

## General assessment

This manuscript applies an unsupervised machine-learning framework, NEMI/UMAP/DBSCAN, to classify Antarctic sea ice regimes based on NEMO-SI<sup>3</sup> sea ice mass-budget components. However, in its current form, I find the manuscript not sufficiently convincing.

1. I have some concerns about the data input first, since the analysis depends entirely on model-derived budget components, yet the model has notable biases in sea ice concentration and especially sea ice thickness. The manuscript states that the model underestimates SIC in the Weddell and Amundsen Seas, overestimates elsewhere, and generally underestimates SIT compared with GIOMAS, but then concludes that the model is adequate for diagnosing sea ice budget components. Then, how the conclusion is reached?
2. The clustering framework is also not sufficiently constrained by physical reasoning. The authors state that UMAP parameters were chosen based on “physical intuition” and stability, and that DBSCAN identifies six clusters with high ensemble stability. Then, how the stability of an unsupervised clustering method guarantee physical validity? A biased model can produce stable but physically misleading clusters. This distinction is not adequately addressed.

I therefore recommend major revision. Here are my major and detailed comments:

## Major comments

1. My first and major concern is: the author acknowledges that the model underestimates SIC in some regions, overestimates it elsewhere, and generally underestimates SIT compared with GIOMAS. And that mean the derived thermodynamic and dynamic mass budget components may also be biased, right? Then, it is possible that resulting clusters may simply reflect model structural errors rather than physically meaningful Antarctic sea ice regimes. And more details questions are: do we expect the biased input can have the physical sea ice processes understanding? How sensitive are the identified regimes to known NEMO-SI3 biases in SIC and SIT? Later in discussion, I saw authors also mention that “Coast” regime is not always along the coasts, could it be a model artifact caused by biases in sea ice motion, thickness, or thermodynamic growth in the NEMO-SI3?
2. My second major concern is about the use of the quantile transformer and how this scaling affects the physical meaning of the input data. The authors mention that they tested several scaling approaches and found that the quantile transformer gave the most robust and consistent results. Why is the quantile transformer the best choice for sea ice mass-budget terms? Can the authors provide specific quantitative evidence to support this choice? For sea ice mass

balance, the absolute magnitude of the budget terms is physically important. By applying a quantile transformer, the original distributions are forced into normal distributions. This may improve the statistical separation of clusters, but it may also distort the physical distances between data points. In other words, weak but spatially variable processes may become overemphasized, while physically dominant processes may become less important in the clustering. The authors state that scaling is only used during the clustering step and that the results are later interpreted in physical units. However, this does not fully solve the problem, because the regime boundaries are already determined in the transformed feature space. Therefore, I think the authors need to provide more evidence for this methodological choice. For example, can they show the clustering results using unscaled variables, standard scaling, robust scaling, and quantile scaling? How different are the regime maps under these different choices? Does the "Coast" regime still appear in the same regions, especially those offshore regions that are not along the Antarctic coast? Is the decreasing trend of the "Coast" regime still significant under other scaling methods?

3. The "Coast" regime is one of the most important results of the paper. The authors state that the regime is mainly found along the coast, but also in offshore regions in the Indian Ocean, Ross Sea, Bellingshausen and Amundsen Seas (maybe too much). They further acknowledge that it does not necessarily represent coastal processes such as landfast ice formation or coastal leads, but rather divergence-dominated dynamics that may occur beyond the coastal zone. This raises a major interpretive problem. If the regime is not necessarily coastal, and if it can occur far from the Antarctic coast, then why it is called "Coast"? This issue becomes more serious in the Discussion, where the authors suggest that shrinking of the "Coast" regime may indicate weakening of new ice formation in coastal polynyas and may have implications for brine rejection and Antarctic Bottom Water formation. This conclusion is not sufficiently supported if the regime includes large offshore divergence regions and is not specifically tied to coastal polynya processes. And one thing I noticed, abstract states that the coastal regime has experienced significant area loss, "potentially reflecting weakened new ice formation." The Discussion makes a similar link to coastal polynyas and Antarctic Bottom Water formation. This is too speculative.
4. I think the authors need to explain more clearly how the UMAP and DBSCAN parameters were selected. For example, what physical criteria were used to tune these parameters? Were the parameters chosen mainly to maximize cluster separation, to maximize stability, or to improve physical interpretability? These are not necessarily the same thing. More specifically, I would like to see sensitivity tests for the main classification parameters. How do the results change if the UMAP neighbourhood size or minimum distance is varied? How do the results change if the DBSCAN epsilon or minimum-sample parameters are

changed? Are the six regimes robust over a broad range of reasonable parameter choices, or do they only appear for a narrow set of parameter values? In particular, does the “Coast” regime still appear in similar locations, including the offshore regions? Does its area trend remain significant? Do the physical budget characteristics of each regime remain similar when these parameters are changed?

### **Specific comments**

Line 30: make sure you have the sea ice or sea-ice in consistency for the whole manuscripts.

Line 33 “we conduct a process-based analysis of sea ice variability..”: This is a good motivation. However, the current analysis uses model-derived budget components whose realism is not sufficiently validated.

Line 82-87: The uncertainty discussed here appears to be mostly clustering uncertainty, not uncertainty in model physics, input variables, or preprocessing. Please clarify what uncertainties are actually quantified.

Line 109: How much is underestimated in SIT compared with GIOMAS?

Lines 146-148: It is quite strange to see the age-income analogy in a geophysical budget analysis. In sea ice mass balance, absolute magnitudes are physically meaningful.

Lines 153: The statement that the quantile transformer "outperformed" other methods requires quantitative evidence. What metrics were used? Cluster separability? Stability?

Lines 158-162 (line 169 as well): Why transfer the data into normal distributions given the above skewed dynamic and thermodynamic distribution. This is risky. The transformation can modify distances, correlations, and relative importance of variables

Lines 196-199: The statement that the model's climatological sea ice mass balance aligns well with established understanding is too broad. Please provide more direct evidence.

Lines 216: "UMAP parameters were chosen based on physical intuition" is too vague. What physical intuition? Which parameters? What ranges were tested? What criteria determined the final choice?

Lines 221-230: The comparison with KMeans and Agglomerative clustering is useful but insufficient. DBSCAN may capture visually appealing structures in UMAP space, but that does not prove physical correctness. Please include sensitivity to DBSCAN parameters and preprocessing choices.

Lines 245-246: The sentence stating that the regime does not necessarily reflect landfast ice or coastal leads undermines later discussion linking this regime to coastal polynyas and new ice formation?

Lines 299-305: The statement that increased divergence is no longer compensated by thermodynamic gain is interesting, but it needs more direct support. Are these trends computed within a moving regime mask? If so, the area and intensity effects may be mixed.

Lines 329-341: I am not sure how much I can speculate from post-2016 change is primarily related to significant area loss in the coastal and pack regimes, given that some concerns above.

Lines 355-364: The claimed advantage over SIC or bathymetry-based classifications needs stronger demonstration in the main text. If the difference is mainly shown in supplementary figures, don't you put that in the main manuscript?