Hi Phoebe,

You're not going to like what comes next. Sorry.

We thank you for your honest and constructive review and believe it will notably strengthen the paper.

Major comments

I. The objective of this manuscript is unclear

Although both the abstract and the introduction are very long, they are also unfocused, so it is unclear whether the aim of this article is to:

- combine satellite, reanalysis, and in-situ data to study the plume;
- demonstrate that satellite data can be used to look at sea surface salinity in the region;
- or determine which reanalysis product is most adapted for this study.

These different objectives would then result in different structures of the manuscript, with the majority of the product comparisons and validations going to the appendix. It would also affect the time period considered. The choice of in-situ data would then be affected as well; see my next point.

The main objective of the paper: "combining satellite, reanalysis, and in-situ data to study the plume" will be clarified and the paper will be restructured to reflect this. The validation component of the paper will be de-emphasized and moved out of the results section.

II. The choice of data, especially their resolution, was surprising

The method section did not clarify much, so I am not sure which time period you worked with. That is, you mention that you use UDASH, but that stopped in 2015 and would not really help with SMAP. The manuscript needs an overarching table that says for all types of products over which time period and at which spatio-temporal resolution they are available (not the resolution at which you use them; their native one).

Coupled with the fact that the objective is unclear, I won't be able to give you a clear direction. But if you want to validate SSS, I would have looked for underway CTD data rather than CTD casts. It will be at approx. 10 m depth, but on most vessels (and especially so on Oden, i.e. for the SWERUS data) the upper 10 m of the CTD casts cannot be used anyway. If you want the upper water column, there might be ITPs nearby, and there should be at least one mooring, but I'm not sure of their time coverage.

The other thing that really surprised me is that you want to investigate a plume dynamics, have a 3-day product available… and downgrade it to monthly resolution. And then regularly show results in the manuscript that you explain with the poor resolution of your monthly product. Use the 3-day version.

And regardless of what you do, you need to say why you do so. Maybe you had a perfectly valid reason for using CTD casts and downgrading everything to one month, but you did not write it so the reader cannot know.

A table similar to that detailing in-situ data used will be added for the reanalysis and satellite products to clarify all products used, their spatial resolutions and time-periods.

The section detailing in-situ data used will be updated to more accurately reflect the range of in-situ data used for validation (not just CTD casts). The SWERUS (Oden) cruise CTD data will no longer used. NABOS underway data will be considered for use instead of CTD casts from the 2018 cruise. Table 1 will be updated to reflect these changes.

ITPs from the Mosaic cruise (from 2020) were considered for use but were >84N (and notably away from the shallow Laptev Shelf and the plume). The NABOS moorings (2013-2015, 2018-2021) were also considered for validation but are realistically too far off the shelf and too deep (minimum depth 30m) for validation of the main plume.

The reasoning behind the choice of data will be better clarified. The shift towards one main objective of understanding plume interannual variability, and the de-emphasis on validation should also help to clarify the choice of data.

III. Causation is not shown

This is my main issue with your manuscript. You do not demonstrate causation. You produce two composites, and declare that the variable you composited against explains the differences.

Let's start with the definition of the composites. I agree that on Figure 2, the circulation is different. However on Figure 4, the uncertainty is so large that some of the strongest years could in fact have a value with either sign. See for example 2012 and 2019. As rotation is involved, a metric based on the curl of the wind, or simply on the sea level pressure, may be more effective and robust.

Anyway, the outcome of the composites is that the SSS looks different. But so do the sea ice and the SST, which could both explain the SSS pattern, and even be responsible for the wind differences. Or maybe wind and SSS are both the result of another variable that is not included in your analysis.

What you are really showing is that the hypothesis that wind drives SSS is not incompatible with the observations. But as your analyses currently are, you do not demonstrate the causality. One option is therefore to just rephrase everything, removing all mentions of "the wind drives" and saying what I just wrote. But that's rather underwhelming a result.

Instead, you could utilise the 3-day product in its full 3-day glory. For each point, or for the overall region, do a lagged (temporal) correlation analysis of the relationship between wind (available at daily resolution; downgrade to 3-day) and the 3-day SSS from BEC. Then see which variable drives the other, based on these values. I personally would push the analysis and perform the same calculation also with the SST and the

sea ice, and have an overall result matrix that shows for each pair of variables for which lag the correlation is maximal and what that correlation value is.

To clarify, the grey overlay in Figure 4 is the maximum and minimum wind stress over the 4 month period (not the uncertainty). However, this comment encouraged us to rethink the 4-month metric we were using.

To better demonstrate causation, a lagged correlation analysis was run between ERA5 eastward turbulent surface stress in individual months and GLORYS SSS in September (see attached figure). The results of this analysis caused us to change to a 3-month metric (July-August) (rather than the 4-month metric previously being used). Given the highly variable nature of eastward wind stress in September, this helps to clarify Figure 4 by decreasing the range of wind stress overlaid in grey and supports that eastward wind stress is the (or at least a) dominant driver of SSS.

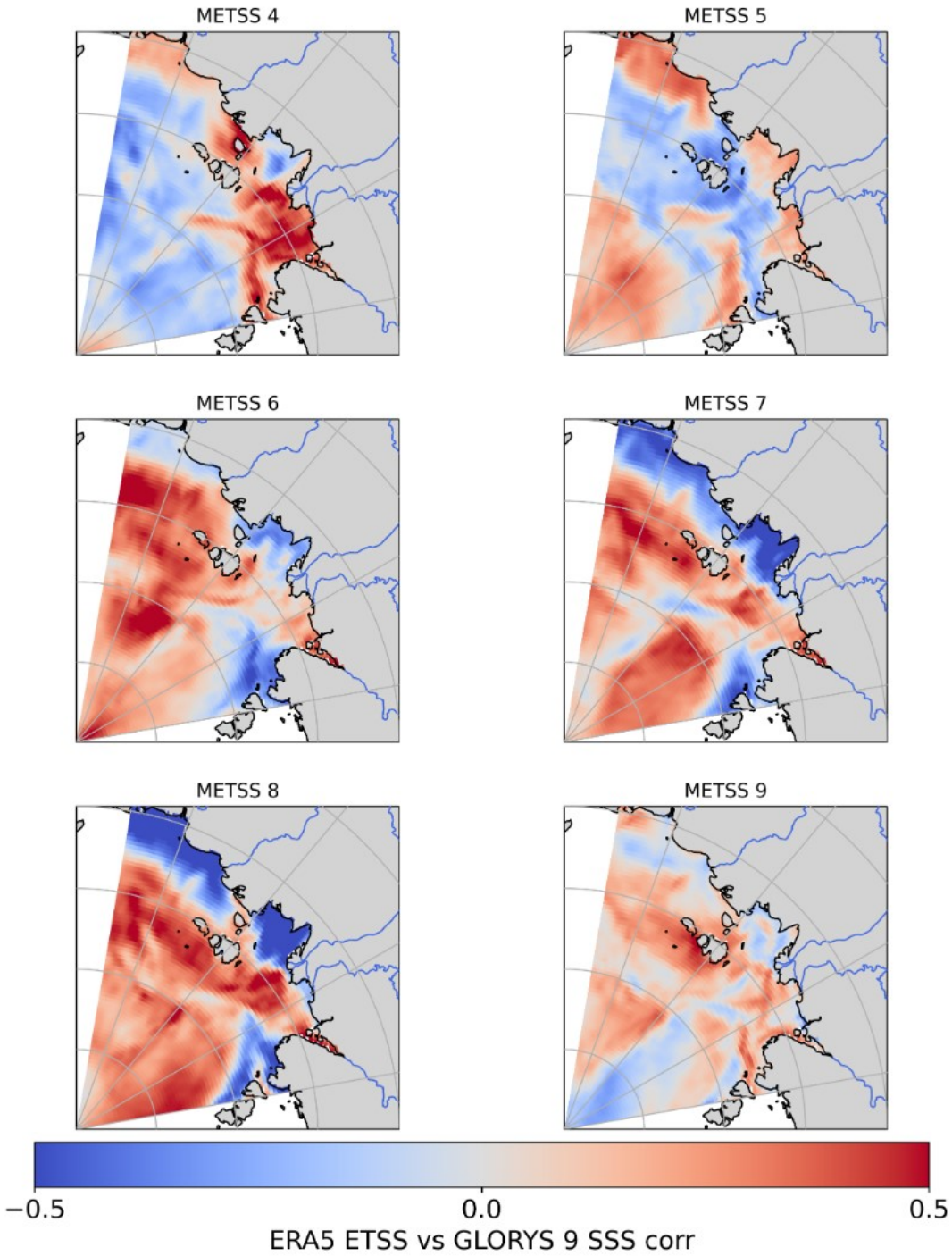

Figure 1: Lagged temporal correlation between ERA5 eastward turbulent surface stress in April-September (4-9) and GLORYS September SSS

Similar lagged-temporal correlation analysis are being conducted between eastward turbulent surface stress, SSS, SST and sea ice concentration.

Less major comments, in order of appearance

Salinity unit: You meant ”psu”or really”pss”? Oceanographers use the absolute salinity in g/kg now.

We will alter the paper to not use any salinity unit.

We avoided the use of “psu” following the guidelines of the TEOS manual ”Note that Practical Salinity is a unit-less quantity. Though sometimes convenient, it is technically incorrect to quote Practical Salinity in “psu”; rather it should be quoted as a certain Practical Salinity “on the Practical Salinity Scale PSS-78”.” (Intergovernmental Oceanographic Commission et al., 2015).

Subsection 2.1.1: not detailed enough. Either there or in the introduction, you need to be more specific about which satellite measures at which band, has which footprint, repeat time, etc. Something similar to the first paragraph of section 2.1 of these people: https://tc.copernicus.org/articles/12/921/2018/tc-12-921-2018.pdf , but for all sensors (at least SMOS and SMAP, and then explanations about how the different products combine them).

The shift in objective away from the focus on validating satellite SSS data also shifts the focus away from the specific concerns/issues raised here. Nonetheless, we will consider how to include additional details in the revised manuscript where necessary and based on the updated content.

Subsection 2.2.1: also not detailed enough; what do you mean by ”correlation” and ”RMSD”? I assume that you took all points available, regardless of location and time, and basically did a regression? Given that the plume is both time and space dependent, as shown on your figure, I would recommend you verify the temporal and spatial accuracy separately. You may need more points for this, agreed, but see Major comment II.

The metrics calculated will be better defined in methods.

Section 3.1: You do not show the RMSD. See my previous comment anyway, but that could be added to the table -which could be shortened once the manuscript is more focused, see Major comment I.

If the RMSD values are mentioned in text, the table containing RMSD values will be included in the Appendix.

Sea ice: Line 377 onwards you give statistics of the sea ice area, without specifying over which region. Overall, and based on the figures shown in the manuscript, the area does not seem to matter as much as the southern / eastern extent. I would rather use such extent, if doing the correlations suggested above. If not, then do not even quantify it; your maps are very clear.

The region over which sea ice area is calculated is specified in the Methods section : “The GLORYS12V1 sea ice area (SIA) in September in the Laptev Sea (defined to be between 120-145 °E and 68-85 °N for the purpose of calculating SIA) is calculated from GLORYS12V1 sea ice concentration for all years used in the (eastward and westward) composite analysis. The mean “eastward” and “westward” SIA is then calculated as the mean of SIA in the three most eastward and westward years respectively.”

Runoff: Line 405, the discussion starts with an analysis of the correlation between runoff and SSS. It is not specified, but I assume that the runoff values are published elsewhere? If so, reproduce them here, and do a proper (lagged) correlation analysis.

Correlations with runoff will be considered for inclusion in the next manuscript version. Interannual variability in GRO runoff values and their (cor)relation to variability in sipatial pattern of SSS was considered early on in this analysis but no strong relation was found. In addition, the GLORYS12 reanalysis is forced with monthly climatological runoff values but manages to replicate variability in the spatial pattern of the plume relatively well, suggestting interannually varying runoff is not needed to replicate the spatial pattern of plume propagation.

Arctic Oscillation: Same comment, no information about where the AOI comes from and the correlation analysis is not shown.

The source of Arctic Oscillation data will be clarified in the methods section.

References:

Intergovernmental Oceanographic Commission, Research, S. C. on O., and Oceans, I. A. for the P. S. of the: The International thermodynamic equation of seawater – 2010: calculation and use of thermodynamic properties. [includes corrections up to 31st October 2015] ., UNESCO, 2015.