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 power-law. 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.

Specific comments (major):
PXLY → Page X Line Y

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.
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).

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).
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.

c) Why is F_frzmlt capped, and why is it important?
P5L12: Please clarify the definition of l_eff. What do you call *"the perimeter of a FSD"*?

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).

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)?

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.

This condition means… It could be worth giving the physical interpretation of this sentence, as not all readers are familiar with FSDs.

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.

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.

However, the brittle fracture-derived mechanisms operate over different timescales and scale with different properties. → This is a very vague sentence.

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?
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?

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.

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.

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 ?

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.

P11L19 *Bateson et al. (2020) demonstrated...*
They did, but so did Roach et al. (2018) and others before (It is reported in Tsamados et al., 2015, which already involve some of the co-authors of this manuscript).

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.

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.
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?

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).

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.

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.

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.

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.

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.

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.

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.

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).

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.

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?).

Abstract:
P1L13→17: The beginning of the abstract would gain from being a bit more synthetic/sharper (until […] "*In this study*".

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.

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.

P2L2: "*here*" is a bit unexpected given the introduction has just started.

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).

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.

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.

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.

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.

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 l_{var} is particularly suitable*"
I find these sentences a bit vague. The link with the rest of the paragraph could be more explicit.

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 l_{eff} rather than l_{var}, since it is possible to calculate exactly how l_{eff} 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.

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

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

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

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.

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

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

P10L25: I am a bit confused by this "*However*,".

P14L5 →P14L13 Therefore… This is interesting, it would gain from being a bit clearer.

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.

Caption of Figure 4: *month(s)*
Why is the "s" between brackets?
"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.

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.

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.

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.

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.