Reviewer 2:

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.

Response: We sincerely thank the reviewer for a thorough assessment of the manuscript. We believe that these comments and suggestions have improved the clarity of the presentation. We respond to the reviewer and describe changes to the manuscript in blue font below.

Major Comments

1. Obviously, the title of the manuscript is too wide, and the authors only conducted analysis in the Beaufort Sea.

Thank you for bringing this up. We changed it to include "Beaufort Sea".

2. The dataset is not well described. For instance, are all four bands of the S2 data used to segment ice floes?

Sentinel-2 has a total of 13 bands, four of which are at ten meter resolution. We apply the IFT to the Sentinel-2 imagery. The IFT uses the red channel for ice-water discrimination. We have added clarification to the description of the dataset.

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?

We now provide further details in the text on the erosion-expansion methodology in the introductory part of the methodology section. The methodology is not the same as Paget et al. (2001). Paget et al. (2001) performs an expansion-erosion algorithm. Stern et al. (2018) and Denton and Timmermans (2022) methods are also based on erosion-expansion operations but add an iterative component to identify different sizes of floes each round. Our algorithm builds on these two algorithms but differs in that the structuring element shape and the ice-water discrimination techniques were redesigned for MODIS imagery.

The text has been edited to highlight the importance of iteration in the expansion-erosion methodology:

"Paget et al. (2001) describes an erosion-expansion algorithm that erodes the boundaries of floes to separate them, subsequently regrowing them to their original shape. Denton and Timmermans (2022) and Stern et al. (2018a) introduce an iterative procedure that cycles through erosion-expansion varying the amount of erosion to identify floes of different sizes.

In this work, we build on the algorithm described in Denton and Timmermans (2022). We rewrite the algorithm in Python, and automate it to process thousands of images consecutively, introducing adaptive thresholds. The algorithm consists of four steps: pre-processing, ice-water discrimination, segmentation, and post-processing (Figure 2)."

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.

Thank you for your comment. We need to keep in mind that the availability of validation imagery is very limited. Sentinel-2 is only available 20 km from the coast. Finding pairs of near-coincident cloud-free images in two different data sets (MODIS and Sentinel-2) is challenging. We present three different ice conditions to demonstrate the ability of the segmentation algorithm to operate in different SIC scenarios. 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. We have added additional text to the manuscript to address uncertainties.

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.

Thank you for bringing up this point. Given the different spatial resolutions, we set the minimum floe area as 5 km^2 for both the Sentinel-2 and the MODIS dataset. Even though more floes are identified in the Sentinel-2 imagery, the floes picked up in both data sets display similar FSDs. This highlights the value of MODIS to analyze FSDs dating back to the early 2000s.

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.

We now provide additional information on the methodology as outlined above. We added a sentence to address that the algorithm performs well on both types of imagery, but does identify more floes in Sentinel-2 due to the higher resolution:
"Although the algorithm is able to identify smaller floes in the Sentinel-2 imagery, the floe sizes retrieved and the floe size distribution agree well between the Sentinel-2 and MODIS imagery."

We developed the algorithm based on the success of previous algorithms (Denton and Timmermans 2022, and Stern et al., 2018). The adaptive threshold is for accounting for variable but minimal low-level fog over the open water, as described in Section 3.2. Stern et al. (2018) also applies an adaptive threshold for ice-water discrimination. This adaptive threshold is different but similar in purpose. Further, as noted in Section 3.5, since the low intensity floes account for less than 1% of floes, removing them does not introduce additional uncertainty.

4. The title of section 4 should be narrowed down.

Thank you for your comment. We apologize if we are misunderstanding the comment, but we believe that the title of section 4 describes exactly what is included in the Section. The title is "Spring to Summer Transition of Floe Characteristics," and we go on to describe how floe area, floe size distribution, and floe orientation evolve from March 1 to October 1.

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).

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

Here we grouped together a few comments regarding the suggestion for more comparison to existing studies.

We discuss the spring to summer evolution seen in our data as well as in other studies. We have added more quantitative comparison with the Stern paper as well. (Section 5.2). We added a figure to show how the cloud fraction evolves over the summer. This figure also includes the number of floes identified in each 10 day running window to address the reviewers comment on more statistical information. The floe area mean, median and 75th percentile are shown in Figure 6. The standard deviation of orientation and alpha value are computed with all the floes in the 10-day running window and thus boxplots and scatters are not applicable. The standard error associated with the alpha value is discussed in Section 5.2.

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.

Thank you for catching this detail. We have revised the text.

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.

We appreciate your attention to detail. We changed to: high-resolution optical satellite imagery.

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

Thank you, these have been revised.

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.
Figure 1c shows the total floe count per year. We also include a figure (Figure 5) and an introduction to Section 4 to discuss these statistics further.

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.

We appreciate this suggestion, but kindly disagree. The purpose of Figure 2 is to illustrate the methodology. We have included more examples of the algorithm applied to a range of ice conditions in Figure 3.

P5 L90 "400-pixel (100 km) neighbourhood of a pixel subtracted by a constant": What is the constant?

Thank you for catching this typo. This has been revised for clarity. This was the description of the python function we used. But the default constant is zero.

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?

The threshold is applied to the masked image, so the NAN values are not included in the calculation for the threshold. We added a sentence: "At this point, the land and cloud pixels are masked and not considered in this step. "

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

We agree with this comment. The justification for using adaptive thresholding is already described in the text: to consider variable pixel values of open water. We added a sentence to clarify this point: "In this way, we are able to account for varying brightness levels of open water given different ice concentrations and atmospheric conditions."

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.

This is a good point. We adapted this iterative erosion-expansion algorithm from Denton and Timmermans 2020 and Stern et al., 2018. Here, we review the original algorithms and focus on the changes we have made to apply similar methodologies to a different dataset at larger scales.

We interpret the second part of this question as why there are so many erosions within a round or why there are so many iterations. Hopefully we can clarify both.

In the first round, the erosion is very aggressive, so that small floes are eliminated, and the large floes are heavily eroded. Then the large floes that remain are tagged, regrown (dilated) the same number of times they were eroded, and then removed from the image. This iterative processes is repeated but with fewer erosions each round, so the size of the floes identified are decreasing in successive rounds.

In each round, the floes can be eroded multiple times since the erosion structuring element is small enough so that the features are not deformed. By removing just the border pixels we are maintaining the shape. And then this process is done multiple times to separate floes.

These methods are consistent with the iterative expansion-erosion algorithm presented in Denton and Timmermans, 2022 and Stern et al., 2018. Figure 3 in Stern et al., 2018 demonstrates the process:

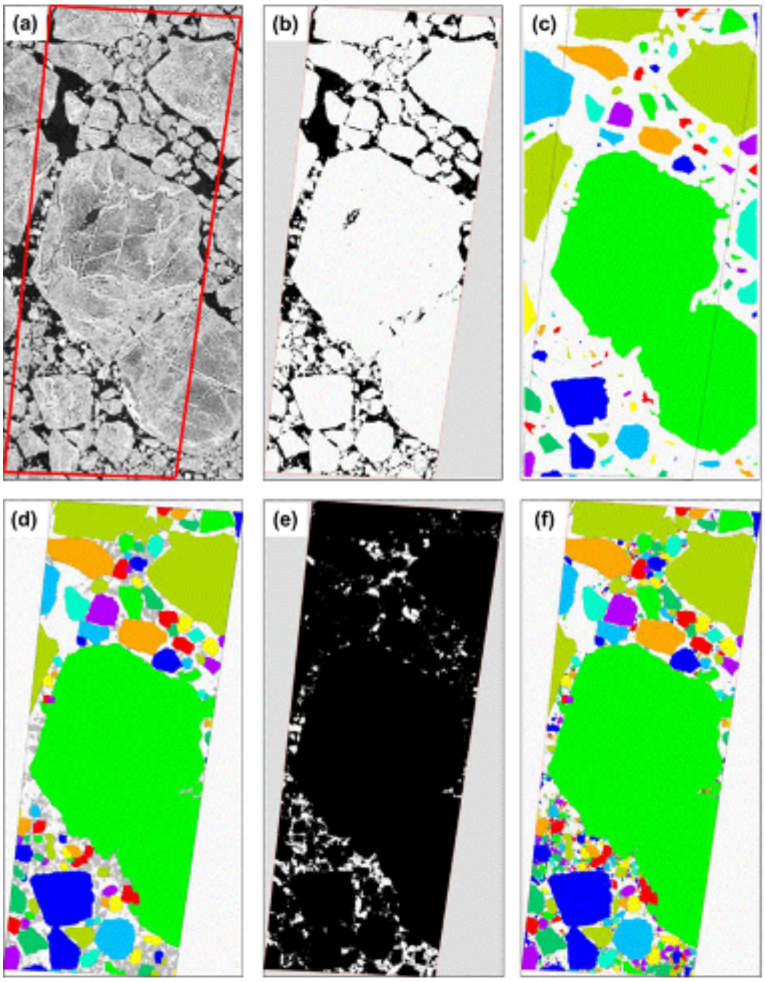

Figure Caption:
MEDEA image processing steps. (a) Subset #1 of July 8, 2014 (within the red outline). See Figure 9 for location. This subset measures 3294 × 6537 pixels (3.3 × 6.5 km). (b) Binary image, first iteration. (c) Eroded and labeled image with holes filled, showing individual ice floes. (d) Re-grown ice floes in color, with unidentified ice floes in gray. (e) Binary image, second iteration, after the largest floes have been tagged and removed. (f) Final labeled image showing all identified floes. DOI: https://doi.org/10.1525/elementa.305.f3

We have text throughout Section 3 that references the methodologies in these studies and includes more statistics of the erosion-expansion algorithm.

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

Of all the images analyzed. This sentence is clarified: "Over the thousands of processed images, on average, 26% of the classified sea ice area is identified as individual floes, with remaining sections consisting of ice filaments, brash ice or pieces of ice smaller than the minimum detectable floe size."

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?

All ice floes in the 10-day running window of images. Added "within the 10-day window" to clarify

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)?

The pixel values are an 8-bit integer from 0-255. We look at the red channel as noted in the manuscript. 150 is determined empirically by testing various thresholds and examining the pixel brightness of floes, brash ice, and mixed pixels. Because of the size of each pixel (250 x 250 m), mixed pixels containing more than one surface type are present. The 150 intensity value ensures that low brightness ice floes are still included but mixed pixels over ambiguous surfaces are not included in this analysis. Also- as mentioned in the manuscript- this only removes <1% of the floes. The further analysis of how floe brightness intensity may represent the surface melt evolution is an interesting question but beyond the scope of this paper.

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.

We choose to retain the information as we think it is important to acknowledge there are multiple ways to describe the sea-ice cover, but have made edits to the paragraph in question. We have reworded this: "Taken together with the floe geometry properties outlined in Section 3.4, the FSD and other parameters commonly used to describe floe fields, such as the sea ice concentration (SIC) and average ice thickness, allow us to study the physical processes that shape the structure and evolution of sea ice."

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.

Thank you for bringing this up. We use Sentinel-2 Level 1C Top of Atmosphere Reflectance- this detail was added to the manuscript. The same algorithm was applied to both datasets, and there is a step to remove low intensity objects.

Fig 3&4: As I mentioned above, I don't think the comparison presents a good result. What is the "xmin"? Also, please present clearer quantitative results in the text.

We added the xmin value to Section 3.5. We have clarified the figure and added discussion on how the high sea ice concentration example exemplifies limitations of the algorithm. We hope that clarifications and additions elsewhere (including the addition of an Appendix deriving the analytical expressions) provide clearer quantitative results.

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.

The comparison of floe areas can only be done with the floes that are identified in both images. The power law distributions shown does include all floes identified in both the images that are in the MODIS range of detectable floe area. Here we are demonstrating that in the floe area range that MODIS can identify floes, there is strong agreement with the Sentinel-2 validation dataset. 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. Here we just assure that MODIS is acceptable for this range of floe sizes.

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.

We do not limit the floe size detected in the Sentinel-2 image. To compare likes, we sample the floes identified in the Sentinel-2 imagery that are in the MODIS size range. The 14% more floes are in the area range that MODIS can detect- so greater than 5 km^2. We have added a sentence to clarify that the Sentinel-2 identification includes floes less than 5 km^2. "as the identification of floes in the Sentinel-2 imagery was not limited to the MODIS range of floe sizes."

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

This text has been modified for clarity. We include this step to show that the power law fit is a robust description of the dataset.

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.

Thank you for this point- we are referring to the availability of SAR data for this study which is a long term study that includes the summer time period. We have clarified the text: "However, SAR data is not as widely available for long-term applications, and has other limitations and complications, such as speckle, granular noise, and ambiguous returns when meltwater is present on the ice surface."

P8 L181-182: "…median floe size exhibits a similar pattern as the mean, with 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.

Thank you for the comment- this image has been modified. A new panel has been added to separate the mean and the median.

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.

Yes, Stern et al 2018 study includes floes as small as ~ 2.6 km^2. We have added a line: "These values are similar and the small differences are likely due to location and time of the observed floes." We acknowledge that this algorithm and dataset have limitations- this is already described in Section 3.6 Validation and Limitations.

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?

"this period" refers to the spring which was described in previous sentences. we replaced "this period" with "the spring" for clarity.

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.

Added for clarity "lower standard deviation of orientation with greater SIC"

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

Confirmed. Rectilinear means straight lines, and this is how we see the ice fracture in the spring.

See also:

Arntsen, A. E., Song, A. J., Perovich, D. K., & Richter-Menge, J. A. (2015). Observations of the summer breakup of an Arctic sea ice cover. *Geophysical Research Letters*, *42*(19), 8057-8063.

where in the abstract they describe the fracturing in the ice pack resulting in rectilinear floes.