: Please note to what degree power law does/doesn't fit observations summarized. Is it a manner of convenience, or do observations support its use?
The opening of the relevant paragraph has been modified to address the above points. It is worth noting that the points raised above are not trivial questions, and indeed whole papers have been written that explore these questions (e.g. Stern et al., 2018b; Horvat et al., 2019).
The opening of the relevant paragraph now reads:
'Observations of the floe size distribution (FSD) show a large ratio of smaller floes to larger floes; this distribution of floe sizes is often summarized using a truncated power law (Rothrock and Thorndike, 1984; Toyota et al., 2006; Perovich and Jones, 2014; Stern et al., 2018b). Studies generally show that a power law produces a reasonable fit to the observations presented, though the validity of using a power law to fit floe size data remains an open question (Stern et al., 2018b), with several studies disputing the extent to which a power law is a good description of the FSD (Herman, 2010; Horvat et al., 2019; Herman et al., 2021).'

P2, L23-27: as noted above, it would be helpful to mention these processes earlier in introduction
The primary purpose of the first paragraph of the introduction is to highlight the important of floe size to the wider Arctic system, and we are concerned that describing processes that influence floe size within the paragraph would detract from this important point.

P2, L28: It would be helpful to include a transition sentence motivating introduction of brittle facture – that it is missing in most models, and may be important. Perhaps something like what is currently L42-43.
In response to comments from the first reviewer, we have moved most of the discussion of brittle fracture previously presented in the introduction to section 2.2.2. The updated reference to brittle fracture in the introduction addresses the above point.
Brittle fracture is now briefly introduced in the introduction in the following way:
'The limited spatial and temporal coverage of floe size observations has prohibited effective evaluation of these models, though there have been recent efforts to develop satellite derived FSD produces to enable such evaluations (Horvat et al., 2019). It is nevertheless anticipated that important processes are not yet represented in these models. For example, thermodynamically-driven break-up of floes along existing cracks and refrozen leads in the sea ice cover (Perovich et al., 2001).'

P3, L3 and L4: replace "represents" with "is in"
'Represents' has been replaced with 'is within'.

P3, L32: Would be helpful to also introduce the ITD, which is referenced in relation to the prognostic model
Following description of ITD has been added to section 2.1.1:

‘The standard sea ice thickness distribution in CICE distributes ice area between five thickness categories, with the spacing increasing for thicker categories. The ice area in a given category evolves in response to dynamic and thermodynamic processes according to a linear remapping scheme (Lipscomb, 2001).’

P4, L11: Is there any possibility for ice-ocean feedbacks, such as albedo feedback, in this setup? Please specifically address in the text

In this study we use a modified version of the prognostic bulk mixed-layer model of Petty et al. (2014) rather than a constant prescribed mixed-layer depth. In this model, mixed-layer temperature, salinity, and depth all evolve in response to sea ice-ocean interactions. A version of this model (with some additional physics related to snow on sea ice) has been used in a previous study to evaluate sea ice-ocean feedbacks in Antarctic shelf seas, including the albedo feedback (Frew et al., 2019).

The introduction to the mixed-layer model in section 2.1.2 has been modified accordingly:

‘Here, a modified version of the prognostic bulk mixed-layer model of Petty et al. (2014) is used rather than a constant prescribed mixed-layer depth, to better represent sea ice-mixed layer interactions and feedbacks (e.g. the ice-ocean albedo feedback) without the complexity and computational expense of a full ocean model (e.g. Frew et al., 2019).’

Frew, R. C., Feltham, D. L., Holland, P. R., and Petty, A. A.: Sea ice – Ocean Feedbacks in the Antarctic Shelf Seas, J. Phys. Oceanogr., 49, 2423–2446, https://doi.org/10.1175/JPO-D-18-0229.1, 2019.

P4, L24: It would be helpful to be more clear in the description of CICE and general model that this is being used in a standalone setup.

Section 2.1.1 now opens as follows:

‘In this study, we model the Arctic sea ice cover using a local version of the CICE sea ice model in a standalone setup.’

P4, L39: What is the L used in standard implementation? Is it 300 m, as used in lateral melt? Please define.

In the Tsamados et al. (2014) implementation, L is defined as a function of local sea ice concentration. We do not want to explicitly define this parameterisation here since this parameterisation is not used in this study. Instead, the following clarification will be added.

We have added the following clarification to the text: ‘which in Tsamados et al. (2014) is calculated as a function of sea ice concentration as per the parameterisation outlined in Lüpkes et al. (2012).’

P5, L19: In the introduction you present the WIPoFSD before the prognostic model. It would be helpful to be consistent about the order throughout manuscript. I suggest presenting WIPoFSD, then prognostic model, the brittle fracture scheme.

We have reviewed the manuscript to ensure the two FSD models are presented in a consistent order.

P6, L31: change to “spatial and temporal scales” or “spatial scale and timescale” or similar

Changed as suggested.

P7, L1, 11, etc.: I’m not really sure I understand the physical implications of l_var. What is it intended to represent? What are the implications for observational comparisons?

Modifications have been made to section 2.3 to better address these points. Bateson et al. (2020) suggested that l_var can be taken as representing the history of a given area of sea ice in terms of physical processes that affect the FSD. It is effectively a model tool to enable spatial and temporal variability of the FSD driven by relevant processes within the constraints of a fixed power law shape. As such there is no exact corresponding observable parameter to the model concept l_var, but it

roughly corresponds to the largest possible floe size that is likely to exist within an area of sea ice after that sea ice area has experienced a series of processes that change floe size.
The following clarification for l_var has been added:
'Bateson et al. (2020) suggested that l_var can be taken as representing the history of a given area of sea ice in terms of physical processes that affect the FSD.'
Further modifications have been made to section 2.3 to provide a clearer explanation of the physical implications of l_var.

P7, L36: Is there anything that can be referenced to demonstrate that ERA-Interim wave product is reasonable to use for the Arctic?
The following details have been added to section 3.1:
'The ERA-Interim reanalysis has been selected for the ocean surface wave field forcing as this dataset has generally been found to perform well in comparison to other reanalyses against observations of wind speed and wind speed profile in the Arctic during summer months (e.g. Jacobson et al., 2012; deBoer et al., 2014).'
What is the treatment for waves in sea ice?
The following clarification has been added to section 2.2.1 to describe the wave-in-ice treatment for the prognostic model:
'Unlike Roach et al. (2019), we do not use a separate wave model coupled to CICE to calculate the necessary wave properties within the sea ice-covered grid cells for use with the wave-dependent floe formation parameterisation. Instead, we adapt the scheme used in Roach et al. (2018), which calculates in-ice wave properties using an extrapolation approach from forcing external to the sea ice cover to calculate the necessary in-ice wave properties.'
For the WIPoFSD model, a summary is provided in Appendix A and full details are available in Bateson et al. (2020). The following line in section 2.3 has been modified to clarify this:
'A full description of how these processes are represented within the WIPoFSD model, including a description of the advection scheme for waves in sea ice, is available in Bateson et al. (2020); a summary has also been provided here in Appendix A.'

P8, L9: Why 'best'? Unless you plan to show others runs involved in selection process, I suggest use of just 'WIPoFSD' and 'prog' for simplicity and clarity
There are a couple of reasons we use 'best'. Firstly, to ensure these simulations are clearly distinguished from the simulations using 16 floe size categories. Secondly, this naming has been retained from Bateson (2021) to ensure consistency.

P8, L27: Perhaps simply "FSD observations" as title?
Corrected as suggested.

P8, L29-30: I believe since this is included in contributions and acknowledgements, it is not necessary to include funding or names of contributors here.
Corrected as suggested.

P8, L31: replace 'samples' with images
Corrected as suggested.

L31: remove "three months"
Corrected as suggested.

L39-40: What is the impact of this (as well as lower cutoff, L43+)? It seems like you could instead include the largest floes in the largest category, which may sometimes show a FSD with the uptick demonstrated by the prognostic model.

In the text the following explanation is provided: 'This step is necessary because the presence of a single large floe, comparable to the image size, can cause a large perturbation across the distribution reported for that location. Instead, only floe size categories that are small enough to consistently be populated by multiple floes across all sampled images are retained.' i.e. we do not include larger floes in the comparison due to the finite size of the images. Another way to consider this is the following. The prognostic model does not simulate the evolution of individual floes but instead floe area; what emerges from each grid cell is a statistical description of the FSD. In comparison, observations effectively produce a sample from that statistical population. If we compare these directly, we would not be comparing the same thing and the comparison would not be valid or useful. The FSD sample and statistical description will however converge for floe size categories in the sample that are sufficiently populated. Due to the power law shape of the FSD this is true for smaller floe size categories but not larger floe size categories, hence by applying an upper cut-off a useful and valid comparison can then be made. In terms of the lower cut-off, this is necessary since floes smaller than 100 m are not well-resolved in the observations and therefore there is no valid comparison that can be made between model output and observations for floes smaller than 100 m.
The text in this section has been updated to better address this point:
'The first step of processing the raw floe size data, consisting of a list of individual floe sizes, is to sort them into the Gaussian-distributed floe size categories used within the prognostic model for ease of comparison. Any floes that exceed the upper diameter cut-off of the largest category, 1892 m, will be discarded from the analysis. This step is necessary because the two models simulate the full FSD, and not individual floes. Floes large compared to the image size are inadequately sampled in observations to construct the full FSD. For example, the presence of a single large floe, comparable to the image size, can cause a large perturbation across the distribution reported for that location. Instead, only floe size categories that are small enough to consistently be populated by multiple floes across all sampled images are retained. A lower floe diameter cut-off of 104.8 m is also applied to this analysis, taken to be the smallest floe size that can be reliably resolved from the observations for the methodology and resolution used. The limiting factor on the smallest resolved floe size is the ability to resolve gaps between floes.'

P9, L11: change "is not novel" to "has been used previously"
Corrected as suggested.

P9, L27: Is there an appropriate reference for this statement?
Appropriate references have been added for this statement (and this section more generally):
'Whilst the PIOMAS volume product is a reanalysis and does not incorporate direct observations of the sea ice thickness, it has been evaluated using available observations of sea ice thickness (e.g. Schweiger et al., 2011). This product is often used to test model performance in simulating the total Arctic sea ice volume (Schröder et al., 2019) due to the challenges in estimating sea ice thickness from radar altimetry and limited availability of in-situ thickness measurements (e.g. Massonnet et al., 2012; Stroeve et al., 2012; Ridley et al., 2018).'

P10, L8: Perhaps simply "Comparison of sea ice extent and volume" might be a clearer heading
Given the later section 4.3.1 compares model output in terms of extent and volume, it is important to be clear that this section is explicitly a comparison of model output to the observed extent and volume.

P10, L1-3: It would be helpful to show some results of when and where the brittle fracture scheme is implemented. Where is it most necessary? What does this suggest about what it corresponds to physically?
The results in Fig. 4 provide case studies of the impacts of this scheme over different locations. We anticipate that the brittle fracture scheme is most necessary in regions that are part of the pack ice

in winter (i.e. subject to in-plane brittle fracture) but then transition to being within the MIZ during the melt season and therefore likely to be subject to significant thermodynamic-weakening of the sea ice cover.
Existing discussions in section 5.1 have been edited an extended to better address the above points. The 2nd paragraph in this section provides a more complete discussion of the different impacts of brittle fracture scheme for the case studies presented in Fig. 4. Paragraph 3 has been updated to give a clearer physical interpretation of the results presented in Fig. 4.

P10, L 18-20: I think it is necessary to clarify here that these runs are done in a standalone setting, and that the possibility for feedbacks in a fully coupled climate model may give different results.
The following clarification has been added to the end of the 2nd paragraph in section 4.2:
'though this conclusion does not necessarily extend to climate simulations where FSD impacts on sea ice feedbacks with the ocean or atmosphere could produce larger changes to the sea ice state.'

P10, L23-24: It's not clear what negative trends in the percent difference suggest. Does this suggest some sort of feedback in model?
The line being referred to is just highlighting that the Arctic sea ice extent and volume has decreased over the period 1990 – 2014, as expected. This point is perhaps unnecessary to make and the section has been modified appropriately.
The line, 'Clear negative trends in the March volume and September volume and extent can be seen.' has been removed.

P13, L14-16: Maps show more substantial changes in representation of sea ice state. Do these suggest improvements?
The following section later in the same paragraph was intended to address this point:
'Nevertheless, significant biases have been identified in coupled climate models in simulating the sea ice concentration (Ivanova et al., 2016) and CICE, in particular, has been shown to overpredict the sea ice concentration at the sea ice edge and underpredict the concentration within the pack ice (Schröder et al., 2019). In Bateson et al. (2020), the WIPoFSD model was found to provide a limited correction to this model bias. Similarly, Fig. 10 shows that the prognostic model produces a stronger correction to this model bias, driving reductions in sea ice area fraction in the MIZ and small increases in area fraction in the pack ice.'
Limitations in the accuracy of sea ice concentration and thickness data obtained from satellites preclude a more detailed comparison.
The following sentences have been added to the relevant paragraph to better address this point:
'The accuracy of sea ice concentration measured using passive microwave data can be as low as ± 20% in summer or the MIZ (Meier and Notz, 2010). Measurements in sea ice thickness from radar altimetry can also have high uncertainty, with snow depth and density being the primary source of error (Tilling et al., 2018).'

P14, L34: It may be worth mentioning that this is particularly relevant in a standalone sea ice model, as run here. In a coupled context (for which climate models are often used) the sea ice model is typically a small component of the total cost, and so the additional cost from the FSD is relatively not substantial.
We are inclined to disagree with this comment. Whilst the inclusion of an FSD model would clearly represent a proportionally smaller increase in run time for a climate model compared to a sea ice model, the pressure on model efficiency is also much higher for climate models than standalone sea ice models due to the existing bulk of the model.

P15, L6: What is meant by 'in-ice wave scheme'? Is it more accurate to say that the waves are forced with reanalysis?

The forcing defines the wave properties external to the sea ice cover, but both models have an internal in-ice wave scheme to determine the wave properties within the sea ice cover.
The text 'in-ice wave scheme' has been replaced with 'external forcing combined with in-ice wave scheme'.

P15, L19: Future work to address the impact on Antarctic sea ice representation and comparison with observations may also be useful
Following comment has been added to the final paragraph of the conclusions: 'In addition, it would also be beneficial to evaluate whether the conclusions reached in this study extend to the Antarctic.'

Figure 1: I find this figure really hard to interpret currently. A few reasons/suggestions… It might be better to use more realistic 'floe diameter' bounds, or to remove numbers from y axis, as it currently is hard to interpret these as actual bins. Only 2 examples are needed showing where redistribution is applied and where it isn't (for example, far left and far right). For one where redistribution is applied, show the new floe size distribution resulting more clearly. It would also be helpful to add lines showing actual density gradient for comparison to dashed purple line.
Figure 1 has been modified based on the above comments. Numbers have been removed both axes since these can produce an incorrect interpretation of the figure. One of the three examples has been removed. The redistribution is now shown in a separate panel showing the FSD before and after a brittle fracture event. We have decided not to use lines to show the actual density gradient to avoid adding too many details to the diagram.

Figure 2: I'm not convinced that this is necessarily a "non-physical feature" of the model, as it is simply capturing floes that are potentially beyond the bounds here, and is not reported as such in Roach et al., 2018. A comparison to observations without largest floes removed may be helpful to show if this is ever observed in observations. Additionally, please add a label to this figure demonstrating that it is only for areas of SIC 15-80%
By 'non-physical feature', what we mean here is that a sudden, discontinuous change in gradient from negative to positive in the perimeter density does not represent a physical behaviour of the FSD, and instead is a feature that emerges due to how the model is designed (either due to missing fragmentation processes or from imposing a fixed maximum floe size). In terms of comparing this to observations, as mentioned above, we exclude larger floes from the comparison due to finite-image size effects.
We have removed the 'non-physical feature' and replaced this with a more complete explanation as to why the uptick is described as 'artificial':
'Also highlighted in the figure by a blue transparent box is an artificial 'uptick', a non-physical feature of the model also reported by Roach et al. (2018a).'
Is replaced with:
'Also highlighted in the figure by a blue transparent box is an artificial 'uptick', a feature of the model also reported by Roach et al. (2018a) that results from prognostic model design and structure (e.g. missing fragmentation processes, upper limit on floe size) and does not represent a physical behaviour seen for the FSD.'
It is has been clarified in the caption that model output is averaged over areas with between 15 – 80% sea ice concentration.

Figure 3: Are these exact bounds of model areas? If not, a different symbol may better communicate that, as the boxes suggest that this is the exact selection of grid cells.
Yes, these are the exact bounds of the model areas.

Figure 4: Again, not necessary to name co-authors in the text.
Corrected as suggested.

Figure 5: The change of prognostic models being compared is a bit confusing. Is it worth including other prognostic models somewhere as well, to show if/that there is little difference in sea ice state? Also, as the 'prog-best' doesn't include brittle fracture (right?) it would be helpful to note that in the short name for clarity.

prog-best does include brittle fracture (it is identical to prog-16 apart from the number of floe size categories, as discussed in section 3.1). The purpose of using 12 floe size categories for the model comparison rather than 16 is due to the significant computational cost associated with simulating the additional categories (and this cost increases non-linearly with the number of floe size categories). The impacts of the FSD on sea ice cover via lateral melting and form drag are determined by the proportion of sea ice taken up by smaller floes (e.g. Steele, 1992; Tsamados et al., 2015), and therefore improving the resolution of the FSD shape for larger floes will not have a significant impact on sea ice state.

Table 1 has been updated to clarify that both *prog-16* and *prog-best* include brittle fracture. The following explanation has been added to section 3.1 to explain why 12 floe size categories are used in *prog-best* rather than 16:

'Whilst 16 floe size categories could also be retained for the prog-best simulation, the increase in model run time increases non-linearly with increasing number of categories. In addition, the improved resolution of the shape of the distribution for floes of a size of 1 km or larger is not significant when considering the impact of an FSD on sea ice via the floe edge contribution to form drag and lateral melt rate, which both scale to the inverse of floe size. Therefore, 12 floe size categories represents a more practical choice for the prognostic FSD model.'

Figure 5-7: This feels like a lot of plots to show for almost no change between any of the models. Can this be simplified to one or two key panels, and then state there is no observable change in others?

We appreciate the point being made here, but a key aim of this study is to establish the contexts where the impacts of the two different FSD models are either significant or can be distinguished. In order to do so, it is also important to discuss and give due prominence to contexts where the impacts are not significant i.e. a 'negative result'. However, given the present length and large number of figures within the paper, we recognise the need to remove at least one figure. As such, we have decided to removed Fig. 6 from the manuscript, since we believe that this figure offers the least additional insight of Figs 5-7.

Figure 6 and any references to this figure have been removed from the manuscript. Figure numbering has been updated accordingly.

Figure 8: I might suggest to swap these plots around to show both models and same subplot, with top for sea ice extent, bottom for volume. This would then allow to show some comparison in difference of observations from reference. (e.g., Are model changes moving it in the right direction?)

Figure corrected as suggested.

Figure 9: I am unclear how this figure is different from what is shown in Fig. 6, in terms of the take-aways. How do we know if this is improving the comparison if the scale of change is not comparable to the difference from observations?

As mentioned above, Fig. 6 has now been removed from the manuscript. Note that the purpose of Fig. 9 is to make a direct comparison between the two FSD models in terms of impacts on the sea ice cover.

Figure 11: I think this is the most useful and interesting plot! But, I'm quite confused how what appears to be substantial changes in sea ice thickness in A(f) agree with what is in Fig. 5 – where almost no change is observed.

This issue has been addressed in an earlier comment.

Some thoughts: Could the difference in fractional ice area/thickness compared to observations also be shown?
As mentioned above, the high uncertainty in observed sea ice concentration and thickness at grid cell scale would make reaching useful conclusions from such a plot challenging.

It would be helpful to place the effective floe size upfront (at the top) to set it apart from differences, and also make this more clear in the figure caption (meaning, that floe sizes are NOT a difference).
Suggested modifications have been made to figure and captions.

Figure 12: I'm not sure what to take from these plots, based on the units show. Would it be more helpful to show standard deviation as a percent of mean value?
The difficulty in presenting the standard deviation as a percentage of the mean is the dominant resulting signal will then just be where the ice is thin; it does not enable us to identify regions of high sea ice variability away from the sea ice edge. By presenting absolute changes in standard deviation alongside the standard deviation in the reference case, it is possible to identify both where notable changes in variability do occur away from the sea ice and whether this change represents a fundamental change in behaviour of sea ice in that region of the Arctic.
The following additional explanation has been added to explain why Fig. 12 is important to consider: 'Furthermore, a small change in mean sea ice state may disguise a much larger change in sea ice variability.'

Also, it would be nice to be consistent with the months shown in Fig. 11.
Whilst we agree it would be nice to include June in addition to March and September in Fig. 12, this would result in a total of 24 panels (rather than the current 16, and also much larger than the 18 panels included in Fig. 11). We believe that the size of panels in such a figure would be too small for effective interpretation, and as such we reluctantly made the decision to exclude results for June.

Table 1: "CPOM-CICE" is not needed in model description, as all are the same. It might be helpful to separate technical details into finer resolution categories, such as brittle fracture (yes/no), # floe size categories; d_min/d_max (where applicable)
We do not think it makes sense to add additional categories here given d_min / d_max and fixed l_eff are only applicable for one out of the five simulations described. Whilst 3/5 use the prognostic model, the differences are straightforward to describe (i.e. 12 or 16 floe size categories and whether or not the setup includes brittle fracture) and do not necessitate additional columns in the table.
CPOM-CICE has been removed as suggested. It has been clarified for all prognostic simulations whether or not they include the brittle fracture scheme.

References:
Keen, A., Blockley, E., Bailey, D. A., Boldingh Debernard, J., Bushuk, M., Delhaye, S., ... & Wyser, K. (2021). An inter-comparison of the mass budget of the Arctic sea ice in CMIP6 models. The Cryosphere, 15(2), 951-982.

Response to editor

Whilst we are responding to referee comments, we also wanted to take this opportunity to address some of the concerns raised by the editor during the initial quality control. We would like to thank the editor for taking the time to provide these clear and constructive comments on our manuscript, and these comments alongside those provided by the referees have been invaluable in helping us to substantially improve the manuscript.

We would also like to highlight that we have updated the 'competing interests' section of the manuscript to follow the format suggested by TC policy (https://www.thecryosphere.net/policies/competing_interests_policy.html):

'Daniel Feltham, David Schröder, and Yevgeny Aksenov are members of the editorial board of The Cryosphere. The peer-review process was guided by an independent editor, and the authors have also no other competing interests to declare.'

1. ORIGINALITY (Novelty): 2/3
The authors present two strategies to consider the floe size distribution (FSD) in a continuum sea ice model which directly affect the evolution of the sea-ice fraction as well as the simulated form drag. The first strategy is purely diagnostic and infers the FSD from the total sea-ice area via an empirical power law. For the second strategy, the FSD is determined prognostically considering processes that transiently change the flow size (e.g., lateral melting, breakup, welding). The performance is assessed in selected study sites in the Arctic Ocean, where recently declassified high-resolution imagery was analysed. The results suggest that the consideration of floe sizes has a negligible effect on the seasonal and inter-annual sea-ice extent and cannot resolve the difference to observations. Yet the seasonality in the sea-ice evolution is better presented in the process-based prognostic model as compared to the diagnostic approach. Considering observed floe sizes, the prognostic approach strongly benefits from an empirical restoring to a theoretical size distribution for brittle fracturing. Moreover, the prognostic approach exhibits higher inter-annual variability in the effective floe size. In terms of novelty, the authors combine existing theory and methods and evaluate their performance against 'novel' observations. So as a non-expert, I consider these a discrete methodological advance in sea-ice modelling. Yet the effect of these refinements on the overall results, together with the drawn conclusions, moderate my assessment on the originality.
Thank you for your comments above. Regarding the originality we would like to highlight that: (a) this is first study to incorporate an explicit treatment of brittle fracture into FSD models; (b) this is the first study that has compared different approaches to modelling the FSD in simulating the observed FSD shape; (c) this is the first study to present a direct comparison of the impacts on the sea ice cover of two alternative paradigms to modelling the FSD. In addition, the 'negative' result that the FSD has a limited impact on sea ice extent and volume is an important result for climate modellers since it assuages concerns of this source of structural uncertainty in climate models. In our revised version we highlight these novel aspects more clearly.

2. SCIENTIFIC QUALITY (Rigour): 1
In this regard, I want to congratulate the authors on having presented their methodology and results in a very concise way. I could very well follow the general idea and nuances in the methods and the experimental setup. However, there are a couple of points that I have identified which might be picked up during the review.

Terminology

— I am somewhat confused by how you address the two model types. While you are rather consistent in the usage of the term ‘prognostic’ model, you are less stringent in your reference to the WIPoFSD model. You also refer to it as the power-law variant, etc. It took me a while to get my head around this. Yet the brittle-fracture restoring in the prognostic model is also a power law. In my view, the key difference is the either the ‘prognostic’ or the ‘diagnostic’ character to determine the FSD. Couldn’t the latter label ‘diagnostic’ be used instead of WIPoFSD or power-law variant.

Thank you for highlighting the inconsistent labelling of the WIPoFSD model. We have reviewed the manuscript to ensure this model is referred to in a consistent manner. We have also included a new table to provide definitions for key terms used in this manuscript (e.g. WIPoFSD model, power-law fit). Regarding the term ‘diagnostic’, we are not sure that this would be an accurate way to describe the WIPoFSD model, since diagnostic implies a parametrisation where the FSD state is determined from co-temporal sea ice / atmosphere / ocean conditions, which is not the case for the WIPoFSD model. We think it is best to retain the WIPoFSD name for consistency with previous publications.

All references to the WIPoFSD model have been reviewed and modified where appropriate to ensure consistent naming. We have also introduced a new table (table 1), to provide clarity on different terms used within this manuscript.

— You use a lot of quantities for the FSD like number density, perimeter density, perimeter density distribution, thickness probability distribution, fragment size distribution, etc. As they are introduced for the respective modules where they are applied, I got quite confused. To facilitate the accessibility, you could optionally present all of them and their relations in a dedicated paragraph. We would like to the thank the editor for this suggestion. Since there are several terms related to the FSD that appear throughout the manuscript, we have decided to summarise these with a table that can easily be referred to.

A new table (table 1) has been introduced to provide clarity on key terms used within this manuscript.

Floe Size Variability
In figure 12, you present the floe size variability. As I understand it it is an inter-annual variability. If so, I have two question:

— In P12L18-19: You mention that the measurement of floe size variability might help you to differentiate between a ‘prognostic’ and a ‘diagnostic’ FSD model. Your observations in Fig.4 each cover 2 distinct years which might allow you to infer some effective floe size differences. You might need to come up with a strategy how to compute this metric from the observations (e.g., Horvat et al., 2019). Then you could give a first answer on which model is better suited. Otherwise explain why this might not be possible.

We believe this should be fairly straightforward to do given Horvat et al., 2019 calculated their observable floe size metric from linear floe size statistics, which corresponds well with the concept of effective floe size.

We have added the following clarification to the conclusion:

‘especially since the methodology of Horvat et al. (2019) involved collecting linear statistics of floe size, and effective floe size is a linearly averaged representation of the FSD.’

— You further mention in the introduction that the floe size distribution shows a seasonal dependence which is partially visible from Fig. 4. Would the 'diagnostic' power-law approach benefit from a seasonally changing exponent linked to climatic conditions. Could this help to increase the variability in the effective floe sizes?

This is an interesting idea, and in fact a seasonally changing exponent has previously been explored within the WIPoFSD model by Bateson et al. (2020). In this case, we wanted to determine WIPoFSD model parameters according to the observations presented in Fig. 4. Whilst these observations may hint at a seasonal evolution in the exponent, the variability between co-temporal datasets is sufficiently high compared to any trend in the exponent that we cannot be confident in this result. It is definitely a theme to explore in future however, as more observations are made available.

Brittle fracturing

In the prognostic model, you employ the brick fracture scheme (BF) which redistributes floes sizes towards smaller categories. Yet in Fig. 4, you show that after activation of this BF scheme, the perimeter density stays roughly the same for low-end floe sizes (around 100 m) but it significantly increases for the larger floe sizes. Why is that? Explain.

This point is addressed by the 2nd paragraph of section 4.1 within the manuscript (note this section has been edited in response to reviewer comments to provide further clarity to the points made here):

'It is worth commenting briefly on how the brittle fracture scheme can improve model performance compared to observations, given it is a counterintuitive result that increasing floe break-up would produce a shallower slope in perimeter density. As discussed in section 3.2, the largest floe size categories in the prognostic model are excluded from the comparison to observations to exclude the non-physical 'uptick' that forms (Fig. 2). Whilst a reduction in ice area fraction in the largest category and an increase in the smallest category can be expected, the change in ice area fraction in the remaining categories depends precisely on the balance between ice area fraction lost from that category (sink) and ice area fraction gained from the adjacent larger category (source). In this case, the presence of the 'uptick' shown in Fig. 2 for the prognostic model without brittle fracture results in the source of floe area being larger than the sink for most floe size categories and a net reduction in gradient overall from including brittle fracture.'

Effectively, larger increases in perimeter density are seen for the larger floe size categories in Fig. 4 due to their adjacency to the categories spanned by the 'uptick' in the model.

Uptick

You mention this spurious increase in perimeter density for the prognostic model. You try to circumvent this problem for the validation agains observation by increasing the number of size categories. Fair enough. I further understand that the brittle fracture scheme might alleviate this problem to some degree. Moreover, you mention that the an increase in the radius' r_max' (Eq. 9) could help, which is set to 1'700m. When comparing to the 'diagnostic' power-law model, a maximum diameter ('d_max') of 30'000 m is specified. The difference is a factor 10. Why not simply adjust 'r_max'. Explain.

In the prognostic model, the largest floe size category effectively acts as a repository for large floes. In order to resolve the shape of the distribution for these larger floes, more floe size categories are needed. This could be achieved in two ways, both of which have a significant associated cost. The first is to reduce the resolution of smaller floe size categories to improve resolution of larger floes, however higher resolution of smaller floes is more important here since these are most important in terms of FSD impacts on the sea ice cover. The alternative strategy is to increase the number of floe size categories, but the computational cost of the prognostic model scales non-linearly with the number of categories and this cost very quickly become prohibitive.

3. SIGNIFICANCE (Impact): 3
With my external perspective, I wonder about the key conclusions.

Can we ignore floe sizes if the interest was not in the seasonality but rather in general min/max sea-ice extents?

Based on the results presented, this is a reasonable conclusion, though there are limitations to this conclusion (see section 5.4 of the manuscript). As mentioned above, this in itself is an important result for climate modellers who must balance reducing structural uncertainty in climate models with maintaining computational efficiency.

We have updated the manuscript to ensure the point made above is clearly addressed in the discussion and conclusions. In particular, at the start of section 5.2, we make the following comment (regarding the impacts of FSD models on a pan-Arctic scale):

'This is an important result for climate modellers since it assuages concerns that the FSD represents a source of structural uncertainty in climate models.'

I wonder if the gain in the seasonal evolution is really significant by either FSD model.

Whether or not the impacts of the two FSD models on seasonal sea ice retreat are significant strongly depends on the research question. For example, if you are a climate modeller, the size of the impact of either FSD model on seasonal retreat of the sea ice is likely insufficient to justify the increased computational cost of the FSD model. Although, as noted above, this in itself is a useful result. Alternatively, if you are interested in regional sea ice modelling, especially the seasonal evolution of sea ice in the Greenland and Barents Sea, the results presented here do in fact indicate that FSD models have a significant role to play in the evolution of sea ice in these regions. The impact of both FSD models is particularly significant on the melt evolution (specifically the ratio of lateral to basal melt) and this can impact open water formation during the melt season (Smith et al., 2022). It is also important to note that we use a standalone sea ice model in this study and, as we note in section 5.4, larger FSD impacts may be found via feedbacks between the sea ice and ocean / atmosphere.

We have updated the manuscript to more clearly highlight the points discussed above in the discussion and conclusions. In particular, we make the following point at the end of the 2nd paragraph in the conclusion:

'These results are important for climate modellers as they suggest that the FSD is not a significant source of structural uncertainty in climate models. FSD processes will, however, be of importance for several key applications and research questions such as regional sea ice modelling and the formation of open water during the melt season (e.g. Smith et al., 2022).'

What are the main reason why the regional sea-ice area/volume are note captured by the continuum mode, if not the floe sizes?

This is a question that preoccupies many authors in this field and one which is very much beyond the scope of this paper (see for example Notz and the SIMIP Community, 2020). It is worth noting, however, that it is not necessarily internal model physics that limits sea ice model performance against observations but the representation of the atmosphere or ocean (in this case the model forcing).

Notz, D., & SIMIP Community (2020). Arctic sea ice in CMIP6. Geophysical Research Letters, 47, e2019GL086749. https://doi.org/10.1029/2019GL086749

How transferable are your methods. Are they very specific to the CICE implementation?

The methods are not specific to the CICE implementation and will be straightforward to incorporate into other continuum sea ice models (in fact the intention will be to include at least one of the two FSD models into the sea ice model SI^3 in the near future).

The first sentence in the 2nd paragraph in the conclusions now reads as follows:

'Simulations were completed using the two FSD models within a standalone setup of the sea ice model CICE, though it should be noted that both FSD models can easily be implemented into any continuum sea ice model.'

Admittedly, I am not an expert on this topic but these questions moderate my evaluation of the significance of this study.

4. PRESENTATION QUALITY: 2
The paper is well written and structured. Findings are well supported by figures of mostly good quality. Yet I want to already suggest that you re-structure sections 2 and 3. In my view, section 2 is the methodology while section 3 is the experimental design and the cal/val data. The data section might be presented separately. Eventually, Before the experimental design. The figure amount is rather high and I sense that some could be transferred to the appendix.

Regarding the paper structure, we have used a similar approach to other FSD modelling studies (Bateson et al., 2021; Roach et al., 2018). In both referenced studies, section 2 is used specifically to describe the model. The simulations actually performed are then described in section 3 i.e. separately to the model description. As such, our preference is to retain the same paper structure here. We also would prefer to retain the description of model simulations and how the observations are processed to compare against model output in the same section since choices made for the former influence the latter and vice versa; presenting them together highlights this relationship.

In terms of the number of figures, both reviewers queried whether Figs 5-7 were all needed within the manuscript or whether one or two could be removed. As such, we have decided to remove Fig. 6 from the manuscript, since we believe that this figure offers the least additional insight of the three. We have decided to retain Figs 5 and 7 since we believe that they are important components in presenting the full picture of impacts of FSDs on the sea ice cover.

Figure 6 and any references to this figure have been removed from the manuscript. Figure numbering has been updated accordingly.

Sea ice floe size: its impact on pan-Arctic and local ice mass, and required model complexity

Adam W. Bateson[1], Daniel L. Feltham[1], David Schröder[1,2], Yanan Wang[3], Byongjun Hwang[3], Jeff K. Ridley[4], Yevgeny Aksenov[5]

[1]Centre for Polar Observation and Modelling, Department of Meteorology, University of Reading, Reading, RG2 7PS, United Kingdom
[2]British Antarctic Survey, Cambridge, CB3 0ET, United Kingdom
[3]School of Applied Sciences, University of Huddersfield, Huddersfield, United Kingdom,
[4]Hadley Centre for Climate Prediction and Research, Met Office, Exeter, EX1 3PB, United Kingdom
[5]National Oceanography Centre Southampton, Southampton, SO14 3ZH, United Kingdom

Correspondence to: Adam W. Bateson (a.w.bateson@pgr.reading.ac.uk)

Abstract

Sea ice is composed of discrete units called floes. Observations show that these floes can adopt a range of sizes spanning orders of magnitude, from metres to tens of kilometres. Floe size impacts the nature and magnitude of interactions between the sea ice, ocean, and atmosphere including lateral melt rate and momentum and heat exchange. However, large-scale geophysical sea ice models employ a continuum approach and traditionally either assume floes adopt a constant size or do not include an explicit treatment of floe size. In this study we apply novel observations to analyse two alternative approaches to modelling a floe size distribution (FSD) within the state-of-the-art CICE sea ice model. The first model considered is a prognostic floe size-thickness distribution where the shape of the distribution is an emergent feature of the model and is not assumed a priori. The second model considered, the WIPoFSD (Waves-in-Ice module and Power law Floe Size Distribution) model, assumes floe size follows a power law with a constant exponent. We demonstrate that a parameterisation of in-plane brittle fracture processes enables the prognostic model to achieve a reasonable match against the novel observations. While neither FSD model results in a significant improvement in the ability of CICE to simulate pan-Arctic metrics in a stand-alone sea ice configuration, larger impacts can be seen over regional scales in sea ice concentration and thickness. We find that the prognostic model particularly enhances sea ice melt in the early melt season, whereas for the WIPoFSD model this melt increase occurs primarily during the late melt season. We then show that these differences between the two FSD models can be explained by considering the effective floe size, a metric used to characterise a given FSD. Finally, we discuss the advantages and disadvantages to these different approaches to modelling the FSD. We note that the WIPoFSD model is less computationally expensive than the prognostic model and produces a better fit to novel FSD observations derived from 2-m resolution MEDEA imagery, possibly making this a stronger candidate for inclusion in climate models, but is unable to represent potentially important features of annual FSD evolution seen with the prognostic model.

1 Introduction

The Arctic sea ice cover consists of contiguous pieces of sea ice referred to as floes (WMO, 2014). Floe size has a direct impact on several processes that are important to the evolution of the sea ice, including lateral melt rate (Steele, 1992; Bateson et al., 2020); momentum exchange between the sea ice, ocean, and atmosphere (Lüpkes et al., 2012; Tsamados et al., 2014); surface moisture flux over sea ice (Wenta and Herman, 2019); sea ice rheology i.e. the mechanical response of sea ice to stress (e.g. Shen et al., 1986; Wilchinsky and Feltham, 2006; Rynders et al., 2022); and the clustering of sea ice into larger agglomerates (Herman, 2012). Historically, continuum sea ice models such as CICE (Hunke et al., 2015) have assumed that floes are of a uniform size or do not explicitly consider floe size at all when evaluating sea ice thermodynamics (Bateson et al., 2020; Keen et al., 2021) or dynamics (Tsamados et al., 2014). In contrast, observations show that floe sizes can span a large range, from metres to tens of kilometres (Stern et al., 2018a). Model studies suggest that floe size has a non-negligible impact on sea ice extent and volume through changing lateral and total sea ice melt, particularly in areas where the sea ice cover largely consists of small floes (Bateson et al., 2020; Bateson, 2021a). Floe size has been found to be particularly important in the Marginal Ice Zone (MIZ); a region of sea ice cover influenced by waves and swell penetrating from the open ocean (Aksenov et al., 2017; Roach et al., 2019; Bateson et al., 2020). The MIZ is taken in this study as regions with sea ice concentration between 15 – 80 %, a definition commonly used due to an absence of observations of ocean surface waves in sea ice over the necessary spatial scales and timescales (Strong et al., 2017; Horvat et al., 2020).

Observations of the floe size distribution (FSD) show a large ratio of smaller floes to larger floes; this distribution of floe sizes is often summarized using a truncated power law (Rothrock and Thorndike, 1984; Toyota et al., 2006; Perovich and Jones, 2014; Stern et al., 2018b). Studies generally show that a power law produces a reasonable fit to the observations presented, though the validity of using a power law to fit floe size data remains an open question (Stern et al., 2018b), with several studies disputing the extent to which a power law is a good description of the FSD (Herman, 2010; Horvat et al., 2019; Herman et al., 2021). The exponents of power laws fitted to observations of the non-cumulative floe number density show a large amount of variability, from -1.9 to -3.5 (see summary of observations in Stern et al., 2018a). Observations show spatial and temporal variability of the FSD. Stern et al. (2018b) analysed satellite imagery collected over the Beaufort and Chukchi seas and reported an approximately sinusoidal seasonal cycle in the exponent with a minimum exponent of about -2.8 in August and a maximum exponent of about -1.9 in April in both 2013 and 2014 for floes larger than 2 km. Perovich and Jones (2014) also found evidence of seasonal variation in the exponent; aerial photographic imagery was analysed from the Beaufort Sea over the period June to September 1998 for floes between 10 m to 10 km in size. They noted a change in exponent from -3.0 over June and July to -3.2 in late August, coinciding with a high wind speed event driving fragmentation of floes under wind and ocean stress. The exponent then increased to over -3.1 by September due to sea ice freeze-up and floe welding.

Modelling studies have been used to understand how the observed FSD shape and behaviours could emerge from relevant processes. These FSD models can be roughly divided into two different classes: (i) models where the shape of the FSD emerges from the constituent sea ice dynamical-thermodynamical processes (e.g. Zhang et al., 2016; Roach et al., 2018, 2019); (ii) models where the general shape of the FSD is fixed (e.g. Bennetts et al., 2017; Bateson et al., 2020). Hybrid approaches have also been proposed, e.g. Boutin et al. (2020) allows the shape of the FSD to evolve in response to processes such as lateral melting, but resets the distribution to a power law after a wave break-up event. These modelling studies have incorporated one or several processes that have been observed to influence floe size: lateral melting and growth at the edges of floes (e.g. Perovich and Jones, 2014; Roach et al., 2018); break-up of sea ice floes into smaller pieces from ocean waves (e.g. Kohout et al., 2014); floes welding together in ocean freeze-up conditions (Roach et al., 2018); the formation mechanism of new floes (Roach et al., 2018); and rafting and ridging of floes during floe collisions (Horvat and Tziperman, 2015). The limited spatial and temporal coverage of floe size observations has prohibited effective evaluation of these models, though there have been recent efforts to develop satellite-derived FSD products to enable such evaluations (Horvat et al., 2019). It is nevertheless anticipated that important processes are not yet represented in these models. For example, thermodynamically-driven break-up of floes along existing cracks and refrozen leads in the sea ice cover (Perovich et al., 2001).

In this study we will consider the prognostic FSTD (Floe-Size-Thickness distribution model) of Roach et al. (2018, 2019) and the WIPoFSD model (Waves-in-Ice module and Power law Floe Size Distribution model) of Bateson et al. (2020). The prognostic model is within the model class where the shape of the distribution emerges primarily from parameterisations at the process level. The WIPoFSD model is within the class of models where the shape of the FSD in the model is actively constrained, in this case by approximating the FSD as a power law. These models present useful case studies to examine the advantages and disadvantages of different approaches to modelling the FSD and its impacts on sea ice. We introduce a new quasi-restoring brittle fracture scheme into the prognostic model, which crudely accounts for in-plane fracture processes in winter and thermodynamically-driven break-up of floes along existing cracks and other linear features over the subsequent melt season (Bateson, 2021a). We complete simulations including the FSD models within a version of CICE where floe size impacts both lateral melt volume and momentum exchange between the sea ice, ocean, and atmosphere. We compare the performance of the prognostic model both with and without the brittle fracture scheme in simulating the shape of novel observations of the FSD and also assess the accuracy of a power-law fit to these observations. By examining the impact of the two FSD models on the sea ice mass balance, we consider whether either FSD model can improve the performance of CICE in our model configuration. The impact of both FSD models on key sea ice and MIZ metrics is investigated, including their interannual variability and spatial differences. Finally, we explore how differences in the impacts of the two models emerge and consider the implications of the results presented here for different strategies in modelling the FSD.

This paper is structured as follows. Section 2 describes the CICE model setup used in this study, the two FSD models, and new model components and modifications introduced in this study. Section 3 describes the methodology for this research, including a description of novel observations of the FSD and an overview of model experiments. Section 4 presents the results of the analysis and simulations, divided into 3 sub-sections: a comparison of the modelled FSD to observations; a comparison of model output to the observed sea ice extent and volume reanalysis; and a pan-Arctic comparison of the impacts of the two FSD models. Section 5 discusses the results and section 6 presents conclusions and summarises the study. This study uses several key terms related to the sea ice floe size distribution. To ensure clarity, Table 1 summarises and defines these key terms.

2 Model description

Here we will use the CPOM (Centre for Polar Observation and Modelling) version of the Los Alamos Sea Ice model v5.1.2, known as CICE (Hunke et al., 2015). In section 2.1.1, we will outline key details of CICE that are pertinent to this study. Within the CPOM-CICE setup, we use the prognostic mixed-layer model of Petty et al. (2014), and the form drag scheme of Tsamados et al. (2014). An overview of each of these model components are provided in sections 2.1.2 and 2.1.3, respectively. In section 2.1.4, we outline how the treatment of lateral melt and form drag in the CPOM-CICE setup has been modified for use with FSD models. In section 2.2 and 2.3 we will provide an overview of the two FSD models considered here: a modified version of the prognostic FSD model of Roach et al. (2018), and a modified version of the WIPoFSD model of Bateson et al. (2020).

2.1 Description of Standard Model Physics

2.1.1 Standard CICE model

In this study, we model the Arctic sea ice cover using a local version of the CICE sea ice model in a standalone setup. CICE is a continuum numerical model of sea ice that has been designed for use within fully coupled climate models. The model consists of several different components designed to simulate the evolution of sea ice on the geophysical scale including sea ice and snow thermodynamics, sea ice dynamics, a sea ice thickness distribution, and advection. The standard sea ice thickness distribution in CICE distributes ice area between five thickness categories, with the spacing increasing for thicker categories. The ice area in a given category evolves in response to dynamic and thermodynamic processes according to a linear remapping scheme (Lipscomb, 2001). Sea ice melt is subdivided into three separate components within CICE: melt from the upper surface of the sea ice floe (top melt), melt from the bottom surface of the floe (basal melt), and melt from the sides of the floe (lateral melt). Full details of CICE can be found within Hunke et al. (2015). Below we provide an overview of features of standard CICE that are pertinent to the evaluation of the lateral melt volume. In section 2.1.4 we will explain how this standard treatment is adapted for use with an FSD model.

The lateral melt volume is explicitly calculated within CICE:

$$\frac{1}{A}\frac{dA}{dt} = \frac{\pi}{\alpha_{shape} L} w_{lat.} \tag{1}$$

A represents the sea ice area fraction, such that the term on the left-hand side, $\frac{1}{A}\frac{dA}{dt}$, represents the fractional rate of sea ice area loss due to lateral melt (units of s^{-1}). The rate of sea ice volume loss from lateral melt is the product of $\frac{1}{A}\frac{dA}{dt}$ with the sea ice area and mean thickness. α_{shape} and L are the constant floe shape and diameter parameters, set to 0.66 and 300 m in standard

CICE. w_{lat} is the lateral melt rate (units of ms^{-1}); it is a function of the elevation of the sea surface temperature above freezing for standalone CICE. The lateral heat flux, F_{lat}, is calculated from the volume of lateral melt (all fluxes have units of $Jm^{-2}s^{-1}$). The melting or freezing potential at the sea ice-surface ocean interface, F_{frzmlt}, is calculated as:

$$F_{frzmlt} = \frac{\Delta T c_{p,oc} \rho_w h_{mix}}{\Delta t}, \quad (2)$$

where ΔT is the difference in sea surface temperature from freezing (units of K), $c_{p,oc}$ is the specific heat capacity of the surface ocean ($Jkg^{-1}K^{-1}$), ρ_w is the density of seawater (kgm^{-3}), h_{mix} is surface mixed-layer depth (units of m), and Δt is the model timestep. The magnitude of F_{frzmlt} is capped at a fixed value. Following CICE sign conventions, a negative value for F_{frzmlt} corresponds to melting of sea ice. The following condition applies to the lateral and basal flux during periods of melt:

$$|F_{bot} + F_{lat}| \leq |F_{frzmlt}|. \quad (3)$$

Here F_{bot} is the net downward heat flux from the sea ice to the ocean. Where the melting potential is exceeded, F_{bot} and F_{lat} are both reduced by a common factor such that the condition set by Eq. (3) is satisfied.

2.1.2 The mixed-layer model

Ocean mixed-layer properties are important in determining lateral and basal melt rates, which are both relevant for evaluating the impact of floe size on the sea ice cover (e.g. Bateson et al., 2020). Here, a modified version of the prognostic bulk mixed-layer model of Petty et al. (2014) is used rather than a constant prescribed mixed-layer depth, to better represent sea ice-mixed layer interactions and feedbacks (e.g. the ice-ocean albedo feedback) without the complexity and computational expense of a full ocean model (e.g. Frew et al., 2019). In this model, the mixed-layer temperature, salinity, and depth are all evaluated prognostically. The deep ocean below the mixed layer is restored to observations, and the model is zero-dimensional i.e. defined for each model grid cell without lateral interactions between grid cells. Full details of the original scheme are available in Petty et al. (2014). The original Petty et al. (2014) mixed-layer model was set-up and tested for the Southern Ocean where the stronger winds and waves, weaker upper ocean stratification, and larger extent of the MIZ enables a high wind power input leading to a much deeper mixed-layer when compared to the Arctic. Here we adopt several adjustments to the mixed-layer model made by Tsamados et al. (2015) to ensure reasonable performance of the mixed-layer model in the Arctic. The three-component model of surface layer, mixed layer, and deep ocean is replaced with a two-component model, with just a mixed layer and deep ocean. In addition, the mixed-layer temperature and salinity are restored to the 10 m depth temperature and sea surface salinity from a monthly climatology reanalysis dataset.

2.1.3 Form drag scheme

Recent versions of CICE include an implementation of the form drag scheme (Hunke et al., 2015) following Tsamados et al. (2014), which aims to better describe the turbulent momentum and heat exchange between the sea ice, ocean, and atmosphere by accounting for the topography of sea ice. The scheme of Tsamados et al. (2014) replaces the constant drag coefficients in CICE with explicit representations of both form drag and skin drag terms. C_a, the updated expression for the atmospheric neutral drag coefficient, can be calculated in terms of contributions from specific spatial features of the sea ice cover:

$$C_a = C_a^{skin} + C_a^{f,rdg} + C_a^{f,floe} + C_a^{f,pond}. \quad (4)$$

C_w, the updated expression for the ocean neutral drag coefficient, can similarly be calculated as:

$$C_w = C_w^{skin} + C_w^{f,rdg} + C_w^{f,floe}. \quad (5)$$

Here C^{skin} refers to the skin drag term, and $C^{f,rdg}$, $C^{f,floe}$ and $C^{f,pond}$ refer to form drag terms for ridges and keels, floe edges, and melt pond edges respectively. Tsamados et al. (2014) outline the following expression for $C^{f,floe}$ in the case of surface momentum exchange over the sea ice-atmosphere interface with a reference height of 10 m:

$$C_a^{f,floe} = \frac{1}{2}\frac{c_{fa}}{\alpha_{shape}} S_c^2 \frac{H_f}{L} A \left[\frac{\ln\left(\frac{H_f}{z_{0w}}\right)}{\ln\left(\frac{10}{z_{0w}}\right)}\right]^2. \quad (6)$$

Here c_{fa} is a local form drag coefficient, taken to be constant. α_{shape} is a geometrical parameter to account for the shape of the floes. The ratio $\frac{c_{fa}}{\alpha_{shape}}$ takes the value 0.2. L is the average floe diameter, which in Tsamados et al. (2014) is calculated as a function of sea ice concentration as per the parameterisation outlined in Lüpkes et al. (2012). z_{0w} is the roughness length of water upstream of the floe, given by 3.27 x 10^{-4} m (Hunke et al., 2015). H_f is the freeboard of the floe i.e. the distance between the upper surface of the floe and the sea surface. To calculate $C_w^{f,floe}$, the form drag of sea ice floes at the sea ice-ocean interface, H_f in Eq. (6) is replaced with D, the draft. D is defined as the distance between the lower surface of the floe and the sea surface. S_c is the sheltering function and is calculated as a function of sea ice area fraction, A, using an approximation from Lüpkes et al. (2012):

$$S_c = 1 - e^{-s_{lf}(1-A)}. \quad (7)$$

s_{lf} is the floe sheltering attenuation coefficient, with $s_{lf} = 11$ as per Lüpkes et al. (2012).

2.1.4 Modifications to standard CICE to incorporate FSD effects

The CPOM-CICE setup has been adapted to represent the impact of the FSD on the sea ice cover via both the lateral melt rate and floe edge contribution to form drag. Eq. (1), used to calculate the fraction of sea ice area lost due to lateral melting, is modified to:

$$\frac{1}{A}\frac{dA}{dt} = \frac{\pi}{\alpha_{shape} l_{eff}} w_{lat}. \quad (8)$$

L, the constant floe diameter, has been replaced by l_{eff}, the effective floe size. l_{eff} is the diameter of the set of identical floes that has the same total perimeter as a set of floes of variable size with the same total ice area (Bateson et al., 2020). l_{eff} is applicable here because the lateral melt volume is proportional to the total floe perimeter. Eq. (6), the expression for the floe edge contribution to form drag at the sea ice-atmosphere interface, has also been modified:

$$C_a^{f,floe} = \frac{1}{2}\frac{c_{fa}}{\alpha_{shape}} S_c^2 \frac{H_f}{l_{eff}} A \left[\frac{\ln\left(\frac{H_f}{z_{0w}}\right)}{\ln\left(\frac{10}{z_{0w}}\right)}\right]^2. \quad (9)$$

The equivalent expression at the sea ice-ocean interface is modified similarly. l_{eff} is used here since it characterises the average floe length scale.

2.2 The prognostic FSD model

2.2.1 Prognostic model overview

The prognostic FSD model used here has been adapted from the version presented by Roach et al. (2018). At the core of the prognostic FSD model is the joint floe size-thickness probability distribution (FSTD), $f(r,h)drdh$. This describes the fraction of a grid cell covered by floes with a radius between r and $r + dr$ and thickness between h and $h + dh$. Processes that change floe size represented in this model include lateral melting and freezing, wave-induced breakup of floes, and welding together of floes. The model also allows the formation of new floes, complete melt out of existing floes, and advects the FSTD between grid cells. The parameterisation introduced in Roach et al. (2019) to determine the size of newly formed floes from the local wave conditions has also been included in the setup used here. A full description of the original prognostic floe size-thickness distribution model is presented in Roach et al. (2018) with details of the wave-dependent floe formation parameterisation available in Roach et al. (2019). Unlike Roach et al. (2019), we do not use a separate wave model coupled to CICE to calculate the necessary wave properties within the sea ice-covered grid cells for use with the wave-dependent floe formation parameterisation. Instead, we adapt the scheme used in Roach et al. (2018), which calculates in-ice wave properties using an extrapolation approach from forcing external to the sea ice cover to calculate the necessary in-ice wave properties. We also introduce a novel treatment of brittle fracture to the prognostic model. This brittle fracture scheme is described in section 2.2.2. For the prognostic FSD model $l_{eff,n}$, the effective floe size for the n^{th} sea ice thickness category, is calculated in terms of $L(r,h)$, the modified areal FSTD, where the integral of $L(r,n)$ over the range r_{min} to r_{max} is 1 i.e. the distribution is normalised per thickness category:

$$l_{eff.n} = \frac{2}{\int_{r_{min}}^{r_{max}} r^{-1} L(r,n)\, dr}. \quad (10)$$

A representative l_{eff} for the full FSTD is then calculated as the area-weighted average of $l_{eff,n}$ across the thickness categories.

2.2.2 The brittle fracture scheme

It will be shown in section 4.1 that the prognostic model struggles to capture the shape of the observed FSD for mid-sized floes. Sensitivity studies show that it not possible to modify existing parameterisations in the prognostic FSD model to substantially improve model performance against observations (Bateson, 2021). This suggests there are important processes currently not represented within the prognostic model. A leading candidate for the latter is brittle fracture and associated processes. Satellite imagery of the Arctic sea ice cover, especially over the winter pack ice, shows linear features such as leads and fractures referred to as slip lines or linear kinematic features existing at scales of kilometres (Kwok, 2001; Schulson, 2001). These linear features have been found to intersect at acute angles, from scales of millimetres to kilometres, creating individual diamond shaped regions and floes over the sea ice cover (Weiss, 2001; Schulson, 2004). The similarity of these linear features to fracture patterns formed in laboratory studies of the shear rupture mechanism, where a crack forms once a large enough shear stress is imposed, has been used to argue that the shear rupture mechanism is responsible for the linear features seen in the pack ice (Weiss and Schulson, 2009). A discrete element model of the sea ice incorporating compressive, tensile, and shear rupture failure mechanisms acting under wind stress has been shown to produce distributions of fractures that are comparable to the distribution of linear features seen in the Arctic pack ice (Wilchinsky et al., 2010). The existence of this brittle fracture behaviour in sea ice from microscopic to macroscopic scales makes it an interesting candidate to consider in terms of FSD evolution, though the scaling of brittle fracture remains an area of ongoing research (e.g. Weiss and Schulson, 2009; Hutchings et al., 2011; Weiss and Dansereau, 2017).

Clearly, brittle fracture events can have a direct impact on floe size. However, plausible indirect mechanisms also exist. Perovich et al. (2001) observed that summer floe break-up of sea ice in the central Arctic in 1998 was driven by thermodynamic weakening of cracks and refrozen leads in the sea ice cover during a period when the dynamic forcing and internal sea ice stress was expected to be small. More recent observational studies have also suggested a link between sea ice melt and floe break-up (Arntsen et al., 2015; Hwang et al., 2017). This suggests that linear features in the sea ice that form from the brittle fracture and subsequent refreezing of sea ice in winter can then influence sea ice break-up in summer as the sea ice thins and weakens. A full physically-derived parameterisation of the impacts of brittle fracture on the sea ice cover, both directly and indirectly via thermodynamic weakening, requires additional direct observations of these processes within the sea ice cover and is therefore beyond the scope of this study. It is nevertheless possible to use theoretical models of brittle fracture processes to explore the potential impact of brittle fracture related mechanisms on FSD shape. In a brittle fracture event cracks can propagate and, where they exceed a critical speed, become unstable and branch. Individual branches and fractures can also merge, with the lifetime of the fracture determining the size of the subsequent fragment that forms. The branching results in a hierarchical process, with several levels of branches forming from the same central fissure (Åstrom et al., 2004; Kekäläinen et al., 2007). Idealised models of brittle fracture show that the fragment size distribution adopts a power law with an exponent of -2 and an upper cut-off determined by an exponential in the square of the fragment size (Gherardi and Lagomarsino, 2015).

In order to investigate the potential impact of in-plane brittle fracture processes on the FSD, the prognostic model has been modified to include a quasi-restoring brittle fracture scheme, which applies a conditional restoring to the FSD towards the theoretical distribution produced by idealised models of brittle fracture i.e. a power law with an exponent of -2. In this scheme brittle fracture can transfer sea ice area fraction from a larger floe size category to the adjacent smaller category. In addition, the following condition must be fulfilled:

$$\frac{\ln n_i - \ln n_{i-1}}{\ln d_i - \ln d_{i-1}} > -2. \quad (11)$$

Here n and d refer to the floe number density and diameter at the midpoint of category i respectively. This condition means that the restoring scheme only applies where the slope between adjacent categories in log-log space is greater (more positive) than -2 i.e. only when the ratio of larger floes to smaller floes exceeds a given value. The sea ice area fraction transferred in a single timestep between two adjacent categories is $C_{bf} a_i$ where a_i is the area fraction of the larger category. C_{bf} the restoring constant, is calculated as:

$$C_{bf} = \frac{\tau}{\Delta t}. \quad (12)$$

Here τ is the restoring timescale, and Δt is the model timestep. Figure 1 provides a visual summary of the quasi-restoring scheme. The motivation for this scheme is to impose a restoring tendency on the FSD to the predicted shape of the distribution if it were acting only under brittle fracture. The transfer of sea ice area fraction is only allowed in one direction from larger to adjacent smaller categories since floes cannot unfracture. This process is conservative in sea ice area i.e. the reduction in floe area in the larger category will be matched by an increase in floe area in the smaller category.

A value for the restoring timescale, τ, needs to be determined. Both direct and indirect mechanisms have been discussed above describing how brittle fracture can impact the sea ice cover. Fracture events occur regularly through autumn, winter and spring within the pack ice to break up floes and form features such as leads, though these generally freeze up again. The result of these fracture events is to create a network of linear features that define weaker regions of ice interspersing stronger ice. Idealised models of brittle fracture suggest that the size distribution of the stronger regions of ice follow a power law with an exponent of -2. The linear features are then vulnerable to increased thinning and melting, increasing the likelihood of break-up along these features during late spring and summer as the sea ice retreats. This effectively 'releases' the floe size distribution defined during brittle fracture events outside of the melt season. It is this second mechanism that is of more relevance when considering the impacts of the FSD on the seasonal retreat of the Arctic sea ice. In this context, τ, the restoring timescale, refers to the timescale for the sea ice to thin sufficiently that the sea ice is vulnerable to in-plane fracture events along existing weaknesses. Sea ice thickness away from the ice edge at the start of the melt season is generally in the range of 1 – 3 m. Vertical melt rates are of the order of 5 – 15cm day^{-1}. Therefore, significant thinning can generally occur over timescales as short as a week up to a couple of months. For simplicity, τ is here set to 30 days.

There are clear limitations associated with this approach. The use of a fixed timescale makes it difficult to capture both the direct mechanism of brittle fracture impact on floe size, which dominates outside of the melt season, and the indirect mechanism via thermodynamic weakening, which is more important within the melt season. The latter mechanism has been prioritised in this case in determining the timescale given it is the FSD state in the melt season that is of primary importance for understanding FSD impacts on the Arctic sea ice (Bateson et al., 2020). Just considering the thermodynamic weakening mechanism, the use of a fixed timescale is still a simplification given the significant spatial and temporal variability of relevant factors to this mechanism such as melt rates, ice strength, and dynamic forcing. In addition, the brittle fracture approach assumes transfer of sea ice area fraction only between adjacent categories, whereas physically a larger floe can break down into floes of any smaller size. Nevertheless, this new scheme remains a useful way to approximate the impact of brittle fracture on FSD shape. Results will be presented in section 4.1 to demonstrate that the inclusion of this new brittle fracture scheme significantly improves prognostic FSD model performance against observations in simulating FSD shape for mid-sized floes.

2.3 The WIPoFSD model

The WIPoFSD model used in this study has been adapted from the version presented Bateson et al. (2020), which in turn was based on the coupled ocean–waves–in–ice model NEMO–CICE–WIM developed at the UK National Oceanography Centre (NOC). The NEMO–CICE–WIM model approximates the shape of the FSD as a multiple-exponent truncated power law with coefficients depending on ice fraction; in this study we use a constant exponent as per Bateson et al. (2020). The WIPoFSD model fits the number-weighted FSD, $N(x)$, where x is the floe diameter, to a power-law distribution:

$$N(x|d_{min} \leq x \leq l_{var}) = Cx^{\alpha}. \tag{13}$$

N has units of reciprocal metres, all floe size variables have units of metres, and α, the WIPoFSD model exponent, is dimensionless. l_{var}, the variable FSD tracer, evolves in each grid cell as a function of physical processes between the upper and lower floe diameter cut-offs, d_{max} and d_{min} respectively. Bateson et al. (2020) suggested that l_{var} can be taken as representing the history of a given area of sea ice in terms of physical processes that affect the FSD. The model is initiated with l_{var} set to d_{max} in all grid cells where sea ice is present. C is calculated such that the total floe area is equal to the total sea ice area. The model parameterises the role of four processes in the evolution of the FSD: lateral melting, wave-induced break-up, winter growth, and advection. A full description of how these processes are represented within the WIPoFSD model, including a description of the advection scheme for waves in sea ice, is available in Bateson et al. (2020); a summary has also been provided here in Appendix A.

The version of the WIPoFSD model used here includes a modified lateral melting scheme to that presented in Bateson et al., (2020). In the WIPoFSD model, processes are parameterised in terms of how they impact l_{var} and useful properties such as l_{eff} can easily be calculated from l_{var}. The appeal of this approach is that it is both simple and enables an exploration of the broader impacts of a power-law distribution on the sea ice cover whilst retaining spatial and temporal variability in l_{eff}. For mechanical processes such as wave break-up, the use of l_{var} is particularly suitable; 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. However, it is not possible to define two clear regimes of how floe size changes in response to lateral melting; instead, floes across the distribution reduce in diameter by the same magnitude in response to a lateral melting event. Here, we have modified the lateral melting scheme to calculate the change in l_{eff} rather than l_{var}, since it is possible to calculate exactly how much the total floe perimeter, and therefore l_{eff}, would increase or decrease in response to any change in the FSD. Whilst it is not possible to exactly capture cumulative changes to the FSD over several timesteps due to the constraints of having a fixed exponent power-law distribution with a lower floe size limit, it is possible to calculate exactly how the effective floe size, l_{eff}, will change over one timestep by integrating across the updated FSD after a lateral melting event. The updated effective floe size, $l_{eff,new}$, can then be calculated as follows:

$$l_{eff,new} = \frac{[l_{var}^{3+\alpha} - d_{min}^{3+\alpha}]A_{new}}{(3+\alpha)\, A_{old}} \left(\frac{[l_{var}^{2+\alpha} - d_{min}^{2+\alpha}]}{(2+\alpha)} - \frac{2\Delta l[l_{var}^{1+\alpha} - d_{min}^{1+\alpha}]}{(1+\alpha)} \right)^{-1}. \tag{14}$$

Here Δl is the length of lateral melting experienced by each floe edge and A_{old} and A_{new} refer to the sea ice area fraction before and after the lateral melting event, respectively. $l_{var,new}$ can then be calculated from $l_{eff,new}$ using a Newton-Raphson iterative scheme. A full derivation of Eq. (14) and description of the iterative scheme can be found in Appendix B.

3 Methodology

3.1 Sea ice simulations

All simulations in this study are initiated with an ice-free Arctic on 1st January 1980 and evaluated over a 37-year period until 31st December 2016. The first ten years of these simulations are taken as spin-up, with timeseries presented over the period 1990-2016. Averages are calculated over the period 2000-2016, taken as representative of the current climatology. The CPOM version of CICE is run over a pan-Arctic domain with a 1° tripolar (129×104) grid. Sections of the Hudson Bay and Canadian Arctic Archipelago are not included within the model domain. Surface forcing is obtained from 6-hourly NCEP-2 reanalysis fields (Kanamitsu et al., 2002). The mixed-layer properties are restored over a timescale of 5 days to a monthly climatology reanalysis at 10 m depth taken from the MyOcean global ocean physical reanalysis product (MYO reanalysis; Ferry et al., 2011). The deep ocean is restored after detrainment from the mixed layer over a timescale of 90 d to the winter climatology (for this we take the mean conditions on 1 January from 1993 to 2010) from the MYO reanalysis. The minimum mixed-layer depth is set to 10 m. H_s, the significant wave height (m), and T_p the peak wave period (s), of ocean surface wave fields are obtained from the ERA-Interim reanalysis dataset (Dee et al., 2011). The forcings are updated at 6 h intervals, but only for locations where the sea ice is at less than 1 % sea ice concentration. The ERA-Interim reanalysis has been selected for the ocean surface wave field forcing as this dataset has generally been found to perform well in comparison to other reanalyses against observations of wind speed and wind speed profile in the Arctic during summer months (e.g. Jacobson et al., 2012; de Boer et al., 2014).

The general CPOM-CICE setup used for all simulations in this study has several further differences to CICE version 5.1.2, in addition to those described in section 2, based on recent work by Schröder et al. (2019). The maximum meltwater added to melt ponds is reduced from 100 % to 50 %. This produces a more realistic distribution of melt ponds (Rösel et al., 2012). Snow erosion, to account for a redistribution of snow based on wind fields, snow density, and surface topography, is parameterised based on Lecomte et al. (2015) with the additional assumptions described by Schröder et al. (2019). The "bubbly" conductivity formulation of Pringle et al. (2007) is also included, which results in larger thermal conductivities for cooler ice. The longwave emissivity is increased from 0.95 to 0.976. The following parameters are modified from the default values used in Tsamados et al. (2014): atmospheric background drag coefficient, ocean background drag coefficient, ridge impact parameter, and keel impact parameter of the form drag parameterisation. They are set to 0.001, 0.0005, 0.1 and 0.5 respectively. Schröder et al. (2019) discuss how these changes increase ice drift over level ice and reduce it over ridged ice leading to more realistic ice drift patterns.

A total of five different simulations are used here; a summary of these simulations is presented in Table 2. The reference simulation, *ref*, sets l_{eff} to a fixed value of 300 m. 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). Apart from the modifications outlined in section 2.2.2 – 2.2.4, prognostic model setup and parameter choices are identical to Roach et al. (2018). The simulation with the WIPoFSD model, *WIPo-best*, uses an identical setup to that in Bateson et al. (2020), but now incorporating the updated lateral melting scheme described in section 2.3. For the WIPoFSD model parameters, d_{max} is set to the standard value used by Bateson et al. (2020) of 30 km. The minimum floe size that can be resolved in *prog-best* is 5.375 m and hence d_{min} will be set to the same value. α is set to -2.56, the average value across the three locations represented in the novel FSD observations that will be discussed in Sect. 3.2. This means that parameter or model choices for both *WIPo-best* and *prog-best* have been selected to produce a best fit to the same set of observations. Two additional simulations are performed using the prognostic model to compare against observations of the FSD. These two simulations will include a total of 16 floe size categories using Gaussian spacing rather than the standard 12. The prognostic model produces an unphysical increase or 'uptick' in the largest few categories, at least partially a result of having a fixed maximum floe size in the model (Roach et al., 2018). By using 16 floe size categories rather than 12, the largest 4 floe size categories that include this 'uptick' will fall outside the range of floe sizes included in the comparison to observations. The first of these additional prognostic simulations, *prog-16*, is otherwise identical to *prog-best*. The second, *prog-16-nobf*, excludes the brittle fracture scheme described in section 2.2.3. Whilst 16 floe size categories could also be retained for the *prog-best* simulation, the increase in model run time increases non-linearly with increasing number of categories. In addition, the improved resolution of the shape of the distribution for floes of a size of 1 km or larger is not significant when considering the impact of an FSD on sea ice via the floe edge contribution to form drag and lateral melt rate, which both scale to the inverse of floe size. Therefore, 12 floe size categories represents a more practical choice for the prognostic FSD model. Figure 2 presents an example of the model output from *prog-16-nobf*, showing the perimeter density distribution within the MIZ for April, June, and August, averaged over 2000-2016. Figure 2 demonstrates how the 'uptick' is confined to the largest 3-4 floe size categories.

3.2 FSD observations

To assess the performance of the two alternative FSD models, we consider a new observational dataset that has not been used to motivate the development of either FSD model. The observations consist of 41 separate images over the period 2000-2014, covering May – July, and collected from three regions: the Chukchi Sea (70 N, 170 W); the East Siberian Sea (82 N, 150 E);

and the Fram Strait (84.9 N, 0.5 E). The raw floe size data has been retrieved using the algorithm described in Hwang et al. (2017) from the GFL HRVI (Global Fiducials Library high-resolution visible-band image) imagery that has been declassified by the MEDEA group (Kwok and Untersteiner, 2011). This has been made available publicly as LIDPs (Literal Image Derived Products) at 1 m resolution (available at http://gfl.usgs.gov/). The total image size varies between observations, but generally has length dimensions of 5 – 20 km. The resolution of the imagery was reduced from 1 m to 2 m prior to processing by the algorithm.

The first step of processing the raw floe size data, consisting of a list of individual floe sizes, is to sort them into the Gaussian-distributed floe size categories used within the prognostic model for ease of comparison. Any floes that exceed the upper diameter cut-off of the largest category, 1892 m, will be discarded from the analysis. This step is necessary because the two models simulate the full FSD, and not individual floes. Floes large compared to the image size are inadequately sampled in observations to construct the full FSD. For example, the presence of a single large floe, comparable to the image size, can cause a large perturbation across the distribution reported for that location. Instead, only floe size categories that are small enough to consistently be populated by multiple floes across all sampled images are retained. A lower floe diameter cut-off of 104.8 m is also applied to this analysis, taken to be the smallest floe size that can be reliably resolved from the observations for the methodology and resolution used. The limiting factor on the smallest resolved floe size is the ability to resolve gaps between floes. Once floes outside the range of 104.8 m to 1892 m in diameter have been discarded, the total area of remaining floes is calculated and taken to be the total sea ice area for normalising the reported perimeter density (perimeter per unit sea ice area). The average normalised perimeter density for each floe size category is then reported at the mid-point of that category. The floe perimeter density distribution, $\rho_{FSD}(x)$, is considered in this study rather than the floe number or area distribution since it is the perimeter that has been identified as most relevant to the impact of the FSD on sea ice when considering lateral melt (Bateson et al., 2020). It is defined here as:

$$P_{FSD} = \int_{x_{min}}^{x_{max}} \rho_{FSD}(x)\, dx \qquad (15)$$

Here x_{min} and x_{max} refer to the minimum and maximum floe diameters respectively within an FSD. P_{FSD} is the perimeter density across the whole distribution, calculated as the total perimeter divided by the total area of all floes in the distribution. The concept of perimeter density has been used previously to this study e.g. Roach et al. (2019), Bateson et al. (2020).

To compare model output to the observations of the FSD, two sample years will be selected for each location: Chukchi Sea, May – June 2006 (4 LIDPs), May 2014 (4 LIDPs); East Siberian Sea, June 2001 (3 LIDPs), June – July 2013 (2 LIDPs); Fram Strait, June 2001 (6 LIDPs), June 2013 (2 LIDPs). These specific years have been selected as they all include at least two separate LIDP-derived floe size observations. Perimeter density distributions from the prognostic model are reported as an average over one or two months for the relevant region. The months selected for this average are chosen to minimise the difference between the mean day of collection for observations and median day of the model output. Figure 3 shows the specific areas over which the FSD is averaged. Each case study area consists of a set of 5 x 5 grid cells that includes the location where the observations were drawn from.

3.3 Observations of sea ice extent and volume

We use sea ice concentration products from the Bootstrap algorithm version 3 (Comiso, 1999) and NASA Team algorithm version 1 (Cavalieri et al., 1996). Both datasets have a spatial resolution of 25 km x 25 km and a temporal resolution of 1 day. Comparisons will focus on the pan-Arctic extent rather than the spatial distribution of sea ice concentration due to the high uncertainty in summer and MIZ of satellite-derived concentration products (Meier and Notz, 2010). We use the sea ice volume product from PIOMAS, the Pan-Arctic Ice Ocean Modelling and Assimilation System (Zhang and Rothrock, 2003). Whilst the PIOMAS volume product is a reanalysis and does not incorporate direct observations of the sea ice thickness, it has been evaluated using available observations of sea ice thickness (e.g. Schweiger et al., 2011). This product is often used to test model performance in simulating the total Arctic sea ice volume (Schröder et al., 2019) due to the challenges in estimating sea ice thickness from radar altimetry and limited availability of in-situ thickness measurements (e.g. Massonnet et al., 2012; Stroeve et al., 2012; Ridley et al., 2018).

4. Results

4.1 A comparison to observations of the FSD

The observations described in section 3.2 can be used to test how well both the prognostic FSD model and a power-law fit capture the shape of the observed FSD for mid-sized floes. Figure 4 compares FSD observations, a power-law fit, and prognostic model output from both the *prog-16* (with brittle fracture) and *prog-16-nobf* (without brittle fracture) simulations for each selected case study region and time period described in section 3.2. When comparing observations across the sites considered in Fig. 4 there is clear variability between the different case studies, but this variability is substantially smaller than the differences between the prognostic model without brittle fracture and the observations across all the case studies considered. It cannot be expected that an FSD model can precisely replicate an observed FSD given other differences will exist between the model and the observed sea ice state such as ice thickness and concentration, but it can be expected that a simulated FSD should be within the variability in FSD shown by observations if an FSD model is accurately capturing the relevant processes. Figure 4 therefore suggests that *prog-16-nobf* performs poorly in capturing the behaviour of the FSD for mid-sized floes. The perimeter density distribution predicted by the *prog-16-nobf* for each category is in general multiple orders of magnitude from the observed value. In particular, the slope of the distribution is much steeper (more negative) for the model output than observations i.e. the model predicts smaller floes within the range 104.8 m - 1892 m take up a much larger proportion of the total sea ice area than is suggested by observations. Figure 4 also shows that *prog-16* significantly improves the shape of the emergent perimeter density distribution for the floe size range considered compared to *prog-16-nobf*. The updated model performs particularly well in the East Siberian Sea and Fram Strait (panels c-f in Fig. 4) but less well in the Chukchi Sea (panels a-b in Fig. 4). Overall, the inclusion of the quasi-restoring brittle fracture scheme represents a significant improvement in the ability of the prognostic model to capture the shape of the FSD for mid-sized floes over the period May – July. The purpose of this comparison against floe size observations is to ensure that the WIPoFSD and prognostic model setups used in this study perform comparably well to the same dataset. There are limitations to this evaluation of model performance, however. In particular, floes smaller than 100 m or larger than 2 km are not considered for the reasons outlined in section 3.2, and the former are especially significant for determining the impact of a given FSD on sea ice concentration and thickness.

It is worth commenting briefly on how the brittle fracture scheme can improve model performance compared to observations, given it is a counterintuitive result that increasing floe break-up would produce a shallower slope in perimeter density. As discussed in section 3.2, the largest floe size categories in the prognostic model are excluded from the comparison to observations to exclude the non-physical ‘uptick’ that forms (Fig. 2). Whilst a reduction in ice area fraction in the largest category and an increase in the smallest category can be expected, the change in ice area fraction in the remaining categories depends precisely on the balance between ice area fraction lost from that category (sink) and ice area fraction gained from the adjacent larger category (source). In this case, the presence of the ‘uptick’ shown in Fig. 2 for the prognostic model without brittle fracture results in the source of floe area being larger than the sink for most floe size categories and a net reduction in gradient overall from including brittle fracture.

4.2 A comparison to observations of sea ice extent and volume

In this section the CICE simulations *prog-best* and *WIPo-best*, summarised in Table 2, will be compared against observations or reanalysis of the sea ice cover. Both of these simulations have been optimised against the FSD observations presented in section 4.1. The *prog-best* simulation includes the brittle fracture scheme described in section 2.2.2 and the exponent selected for the *WIPo-best* simulation is the average fitted exponent across the FSD observations. These two simulations, alongside the *ref* simulation that applies a constant floe size, will be tested against observations by considering the following metrics: the performance of the simulations in capturing the annual and interannual variability of the sea ice extent and volume; the performance of the simulations in capturing interannual trends in the sea ice extent and volume; and whether the inclusion of either FSD model reduces known model bias in the sea ice area fraction.

Figure 5 shows timeseries for the total Arctic sea ice extent and volume in March and September over 1990 – 2016 for *ref*, *WIPo-best*, *prog-best* alongside observations (extent) or reanalysis (volume) over the same period. The differences between the simulations are significantly smaller than the difference between *ref* and the observations / reanalysis in both March and September. This plot shows that the inclusion of either FSD model does not produce a significant improvement in the ability of CICE to simulate pan-Arctic sea ice extent and volume or the variability in these metrics, though this conclusion does not necessarily extend to climate simulations where FSD impacts on sea ice feedbacks with the ocean or atmosphere could produce larger changes to the sea ice state.

Previous studies, e.g. Bateson et al. (2020) and Roach et al. (2018), show that the largest impacts of including an FSD model occur within the MIZ. Figure 6 compares timeseries from 1990 – 2016 for the MIZ and pack ice extent in both March and September for each of *ref*, *prog-best*, and *WIPo-best* in addition to the satellite-derived observations. Figure 6 shows that all three simulations generally simulate both the MIZ and pack ice extent within observational uncertainty, though this is partly due to the large differences in MIZ and pack ice extent between the two observational products. The simulations are unable to replicate a negative trend in March pack extent shown by the observations. *prog-best* produces both a higher MIZ extent and variability on average in March compared to *ref* but a lower extent in September. In comparison, *WIPo-best* shows a reduced MIZ extent throughout the year compared to *ref*. In September, all three simulations produce very similar pack extents, but in March there is a moderate reduction for *prog-best* and a small reduction for *WIPo-best* relative to *ref*. Overall, inclusion of FSD processes within CICE results in changes to extent metrics of order 1 x $10^5\,km^2$.

4.3 Comparing the two FSD models

In this section, the *prog-best* and *WIPo-best* simulations will be compared directly, considering several metrics including total sea ice extent and volume, and spatial difference plots for area fraction, thickness, and l_{eff}. The aim of this comparison is to understand the differences in the impacts of the two alternative FSD models and how these differences emerge.

4.3.1 Pan-Arctic extent and volume

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 prognostic model produces a mean reduction in sea ice extent of just under 2% in June, compared to less than a 1% reduction with the WIPoFSD model; this reduction is just under 2% for both models in August. The average reduction in sea ice volume in September is 2.5% and 4% for *WIPo-best* and *prog*-best respectively relative to *ref*. The minimum reduction for the prognostic model is 1.5% in the spring months, compared to just 0.5% with the WIPoFSD model. The prognostic model also shows a larger interannual variability (indicated by the width of the ribbon) compared to the WIPoFSD model.

Figure 7 considers differences in the mean behaviour only, however the mean behaviour may obscure important trends. Figure 8 shows the percentage difference in total Arctic sea ice extent and volume for both *prog-best* and *WIPo-best* compared to *ref* from 1990 – 2016 in March and September. The differences are consistent with Fig. 7, with *prog-best* generally showing larger reductions than *WIPo-best* relative to *ref* other than for September extent. There is a possible positive trend for the difference in the March sea ice extent and a negative trend for the September sea ice extent for both *prog-best* and *WIPo-best* relative to *ref*, but this is inconclusive due to high interannual variability relative to the strength of the trend. More robust trends can be seen in the sea ice volume e.g. *prog-best* produces an average reduction in September volume of 2% compared to *ref* in the 1990s increasing to about a 5% reduction in the 2010s. A similar but weaker trend can be seen for *WIPo-best* relative to *ref*. The reduction in the March volume changes from about 1.1% in the 1990s to about 1.5% in the 2010s for *prog-best* relative to *ref*, whereas there is no evidence of any trend for *WIPo-best* relative to *ref*. Figure 8 shows that the interannual variability shown in Fig. 7 can be partly explained by long term trends, particularly for *prog-best* relative to *ref*.

4.3.2 Sea ice melt components

In *WIPo-best* and *prog-best*, floe size impacts the sea ice via two model components: form drag and lateral melt volume.

Several previous studies, including Tsamados et al. (2015) and Roach et al. (2018), found that increases in the lateral melt volume resulting from higher floe perimeter were compensated by a reduction in the basal melt. Bateson et al. (2020) demonstrated that this compensation effect was shown to primarily be a result of the physical reduction of sea ice area in locations of high basal melt. Figure 9 explores whether the same basal melt compensation effect is produced by the prognostic model. Figure 9 shows annual timeseries of the difference in the cumulative top, basal, lateral, and total melt for both *prog-best* and *WIPo-best* relative to *ref* averaged over 2000 – 2016. For both models a significant increase in lateral melt is compensated by a reduction in basal melt of similar magnitude, leading to only a small net increase in total melt. Whilst the increase in the lateral melt for *prog-best* is higher than for *WIPo-best*, both show an increase in the total melt of a small and similar magnitude. This suggests that any feedbacks on the total melt resulting from the decrease in area from enhanced lateral melt, such as the albedo feedback, are weak even for the *prog-best* simulation. The similar magnitude of change in the total melt also means that the results shown in Figs 7 and 8, where the sea ice volume is lower in both September and March for *prog-best* compared to *WIPo-best*, are unlikely to be driven by an increase in the total melt. This point will be discussed further in section 5.2. Also shown in Fig. 9 is the difference in melt components for *prog-best* compared to *WIPo-best*. The difference in cumulative total melt peaks in July and then decreases and switches sign. This is consistent with Fig. 7, where *prog-best* shows a stronger reduction in sea ice extent in the early melt season compared to *WIPo-best*, but the reduction in extent in August is comparable.

4.3.3 Spatial distribution of ice area fraction, thickness, and effective floe size

Previous studies e.g. Bateson et al. (2020) and Roach et al. (2018), have shown large FSD model impacts locally even where pan-Arctic impacts are small. Figure 10 shows maps of differences in sea ice area fraction and thickness for both *prog-best* and *WIPo-best* relative to *ref*, and spatial distribution plots of l_{eff} for both *prog-best* and *WIPo-best*. Results are presented for March, June, and September. The spatial pattern of the reduction in area fraction is similar for both *prog-best* and *WIPo-best* relative to *ref* but the magnitude is larger for the former in the early melt season. The region where l_{eff} drops significantly below 280 m in *WIPo-best* is generally confined to the outer MIZ. For the prognostic model, l_{eff} is generally well under 100 m across the MIZ. The distribution in l_{eff} corresponds to the regions of largest reduction in sea ice area fraction for *prog-best* relative to *ref* in the early melt season. The reduction in sea ice area fraction in the September MIZ is comparable in magnitude for both *prog-best* and *WIPo-best*, which is consistent with the results presented in Figs 7 and 9. For *prog-best*, l_{eff} is shown to increase within the MIZ over the course of the melt season from March to September, which can explain the different results in the early and late melt season. *prog-best* shows an increase in the sea ice area fraction across much of the pack ice in September, with a particularly strong response in the central Beaufort Sea. This response is not seen with *WIPo-best* because the maximum l_{eff} is 300 m for the selected model parameters i.e. the same as the fixed floe size in *ref*, whereas for *prog-best* it can be as high as 1700 m. For *prog-best* relative to *ref*, reductions in sea ice thickness persist through March and June across the central Arctic, but for *WIPo-best* differences only persist in locations that become marginal for at least some of the year and along the Canadian Archipelago. In September, the reduction in thickness spans the full Arctic for *prog-best*, whereas differences are mostly confined to the outer MIZ for *WIPo-best*.

Figure 10 shows much higher spatial variability in l_{eff} for *prog-best* compared to *WIPo-best*. A good case study is the relatively low l_{eff} seen in the Chukchi Sea during March and June. The floe formation mechanism is important to FSD evolution in this region of the Arctic since it experiences ice-free conditions for at least part of the year. Higher wave activity is also expected in this region due to an increased fetch via the Bering Strait, and this will increase the proportion of floes that form in smaller floe size categories. Other regions that experience ice-free conditions are generally more sheltered from wave activity due to adjacency to continental land mass. The only comparable regions in terms of wave exposure are the Greenland Sea and Barents Sea, where lower values of l_{eff} can also be seen. Further analysis (Bateson, 2021) indicates more generally that the high spatial variability in l_{eff} for the prognostic model cannot easily be attributed to a single process but is particularly sensitive to the floe formation mechanism, brittle fracture scheme, and welding, all processes not explicitly represented in the WIPoFSD model. The processes included in the WIPoFSD model, such as wave break-up of floes and lateral melt, are not found to have a large impact on the spatial distribution of l_{eff} within the prognostic model. One point of interest here is the regions of reduced l_{eff} shown for the WIPoFSD model appear to correspond well with the MIZ defined using wave activity presented in Horvat et al. (2020), which suggests that a possible application of FSD models would be an alternative way of defining the MIZ compared to the sea ice concentration-derived definition.

4.3.4 Standard deviation of sea ice area fraction, thickness, and effective floe size

Figure 10 is useful to understand the spatial distribution of the pan-Arctic changes in sea ice state shown in Fig. 7, but the inclusion of an FSD model may not only act to change the mean state of the sea ice but also the interannual variability. Furthermore, a small change in mean sea ice state may disguise a much larger change in sea ice variability. Figure 11 shows the standard deviation in sea ice area fraction and thickness for *ref*, the difference in standard deviation for area fraction, thickness, and l_{eff} for *prog-best* and *WIPo-best* relative to *ref*. Changes to the standard deviation in area fraction have a low magnitude and are isolated rather than part of a more systematic change in behaviour. For the standard deviation in thickness, differences are again small and isolated in March, but larger changes can be seen within the September MIZ of up to 10 – 20%. These changes in thickness variability correspond to where the largest differences in sea ice thickness can be seen in Fig. 10 and are consistent with the high interannual variability of the reduction in sea ice volume suggested in Fig. 8. For *WIPo-best*, variability in l_{eff} is generally only seen within the MIZ as l_{eff} remains close to the maximum value within the pack ice. For *prog-best*, the standard deviation in l_{eff} broadly correlates to the magnitude of l_{eff} seen in Fig. 10. High interannual variability can be seen across the pack ice in both March and September for *prog-best*, suggesting that all locations experience some differences in the contributing processes to the emergent FSD year on year. These distinct patterns in interannual variability of l_{eff} for *prog-best* and *WIPo-best* may therefore be a useful metric to measure in the Arctic to discriminates between the different approaches to modelling the FSD.

5. Discussion

5.1 Inclusion of brittle fracture in FSD models

Observations of the sea ice cover suggest brittle fracture processes have a role in the evolution of the FSD via both the direct impact of a fracture event on floe size and indirectly via thermodynamic weakening and subsequent breakup along weaknesses in the sea ice cover resulting from prior fracture events that subsequently froze up (e.g. Perovich et al., 2001, Arntsen et al., 2015; Hwang et al., 2017). This motivated the inclusion of a scheme to approximate the impact of brittle fracture processes on the FSD within the prognostic model based on idealised models of brittle fracture.

The inclusion of the quasi-restoring brittle fracture scheme into the prognostic FSD model significantly improved the simulated shape of the FSD for mid-sized floes of 100 m – 2000 m (Fig. 4). This was particularly true when comparing model output to observations in the Fram Strait and East Siberian Sea (panels c-f in Fig. 4), however, improvements were less impressive for the Chukchi Sea site (panels a-b in Fig. 4). To understand the difference in model performance at these sites, consider their locations shown in Fig. 3. Brittle fracture acts on the FSD at two of the three sites throughout the year, but only for part of the year for the Chukchi Sea since it fully transitions to an ice-free state over the melting season, unlike the other sites. The restoring timescale for the brittle fracture scheme, τ, only determines the timescale for two neighbouring floe size categories to reach an equilibrium state rather than the entire distribution. The prognostic FSD model consists of 16 categories for the simulations considered in Fig. 3 and it would take several months for a starting state with all sea ice area in the largest floe size category to reach equilibrium across all floe size categories, a long enough lag to explain the different prognostic model performance for the Chukchi Sea. A physical interpretation for τ has previously been given as the timescale for sea ice to thin sufficiently that the sea ice is vulnerable to in-plane fracture events along existing weaknesses. It was discussed in section 2.2.2 that assuming a fixed value for τ is a significant approximation given the significant spatial and temporal variability of dependent factors such as sea ice thickness and melt rates. The difference in performance between sites considered in this study can therefore be considered a result of this approximation.

In order to understand the implications of the results presented in Fig. 4 for prognostic floe size modelling it is useful to consider why the brittle fracture scheme is able to improve model performance against observations. Considering the distributions presented in Fig. 4 for the standard prognostic model without brittle fracture, it does not obviously follow that a redistribution of sea ice area from larger categories to smaller categories would improve the shape of the distribution compared to the observations given the gradient is already too negative. However, the largest floe size categories in the prognostic model are excluded from the comparison to observations to exclude the non-physical 'uptick' that forms, as demonstrated in Fig. 2. The inclusion of brittle fracture acts to reduce the size of the uptick and redistributes sea ice area to mid-sized floe categories. There are two plausible factors that can produce this uptick: the truncation of the maximum possible floe size such that sea ice area accumulates in the largest category that would otherwise be distributed over several larger categories; and missing floe fragmentation processes in the prognostic model. Observations show that floes can exceed a size of 10 km (Stern et al., 2018 a), providing evidence that the former factor contributes to the uptick, but the results presented in Fig. 4 suggest that missing floe fragmentation processes also contribute, with brittle fracture related mechanisms being a leading candidate.

We do not suggest that our brittle fracture scheme is taken to be the final solution to representing brittle fracture processes in the prognostic model. The current scheme makes several significant simplifications including only allowing area transfer between adjacent categories and using a fixed restoring timescale. However, sensitivity studies show that the brittle fracture scheme does not dominate the shape of the emergent FSD and other processes continue to be important in the evolution of the FSD, particularly winter growth processes such as floe formation and welding (Bateson, 2021). Despite the limitations with the quasi-restoring brittle fracture scheme, the results shown in Fig. 4 strongly suggest that brittle fracture or related fragmentation processes are required to capture the shape of the FSD for mid-sized floes, motivating the need to develop a physically derived brittle fracture parameterisation for FSD models.

An interesting comparison can be made between the treatment of brittle fracture within the prognostic model presented here and recent developments introducing the concept of 'damage' to the treatment of rheology within sea ice models (Dansereau et al., 2016). One such sea ice rheology, named the Maxwell-elasto-brittle (Maxwell-EB) rheology, has been applied within the continuous and fully Lagrangian sea ice model neXtSIM (Rampal et al., 2019). This new rheology retains a 'memory' of any fracture events, effectively tracking how 'damaged' the sea ice cover is, and modifies the sea ice properties accordingly. This concept of 'damage' has clear parallels with the mechanism discussed above of how winter in-plane brittle fracture events can determine how the sea ice breaks-up in summer and may therefore present a useful basis for the development of a full parametrisation of brittle fracture processes for use in FSD models. Boutin et al. (2021) also demonstrated that the Maxwell-EB rheology can be combined with an FSD model in order to explore how wave break-up of floes can impact sea ice dynamics, highlighting an application of floe size modelling not considered in this study.

5.2 Differences in the impacts of the FSD models and how they emerge

Focusing first on a pan-Arctic scale, Figs 5-6 do not provide any evidence that the inclusion of an FSD model improves the performance of CICE in simulating the aggregated Arctic sea ice extent and volume behaviours and trends against observations or reanalyses. This is an important result for climate modellers since it assuages concerns that the FSD represents a source of structural uncertainty in climate models. This also does not preclude either FSD model from being an important improvement to sea ice models; these improvements may be at a regional scale rather than at a pan-Arctic scale. The accuracy of sea ice concentration measured using passive microwave data can be as low as ± 20% in summer or the MIZ (Meier and Notz, 2010). Measurements in sea ice thickness from radar altimetry can also have high uncertainty, with snow depth and density being the primary source of error (Tilling et al., 2018). It is therefore non-trivial to obtain high accuracy observations of the spatial distribution of sea ice concentration and thickness, and the use of the latter to validate models requires a careful use of case studies such as demonstrated by Schröder et al. (2019). Nevertheless, significant biases have been identified in coupled climate models in simulating the sea ice concentration (Ivanova et al., 2016) and CICE, in particular, has been shown to overpredict the sea ice concentration at the sea ice edge and underpredict the concentration within the pack ice (Schröder et al., 2019). In Bateson et al. (2020), the WIPoFSD model was found to provide a limited correction to this model bias. Similarly, Fig. 9 shows that the prognostic model produces a stronger correction to this model bias, driving reductions in sea ice area fraction in the MIZ and small increases in area fraction in the pack ice.

Figures 7-10, described in section 4.3, all highlight a key difference in the impact of the two FSD models. *prog-best* produces a stronger reduction in sea ice area fraction relative to *ref* in the early melt season but a more comparable reduction by August compared to *WIPo-best*. This difference can be attributed to the different treatment of floe formation and growth processes between the two models. The WIPoFSD model uses a simple restoring approach that operates over a short timescale of 10 days, which is applied during conditions of freezing. This means that over winter l_{eff} will be at or close to its maximum value uniformly across the sea ice, except in locations at the outer sea ice edge that are exposed to wave break-up, as shown in Fig. 10. In comparison, the prognostic model aims to represent physically floe formation and growth processes such that the homogeneity produced after the freeze-up season with the WIPoFSD model is not seen with the prognostic model in Fig. 10. In the early melt season l_{eff} is particularly low across the MIZ due to wave activity in this region causing existing floes to fragment and new floes to form in the smallest size category. As the melting season proceeds, floes in the smallest floe size categories preferentially lose surface area and melt out in response to lateral melting, a behaviour that is also visible in Fig. 10 e.g. l_{eff} increases in the Fram Strait between March and June. This behaviour is not possible with the WIPoFSD model since it has a fixed exponent and minimum floe size. These results show the value of l_{eff} in being able to characterise and understand how the inclusion of either FSD model impacts the sea ice cover and in understanding how differences in these impacts emerge. These results also show the potential limitations of using a simplified FSD model such as the WIPoFSD model; even though a power law might in general be a good fit to the FSD over the melt season, there could still be important mechanisms and features of FSD impacts that it fails to properly capture.

In section 4.3.2 it was noted that the total cumulative melt is slightly higher for *WIPo-best* compared to *prog-best* despite the larger reduction in volume for *prog-best* compared to *WIPo-best* relative to *ref* shown in Figs 7 and 8. Figure 9 also shows the lateral to basal melt ratio is higher for *prog-best* compared to *WIPo-best*; the change in this ratio may present an alternative explanation to explain the larger volume reduction produced by *prog-best*. Two floes with the same diameter but different thickness will, under identical conditions, contribute the same volume to the total basal melt (provided the thinner floe does not melt out) whereas the thicker floe will contribute a greater volume to the total lateral melt, since lateral melt volume is proportional to thickness. This means that increasing the lateral melt contribution to the total melt increases the loss of thick ice in a given melt season. Vertical sea ice growth rates are inversely proportional to the sea ice thickness. Therefore, whilst the moderate reductions in thickness across large areas of sea ice from basal melt can be recovered within a single freeze-up season, the recovery of thick ice that has completely melted out from lateral melt will take several seasons of freeze-up to recover despite being over a smaller area. The reason for larger reductions in sea ice volume for *prog-best* compared to *WIPo-best* may therefore be a result of a changes to the ice thickness distribution that emerge due to the higher lateral to basal melt ratio for *prog-best* compared to *WIPo-best*. This result shows that the inclusion of an FSD model can have important mechanistic impacts on sea ice evolution, even if the immediate change in pan-Arctic properties are small. Smith et al. (2022) recently demonstrated that the partitioning between lateral and basal melt can also have a large impact on open water formation, with implications for albedo feedback.

5.3 Advantages and disadvantages of FSD models

We have examined examples of two broad categories of FSD models where either the FSD shape is imposed or where the FSD shape emerges from parametrisations at process level. A factor that must always be considered when introducing new components to sea ice and climate models is computational efficiency. A simple, low-cost FSD plugin would in general be preferential for use in climate models, but only if it is able to capture how the FSD will behave in future climate scenarios. A high level of uncertainty around model parameters and parameterisations would also preclude the use of a given FSD model in climate simulations. The WIPoFSD model is not computationally expensive; including the WIPoFSD model within CICE increases model run time by 30%. In comparison, the prognostic model is computationally expensive and is data intensive. The use of 12 floe size categories with the standard 5 thickness categories introduces a total of 60 floe size-thickness outputs to the model and simulation times increase by a factor of 2.1. Extending this to 16 floe size categories leads to a total of 80 categories and a further increase in simulation run time. The number of floe size-thickness categories can also make it difficult to diagnose and understand how changes to the sea ice state emerge in response to prognostic model processes. In comparison, identifying the mechanisms driving changes in sea ice state is more straightforward with the WIPoFSD model. Development of the prognostic model can also be more time intensive since each process requires either observations or lab-based studies to determine a suitable parameterisation to describe the physical process in the model. It is worth noting that future advancements in modelling techniques may reduce or mitigate the computational expense or complexity of either model e.g. Horvat and Roach (2022) presented a machine-learning-based parameterisation to simulate wave break-up of sea ice floes that can replace the existing treatment of wave break-up in the prognostic model. The study found that CICE simulations including the prognostic model with this new parameterisation have an approximately 40% longer run time than CICE simulations without the prognostic model i.e. a comparable cost to the WIPoFSD model.

Another important test for any FSD model is whether it simulates realistic FSD shape and variability. Figure 4 shows a power law produces a strong fit to observations of floes over a mid-sized (100 m – 2 km) range. The prognostic model with brittle fracture performed comparably to the power-law fit except in the Chukchi Sea. However, this comparison only included observations covering one quarter of the year and excluded floes smaller than about 100 m and larger than 2 km. Floes smaller than 100 m are particularly important for determining the impact of a given FSD on sea ice evolution (Bateson et al. 2020) and there is growing evidence that the power law may not hold across all floe sizes (Horvat et al., 2019) and that the power-law exponent changes significantly over an annual cycle (Stern et al., 2018b). In Bateson et al. (2020), it was found that imposing the annual cycle reported by Stern et al. (2018b) on the exponent only had a small impact on the sea ice state. The annual cycle imposed was taken as the mean from the Chukchi and Beaufort Seas only, so it is not sufficient evidence to conclude that a fixed exponent is a reasonable assumption.

A key advantage of the prognostic model approach is the shape of the FSD is an emergent feature of the model rather than imposed, avoiding the need to make assumptions about FSD shape or variability, notwithstanding the newly introduced brittle fracture scheme, which, as discussed above, requires further development. This also means the prognostic model can be used to understand the role of individual processes in determining the emergent FSD and can respond to future changes in the behaviour or strength of these processes. In section 5.2 it is shown how behaviours seen within the prognostic model such as the preferential melt out of smaller floes from the distribution or explicit simulation of floe formation and growth processes have impacts on the evolution of the sea ice cover. These impacts are not seen within the WIPoFSD model due to the restrictions of assuming a fixed FSD shape. Furthermore, the WIPoFSD model effectively operates by tuning model parameters to best capture observations of the FSD, however this tuning may not be appropriate over the full timescale of a simulation.

5.4 Limitations of these results

A significant limitation originates from the limited availability of observations to constrain FSD model parameters and parameterisations. Whilst the parameters selected for the standard setup of the WIPoFSD model considered here were motivated as a best fit to observations, Bateson et al. (2020) demonstrated high sensitivity in the model response within the observational uncertainty of these parameters. For the prognostic model, the uncertainties are primarily associated with both the representation of existing processes and important processes not yet represented in the model. For example, Roach et al. (2019) demonstrated that using a full wave model coupled to CICE rather than the in-ice wave scheme used here approximately doubled the total lateral melt, though this was compensated by a reduction in basal melt of comparable magnitude.

A second limitation emerges from the use of a standalone sea ice model since this prevents the full representation of sea ice-ocean or atmosphere feedbacks. For example, Roach et al. (2019) found with a coupled sea ice-ocean model that their prognostic FSD setup produced an increase in lateral melt in the Arctic about 3-4 times higher than the reduction in basal melt, resulting in an approximately 20% increase in the total lateral and basal melt. Roach et al. (2019) do not identify the mechanism responsible for this increase, so it is not clear whether the FSD model setups used here would produce a similar response in a coupled CICE-NEMO setup.

In addition, this study has not explored the role of the FSD in sea ice evolution in the Southern Ocean. Several observational studies e.g. Alberello et al. (2019) show the presence of FSDs dominated by pancake ice floes smaller than 10 m in the Southern Ocean. Sensitivity studies presented in Bateson et al. (2020) suggest that distributions dominated by such small floes can result in a significant increase in total melt and a large corresponding reduction in sea ice volume. Therefore, conclusions regarding the role of the FSD in Arctic sea ice evolution do not necessarily extend to the Antarctic. For example, Roach et al. (2019) apply a version of the prognostic model to both the Arctic and Antarctic (including a coupled wave model but not the brittle fracture scheme) and demonstrate a pan-Antarctic reduction in sea ice volume whereas in the Arctic there are regions of volume increase and decrease, with the latter found primarily in the MIZ.

6. Conclusion

In this study we have evaluated and compared two alternative approaches to modelling the sea ice floe size distribution: a prognostic model where the shape of the FSD emerges from the model physics (Roach et al., 2018, 2019), and the WIPoFSD model where the shape of the FSD is constrained to a power law with a fixed exponent (Bateson et al., 2020). New, high-resolution observations of the FSD were used to assess model performance in simulating the FSD shape for mid-sized floes and to determine FSD model parameters for the comparison. The prognostic model was unable to simulate the observed FSD shape, however the introduction of a new scheme to approximate the effects of brittle fracture related processes significantly improved prognostic model performance.

Simulations were completed using the two FSD models within a standalone setup of the sea ice model CICE, though it should be noted that both FSD models can easily be implemented into any continuum sea ice model. Whilst impacts of both FSD models were small over a pan-Arctic scale, larger impacts could be found regionally. Changes to the spatial distribution in sea ice area fraction were consistent with known model biases, particularly for the prognostic model. Clear differences were found between the two models in the strength of the early melt season and long-term trends in the sea ice volume. The faster retreat of sea ice in the early melt season for the prognostic model compared to the WIPoFSD model was attributed to the different model treatments of floe formation and growth in winter. The slower retreat in extent during the later melt season for the prognostic model was found to be a result of melt out of smaller floes, a feedback process not possible with the WIPoFSD model due to the restrictions on FSD shape. It was also highlighted how changes to the lateral to basal melt ratio can indirectly impact the volume of winter sea ice growth via the ice-thickness distribution. These results are important for climate modellers as they suggest that the FSD is not a significant source of structural uncertainty in climate models. FSD processes will, however, be of importance for several key applications and research questions such as regional sea ice modelling and the formation of open water during the melt season (e.g. Smith et al., 2022).

We discussed advantages and disadvantages of the two approaches to modelling the FSD. The WIPoFSD model is more computationally efficient and a simple model to interpret. In addition, a power law fit using a single exponent averaged across all FSD observations produced a good fit to the observed FSD across all locations in the dataset considered here. However, some behaviours seen with the prognostic model could not be replicated by the WIPoFSD model due to restrictions on FSD shape. Furthermore, the prognostic model is better able to respond to regime change of the processes that determine the FSD shape and new modelling techniques could significantly mitigate the computational cost (e.g. Horvat and Roach, 2022).

Future work should focus on the development of a full physical treatment of the impact of brittle fracture on the FSD. Whilst the quasi-restoring scheme presented here is a useful tool to improve prognostic model performance and based on idealised models of brittle fracture, its current formulation relies on significant approximations. The concept of defining the 'damage' of a given area of sea ice such as used within the Maxwell-EB rheology (Dansereau et al., 2016) presents a promising basis for future developments of the brittle fracture scheme. In addition, it would also be beneficial to evaluate whether the conclusions reached in this study extend to the Antarctic. Finally, the results presented here highlight the need to collect further observations of the FSD. Horvat et al. (2019) demonstrated that it is possible to estimate the area-weighted floe size from satellite imagery. It is plausible that l_{eff} could also be estimated from satellite imagery especially since the methodology of Horvat et al. (2019) involved collecting linear statistics of floe size, and l_{eff} is a linearly averaged representation of the FSD. This would establish a way to observationally establish the spatial and temporal variability of the FSD. These observations can provide further constraints for FSD models, which have previously been demonstrated to have high sensitivity to FSD parameters (Bateson et al., 2020) and how individual processes are represented or parameterised (Roach et al., 2019).

Appendices

Appendix A – Description of processes represented in the WIPoFSD model

The WIPoFSD model presented in Bateson et al. (2020) includes four mechanisms that can change l_{var} and therefore the FSD. The first of these, lateral melting, is treated by assuming the reduction in l_{var}^2 from lateral melting is proportional to the reduction in the sea ice area fraction from lateral melting, ΔA_{lm}:

$$l_{var,final} = l_{var,initial}\sqrt{1 - \frac{\Delta A_{lm}}{A}}. \quad \text{(A1)}$$

The second mechanism is the break-up of floes by waves. If a wave break-up event is identified, l_{var} is then updated according to the following expression:

$$l_{var} = \max\left(d_{min}, \frac{\lambda_W}{2}\right). \quad \text{(A2)}$$

Here λ_W is a representative wavelength, in units of metres. In order to calculate the value of λ_W and to identify where the ocean surface conditions are sufficient to drive wave break-up, the WIPoFSD model uses a wave attenuation and floe breakup scheme adapted from the waves-ice model of the Nansen Environmental and Remote Sensing Centre (NERSC) Norway, details are given by Williams et al. (2013a, 2013b). The WIPoFSD model also uses a wave advection scheme developed by NOC in the NEMO-CICE-WIM model. Further details of how these schemes have been incorporated into the WIPoFSD model, and how break-up events are identified, are available in Bateson et al. (2020).

The WIPoFSD model treats floe growth using a simple floe restoring scheme. During periods where CICE identifies frazil ice growth i.e. when the freezing / melting potential is positive, l_{var} is restored to its maximum value:

$$l_{var,final} = \min\left(d_{max}, l_{var,initial} + \frac{d_{max}\Delta t}{T_{rel}}\right). \quad \text{(A3)}$$

T_{rel} is a relaxation timescale that represents how quickly floes would be expected to grow to cover the entire grid cell area. It is set to 10 days, taken as representative of the rapid increase in sea ice concentration during the early freeze-up season. In grid cells that newly transition to having a sea ice cover from an ice-free state, l_{var} is initiated with the value d_{min}.

The fourth and final mechanism treated by the WIPoFSD model is advection. l_{var} is transported using the horizontal remapping scheme with a conservative transport equation, the standard within CICE for ice area tracers (Hunke et al., 2015). Since l_{var} is not defined independently for each thickness category, the change in l_{var} after advection and subsequent mechanical redistribution is calculated independently for each thickness category, with the net change in l_{var} taken as the average across all the thickness categories.

Appendix B – The updated lateral melting scheme

A derivation is presented in this section of Eq. (14), the updated lateral melting scheme within the version of the WIPoFSD model used in this study. The floe number distribution can be written explicitly according to Eq. (13) and by evaluating the constant, C, according to its definition as described in section 2.3:

$$N(x \,|d_{min} \le x \le l_{var}) = \frac{(3+\alpha)Al^2}{\alpha_{shape}} \frac{x^{\alpha}}{[l_{var}^{3+\alpha} - d_{min}^{3+\alpha}]}. \tag{B1}$$

Bateson et al. (2020) derives an expression of l_{eff} for this distribution:

$$l_{eff} = \frac{(2-\alpha)[l_{var}^{3-\alpha} - d_{min}^{3-\alpha}]}{(3-\alpha)\,[l_{var}^{2-\alpha} - d_{min}^{2-\alpha}]}. \tag{B2}$$

It is also possible to derive an expression for l_{eff} after lateral melting. If floes experience an amount Δl of lateral melting on each edge, the total diameter of each floe must decrease by $2\Delta l$. This changes the size of the floes but does not impact the shape of the number distribution i.e. floes of diameter G prior to lateral melting and floes of diameter $G - 2\Delta l$ after lateral melting have the same number density, $N(G)$, where $N(x)$ is the number distribution prior to lateral melting. This description is true provided $d_{min} > 2\Delta l$ i.e. no floes are completely lost from the distribution due to lateral melting. The total perimeter after the lateral melting event, P_{lm}, can therefore be calculated as:

$$P_{lm} = \int_{d_{min}}^{l_{var}} \pi(x - 2\Delta l) N(x)\, dx. \tag{B3}$$

$N(x)$ in Eq. (B3) is the number FSD prior to lateral melting, and this equation holds provided $d_{min} > 2\Delta l$. This can then be evaluated as:

$$P_{lm} = \frac{\pi(3+\alpha)A_{old}l^2}{\alpha_{shape}\,[l_{var}^{3+\alpha} - d_{min}^{3+\alpha}]}\left(\frac{[l_{var}^{2+\alpha} - d_{min}^{2+\alpha}]}{(2+\alpha)} - \frac{2\Delta l[l_{var}^{1+\alpha} - d_{min}^{1+\alpha}]}{(1+\alpha)}\right). \tag{B4}$$

The subscript for A_{old} indicates that this is the sea ice area fraction before lateral melting. An expression for the total perimeter in terms of the new effective floe size, $l_{eff,new}$, can also be written using the updated sea ice area fraction after lateral melting, A_{new}:

$$P_{leff} = \frac{A_{new}l^2\pi}{\alpha_{shape}l_{eff,new}}. \tag{B5}$$

The two expressions for total perimeter after lateral melting can then be equated to give the updated effective floe size, $l_{eff,new}$:

$$l_{eff,new} = \frac{[l_{var}^{3+\alpha} - d_{min}^{3+\alpha}]A_{new}}{(3+\alpha)\,A_{old}}\left(\frac{[l_{var}^{2+\alpha} - d_{min}^{2+\alpha}]}{(2+\alpha)} - \frac{2\Delta l[l_{var}^{1+\alpha} - d_{min}^{1+\alpha}]}{(1+\alpha)}\right)^{-1}. \tag{B6}$$

It is possible to calculate an analytical result for A_{new} as a result of lateral melting of floes across the distribution, however CICE already accounts for changes to the sea ice area fraction. For internal model consistency, it is this internal CICE A_{new} that will be used.

In order to parameterise processes in terms of l_{eff}, a method is needed to calculate l_{var} from l_{eff}. There is no analytical solution to this problem; instead a numerical approach must be used such as Newton-Raphson iteration:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{B7}$$

Here x is l_{var} and the function to solve is derived from the expression to calculate l_{eff} for the WIPoFSD model i.e.

$$f(l_{var}) = 0 = \frac{(2+\alpha)[l_{var}^{3+\alpha} - d_{min}^{3+\alpha}]}{(3+\alpha)\,[l_{var}^{2+\alpha} - d_{min}^{2+\alpha}]} - l_{eff}. \tag{B8}$$

The iterative scheme can then be evaluated as:

$$l_{var,n+1} = l_{var,n} - \frac{\left(\frac{[l_{var.n}^{3+\alpha} - d_{min}^{3+\alpha}]}{(3+\alpha)} - \frac{l_{eff}\left[l_{var,n}^{2+\alpha} - d_{min}^{2+\alpha}\right]}{(2+\alpha)}\right)}{l_{var}^{1+\alpha}(l_{var.n} - f(l_{var.n}))}. \tag{B9}$$

Note that, for simplicity, where α = -1, -2 or -3, a value of 0.001 will be taken off. Whilst an exact solution is possible for these cases, this adds additional and unnecessary complexity to a scheme that is already an approximation. This scheme is evaluated until either $l_{var,n+1}$ - $l_{var,n}$ is less than 0.01% of the change in l_{eff} over a timestep or until a maximum of 50 iterations are complete. In general, the threshold for convergence is achieved within 10 iterations, however where l_{eff} and l_{var} are close in value i.e. where l_{var} is within a few metres of d_{min}, convergence can take longer than 50 iterations. These circumstances are associated with conditions of very low sea ice concentration, where the net error in the lateral melt volume calculation associated with the failure to reach the threshold condition for convergence is negligible.

Data Availability

Model output used in this paper is publicly available via the University of Reading research data archive (http://dx.doi.org/10.17864/1947.300; Bateson, 2021b). Output for the simulation *prog-16-nobf* can be found listed under chapter 6 as *cice_cpom_prog_16cat_nobf*. Output for the simulation *prog-16* can be found listed under chapter 7 as *cice_cpom_prog_16cat*. Output for the simulations *ref*, *prog-best*, and *WIPo-best* can be found listed under chapter 8 as *cice_cpom_ref*, *cice_cpom_prog_best*, and *cice_cpom_WIPo_best* respectively. Mask files defining the regions described in Fig. 3 can be found listed under chapter 6. Please contact the corresponding author to discuss access to model code.

Author Contributions

YW and BH produced the novel FSD observations from satellite imagery. AB completed the comparison of the novel FSD observations to model output, with support from YW, BH, DF, and DS. DS adapted the prognostic FSTD model of Roach et al. (2018) into the CPOM CICE stand-alone set-up. AB implemented the novel modifications described in this paper to the original FSD model setups. AB completed simulations and analysis under the supervision of DF and DS and with further support from YW, BH, JR, and YA. DS provided additional technical support. AB composed the paper with feedback and contributions from all authors.

Competing interests

Daniel Feltham, David Schröder, and Yevgeny Aksenov are members of the editorial board of The Cryosphere. The peer-review process was guided by an independent editor, and the authors have also no other competing interests to declare.

Acknowledgements

AB was funded through a NERC industrial CASE studentship with the UK Met Office (NE/M009637/1). DF, DS, YW, and BH were supported by NERC grant (NE/R000654/1). DS was also supported under the NERC projects ACSIS (NE/N018044/1) and UKESM. YW and BH were also supported through NERC grant (NE/S002545/1). JR was supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01 101). YA was supported by the NERC projects ACSIS (NE/N018044/1), “Towards a marginal Arctic sea ice cover" (NE/R000085/1), the NERC Project “PRE-MELT” (NE/T000260/1) and the NERC LTS-S Programme Climate-Linked Atlantic Sector Science (CLASS; NE/R015953/1). This work was also supported by NERC through National Capability funding, undertaken by a partnership between the Centre for Polar Observation Modelling and the British Antarctic Survey. We would like to express our gratitude to Lettie Roach (University of Washington, USA) for her guidance and support in the use of the prognostic FSTD model. Similarly, we would like to thank Lucia Hosekova (University of Reading, UK; Applied Physics Laboratory, University of Washington, USA) and Stefanie Rynders (National Oceanography Centre, Southampton, UK) for their support in the use of the WIPoFSD model. We would also like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the “Mathematics of Sea Ice Phenomena” programme when work on this paper was undertaken. Finally, we would like to thank both the reviewers and the editor, whose detailed and constructive comments have enabled us to significantly improve the manuscript.

A brief note on WIPoFSD model development: The WIPoFSD model code has been derived and modified from a scheme used within the coupled NEMO–CICE–WIM sea ice–ocean–waves interaction model developed by the L. Hosekova and Y. Aksenov at the National Oceanography Centre (NOC) in the EC FP7 project ‘Ships and Waves Reaching Polar Regions’ (SWARP)’ in 2014-2017 (https://cordis.europa.eu/project/id/607476, grant agreement 607476). The physics of waves–ice interactions in NOC–WIM (WIM – stands for waves in ice model) model uses the framework of the waves–in–ice model of the Nansen Environmental and Remote Sensing Center (NERSC, Norway) by Williams et al. (2013a, b). Given the differences between the NERSC Arctic regional model setup (HYCOM ocean model with an early EVP sea ice rheology realisation and

Semtner's sea ice thermodynamics) and the global NOC–NEMO–CICE setup, the coding of the NOC–WIM model and all coupling with ocean and sea ice modules has been done completely from the start by NOC. During the course of the NOC–NEMO–CICE–WIM model development L. Hosekova, advised by Y. Aksenov, has incorporated extra key processes in the model, including up-wind wave spectrum advection in the ice-covered areas, floe sizes evolution due to lateral floes melting, including renormalization algorithm with the Newton-Raphson method to compute mean from maximum floe sizes, freeze-up and floe sizes advection, and an optional choice of the multiple power law exponents for the FSD (Rynders, 2017); for these we also acknowledge contributions from G. Madec (IPSL) and A.J.G Nurser (NOC). The NOC–NEMO–CICE–WIM model includes the novel development of the combined EVP-collisional rheology (EVCP) (Feltham, 2005) coded in CICE by S. Rynders and advised by Y. Aksenov and D. Feltham (Rynders, 2017, Rynders et al., 2022). The NEMOv3.6–CICEv5.1–WIM model code had been shared with the Centre for Polar Observations and Modelling (CPOM) at the University of Reading for the joint research under the UK Joint Marine Modelling Programme (the UK Joint Weather and Climate Research Programme – JWCRP).

References

Aksenov, Y., Popova, E. E., Yool, A., Nurser, A. J. G., Williams, T. D., Bertino, L., and Bergh, J.: On the future navigability of Arctic sea routes: High-resolution projections of the Arctic Ocean and sea ice, Mar. Policy, 75, 300–317, https://doi.org/10.1016/j.marpol.2015.12.027, 2017.

Alberello, A., Onorato, M., Bennetts, L., Vichi, M., Eayrs, C., MacHutchon, K., and Toffoli, A.: Brief communication: Pancake ice floe size distribution during the winter expansion of the Antarctic marginal ice zone, The Cryosphere, 13, 41–48, https://doi.org/10.5194/tc-13-41-2019, 2019.

Arntsen, A. E., Song, A. J., Perovich, D. K., and Richter-Menge, J. A.: Observations of the summer breakup of an Arctic sea ice cover, Geophys. Res. Lett., 42, 8057–8063, https://doi.org/10.1002/2015GL065224, 2015.

Åstrom, J. A., Ouchterlony, F., Linna, R. P. and Timonen, J.: Universal dynamic fragmentation in D dimensions, Phys. Rev. Lett., 92(24), 1–4, doi:10.1103/PhysRevLett.92.245506, 2004.

Bateson, A. W.: Fragmentation and melting of the seasonal sea ice cover, Ph.D. thesis, Department of Meteorology, University of Reading, United Kingdom, 293 pp., 2021a.

Bateson, A. W.: Simulations of the Arctic sea ice comparing different approaches to modelling the floe size distribution and their respective impacts on the sea ice cover, University of Reading [data set], doi:10.17864/1947.300, 2021b.

Bateson, A. W., Feltham, D. L., Schröder, D., Hosekova, L., Ridley, J. K. and Aksenov, Y.: Impact of sea ice floe size distribution on seasonal fragmentation and melt of Arctic sea ice, Cryosphere, 14(2), 403–428, doi:10.5194/tc-14-403-2020, 2020.

Bennetts, L. G., O'Farrell, S. and Uotila, P.: Brief communication: Impacts of ocean-wave-induced breakup of Antarctic sea ice via thermodynamics in a stand-alone version of the CICE sea-ice model, Cryosphere, 11(3), 1035–1040, doi:10.5194/tc-11-1035-2017, 2017.

Boutin, G., Lique, C., Ardhuin, F., Rousset, C., Talandier, C., Accensi, M. and Girard-Ardhuin, F.: Towards a coupled model to investigate wave-sea ice interactions in the Arctic marginal ice zone, Cryosphere, 14(2), 709–735, doi:10.5194/tc-14-709-2020, 2020.

Boutin, G., Williams, T., Rampal, P., Olason, E., and Lique, C.: Wave–sea-ice interactions in a brittle rheological framework, The Cryosphere, 15, 431–457, https://doi.org/10.5194/tc-15-431-2021, 2021.

Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally.: Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1, Natl. Snow and Ice Data Cent., Boulder, CO, available at: http://nsidc.org/data/NSIDC-0051/versions/1.html (last access: 31 December 2016), 1996 (updated 2016).

Comiso, J.: Bootstrap Sea Ice Concentrations From NIMBUS-7 SMMR and DMSP SSM/I, Natl. Snow and Ice Data Cent., Boulder, CO, available at: http://nsidc.org/data/nsidc-0079.html (last access: 31 December 2017), 1999 (updated 2017).

Dansereau, V., Weiss, J., Saramito, P., and Lattes, P.: A Maxwell elasto-brittle rheology for sea ice modelling, The Cryosphere, 10, 1339–1359, https://doi.org/10.5194/tc-10-1339-2016, 2016.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N. and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., doi:10.1002/qj.828, 2011.

de Boer, G., Shupe, M. D., Caldwell, P. M., Bauer, S. E., Persson, O., Boyle, J. S., Kelley, M., Klein, S. A., and Tjernström, M.: Near-surface meteorology during the Arctic Summer Cloud Ocean Study (ASCOS): evaluation of reanalyses and global climate models, Atmos. Chem. Phys., 14, 427–445, https://doi.org/10.5194/acp-14-427-2014, 2014.

Feltham, D. L.: Granular flow in the marginal ice zone, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., 363(1832), 1677–1700, doi:10.1098/rsta.2005.1601, 2005.

Ferry, N., Masina, S., Storto, A., Haines, K., Valdivieso, M., Barnier, B. and Molines, J.-M.: Product user manual global-reanalysis-phys-001-004-a and b, MyOcean, Eur. Comm., Brussels., 2011.

Frew, R. C., Feltham, D. L., Holland, P. R., and Petty, A. A.: Sea ice – Ocean Feedbacks in the Antarctic Shelf Seas, J. Phys. Oceanogr., 49, 2423–2446, https://doi.org/10.1175/JPO-D-18-0229.1, 2019.

Gherardi, M. and Lagomarsino, M. C.: Characterizing the size and shape of sea ice floes, Sci. Rep., doi:10.1038/srep10226, 2015.

Herman, A.: Sea-ice floe-size distribution in the context of spontaneous scaling emergence in stochastic systems, Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys., 81(6), 1–5, doi:10.1103/PhysRevE.81.066123, 2010.

Herman, A.: Influence of ice concentration and floe-size distribution on cluster formation in sea-ice floes, Cent. Eur. J. Phys., 10(3), 715–722, doi:10.2478/s11534-012-0071-6, 2012.

Herman, A., Wenta, M. and Cheng, S.: Sizes and Shapes of Sea Ice Floes Broken by Waves–A Case Study From the East Antarctic Coast, Front. Earth Sci., 9(May), 1–14, doi:10.3389/feart.2021.655977, 2021.

Horvat, C. and Tziperman, E.: A prognostic model of the sea-ice floe size and thickness distribution, Cryosphere, 9(6), 2119–2134, doi:10.5194/tc-9-2119-2015, 2015.

Horvat, C. and Roach, L. A.: WIFF1.0: a hybrid machine-learning-based parameterization of wave-induced sea ice floe fracture, Geosci. Model Dev., 15, 803–814, https://doi.org/10.5194/gmd-15-803-2022, 2022.

Horvat, C., Roach, L. A., Tilling, R., Bitz, C. M., Fox-Kemper, B., Guider, C., Hill, K., Ridout, A., and Shepherd, A.: Estimating the sea ice floe size distribution using satellite altimetry: theory, climatology, and model comparison, The Cryosphere, 13, 2869–2885, https://doi.org/10.5194/tc-13-2869-2019, 2019.

Horvat, C., Blanchard-Wrigglesworth, E. and Petty, A.: Observing Waves in Sea Ice With ICESat-2, Geophys. Res. Lett., 47(10), 1–10, doi:10.1029/2020GL087629, 2020.

Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N. and Elliott, S.: CICE : the Los Alamos Sea Ice Model Documentation and Software User's Manual LA-CC-06-012., 2015.

Hutchings, J., Roberts, A., Geiger, C., and Richter-Menge, J.: Spatial and temporal characterization of sea-ice deformation, Ann. Glaciol., 52, 360–368, 2011.

Hwang, B., Wilkinson, J., Maksym, E., Graber, H. C., Schweiger, A., Horvat, C., Perovich, D. K., Arntsen, A. E., Stanton, T. P., Ren, J., and Wadhams, P.: Winter-to summer transition of Arctic sea ice breakup and floe size distribution in the Beaufort Sea, Elem. Sci. Anth., 5, p. 40, https://doi.org/10.1525/elementa.232, 2017.

Ivanova, D. P., Gleckler, P. J., Taylor, K. E., Durack, P. J. and Marvel, K. D.: Moving beyond the total sea ice extent in gauging model biases, J. Clim., 29(24), 8965–8987, doi:10.1175/JCLI-D-16-0026.1, 2016.

Jakobson, E., Vihma, T., Palo, T., Jakobson, L., Keernik, H., and Jaagus, J.: Validation of atmospheric reanalyses over the central arctic ocean, Geophys. Res. Lett., 39, L10802, doi:10.1029/2012gl051591, 2012.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S. K., Hnilo, J. J., Fiorino, M. and Potter, G. L.: NCEP-DOE AMIP-II reanalysis (R-2), Bull. Am. Meteorol. Soc., 83(11), 1631–1644, doi:10.1175/BAMS-83-11-1631(2002)083<1631:NAR>2.3.CO;2, 2002.

Keen, A., Blockley, E., Bailey, D. A., Boldingh Debernard, J., Bushuk, M., Delhaye, S., Docquier, D., Feltham, D., Massonnet, F., O'Farrell, S., Ponsoni, L., Rodriguez, J. M., Schroeder, D., Swart, N., Toyoda, T., Tsujino, H., Vancoppenolle, M., and Wyser, K.: An inter-comparison of the mass budget of the Arctic sea ice in CMIP6 models, The Cryosphere, 15, 951–982, https://doi.org/10.5194/tc-15-951-2021, 2021.

Kekäläinen, P., Aström, J. A. and Timonen, J.: Solution for the fragment-size distribution in a crack-branching model of fragmentation, Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys., 76(2), 1–7, doi:10.1103/PhysRevE.76.026112, 2007.

Kohout, A. L., Williams, M. J. M., Dean, S. M. and Meylan, M. H.: Storm-induced sea-ice breakup and the implications for ice extent, Nature, 509(7502), 604–607, doi:10.1038/nature13262, 2014.

Kohout, A. L., Williams, M. J. M., Toyota, T., Lieser, J. and Hutchings, J.: In situ observations of wave-induced sea ice breakup, Deep. Res. Part II Top. Stud. Oceanogr., 131, 22–27, doi:10.1016/j.dsr2.2015.06.010, 2016.

Kwok, R.: IUTAM Symposium on Scaling Laws in Ice Mechanics and Ice Dynamics, IUTAM Symp. Scaling Laws Ice Mech. Ice Dyn. SE - 26, 94(November 1996), 315–322, doi:10.1007/978-94-015-9735-7, 2001.

Kwok, R. and Untersteiner, N.: New high-resolution images of summer arctic Sea ice, Eos (Washington. DC)., 92(7), 53–54, doi:10.1029/2011EO070002, 2011.

Lecomte, O., Fichefet, T., Flocco, D., Schroeder, D. and Vancoppenolle, M.: Interactions between wind-blown snow redistribution and melt ponds in a coupled ocean-sea ice model, Ocean Model., 87, 67–80, doi:10.1016/j.ocemod.2014.12.003, 2015.

Lipscomb, W. H.: Remapping the thickness distribution in sea ice models, J. Geophys. Res. Ocean., 106, 13 989–14 000, 2001.

Lüpkes, C., Gryanik, V. M., Hartmann, J. and Andreas, E. L.: A parametrization, based on sea ice morphology, of the neutral atmospheric drag coefficients for weather prediction and climate models, J. Geophys. Res. Atmos., doi:10.1029/2012JD017630, 2012.

Massonnet, F., Fichefet, T., Goosse, H., Bitz, C. M., Philippon-Berthier, G., Holland, M. M., and Barriat, P.-Y.: Constraining projections of summer Arctic sea ice, The Cryosphere, 6, 1383–1394, https://doi.org/10.5194/tc-6-1383-2012, 2012.

Meier, W. and Notz, D.: A note on the accuracy and reliability of satellite-derived passive microwave estimates of sea-ice extent, Climate and Cryosphere Sea Ice Working Group Consensus Document, World Climate Research Program, http://www.arcus.org/files/page/documents/1707/GCW_CliC_Sea_ice_Reliability.pdf, 2010.

Perovich, D. K. and Jones, K. F.: The seasonal evolution of sea ice floe size distribution, J. Geophys. Res. Ocean., 119(12), 8767–8777, doi:10.1002/2014JC010136, 2014.

Perovich, D. K., Richter-Menge, J. A. and Tucker, W. B.: Seasonal changes in Arctic sea-ice morphology, Ann. Glaciol., 33, 171–176, doi:10.3189/172756401781818716, 2001.

Petty, A. A., Holland, P. R. and Feltham, D. L.: Sea ice and the ocean mixed layer over the Antarctic shelf seas, Cryosphere, 8(2), 761–783, doi:10.5194/tc-8-761-2014, 2014.

Pringle, D. J., Eicken, H., Trodahl, H. J. and Backstrom, L. G. E.: Thermal conductivity of landfast Antarctic and Arctic sea ice, J. Geophys. Res. Ocean., 112(C4), doi:10.1029/2006JC003641, 2007.

Rampal, P., Dansereau, V., Olason, E., Bouillon, S., Williams, T., Korosov, A., and Samaké, A.: On the multi-fractal scaling properties of sea ice deformation, The Cryosphere, 13, 2457–2474, https://doi.org/10.5194/tc-13-2457-2019, 2019.

Ridley, J. K., Blockley, E. W., Keen, A. B., Rae, J. G. L., West, A. E., and Schroeder, D.: The sea ice model component of HadGEM3-GC3.1, Geosci. Model Dev., 11, 713–723, https://doi.org/10.5194/gmd-11-713-2018, 2018.

Roach, L. A., Horvat, C., Dean, S. M. and Bitz, C. M.: An Emergent Sea Ice Floe Size Distribution in a Global Coupled Ocean-Sea Ice Model, J. Geophys. Res. Ocean., doi:10.1029/2017JC013692, 2018.

Roach, L. A., Bitz, C. M., Horvat, C. and Dean, S. M.: Advances in Modeling Interactions Between Sea Ice and Ocean Surface Waves, J. Adv. Model. Earth Syst., 11(12), 4167–4181, doi:10.1029/2019MS001836, 2019.

Rösel, A., Kaleschke, L. and Birnbaum, G.: Melt ponds on Arctic sea ice determined from MODIS satellite data using an artificial neural network, Cryosphere, doi:10.5194/tc-6-431-2012, 2012.

Rothrock, D. A. and Thorndike, A. S.: Measuring the sea ice floe size distribution, J. Geophys. Res., 89, 6477–6486, https://doi.org/10.1029/JC089iC04p06477, 1984.

Rynders, S.: Impact of surface waves on sea ice and ocean in the polar regions, University of Southampton, United Kingdom., 2017.

Rynders, S., Aksenov, Y., Feltham, D. L., Nurser, A. J. G. and Madec, G.: Impact of granular behaviour of fragmented sea ice on marginal ice zone dynamics, IUTAM Springer, 2022.

Rynders, S., Aksenov, Y., Feltham, D. L., Nurser, A. J. G. and Madec, G.: Impact of granular behaviour of fragmented sea ice on marginal ice zone dynamics, in: IUTAM Symposium on Physics and Mechanics of Sea Ice, edited by: Tuhkuri, J., Polojärvi, A., Springer, Cham, 261-274, https://doi.org/10.1007/978-3-030-80439-8_13, 2022.

Schröder, D., Feltham, D. L., Tsamados, M., Ridout, A., and Tilling, R.: New insight from CryoSat-2 sea ice thickness for sea ice modelling, The Cryosphere, 13, 125–139, https://doi.org/10.5194/tc-13-125-2019, 2019.

Schulson, E. M.: Brittle failure of ice, Eng. Fract. Mech., doi:10.1016/S0013-7944(01)00037-6, 2001.

Schulson, E. M.: Compressive shear faults within arctic sea ice: Fracture on scales large and small, J. Geophys. Res. C Ocean., doi:10.1029/2003JC002108, 2004.

Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok, 2011: Uncertainty in modeled Arctic sea ice volume. J. Geophys. Res., 116, C00D06, https://doi.org/10.1029/2011JC007084.

Shen, H., Hibler, W., and Leppäranta, M.: On applying granular flow theory to a deforming broken ice field, Acta Mech., 143–160, doi:10.1007/BF01182545, 1986.

Smith, M. M., Holland, M., and Light, B.: Arctic sea ice sensitivity to lateral melting representation in a coupled climate model, The Cryosphere, 16, 419–434, https://doi.org/10.5194/tc-16-419-2022, 2022.

Steele, M.: Sea ice melting and floe geometry in a simple ice-ocean model, J. Geophys. Res. Ocean., 97(C11), 17729–17738, doi:10.1029/92JC01755, 1992.

Stern, H. L., Schweiger, A. J., Zhang, J. and Steele, M.: On reconciling disparate studies of the sea-ice floe size distribution, Elem Sci Anth, 6(1), doi:10.1525/elementa.304, 2018a.

Stern, H. L., Schweiger, A. J., Stark, M., Zhang, J., Steele, M. and Hwang, B.: Seasonal evolution of the sea-ice floe size distribution in the Beaufort and Chukchi seas, Elem Sci Anth, 6(1), 48, doi:10.1525/elementa.305, 2018b.

Stroeve, J. C., Kattsov, V., Barrett, A., Serreze, M., Pavlova, T., Holland, M., and Meier, W. N.: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, Geophys. Res. Lett., 39, L16502, https://doi.org/10.1029/2012GL052676, 2012.

Strong, C., Foster, D., Cherkaev, E., Eisenman, I., and Golden, K. M.: On the definition of marginal ice zone width, J. Atmos. Ocean. Tech., 34, 1565–1584, https://doi.org/10.1175/JTECH-D-16-0171.1, 2017.

Tilling, R. L., Ridout, A., and Shepherd, A.: Estimating Arctic sea ice thickness and volume using CryoSat-2 radar altimeter data, Adv. Space Res., 62, 1203–1225, https://doi.org/10.1016/j.asr.2017.10.051, 2018.

Toyota, T., Takatsuji, S., and Nakayama, M.: Characteristics of sea ice floe size distribution in the seasonal ice zone, Geophys. Res. Lett., 33, 2–5, https://doi.org/10.1029/2005GL024556, 2006.

Tsamados, M., Feltham, D. L., Schroeder, D., Flocco, D., Farrell, S. L., Kurtz, N., Laxon, S. W. and Bacon, S.: Impact of Variable Atmospheric and Oceanic Form Drag on Simulations of Arctic Sea Ice*, J. Phys. Oceanogr., 44(5), 1329–1353, doi:10.1175/JPO-D-13-0215.1, 2014.

Tsamados, M., Feltham, D., Petty, A., Schroder, D. and Flocco, D.: Processes controlling surface, bottom and lateral melt of Arctic sea ice in a state of the art sea ice model, Philos. T. Roy. Soc., A, 17, 10302, doi:10.1098/rsta.2014.0167, 2015.

Weiss, J.: Fracture and fragmentation of ice: A fractal analysis of scale invariance, Eng. Fract. Mech., doi:10.1016/S0013-7944(01)00034-0, 2001.

Weiss, J. and Dansereau, V.: Linking scales in sea ice mechanics, Philos. Trans. Roy. Soc. A, 375, 20150352, https://doi.org/10.1098/rsta.2015.0352, 2017.

Weiss, J. and Schulson, E. M.: Coulombic faulting from the grain scale to the geophysical scale: Lessons from ice, J. Phys. D. Appl. Phys., doi:10.1088/0022-3727/42/21/214017, 2009.

Wenta, M. and Herman, A.: Area-averaged surface moisture flux over fragmented Sea Ice: Floe size distribution effects and the associated convection structure within the atmospheric boundary layer, Atmosphere (Basel)., 10(11), 1–19, doi:10.3390/atmos10110654, 2019.

Wilchinsky, A. V. and Feltham, D. L.: Modelling the rheology of sea ice as a collection of diamond-shaped floes, J. Nonnewton. Fluid Mech., 138(1), 22–32, doi:10.1016/j.jnnfm.2006.05.001, 2006.

Wilchinsky, A. V., Feltham, D. L. and Hopkins, M. A.: Effect of shear rupture on aggregate scale formation in sea ice, J. Geophys. Res. Ocean., 115(C10), doi:10.1029/2009JC006043, 2010.

Williams, T. D., Bennetts, L. G., Squire, V. A., Dumont, D. and Bertino, L.: Wave-ice interactions in the marginal ice zone. Part 1: Theoretical foundations, Ocean Model., 71, 81–91, doi:10.1016/j.ocemod.2013.05.010, 2013a.

Williams, T. D., Bennetts, L. G., Squire, V. A., Dumont, D. and Bertino, L.: Wave-ice interactions in the marginal ice zone. Part 2: Numerical implementation and sensitivity studies along 1D transects of the ocean surface, Ocean Model., 71, 92–101, doi:10.1016/j.ocemod.2013.05.011, 2013b.

WMO: WMO Sea-Ice Nomenclature, Tech. Rep. 259, The Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM), available at: https://library.wmo.int/doc_num.php?explnum_id=4651 (last access: 14 July 2021), 2014.

Zhang, J. L. and Rothrock, D. A.: Modelling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates, Mon. Weather Rev., 131, 845–861, https://doi.org/10.1175/1520-0493(2003)131<0845:MGSIWA>2.0.CO;2, 2003.

Zhang, J., Stern, H., Hwang, B., Schweiger, A., Steele, M., Stark, M. and Graber, H. C.: Modeling the seasonal evolution of the Arctic sea ice floe size distribution, Elem. Sci. Anthr., 4, 000126, doi:10.12952/journal.elementa.000126, 2016.

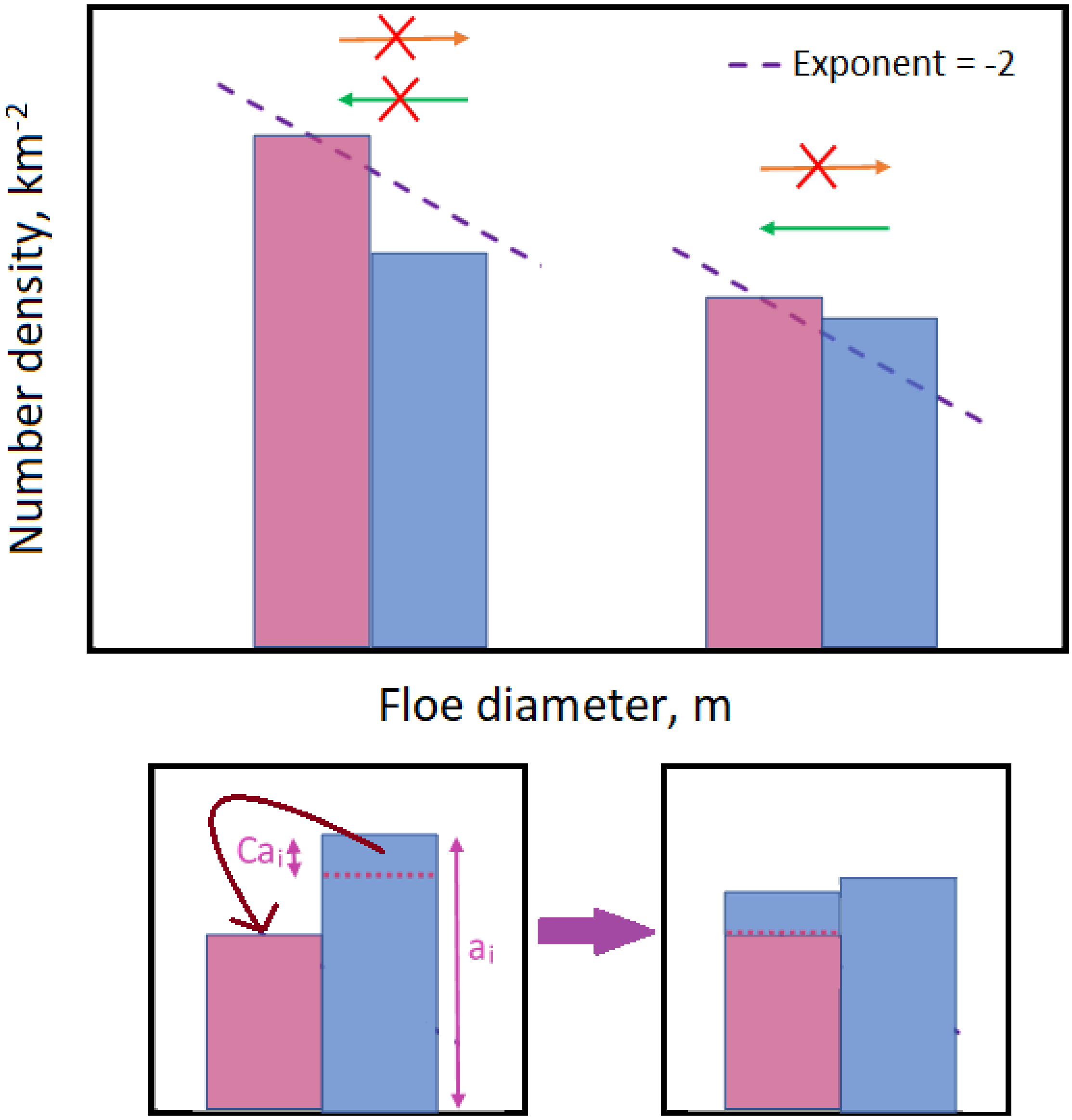

Figure 1: Diagram of the quasi-restoring brittle fracture scheme introduced to the prognostic FSD model. The upper section of the plot shows two adjacent pairs of floe size categories, where blue highlights the larger category and red the smaller category. The brittle fracture scheme only transfers sea ice area fraction from a larger category to the adjacent smaller category and only where the number density gradient between adjacent categories in log-log space is larger (more positive) than – 2. A slope with this gradient is shown by the purple, dashed line. The lower section of the plot shows how the scheme can modify the FSD. The sea ice area fraction transferred from the larger to the smaller floe size category is Ca_i where a_i is the total sea ice area fraction in the larger category and C is the restoring constant.

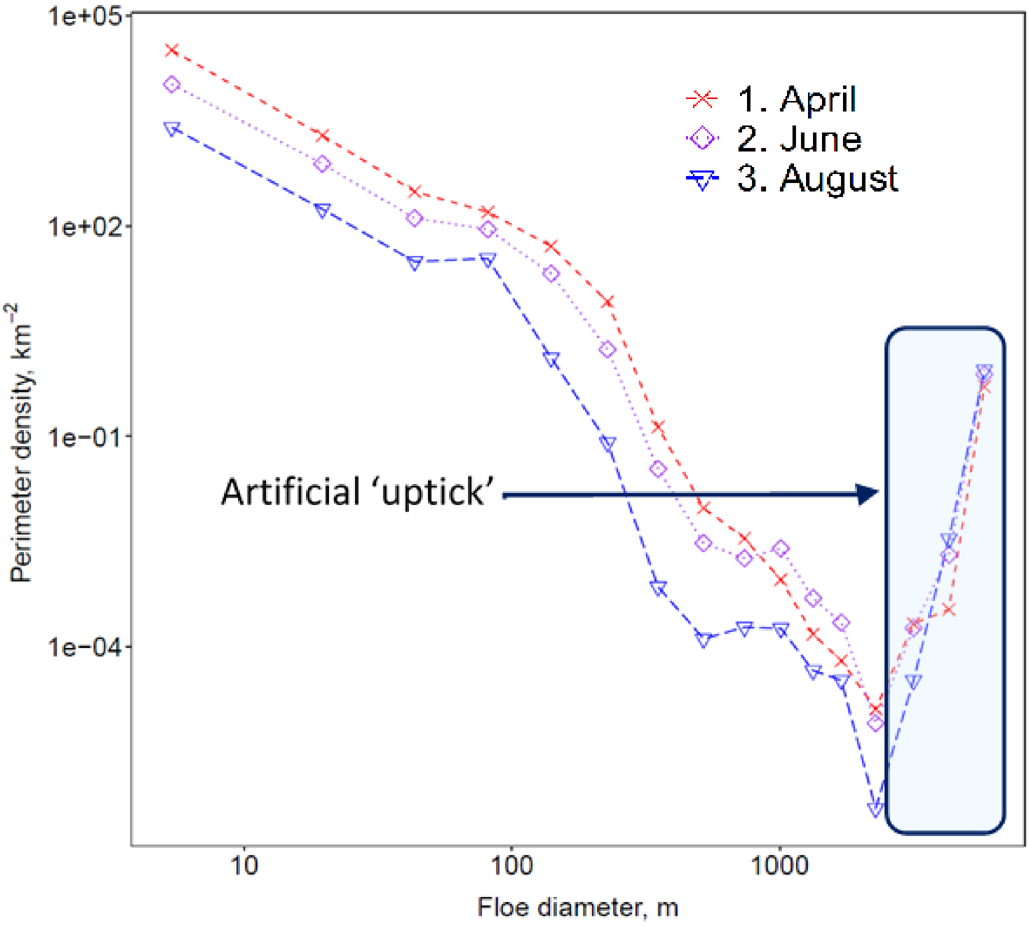

Figure 2: An example of prognostic model output from the *prog-16-nobf* simulation averaged over the MIZ (i.e regions with between 15 – 80% sea ice concentration). Presented in the figure is the perimeter density distribution, km^{-2} for the April (red, cross, dashed), June (purple, diamond, dotted), and August (blue, triangle, long-dash) averaged over 2000 – 2016. Also highlighted in the figure by a blue transparent box is an artificial 'uptick', a non-physical feature of the model also reported by Roach et al. (2018a). Also highlighted in the figure by a blue transparent box is an artificial 'uptick', a feature of the model also reported by Roach et al. (2018) that results from prognostic model design and structure (e.g. missing fragmentation processes, upper limit on floe size) and does not represent a physical behaviour seen for the FSD.

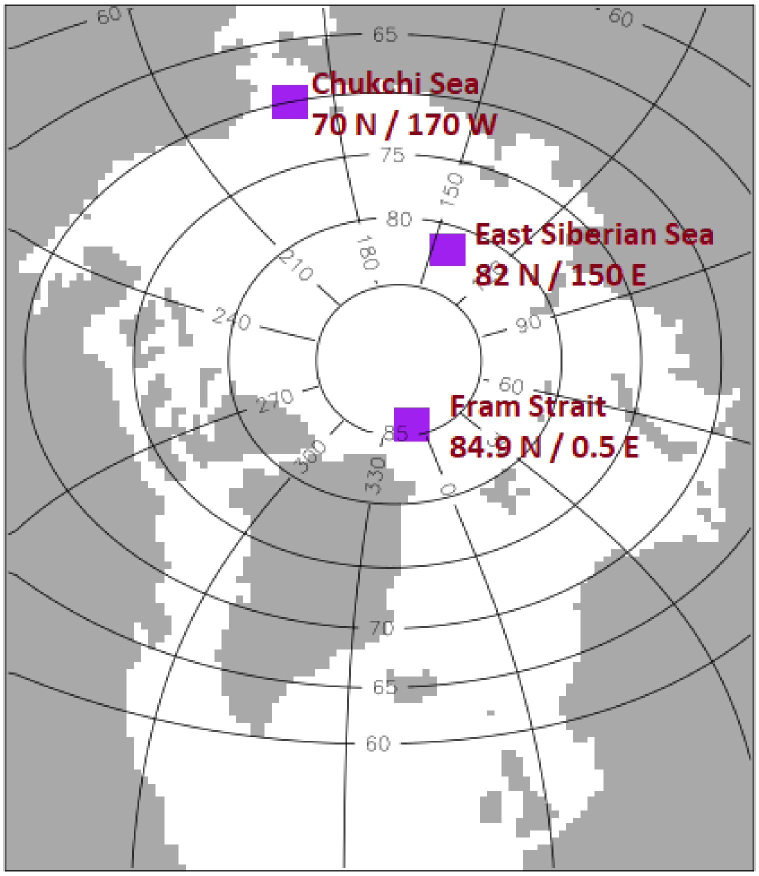

Figure 3: Boxes indicate the areas over which the prognostic model emergent FSD is averaged to represent the three locations included in the observational study. Each case study area spans a set of 5 x 5 grid cells that includes the site stated for collection of observations.

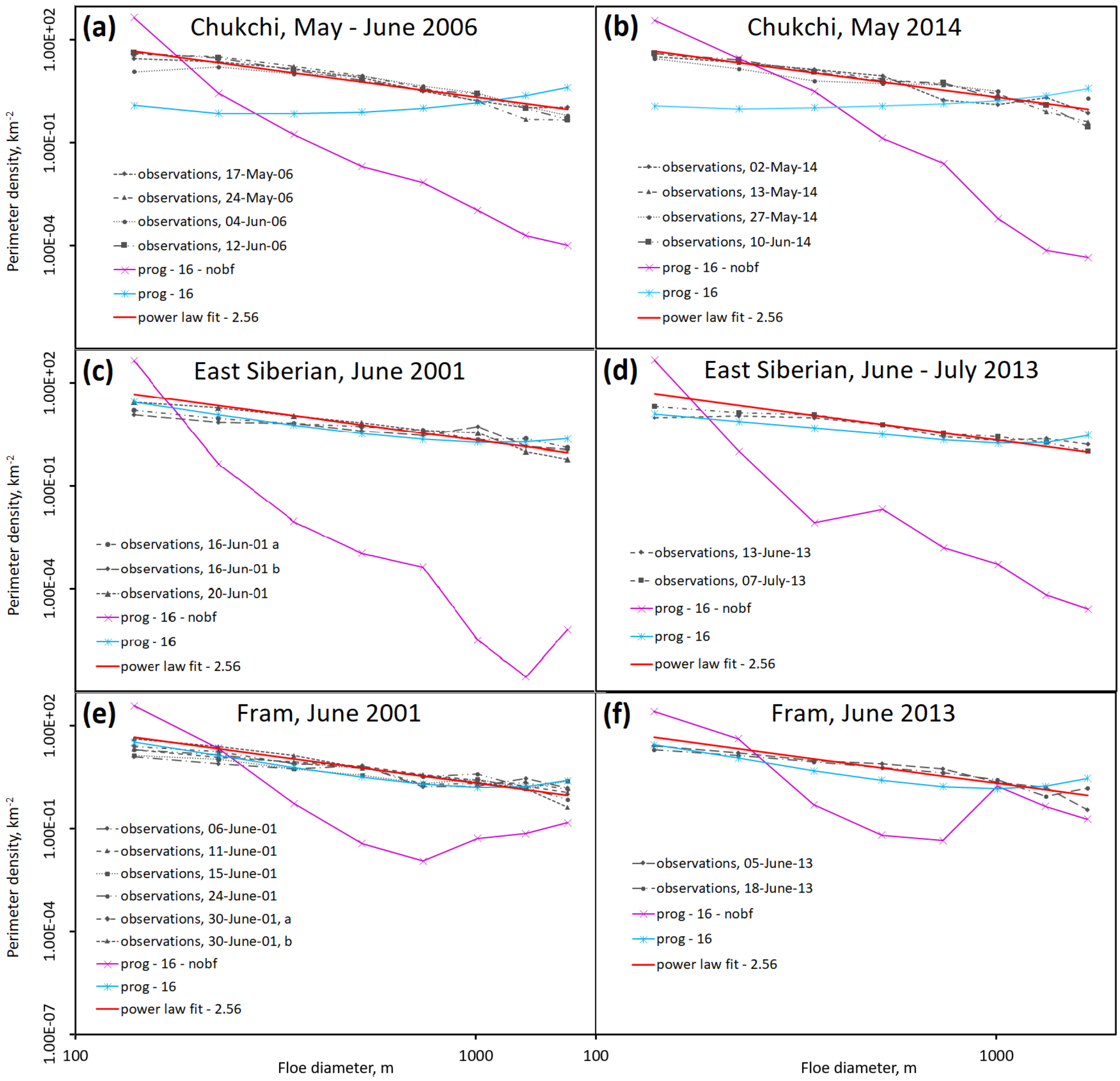

Figure 4: A comparison of the observations and prognostic model output for the perimeter density distributions for the Chukchi Sea in May – June 2006 (a) and May 2014 (b), the East Siberian Sea in June 2001 (c) and June – July 2013 (d), and the Fram Strait in June 2001 (e) and June 2013 (f). Observations are identified with pink or purple dashed lines. Output for *prog-16* (light blue, solid, stars) and *prog-16-nobf* (dark blue, solid, crossed) is averaged across the relevant region identified in Fig. 3 over the stated month(s). The average power-law fit across all locations is also shown (red, solid). The floe size data used within this figure was obtained from imagery using the methodology of Hwang et al. (2017) and the exponent of the power-law fit was calculated according to Virkar and Clauset (2014).

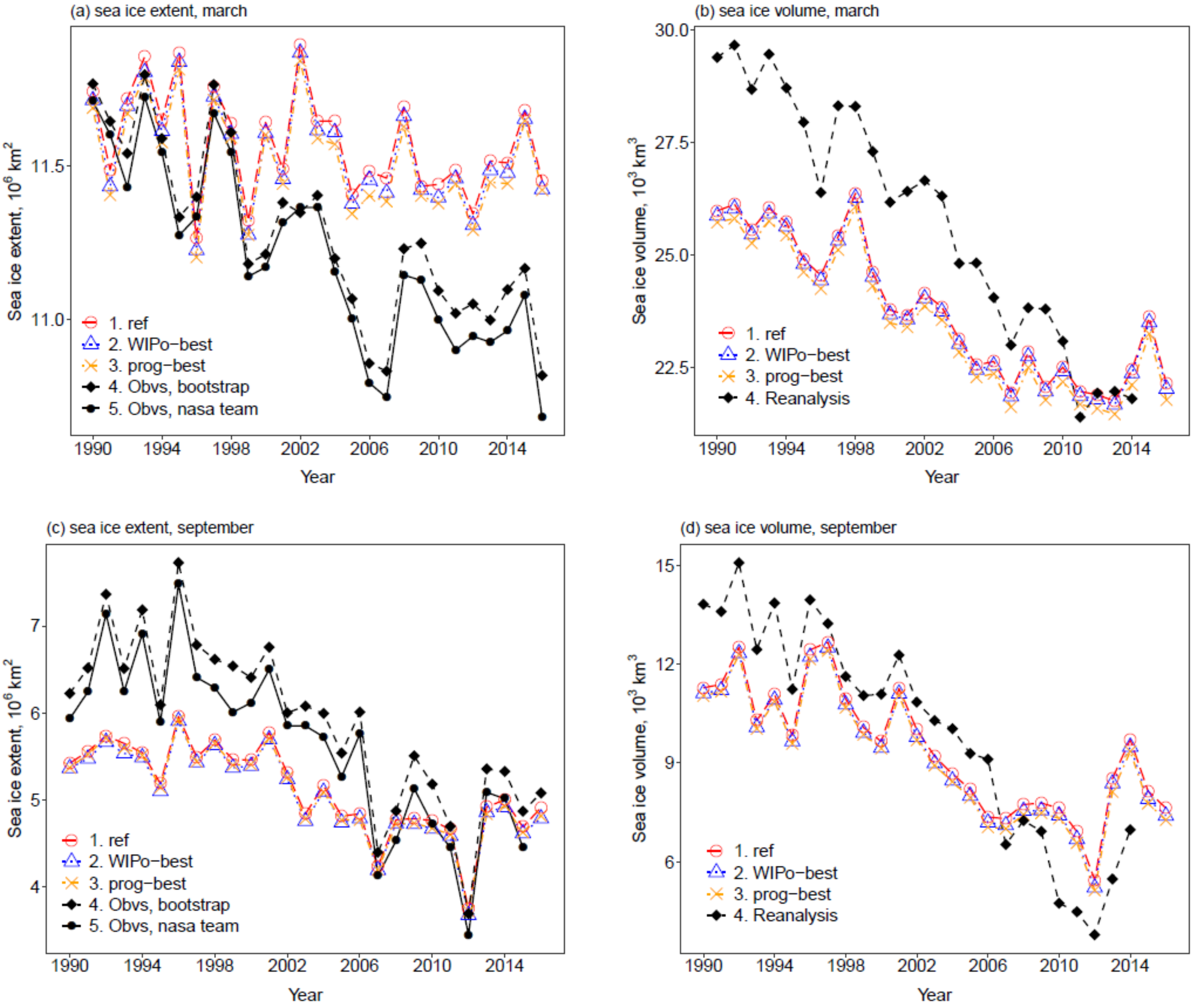

Figure 5: The total Arctic sea ice March extent (a, top left), March volume (b, top right), September extent (c, bottom left), and September volume (d, bottom right) within the model domain over the period 1990 – 2016 for *ref* (red, circles), *WIPo-best* (blue, triangles), *prog-best* (yellow, cross) and observations / reanalysis (black). Sea ice concentration data is obtained from satellites using the Bootstrap (filled diamond, dashed) algorithm version 3 (Comiso, 1999) and the NASA Team (filled circle, solid) algorithm version 1 (Cavalieri et al., 1996). Sea ice volume data (filled diamond, dashed) is taken from PIOMAS (Zhang and Rothrock, 2003).

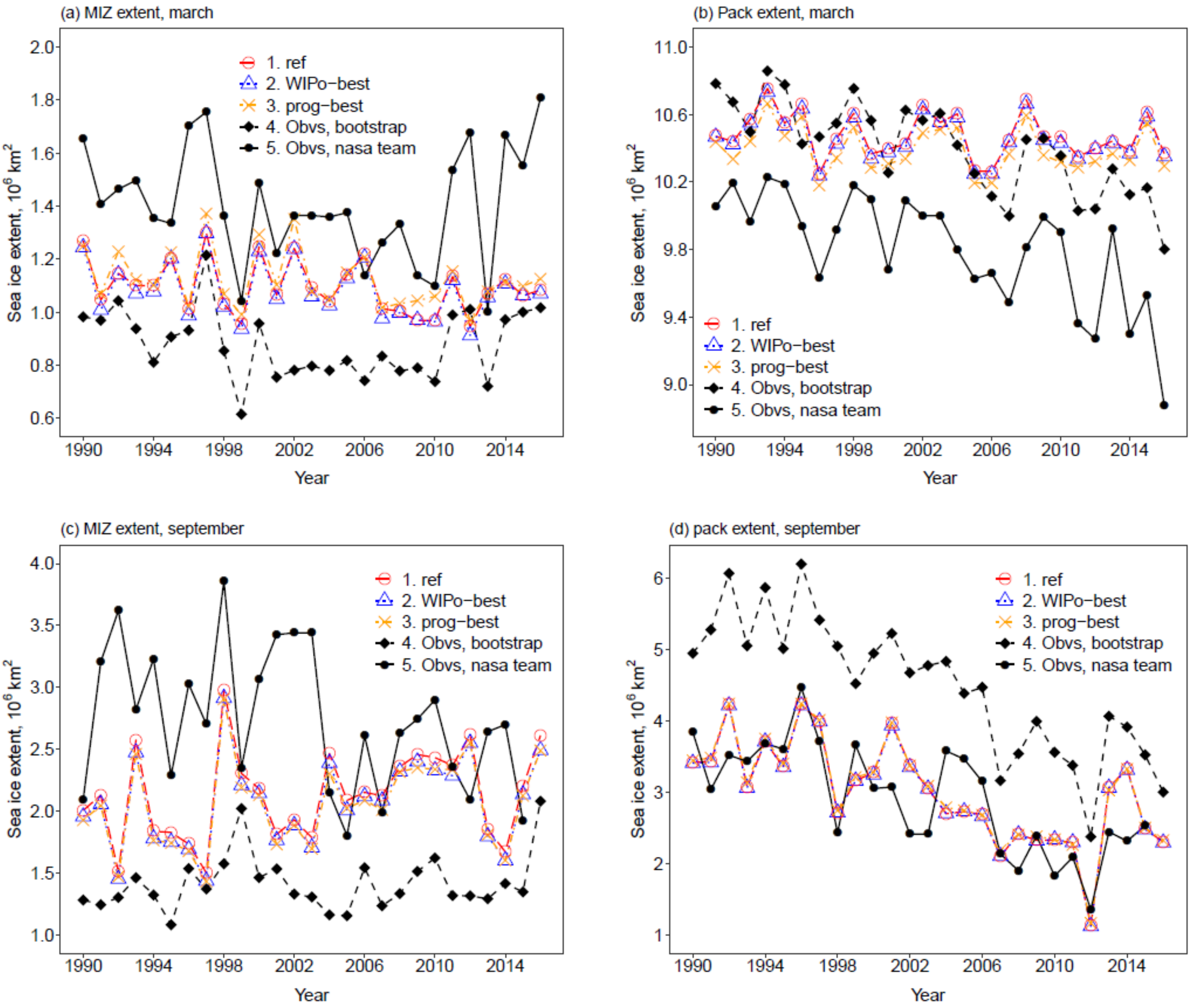

Figure 6: The total Arctic sea ice March MIZ extent (a, top left), March pack extent (b, top right), September MIZ extent (c, bottom left), and September pack extent (d, bottom right) over the period 1990 – 2016 for *ref* (red, circles, long dash), *WIPo-best* (blue, triangles, dotted), *prog-best* (yellow, cross, dot-dash) and observations (black). Sea ice concentration data is obtained from satellites using the Bootstrap (filled diamond, dashed) algorithm version 3 (Comiso, 1999) and the NASA Team (filled circle, solid) algorithm version 1 (Cavalieri et al., 1996). Sea ice volume data (filled diamond, dashed) is taken from PIOMAS (Zhang and Rothrock, 2003). The MIZ is here defined as the region with between 15% and 80% sea ice concentration.

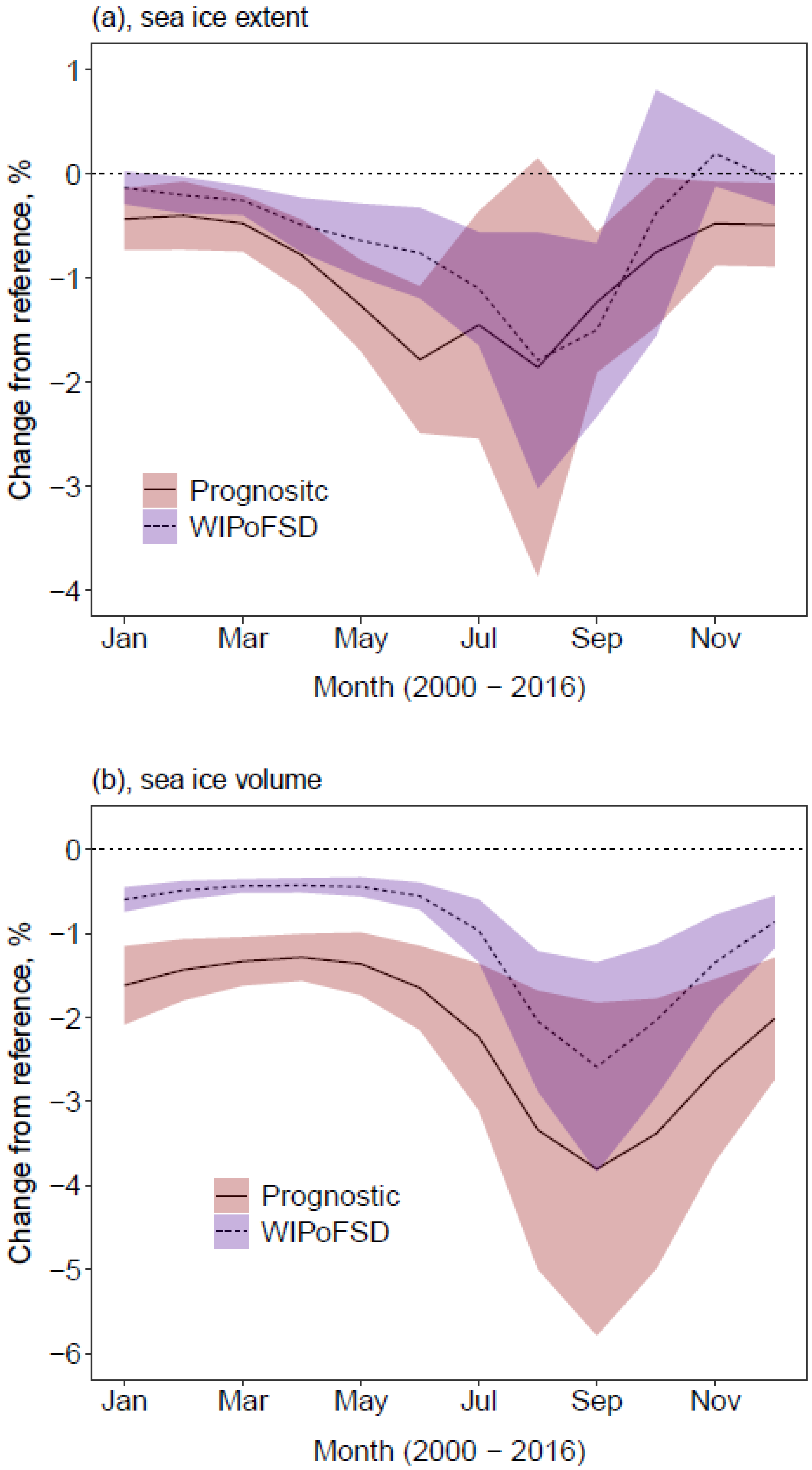

Figure 7: Difference in sea ice extent (top) and volume (bottom) of *prog-best* (solid, red ribbon) and *WIPo-best* (dashed, blue ribbon) relative to *ref* averaged over 2000 - 2016. The ribbon shows, in each case, the region spanned by the mean value plus or minus two times the standard deviation.

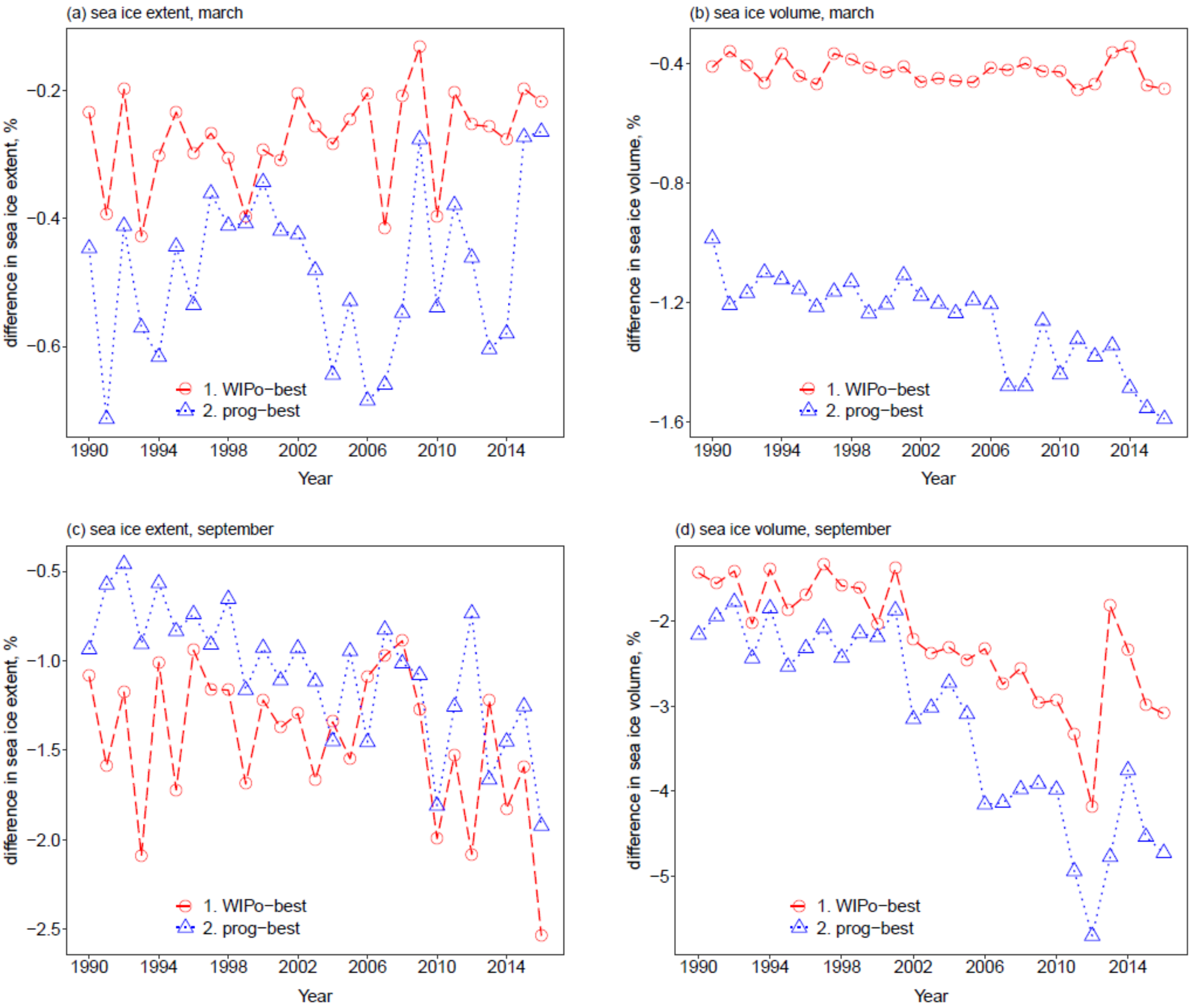

Figure 8: The % difference in the Arctic total sea ice March extent (a, top left), March volume (b, top right), September extent (c, bottom left), and September volume (d, bottom right) over the period 1990 – 2016 for *WIPo-best* (red, circles, dashed), and *prog-best* (blue, triangles, dotted) relative to *ref*.

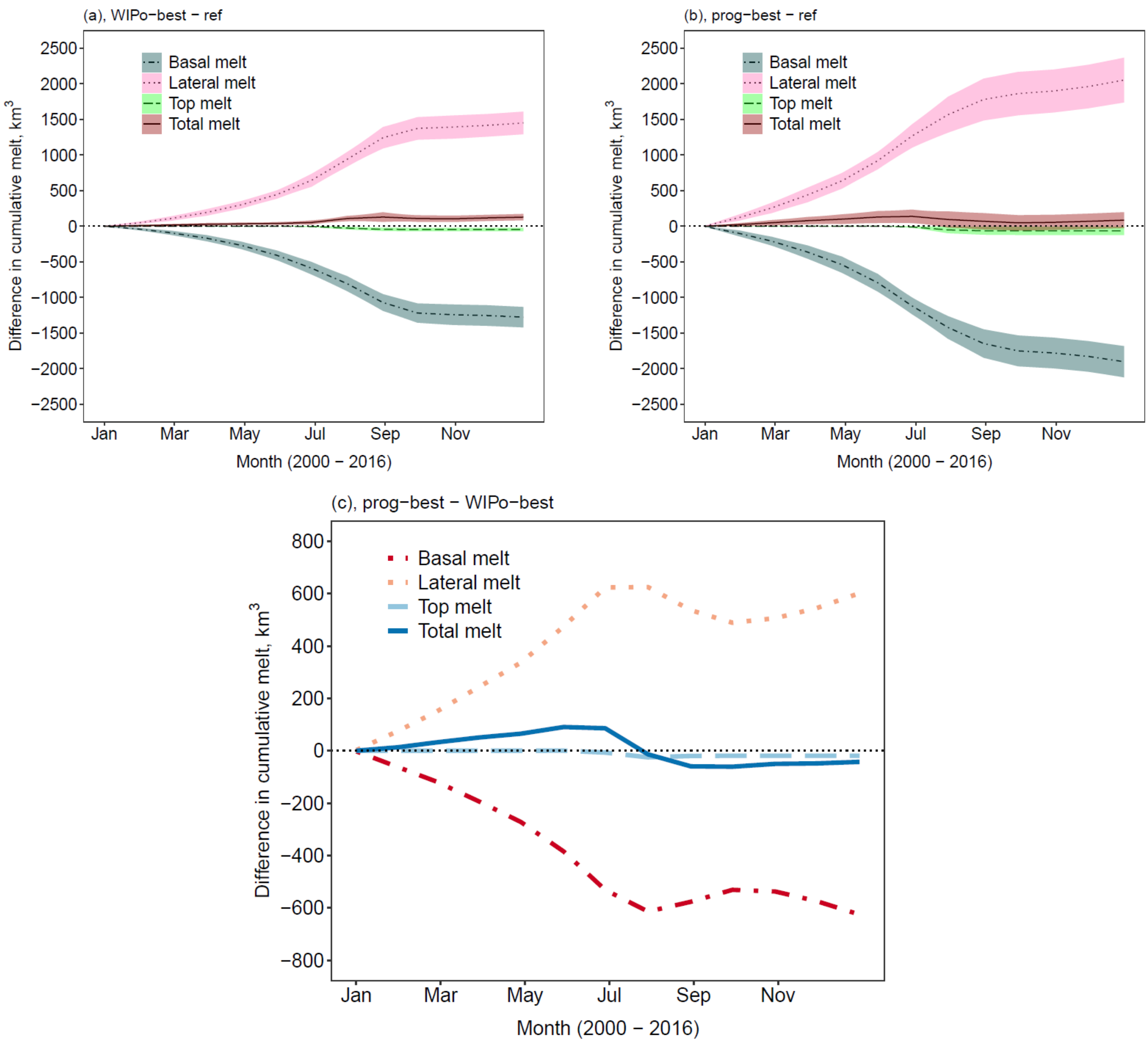

Figure 9: The top two plots show the difference in the cumulative lateral (pink ribbon, dotted), basal (grey ribbon, dot-dashed), top (green ribbon, dashed), and total (red ribbon, solid) melt averaged over 2000–2016 for *prog-best* (a, left) and *WIPo-best* (b, right) relative to *ref*. The ribbon shows, in each case, the region spanned by the mean value plus or minus twice the standard deviation. The bottom plot shows show the difference in the cumulative lateral (orange, dotted), basal (red, dot-dashed), top (light blue, dashed), and total (dark blue, solid) melt averaged over 2000–2016 for *prog-best* relative to *WIPo-best* (c).

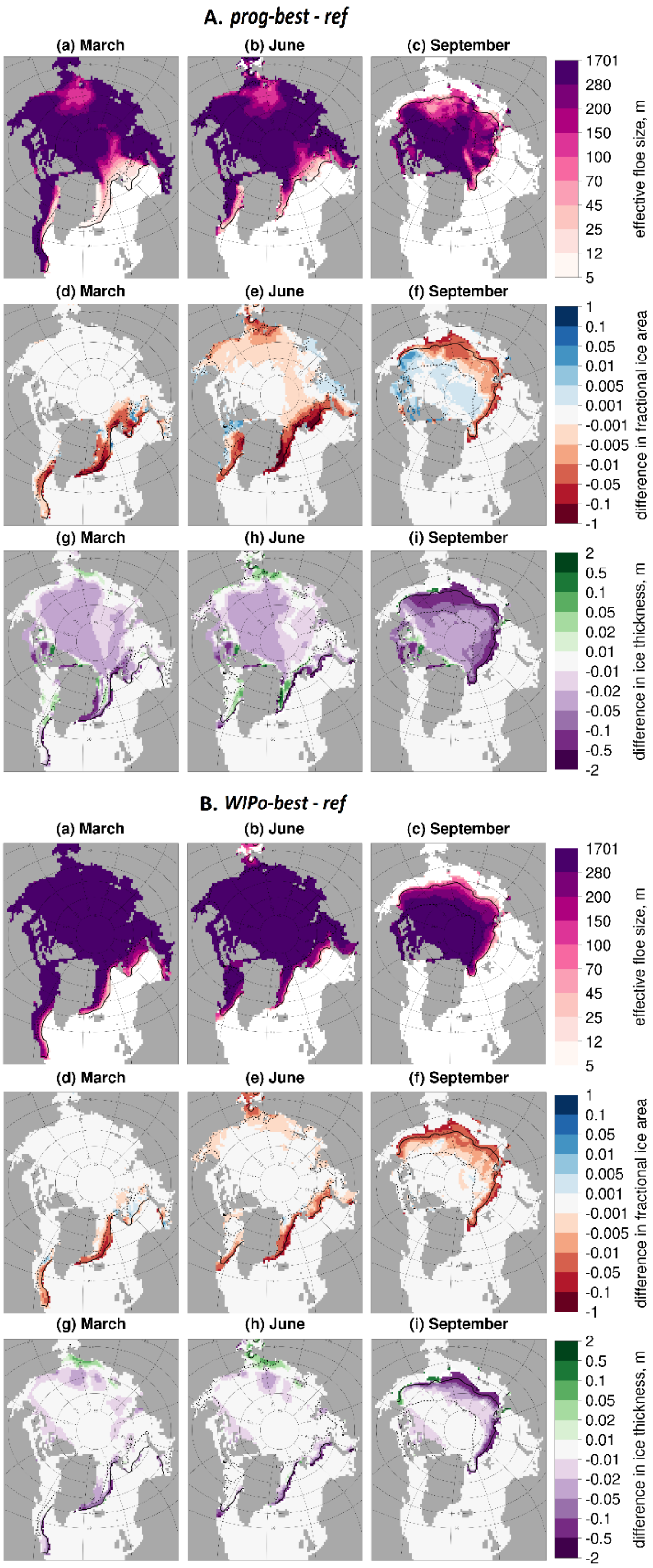

Figure 10: Absolute values of l_{eff} (a–c) for *prog-best* (A, top) or *WIPo-best* (B, bottom) and the difference in sea ice area fraction (d–f) and difference in ice thickness (g–i) between *prog-best* (A, top) or *WIPo-best* (B, bottom) and *ref* averaged over 2000–2016. Results are presented for March (a, d, g), June (b, e, h), and September (c, f, i). Values are shown only in locations where the sea ice concentration exceeds 5 %. The inner (dashed black) and outer (solid black) extent of the MIZ averaged over the same period is also shown.

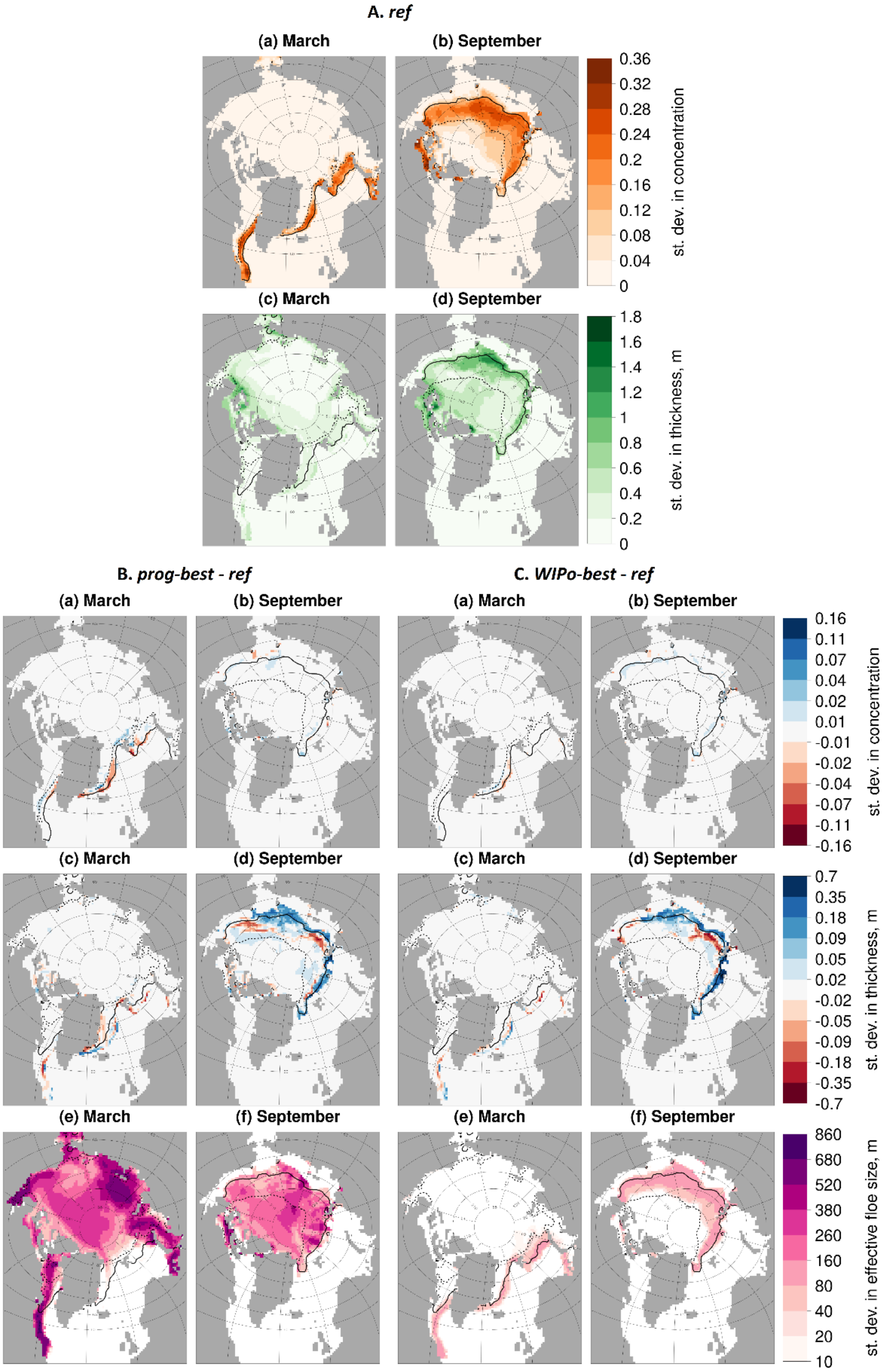

Figure 11: Section A (top) shows the standard deviation of the sea ice area fraction (a, b) and thickness (c, d) in March (a, c) and September (b, d) for *ref*. Section B (bottom left) and C (bottom right) show difference plots in the standard deviation of the sea ice area fraction (a, b) and thickness (c, d) in March (a, c) and September (b, d) for *prog-best* and *WIPo-best* relative to *ref* respectively. In B and C, the standard deviation in l_{eff} is also plotted for both March (e) and September (f). Values are shown only in locations where the sea ice concentration exceeds 5 %. The inner (dashed black) and outer (solid black) extent of the MIZ averaged over the same period is also shown. Plots show that changes to the standard deviation in the sea ice area fraction and thickness are generally localised to the outer edge of the MIZ.

Term	Description
Floe size	The mean diameter of a sea ice floe.
Perimeter density	Floe perimeter per unit sea ice area. Calculated as the total perimeter of an ensemble of floes divided by the total sea ice area.
Effective floe size	The diameter of the set of identical floes that has the same perimeter density as an ensemble of floes of variable size.
Floe size distribution / FSD	General term to refer to the size distribution of an ensemble of sea ice floes. The FSD can be considered in terms of the probability a floe will have a given size (probability distribution), the number of floes with a given size (number distribution), the perimeter of floes with a given size (perimeter distribution), or the area of floes with a given size (area distribution).
Density distribution	When expressing the FSD in terms of a 'real' descriptor i.e. as a number, perimeter, or area distribution, the FSD can be expressed per unit sea ice area i.e. as a density distribution.
Non-cumulative distribution	The value reported for each floe size (category) in a non-cumulative FSD (e.g. number of floes) includes floes of that size only. Some studies alternatively consider a cumulative FSD where the value reported per floe size (category) refers to all floes larger than the given size (category), in addition to those of that floe size (category).
Fragment size distribution	A generic term to describe the size distribution of any system consisting of an ensemble of individual components with a varying size metric. The FSD is a specific example of a fragment size distribution.
Ice-thickness distribution	Describes the proportion of the total sea ice area within discrete floe thickness categories. This study uses the standard CICE formulation for the ice-thickness distribution, as described in Hunke et al. (2015).
Prognostic FS(T)D model	A modified version of the prognostic FSTD (Floe-Size-Thickness distribution model) of Roach et al. (2018, 2019). Details of the version used here are provided in section 2.2.
WIPoFSD model	A modified version of Waves-in-Ice module and Power law Floe Size Distribution model of Bateson et al. (2020). Details of the version used here are provided in section 2.3.
(Truncated) Power-law fit	A power law of the form x^{α} (where x refers to the floe diameter) between lower and upper floe size limits that has been fitted to observations of floe size.
Power-law exponent	The value of α for a power law of form x^{α}. Can also be described as the slope of a power law when plotted using logarithmic axes.

Table 1: A summary of important terms related to the sea ice floe size distribution (FSD) that are used within this study. Note that these terms are defined in the context of this study only.

Simulation	Description of Simulation	Technical details
ref	Reference simulation with fixed floe size and no FSD model.	$l_{eff} = 300\ m$
prog-16	Simulation uses standard prognostic FSD model setup but using 16 floe size categories.	16 floe size categories following Gaussian spacing; includes brittle fracture scheme described in section 2.2.2.
prog-16-nobf	Simulation uses standard prognostic FSD model setup but without brittle fracture and using 16 floe size categories.	16 floe size categories following Gaussian spacing. Brittle fracture scheme, described in section 2.2.2, excluded from model.
prog-best	Simulation uses standard prognostic FSD model setup.	12 floe size categories following Gaussian spacing; includes brittle fracture scheme described in section 2.2.2.
WIPo-best	Simulation uses WIPoFSD model, with WIPoFSD model parameters optimised against observations.	$d_{min} = 5.375\ m, d_{max} = 30{,}000\ m,$ $\alpha = -2.56$

Table 2: A summary of the CPOM CICE simulations described in section 3.1. All simulations are initiated sea ice free on the 1st January 1980 and evaluated until 31st December 2016.