Marion Lebrun

General Comments

This paper introduces a novel method for characterizing the seasonality of Arctic sea ice. The k-means clustering method, which has been successfully applied in other contexts, serves as the foundation for this approach. This study enhances the method by incorporating the Mahalanobis distance, providing a more robust physical representation of sea ice seasonal cycles. By applying this refined technique to sea ice concentration data spanning 1979 to 2023, the authors identify four distinct clusters that effectively describe the seasonality of Arctic sea ice.

Building on these results, the authors analyze the probability of individual seasonal cycles within the dataset belonging to each of these four clusters. They also introduce new diagnostic tools to describe the temporal evolution of sea ice seasonal cycles, as well as to pinpoint the moments when a seasonal cycle transitions from one cluster to another.

This study offers an innovative approach that complements previous research on sea ice seasonality. I commend the authors for the thoroughness of their analysis and the clarity of their well-structured manuscript. Overall, I believe this work represents a significant contribution to the field and, with the revisions suggested below, has the potential for a strong impact upon publication.

While the bibliography section requires attention before submission (as detailed later), most of the comments below are intended to enhance the paper's clarity, precision, and relevance. Authors are encouraged to consider these suggestions and apply them at their discretion.

Major Comments

The only major comment regarding this paper concerns the paragraph between lines 325 and 339. In my opinion, this paragraph requires further clarification or a more detailed presentation of the results to be fully convincing. The results presented in this section and reiterated in the conclusion, showing that ice conditions during the summer (or winter) are strongly correlated with the onset of melting (or freezing), are highly interesting for understanding the mechanisms behind the Arctic sea ice seasonal cycle. However, I find that relying solely on clusters to describe these results is limiting.

If I understand correctly, the seasonal cycles of each grid point tend to group around one (or several, considering the results in Section 3.2) cluster. Therefore, I am not fully convinced that the seasonal cycles associated with a particular cluster (as shown in Figure 4b) behave exactly as summarized in this paragraph.

I believe it would be helpful to visually demonstrate this with supporting evidence to strengthen the argument. For example, you could present the interquartile range around each cluster on Figure 4a, using the data already employed to generate Figure

4b. Alternatively, if you prefer not to overload Figure 4a, you could clearly define the melting onset and freezing onset dates here (using concentration thresholds already applied in other studies) and provide statistics of these diagnostics in each cluster. In my opinion, these revisions would significantly enhance the impact and clarity of these results.

Thank you very much for your relevant comment and nice idea. We have further pushed the diagnostic.

To have an idea of the spread of seasonal cycles around each cluster, we have also plotted the median (solid line) together with quantiles 0.90 and 0.10 (dashed line) for each cluster in a supplementary figure.

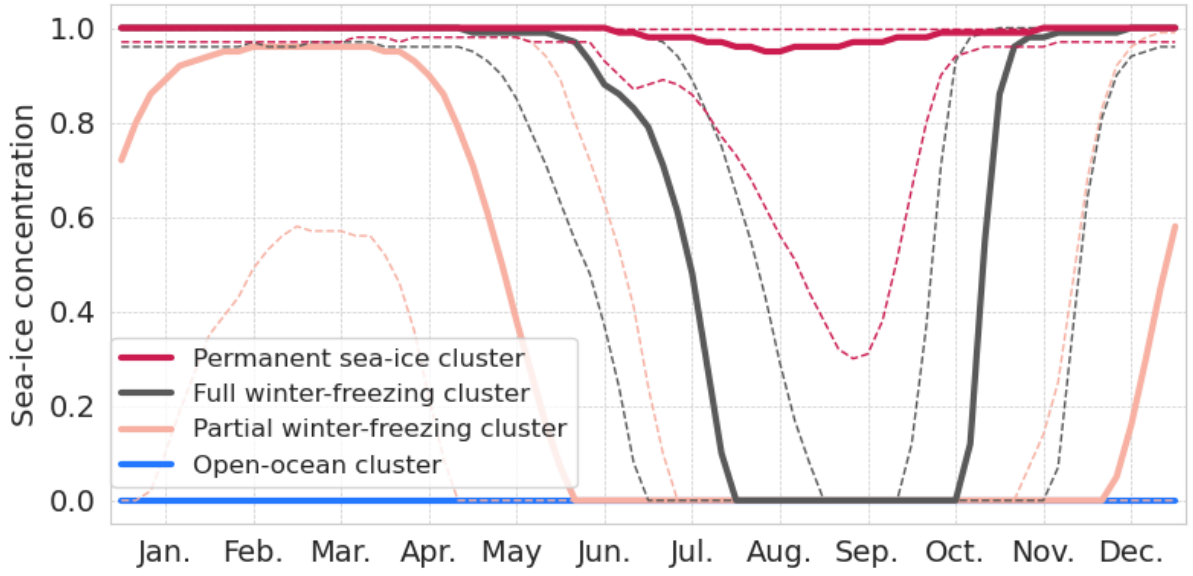

Figure S2: As Fig. 4a, but for the median (solid line) and quantiles 0.10 and 0.90 (dashed line)

In the main text, we have added a new figure and interpretation considering the spread for the date of retreat and date of advance.

We now say:

"... Therefore, it seems that, for ice-free conditions in summer, the first date of freezing is a good predictor for the appearance of full ice conditions in the next winter.

However, this suggestion relies solely on the shape of the four types of seasonal cycles but to properly quantify this, the spread must be taken into account. Figure S2 displays the spread of the seasonal cycle by plotting the quantiles 0.1, 0.5 and 0.9 of each cluster. To verify our hypothesis on sea-ice predictors, we account for the spread of the date of retreat and date of advance for each cluster. To do so, we

calculate the first date of retreat (the first date after the maximum SIC that is below 0.9) for each seasonal cycle experiencing fully ice covered conditions (having at least one value above 0.99 during the year). We also calculate the first date of advance (the first date after the minimum SIC that is above 0.1) for each seasonal cycle experiencing ice-free conditions (having at least one value below 0.01 during the year). For these calculations, seasonal cycles have been temporally filtered using a 15 days sliding window in order to get rid of short-term dynamical ice events, as done in Lebrun et al., (2019). To circumvent the effect of the discontinuity between 31 December and 1 January, we define the origin of time in May for the calculation of the date of advance. We then label each first date of retreat and first date of advance for each seasonal cycle with its corresponding cluster according to our clustering analysis (Figure 4a).

The normalized probability over each cluster of the first date of retreat and first date of advance at each date is shown Figure 5. This figure also displays the total number of the first date of retreat and the first date of advance of all clusters for each date. If the first date of retreat occurs between January and April, there is around 95% of chance to belong to either the open-ocean cluster, the partial winter-freezing cluster or full winter freezing cluster, which all present ice-free duration in the following summer. However, this situation did not often occur, as the total first date of retreat happening in this period is unlikely (solely 5% of first date of retreat for all clusters). The first date of retreat is more likely to occur between the beginning of April and August, as within this period around 90% of the total date of retreat for all clusters exist. A first date of retreats in early June has solely around 10% of chance to belong to the permanent sea-ice cluster which do not present ice-free conditions in summer while a first date of retreat in early July has around 70% of chance to belong to the full winter-freezing cluster which shows ice-free conditions in summer.

The first date of advance is more likely to occur between the beginning of August until the beginning of January, as within this period around 90% of the total date of advance for all clusters exist. A first date of advance in early September has around 95% of chance to belong to the full winter freezing cluster which present fully ice covered condition in the following winter, while a first date of advance in early

November has around 80% of chance to belong to the partial winter freezing or open ocean clusters which do not show fully ice covered conditions in the following winter. Therefore, this simple model suggests that the first date of retreat could be a good predictor for ice-free conditions the following summer and the first date of advance a good predictor for fully ice cover conditions the following winter. “

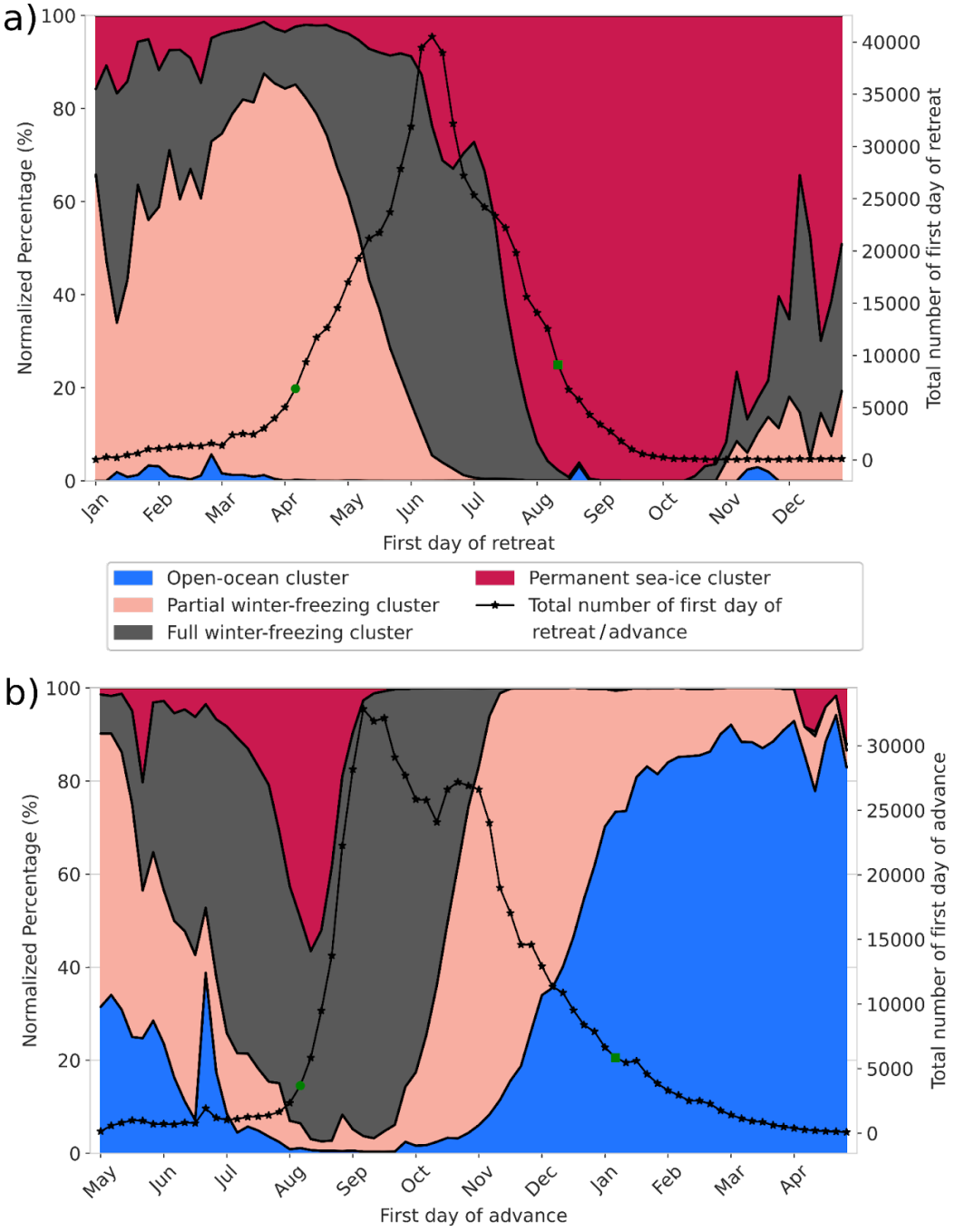

Figure 5: Normalized probability of the first date of retreat (panel a) and first date of advance (panel b) for each cluster. The solid lines with star markers are the total number of first dates of retreat and first dates of advance for each date. The

green circle markers (start date) and green square markers (end date) cover the shortest period where around 90% of the first date of retreat, respectively the first date of advance, for all clusters occurs.

Specific Comments

Lines 137-139 and lines 152-157: I feel, with both sentences, that the authors try to highlight the novelty of their method compared to previous studies, but these sentences appear before the work itself is introduced. I am not sure of the relevance of these sentences at this point. This creates some confusion during a first reading. I suggest either removing these sentences or moving them to a later paragraph, ideally after the authors have introduced their work more clearly.

The lines 137-139 *("Here, in this paper, we describe the evolution of the Arctic by delimiting spatio-temporal regions having a common type of seasonal cycle.")* and the lines 152-157 *("However, the criteria for the boundaries of these proposed regions are hard to determine and somewhat arbitrary. The originality of our analysis also resides in the fact that we regionalize the Arctic based on physical criteria of the dynamics of the sea-ice seasonal cycle, therefore without imposing predefined regions. To do so, we set up a clustering method (unsupervised machine learning") and started* introducing our work in regards to previous work. We think by removing or moving to the last paragraph, the originality of our work will be less easy to follow. In this manner, we can easily and step by step compare the differences/added value of our work to previous studies.

Lines 200-202: This statement could benefit from additional evidence (by including figures or statistics in the supplementary material, for example).

We agree that the lines 200-202 *('We choose this 5-day temporal resolution as similar results are found for a daily temporal resolution whereas a monthly temporal resolution shows small differences in the spatial distribution of clusters.)* could benefit from additional evidence.

We now say: "We choose this 5-day temporal resolution as similar results are found for a daily temporal resolution whereas a monthly temporal resolution shows different numbers of optimal clustering and small differences in the spatial distribution of clusters (Figure S1)."

We have included this new figure and associated paragraph in the supplementary

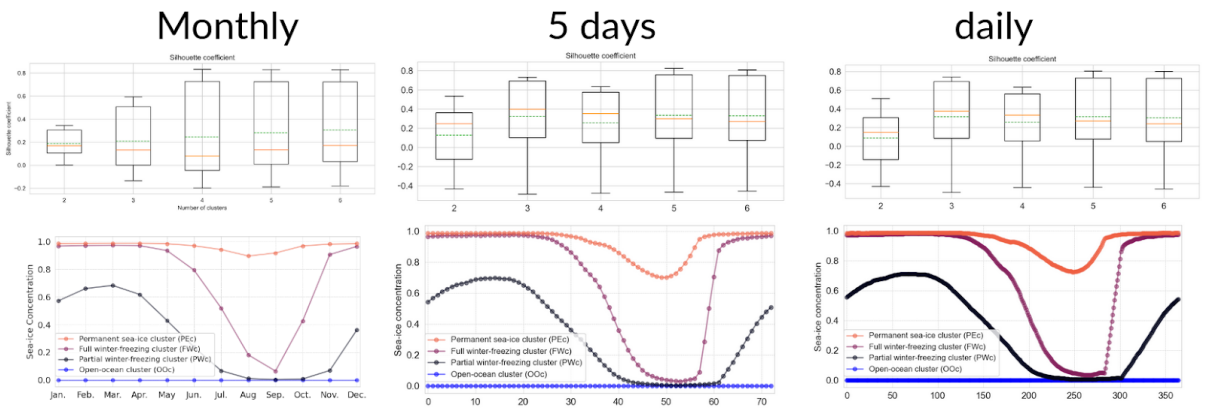

Figure S1: Comparison between monthly (left), 5-day (middle) and daily temporal resolution (right). The Silhouette coefficient for a number of clusters from 2 to 6 (top row), the four types of seasonal cycles (bottom row). In the top row, the box extends from the first quartile (0.25) to the third quartile (0.75) of the Silhouette coefficient. The whiskers indicate the 1st and 99th percentiles. The green-dashed and orange-solid lines indicate the mean and median values, respectively.

The Silhouette coefficient is maximum (so optimal clustering) with 6 clusters with monthly data and for 3 clusters for 5-days and daily data.

Lines 216-217: It could help the reader if you briefly explain what you mean by "*non-zero seasonal cycle*."

We now say: "We consider all oceanic grid cells above 55°N having a non-zero sea-ice seasonal cycle (having at least a non-zero value for SIC throughout the year)"

Lines 353: "*consistent and continuous patterns*", consistent according to what?

We have removed "consistent". We now say: "we retrieve spatially continuous patterns"

Lines 395-397: I find it difficult to discern "*the edge of the 0.3 probability*" in Figure 5 due to the continuous colorbar. Adding a contour line to indicate this boundary could make it clearer.

We think it will unnecessarily overload the plots. We now say: "The edge of the 0.3 probability of belonging to the permanent sea-ice clusters of the period 1979-1993 follows the border of the Marginal Ice Zone (0.8 SIC) located in the Central Arctic (not shown)"

Figure 5: I find it difficult to discern the probability differences between clusters when the colorbar changes for each cluster. While I understand your choice to maintain consistency with other figures, for this particular figure, I suggest using a single colorbar for all four clusters to enhance readability.

The figure now has clear separation between probabilities, as we no longer used the Mahalanobis for the calculation of probability. We use the Euclidean distance for that. We still use the Mahalanobis for the clustering though. We now say:"The

Mahalanobis norm, deriving from a symmetric operator, effectively rotates the original physical phase space (here., date of the annual cycle) to align with the data's natural directions—linear combinations of the physical time axis. This transformation allows centroid detection in a space that reflects the intrinsic structure of the data. Therefore, using the Mahalanobis distance helps the clustering algorithm to follow the direction of the correlation and capture the elongated shapes of clusters. When calculating the probability to belong to one cluster, we do not need to work with the data's natural directions, but rather work in the original physical time space. Therefore we use Euclidean distance for the calculation of probability and the Mahalanobis for the clustering."

The new figure : "

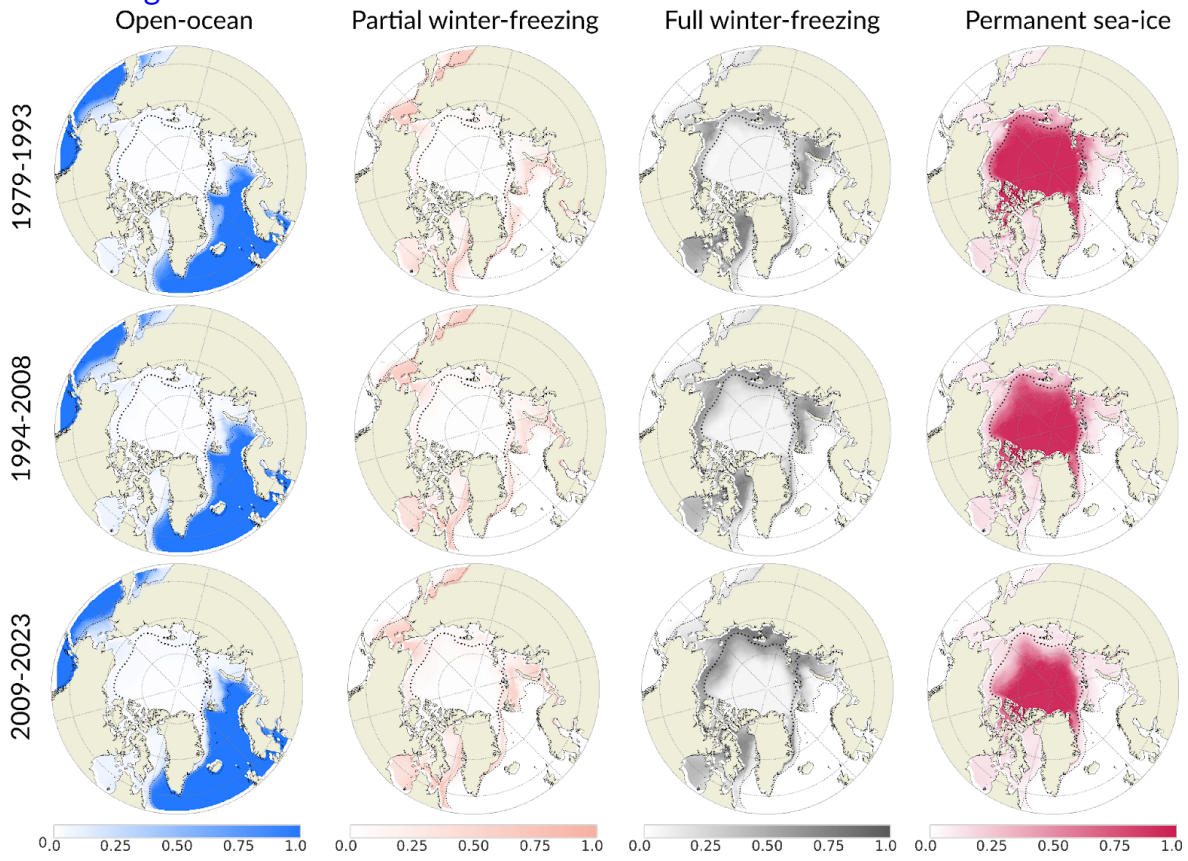

Figure 8: Map of the probability of each cluster: open-ocean (first column), partial winter-freezing (second column), full winter-freezing (third column) and permanent sea-ice (fourth column). Rows correspond to three periods of 15 years: 1979-1993 (top row), 1994-2008 (middle row) and 2009-2023 (bottom row). The dotted thin and thick lines are the mean SIC of 0.15 and 0.8 for the period 1979-2023, respectively. The circle sitting over the north pole is the pole hole (see section 2.1). "

Paragraph 464-476: The reference to Figure 7 is missing here and should be explicitly mentioned for clarity.

Thank you. We now say: " The probability of belonging to the open-ocean cluster is

around 40%, to the permanent sea-ice cluster is around 29% and to the full winter-freezing cluster is around 18 % and to the partial winter-freezing cluster is around 13% (Figure 7)”

Line 468: “ *the trend for the other two clusters are statistically significant*”. Does it mean that the trends for the partial and full winter-freezing clusters are not significant?

Yes, we have now clarified by saying:”All curves show a significant linear trend with a p-value less than 0.05 using a Wald Test with a t-distribution.”

Lines 476: “*and to a smaller extent, of the full winter-freezing cluster*”. I am confused by this statement. Earlier (line 466), it was mentioned that the trend for the full winter-freezing cluster was nearly constant, and line 468 suggests that the trend is not significant. This part seems to lack precision to be clearly understood.

Indeed, it was confusing. We have removed this last part of the sentence.

Lines 502-503: "*while the partial and full winter-freezing clusters remain relatively stable*." I find it unclear in Figure 8a that the total area covered by the partial and, in particular, the full-winter cluster is stable. It might be helpful to add the trend for each cluster, as was done in Figure 7, to make this clearer.

Thank you very much. We have now computed the trend and significance.

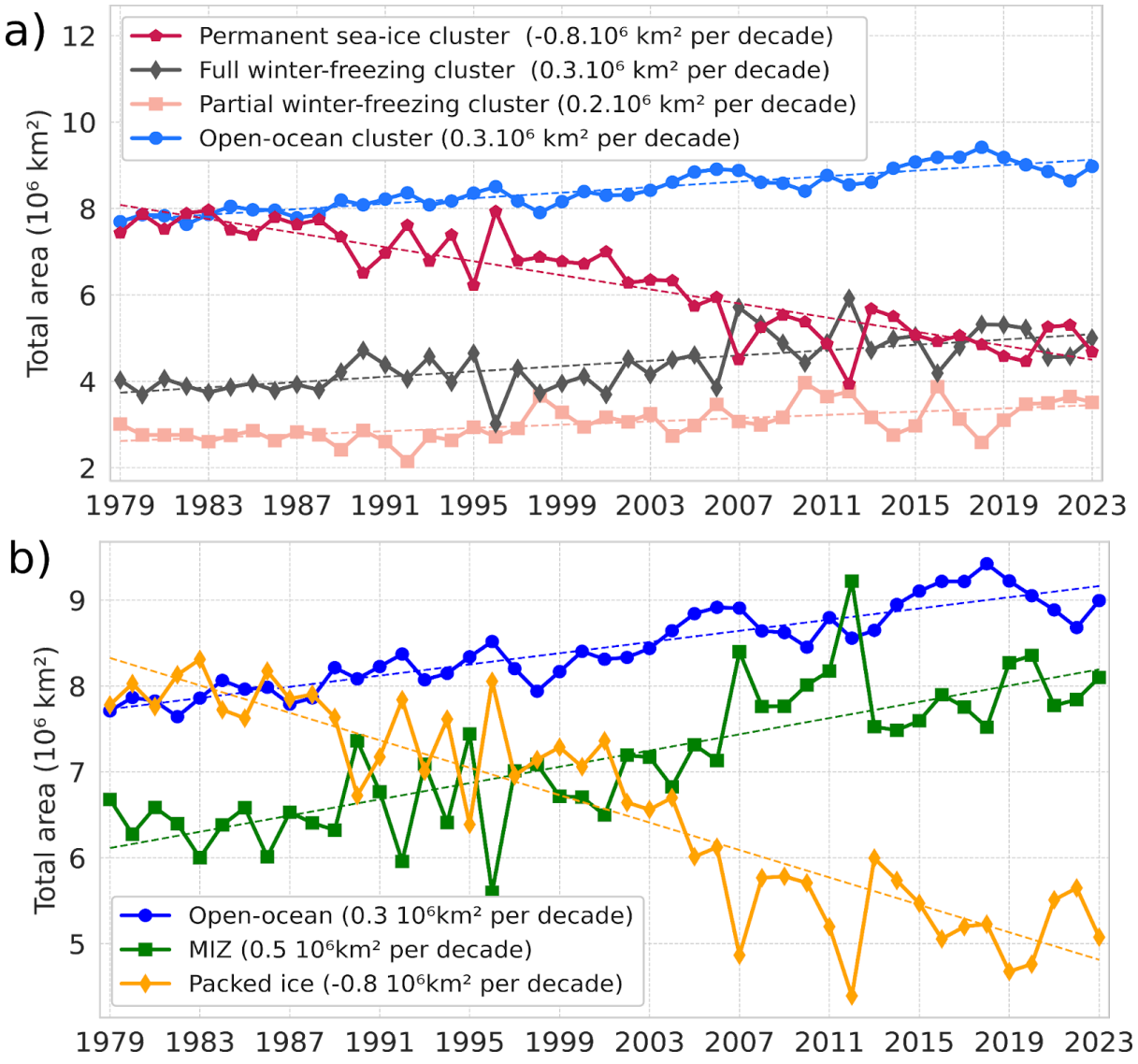

Figure 6: (a) Time series of the total area covered by each of the four clusters. (b) Times series of the area covered by three categories: packed ice (0.8 < SIC < 1), the Marginal Ice Zone (MIZ; 0.15 < SIC < 0.8) and the open-ocean (SIC< 0.15). All curves show a significant linear trend with a p-value less than 0.05 using a Wald Test with a t-distribution.

And we now say:"These two methods (Figure 6a and Figure 6b) both indicate a shift toward more seasonal and ice-free conditions. Indeed, in the clustering method the permanent sea-ice cluster has notably decreased of the same amount than the packed ice in the classical categorization (-0.8 .10⁶km² per decade). Also, the open-ocean cluster follows the same trend of the open-ocean category (0.3 10^6 km² per decade). The increase in the area of MIZ category is around 0.5 10^6 km² per decade and has been demonstrated previously (Cocetta et al., 2024; Song et al., 2025). Therefore, it appears with our clustering that the MIZ is refined into two clusters : the full winter-freezing (0.3 10^6 km² per decade) and the partial

winter-freezing cluster (0.2 10^6 km^2 per decade). This suggests that the tendency is more likely to shift to a more abrupt melting and growth seasonal cycle (full winter-freezing cluster) compared to a quasi-sinusoidal sea-ice seasonal cycle (partial winter-freezing cluster). ”

Lines 505-509: I find it difficult to discern the nuances between these two sentences. Perhaps the last sentence could be omitted, as it might not add significant value.

We agree. We have removed the last sentence.

Lines 518-519: “*Sensitivity tests have been performed on this definition, and the results do not change when we apply small definition changes (i.e., 9 to 11 years minimum length of the same cluster with zero to 2 years of tolerance).* ” It would be valuable to include the results of these sensitivity tests, perhaps in the supplementary material, to provide additional context and support for this statement.

We added the following figure in the supplementary and we now say: “Sensitivity tests have been performed on this definition (Figure S2)”

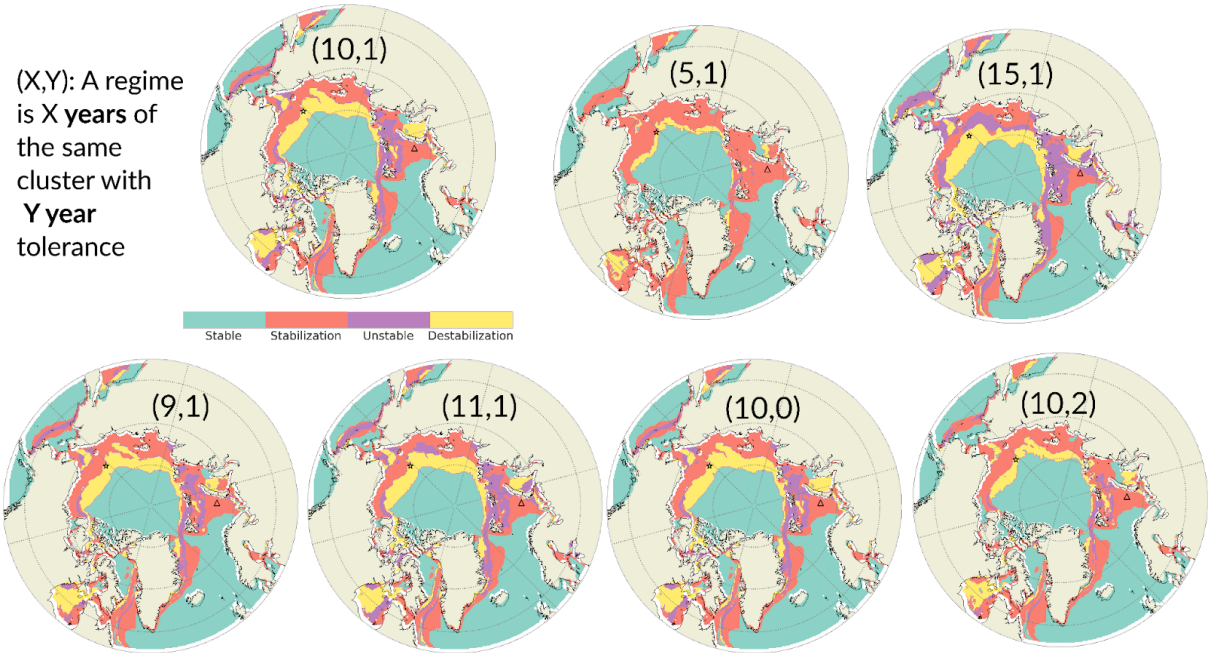

Figure S3: Sensitivity tests on our definition of regime. Same as Figure 11 but with a different set of values for the minimum number of consecutive years and tolerance.

Lines 522-525: A reference to Figure 9 would enhance clarity here, as the figure significantly helps in understanding the definition of these new diagnostics.

Thanks. Figure 9 is now Figure 10. We now say: “Figure 10 illustrates how we define the stabilization and destabilization labels.”

Lines 578-582: I am uncertain about the relevance of this paragraph, as all the information presented here seems to be already covered in the previous paragraph.

We agree that the Figure 11d is more a confirmation from the previous paragraph. However the Figure 11c does as it allows additional spatial information to which seasonal cycle the regions stabilized and informs that it depends on the regions. We keep both figures for consistency among the stabilization and destabilization results. We have reformulated the paragraph.

We now say: ”To describe the stabilization and destabilization regimes, we display the dominant cluster (the cluster having the maximum probability) during the stable phase of these two regimes (early period for the destabilization regime and late period for the stabilization period; Figure 12c and 12d). ”

References section :

I noticed that several references cited in the paper are missing from the reference list. I’ve compiled a list of the missing references I found, but I strongly recommend that the author carefully review this section, as there may be other errors that I may have missed. Additionally, the format of the references is not consistent throughout the section. For example, the publication date is sometimes listed immediately after the authors' names, and in other cases, it appears at the end of the reference. To ensure consistency, I suggest following the EGU standardized citation format.

Thank you very much. We have now completed the missing references and ensure that the year is right after the name’s author in a consistent way.

Missing reference in the references section :

Ardyna and Arrigo, 2020: line 83

Eisenman, 2010: line 439 ; 693

Eyring et al., 2021: line 77

Lebrun et al., 2019: line 125 ; 137 ; 631 ; 648

Meier et al., 2007: line 146

Markus et al., 2009: line 127 ; 629

Maze et al., 2017: line 161

Parkinson and Cavalieri, 2008: line 678

Parkinson et al., 1987: line 304

Parkinson et al., 1999: line 677

Parkinson, 2014: line 127

Peixoto and Oort, 1992: line 683

Peng and Meier, 2018: line 146

Regan et al., 2023: line 119

Lines 803-805: Houghton and Wilson 2020 should appears before Huntington et al., 2017