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

The authors presented a study on sea ice floe size distribution in the Beaufort Sea based on the long-term MODIS dataset.

The method of ice floe segmentation is primarily based on the previously developed method by Paget et al. (2001). MODIS-derived ice floe segmentation was compared with the Sentinel-2 data based on three cases. They applied the method to 4,861 images of MODIS, identifying more than 9.4 million floes over 23 years. By analyzing the FSD in the Beaufort Sea over a long period, they unfolded the seasonal variability, including the decreasing mean floe area and increasing FSD power law slope as the summer progresses.

Some major comments and specific comments are below for reference.

Major Comments

1. Obviously, the title of the manuscript is too wide, and the authors only conducted analysis in the Beaufort Sea.
2. The dataset is not well described. For instance, are all four bands of the S2 data used to segment ice floes?
3. The methodology is too brief to be understood. Is the method exactly the same as that described by Paget et al. (2001)? Are any improvements achieved?
4. Validation of the MODIS-derived sea ice floes is not convincing. On the one hand, only by presenting only three cases for nearly 5000 MODIS images seems to be inadequate. I would suggest that the authors add comparison experiments to provide a more robust argument. On the other hand, the current comparison itself is not convincing. As the three cases shown in Fig.3, there are obvious distinctions (e.g., the N values in the figure) between the two datasets. The MODIS data have a spatial resolution of 250 m, which is much larger than the S2 data of 10 m. Therefore, a threshold should be set to exclude those floes that MODIS cannot observe at all due to its coarse spatial resolution. Moreover, the validation shown in subsection 3.6 fails to "perform equally well". These ice

floes, which are not segmented by MODIS, not only include the clustered loose ice but also some pack ice. Such a discrepancy is too visible to be explained by the different data resolution. The proposed method probably has limitations in terms of accuracy (however, it is hard to judge because the method is described too briefly). It is possible that some steps, such as the use of adaptive thresholds, excessive morphological processing, or the direct discarding of low-intensity floes, lead to unreasonable results.

5. The title of section 4 should be narrowed down. I think that many relevant studies have been conducted with the FSD in the Arctic. Why not compare some previously derived FSD with the present results for further verification of its accuracy? Additionally, the statistical charts should be improved. For Fig. 5, it is fine to apply the "10-day running window", but one should also consider quantitative results such as scatters, boxplots, or upper and lower significance intervals to give the reader a clearer view of the author’s raw statistical results. It would be better to try seasonal statistics, which might work better for the author’s dataset (as too many MODIS data are excluded from analysis due to the cloud effect).

Specific Comments

P2 L25 “floe size distribution”: The full name has already been presented. Similar issues occurred several times in the manuscript, including “SIC”, “SAR”, and “MODIS”. Please revise them.

P2 L26 “see Stern et al. (2018b) for a comprehensive list of FSD studies”: Even without a systematic review, a proper overview and summary of state of the art should be briefly presented here.

P2 L27-28 “... from radar imagery ... high-resolution imagery ....”: The summary of research types is weird. Isn't a SAR image a high-resolution image? Please rewrite this sentence to provide a detailed review of these studies.

P2 L29-L31 “These studies have advanced our knowledge of seasonal evolution of the FSD...”: The authors' review does not serve to summarise this knowledge.

P2 L35 “(Lopez Acosta et al., 2019) demonstrate...” -> “Lopez Acosta et al. (2019) demonstrate...”. Please also note other similar citations.

P3 Fig.1 (c): Rather than showing the annual average ice floe numbers here, I'd be more interested in first finding out the annual use of MODIS data. In particular, long-term statistics need to know the amount of cloud-contaminated data for each year.

P4 Fig.2: The authors presented a very clear case of pack ice. In addition, I would also like to see the algorithm's adaptability to high SIC, melting ice, MIZ, ice-water mixing, etc. After all, the authors aimed to focus on the phenomena related to the transition, which implies a rather complex ice condition.

P5 L90 “400-pixel (100 km) neighbourhood of a pixel subtracted by a constant”: What is the constant? The 400-pixel neighbourhood is a relatively large region. However, the masked MODIS contains many NAN values. How did the author choose the threshold at the edge of the NAN?

Additionally, it would be better for the authors to argue for the rationality and specific benefits of adaptive thresholds where appropriate.

P5 3.3 section: Since erosion-expansion is an important step, it would be better to show the effect before and after morphological processing using the Fig.2 case.

And, what is the necessity of performing multiple iterations? Actually, in my opinion, besides the fact that it does improve the visual effect, too much morphological processing may lead to losing the original sea ice features.

P5 L103-L105 “On average, 26% of the classified sea ice area is identified as individual floes”: Average of what?

P5 L111 “…the variation of the floe orientation”: Which floes are the authors using circularity std to compare their variation? Are they ice floes that are tracked between time-series images, or are they all ice floes in the same image?

P6 L116-119: In fact, the low-intensity ice may be an important ice condition as well (especially in the transition, where it may represent a melting scenario). However, the author removed them outright, which would cause the subsequent results, especially for power law distributions, to be different from previous results.

How can the authors justify this proposed step? Also, what are the units of 150? Can the authors prove <150 to be so-called brash ice rather than other types of sea ice (e.g. grey ice, melting ice)?

P6 L122-123 “that complements other parameters commonly used to describe ice floe fields, such as the sea ice concentration (SIC) and average ice thickness”: I would suggest deleting this sentence.

P7 3.6 section: When using S2 images for segmentation, is there also a step to remove the <150 intensity? And, are the other parameter settings exactly the same (number of erosions and the calculation of adaptive thresholds)? What level of S2 data is used? These details may lead to a different adaptation of the ice floe segmentation algorithm to S2 data.

Fig 3&4: As I mentioned above, I don't think the comparison presents a good result.

What is the “x_{min}”? Also, please present clearer quantitative results in the text.

L149-151 “The areas of the matching 82 floes agree very well, with a correlation of 0.99, and an absolute mean area difference of 0.18 km2 (Figure 4)”: The authors only compared the 81 floes identified by both. I don't think this comparison is fair. The high-resolution S2 results should be treated as a reference, and all the S2-segmented floes should be compared to MODIS results. Obviously, the MODIS results have gross omissions.

P7 L153-154 “...14% more floes are identified in the MODIS imagery compared to the Sentinel-2 imagery.”: How can the author tell the 14%? I can only find that MODIS significantly underestimated the floe numbers.

P7 L163 “We create 1000 bootstrap samples with replacement”: I can't follow the author's point of this step.

P7 L171 “ SAR data is not as widely available”: This sentence is misleading. If it refers specifically to long-term applications, SAR does have limitations. “other limitations or complications, such as speckle or granular noise”. It is not correct. It is just the authors do not know how to deal with the valuable dataset.

P8 L181-182: “…median floe size exhibits a similar pattern as the mean, with a maximum median value of 17 km2 on 10 April...": The variations in the median are not visible at all, and this image should be modified.

P11 L210 "... Stern et al. (2018a) ...": Stern (2018a) included the analysis of small ice floes (<5km2), right? So, it can be different from the authors' results.

In fact, in section 3.6, the alpha estimated from the MODIS is also lower than the S2 result. It is suggested that the proposed algorithm is supposed to have some limitations. I would suggest that the authors explain such reasons in detail.

P12 L223 "…we find a low variability in their orientation (Fig. 5c)": Low variability? It seems to me that Fig. 5(c) displays a clear increasing trend. Doesn't this suggest that as the sea ice melts, the direction of the ice floes becomes more cluttered?

P12 L223-224 "This effect is especially noticeable in areas of high ice concentration, where the ice movement and readjustment to external forces is limited by the surrounding ice ": It's hard to follow. If the authors mean that there is a decreasing tendency for std as SIC increases, it should be expressed more clearly. Otherwise, the most obvious phenomenon on this graph should be the very wide range of std variation for SIC > 0.7.

P12 L227 "large rectilinear floes": This is less common usage, so please confirm.