We sincerely thank the editor and the two anonymous reviewers for the helpful comments and suggestions. We addressed the raised comments in the manuscript, and include our responses to your questions in blue below. These revisions have enhanced the clarity of our paper and we thank the Editor and reviewers for bringing them to our attention.

Editor questions and comments:

Reviewer 1 emphasized the very narrow range of values for fitting a power law. I share these concerns and point to other methods to determine the fit, which represents further uncertainty (Muchow et al., 2021).

Muchow, M., Schmitt, A. U., and Kaleschke, L.: A lead-width distribution for Antarctic sea ice: a case study for the Weddell Sea with high-resolution Sentinel-2 images, The Cryosphere, 15, 4527–4537, https://doi.org/10.5194/tc-15-4527-2021, 2021.

Muchow et al. (2021) use two methods to determine the fit, the linear fit and the maximum likelihood fit as in Clauset et al., (2009).

For the linear fit, they reference Berk (2004), and state "that atypical values have a strong effect on the result." For this reason, we do not use a linear regression and stick to the maximum likelihood approach from Clauset et al. (2009). This method is widely used in many FSD studies as well as in the Muchow et al., (2021) paper. The powerlaw python package (Alstott et al., 2014) uses the statistics described by Clauset et al. (2009) and package the steps of fitting a power law in a reproducible, open methodology. This package has been used in many sea ice and iceberg size distribution studies for example (there are more beyond these):

Shiggins, C. J., Lea, J. M., & Brough, S. (2023). Automated ArcticDEM iceberg detection tool: insights into area and volume distributions, and their potential application to satellite imagery and modelling of glacier–iceberg–ocean systems. *The Cryosphere*, *17*(1), 15-32.

Walter, A., Lüthi, M. P., & Vieli, A. (2020). Calving event size measurements and statistics of Eqip Sermia, Greenland, from terrestrial radar interferometry. *The Cryosphere*, *14*(3), 1051-1066.

Köhler, A., Pętlicki, M., Lefeuvre, P. M., Buscaino, G., Nuth, C., & Weidle, C. (2019). Contribution of calving to frontal ablation quantified from seismic and hydroacoustic observations calibrated with lidar volume measurements. *The Cryosphere*, *13*(11), 3117-3137.

Denton, A. A., & Timmermans, M. L. (2022). Characterizing the sea-ice floe size distribution in the Canada Basin from high-resolution optical satellite imagery. *The Cryosphere*, *16*(5), 1563-1578.

In our work we follow the full methodology of the powerlaw package which includes testing to determine if the data do indeed fit a power law distribution with the log-likelihood ratio of various distribution's fit to the data. We test: power law, exponential, and lognormal distributions.

I find the description of the illustration of the segmentation results not sufficient. Especially the fact that large ice floes are sometimes not separated by the segmentation and form larger objects of the same color. What does this mean for the resulting FSD?

We have added discussion of algorithm limitations to a new Section 3.7 Limitations.

Where does the number of 9.4 million floes come from? Are any of them counted multiple times?

The 9.4 million floes are not unique identifications. This number reports all identified floes, treating each image as an independent observation. We added the clause:
“ \ref{fig:loc}b; note that these are not 9.4 unique floes as each image is taken as an independent observation.”

How is the ice floe tracker IFT used in this study? It was mentioned in the introduction and I expected that it is used for some purpose, e.g. to avoid double counting. Please clarify.

This work was motivated by expanding on the recently developed IFT, which provides unique sea ice floe observations from the MODIS dataset. We agree this is not clearly explained and not essential to include as currently worded. We have reworded the introduction to describe the IFT as a previous study and in the conclusions we suggest incorporating the new segmentation algorithm into the tracking of the IFT pipeline.

What information is stored in the “floe library”? Please describe this library in more detail. Is this a data set? See https://www.the-cryosphere.net/policies/data_policy.html

The floe library dataset has been uploaded and documented on Zenodo- we added this to the data availability section.

Section 3.5: 97% of the floes fall in this size range. Where does this number come from? Does this percentage mean by area coverage? Or by the number of floes? Is this the same for MODIS and S2?

By number of floes. We added a clarification here. This is referring to the MODIS floes. We provide statistics for the Sentinel-2 floes when Sentinel-2 is introduced in the next section.

Section 5: the imagery largest record of Earth to date. Really? What about Landsat? Maybe a question of definition?

Landsat has acquired imagery since 1972, but primarily of Earth’s land surface. The coverage allows for an 8-day repeat cycle of each Landsat scene. MODIS, on the other hand, is the longest continuous daily global satellite observation record ever compiled. We added clarification in the text.

Most references are incomplete, the DOIs are missing. This would result in a long list of complaints for copy editing.

Thank you for pointing this out, we have added the DOIs for all references.

Figure 3cfi: Better plot data points (with error bars) instead connected lines

We have now plotted points for each floe area bin. We have not plotted error bars as it is not obvious what statistic would be optimal here for the binned data. One approach may be to plot the bootstrapped samples, but we hope for the sake of clarity that the uncertainty on the power law fit alpha will suffice.

You can upload your code+data already now on Zenodo with restricted access.

Added code in the data availability section.

Reviewer #1.

The authors have addressed all of my comments. I have a few further comments about the revised manuscript.

Line 87. "on average 58% of an image is covered in opaque clouds..." -- The first draft of this paper said 71%. Why did it change to 58%?

After submission, we noticed the original calculation included the land mask, giving the fraction of the area masked by land or cloud, not just cloud. The 58% is the percentage of the image covered in clouds.

Line 142. The floe areas considered in this work range from 5 km^2 to 300 km^2. That's only a factor of 60, and the range of floe diameters is only sqrt(60) < 8. This is a VERY narrow range of values for fitting a power law -- less than one order of magnitude -- and is necessarily accompanied by a high degree of uncertainty as to whether a power law is indeed the best-fitting function. The authors should acknowledge that the range of floe sizes is very narrow for fitting a power law.

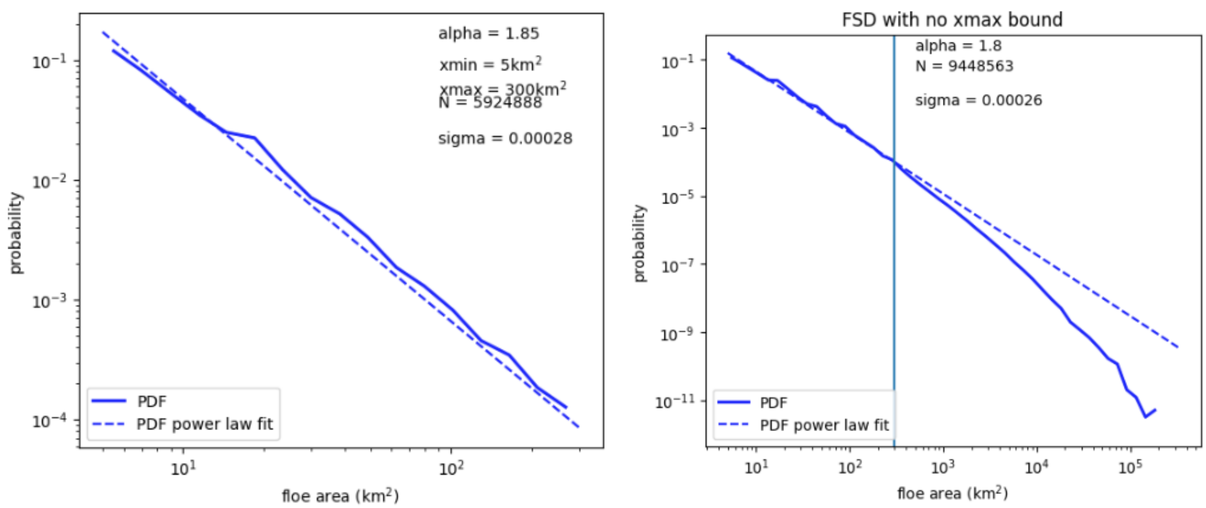
We refer the reviewer to our response to the editor on this point (above). We follow the full methodology of the powerlaw package which includes testing to determine if the data fit a power law distribution with the log-likelihood ratio of various distribution's fit to the data. We test: power law, exponential, and lognormal distributions and determine that the power law distribution is indeed the best fit.

We are limited by the minimum and maximum floe sizes we can accurately detect. The smallest floe size is determined by sensor resolution. To be consistent in fitting the power law throughout the season, we use 300 km^2 as the maximum floe size. The maximum floe size is somewhat constrained by the image size. The larger a floe is, the more likely it is to intersect with the image border, and our algorithm eliminates floes that intersect with the border because the floe area statistics are then inaccurate. Thus, the number of large floes is underestimated.

We provide the low values of the sigma values associated with the power law fit. We have made clear in the paper the range of floe sizes to which this power law fit is applied. A number of other studies include analysis of a small range of floe sizes with the range of floe max:min floe diameters e.g: (Holt and Martin (2001) ratio 11.1, Inuoe et al., (2004) ratio 7.6, Hwang et al., (2017) ratio 6.7).

Data limitations are such that our power law fits are necessarily over a narrow range of values, 5 to 300 km^2. It may be that the fits presented here do not fully represent the sea ice in areas of high ice concentration, but we show a robust seasonal trend in floe size distribution in the floes size 5 to 300 km^2. We have added a discussion of the limitations of the floe size range analyzed in Section 3.7 Data and Algorithm Limitations.

It may be useful to see how the fit differs for different ranges of floe sizes. In the figure below we show the FSD fit to floes with area $5 <= x <= 300$ km^2 (left panel) and floes $x>=5$ km^2 - i.e. all floes because floes with $x<5$ km^2 are not identified (right panel). The maximum floe size is 321, 813 km^2. 97% of identified floes are in the fitted range ($5 <= x <= 300$ km^2). Although the alpha values are similar, in the right panel we see the true FSD (solid line) deviates from the power law fit (dotted) around floes of size 300 km^2 (vertical line).



Equation 3 and following -- is alpha required to be greater than 1? It seems that equation 3 only requires alpha not equal to 1, but equation 4 does apparently require alpha greater than 1, otherwise the uncertainty would be negative.

We added a note that alpha must be greater than 1.

Line 173. "The difference in alpha values ranges from 0.03 to 0.21" -- When I look at Figure 3(c,f,i) I see differences of 0.06 (3c), 0.25 (3f), and 0.17 (3i). Where does "0.03 to 0.21" come from?

Corrected to 0.06 to 0.25. Thank you for noticing that.

Line 191. "bootstrap with replacement" should be "sampling with replacement"

Corrected

Lines 183-194. About the bootstrapping procedure, my understanding is as follows. Start with a randomly selected image that contains N floes. From that collection of floes, select N floes WITH replacement. If the sampling had been WITHOUT replacement, then there would be no difference between the original set of floes and the bootstrapped set. The only randomization in this bootstrapping procedure is the "with replacement" sampling scheme. That hardly seems like a proper bootstrap to me.

Our sampling has been done with replacement as described. This is a powerful approach in data analysis that gives us more confidence in the statistics for a number of reasons. Our large number of bootstrap samples (1000 for each image) gives us additional confidence in alpha.

Furthermore, if the intention is to assess the uncertainty in the exponent alpha, that should come from the fitting procedure itself (most packages include an uncertainty estimate for the estimated parameters) or from a goodness-of-fit test. Why is bootstrapping even needed?

We include the bootstrapping as an additional method of estimating uncertainty. We have also included the uncertainty in the exponent alpha that is provided by the powerlaw package.

Lines 231-232. "The mean and median OF THE FITTED POWER LAW can be expressed analytically."
Corrected.

Line 239. "Fig 6b" should be "Fig 6c"
Corrected.

Line 254. "50 m to 5 km^2" -- Is this supposed to say 50 m^2? Are the units supposed to be diameter or area?
Yes, 50 m^2, this has been corrected.

Line 268. "Fig 6c, shaded" -- I don't see shading in Fig 6c.
Corrected, removed "shaded"

Appendix A. In A1, does alpha have to be greater than 1?
In A2, I think it would be useful to plug in the expression for c from A1 and write E as
((1-a)/(2-a))*((xmax^(2-a)-xmin^(2-a))/(xmax^(1-a)-xmin^(1-a)))
where a=alpha.

Yes, alpha must be greater than 1, now noted. We now provide this expression (Equation A6).

Note than alpha cannot equal 1 or 2. Those are special cases.
We have similarly added after (A6): "for \alpha \ne 2.".

In A3, I think it would be useful to plug in the expression for c from A1 and write x0 as
((1/2)*(xmax^(1-a)+xmin^(1-a)))^(1/(1-a))
Again alpha=1 is a special case.
Thank you, we have added this (Equation A9).

Reviewer 2:

Review on the manuscript "Seasonal evolution of the sea ice floe size distribution from two decades of MODIS data" by Buckly et al.

Thank the authors for detailed responses to my previous review and they made a major revison to this manuscript. Most of the raised questions and comments were addressed. There are still some issues need to be further claified, mainly concerning the floe segmentation method and accuracy validation.

Thank you for your helpful comments, we have added a Section 3.7 to further detail the limitations of the algorithm.

A major comment: The improvement of this floe segmentation method compared to the previous methods is minimal. These fundamental algorithms were proposed based on a relative small amount of data or small coverages. Therefore, the present comparison and validation are still in doubt. Why the improved algorithm can be applied to a large amount of dataset and large spatial coverage are not very convinced. In their responses, it is mentioned that "Also each Sentinel-2 image is 110 km x 110 km, so the analysis of three Sentinel-2 images and coincident MODIS imagery allows for comparison of 36,300 km^2, a sufficient area for validation.", and "We need to keep in mind that the availability of validation imagery is very limited." However, these 36,300 km^{2}is like "a drop in the ocean" compared with the three-years data in the vast Beaufort sea. I still suggest to add some quantitative methods to measure the effectiveness of the proposed segmentation method, rather than relying only on comparisons with the limited Sentinel-2 data.

We recognize there are limitations to this dataset and methodology, such as the splitting of floes in high ice concentration areas, and the inability to detect the smallest floes due to sensor resolution. We hope this has been sufficiently addressed in our added section. It seems somewhat arbitrary that the validation imagery must cover a certain percentage of the total area analyzed. Rather, we ensured an appropriate validation via the use of a selection of varied sea-ice states (i.e., a range of sea-ice conditions in the Sentinel-2 imagery).

Specific comments:
1. P3 Fig.1 (c): Rather than showing the annual average ice floe numbers here, it might be more sutiable to show the data amount of processed MODIS in this figure. In particular, long-term statistics need to know the amount of cloud-contaminated data for each year.

Please see Figure 5 for those statistics.

2. P6 L118: The authors mentioned in the revised article, "The morphological erosion operation is applied to the binary image, removing pixels on the object boundaries with a diamond-shaped structuring element with a radius of 1." Why is this improvement more suitable for the task compared to previous methods? Why is the radius set to 1?

This is the same structuring element used by Denton and Timmermans, 2022; we note this as part of our description of the methodology, not to suggest that this is necessarily an improvement on previous methods.

3. P7 section 3.6, the authors mentioned: "The same algorithm was applied to both datasets, and there is a step to remove low-intensity objects." For floes in the S2 data, is it reasonable to remove floes with an intensity lower than 150, just like in the MODIS data?

We use identical algorithms for both data types for appropriate comparison of the results.

4. Despite the comparison with S2, as shown in Figure 3, the method has produced results with a significant amount of undetected floes. Consequently, conclusions drawn from the analysis based on these segmentation results would be greatly affected by the algorithm's performance. For the large amount of MODIS data that cannot be compared with S2 data, it is impossible to ensure the comprehensiveness of the detected floes. Therefore, the conclusions may be unreliable.

We believe we have sufficiently validated the algorithm while acknowledging its deficiencies. Please see the new Section 3.7 Limitations for an in-depth discussion of the limitations of this algorithm.

5. I am particularly interested in how floes are defined in areas with high sea ice concentration, as shown in Figures 3(g) and (h). How can we determine whether a floe ice is an independent entity rather than having re-grown together with other floes?

This is not only a question about how the algorithm handles this but physically how we define an individual floe, which is subjective to a certain extent. The refreezing of leads and fusion of floes into a single floe is complicated. For example, at what point in the fusion process should we consider multiple floes as a single floe? Further, this question is probably answered differently depending on the end use of the floe size distribution. For example, considering how a floe field reacts to an ocean swell, a weak fusion might as well be two independent floes, because it will break easily. When considering the distribution of upper-ocean solar heating, adjacent floes that are not necessarily fused but do not allow the penetration of light into the ocean, in this case may appropriately be considered as a single floe. This dilemma also bolsters the case for considering floe size distribution in context of other defining ice characteristics such as ice concentration and lead fraction. Any floe segmentation algorithm will be limited by the resolution of imagery and the sensor sensitivity to leads.

We split Section 3.6 Validation and Limitations into 3.6 Validation and 3.7 Limitations. In Section 3.7, we discuss limitations of the algorithm such as:

- Delineation of large floes in high sea ice concentration areas
- The undersampling of large floes due to image size and separation
- Limitations on the power law fit related to the limited range of sea ice floe areas

6. In the authors' response to the specific comments on L149-151 in the first-round review, they stated, "There is another question of the omission of smaller floes that I believe the reviewer may be referring to.

We are carrying out a further error analysis to understand these limitations, but this is beyond the scope of this paper." In addition to the smaller floes, some large floes were also omitted.

Thanks for the comment. We have included discussion in the new Section 3.7 on the large floes that are omitted.