From the response to reviewers, it looked like all my comments had been carefully addressed. This was not the case in the manuscript, which, besides, had many formatting and other obvious errors, leading me to wonder whether I received the correct version. Feel free to ignore some of my comments below if indeed, this is simply an upload/version error.

Apologies, indeed the version submitted was missing multiple applied formatting changes and typo corrections that were lost in version control.

Major comments

Section 2 is still not detailed enough, which makes the entire paper unclear. That is, one cannot reproduce the analyses with the few information provided in section 2. For example, I had to wait until section 4 to understand that the correlation analyses had been performed with the complete time series of GLORYS12V1, not just the 10 years that overlap with SMOS/SMAP – the only time series shown. And I am still unsure whether the correlation AOI – wind stress in section 3.3 was performed on the only 12 points shown or on a longer series.

Solutions: Have all the products in table 1, not just the SSS ones, and make table A1 more detailed. In fact, alongside table A1, consider having a supplementary map or showing the most relevant cruise tracks on Fig 1. Then, for every single figure, table, and result described, say very clearly which time period you are using in the caption / in the panel titles.

The time period over which the lagged correlation analyses are calculated (the full GLORYS timeseries) has been clarified in the methods section, when the lagged correlation analysis is first mentioned.

The time period used to calculate correlations between spring runoff, eastward turbulent surface stress and the AOI has also been clarified in the methods sections. Previously, the correlation coefficients quoted were calculated over the satellite (2010-2022) timeseries. They are now calculated over both the satellite timeseries (2010-2022) and over a longer timeseries (1993-2022).

Where they were not already present, the time periods relevant to each figure have been added in the associated figure caption. When discussing results relevant to a certain time period in text, the relevant time period used has been re-stated.

A table detailing all the reanalysis products used has been added to the methods section detailing reanalysis products.

A map depicting the in-situ data used for validation of satellite and reanalysis products has been included in the appendix.

Also, now that you show more clearly the performances of the various reanalyses: Why GLORYS12V1? In tables A2 and A3, GloSea5 and ORAS5 clearly outperform it over the two satellite time periods.

GLORYS12V1 was primarily chosen as it has been widely used in previous studies and hence its properties are well established in the Arctic and North Atlantic (Lellouche *et al.*, 2021; Verezemskaya *et al.*, 2021; Huang *et al.*, 2023).

One of the primary drivers of comparison with reanalysis was to be able to investigate stratification dynamics to compliment the view of the surface obtained from satellite data. GLORYS12V1 has previously been used for mixed layer studies (Hordoir *et al.*, 2022) and was the only model considered which regularly had mixed layer depths <10m (and therefore was not entirely well-mixed everywhere on the shelf).

Section 3.2 is better (even though that is not what I meant by lagged correlation, but never mind). The typos in the figure captions confused me for a moment: Figure 5, second sentence should read SST, not SSS (same on A2, appendix B); Figure 6, SIC, not SST. And latest in that section, but ideally already in section 3.1, you ought to show GLORYS's SST. You describe it a lot, use it regularly, even have section 3.4 dedicated to it – show, similar to Figure 3, how it performs in 3D.

These typos have now been corrected.

A similar figure to Figure 3 for SST has been added in the appendix. A more in depth analysis of SST is not done as SST is highly variable in this region at this time of year (of over 5 degrees between September and October). Therefore, as is mentioned in text "SST, and in turn stratification in temperature also vary considerably over the course of September, so a higher temporal resolution analysis would be needed for investigating temperature stratification dynamics."

One last analysis comment: Especially relevant based on what you discuss in section 4.4, why do you not perform any water mass analysis? Or at least properly verify the correlation between SST and SSS instead of just discussing similarities in the maps? This is the Ocean Science journal after all: Show us a T-S diagram, colour-coded by years.

Whilst some of this analysis has been conducted (correlations between SSS and SST were calculated and a T-S diagram was plotted), given the complexity of the relationship between SSS and SST in this region, further analysis is needed to interpret these, which is beyond the scope of this paper.

Minor comments

Overall, the paper remains too long and repetitive. With the exception of section 2 (see major comment), every sentence should be examined, whether it brings anything new and important to the story assessed, and if necessary be deleted. You could start with the undeserved superlatives, such as:

Line 245: 0.79 is not "notably" lower than 0.86, it is lower.

Lines 480 and 482, you do not provide a "complete" analysis; you are looking at the interannual variability of the monthly September values over a short time period.

Line 504: 0.49 is not "strongly" correlated.

Line 553: the gradient is not "clearly" visible (since you do not show the SST).

Line 616: the difference is not "considerable", at least not in the common usage of the term.

The superlatives directly mentioned (and a number of others) have been removed. Several sentences have been shortened / removed to try and decrease repetition and help shorten the paper, whilst still ensuring the clarifications raised in review are also accommodated.

Other minor comments, in order of appearance:

Line 106: Your introduction remains long and a bit unfocused. Help the reader by stating here "In this manuscript, we [objectives]"

This phrasing has been added where the objectives are detailed, at the end of the introduction.

Line 124: Since you already refer to the in-situ data there, consider swapping the subsections and starting with 2.1.3.

The data products section now begins by describing the in-situ data, then the reanalysis data then the satellite data.

Line 126: As it currently reads, it still feels like one of the objectives of the manuscript is to determine which reanalysis is best. Clarify this better, and refer to the appendix.

This line has been altered to reflect the aim of comparing in-situ and satellite SSS with reanalyses, rather than a reanalysis intercomparison.

Line 147 and throughout the manuscript (the usage of pss): I see your response, and would like to redirect you to the instructions you say you are following. Put more clearly: When working with practical salinity, option 1 (less correct) is to write psu as a unit; option 2 (most correct but rarely done) is to say at the very beginning of the manuscript that "salinities are quoted on the practical salinity scale" and then not write any unit. So on line 147 this would read as "SSS uncertainty is lower than 1".

As per the response to both reviews, "We tried altering the paper to not use any salinity unit but found this to be unclear. Whilst the difference between absolute and practical salinity is negligible, especially compared to the uncertainty in salinity in this region, both satellite and GLORYS12V1 SSS are provided as practical salinities. Therefore, we choose to use units for practical salinity rather than absolute salinity. 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)."

Line 228: "anomalous" is the wrong word. They would be anomalous if they were both compared to the mean of figures A3+A4… which maybe is what you mean, but in this case you have to say it and say over which time period considered. I rather suspect you mean that they are clearly different from each other.

The wording has been changed to "notably different" rather than "anomalous" to reflect that this sentence is comparing the two years from each other rather than differences from a climatology.

Line 230: "interannual variability" is not the correct word – again, I suspect you simply mean differences, unless you meant to add a reference to figures A3 and A4.

The wording here has been changed.

Line 237-241: Same comment as on the previous version. You have a 3-day product. You can verify this hypothesis for the satellites vs in-situ data.

The wording of this line has been changed to reflect that satellite data clearly aligns more closely with in-situ data when using a daily satellite product for the relevant day. This is very clear in 2019 (where there is sufficiently abundant in-situ data which nicely captures the entire salinity front) and visible, but less clear, in 2016.

Line 245 and Table A2: There are some very high correlations in there. Did you verify that not outlier is skewing the correlations?

Yes, scatterplots were generated to ensure correlations are not influenced by the presence of outliers. In the case of BEC, the correlation deteriorates for values <25 pss.

Line 255-259: Add a reference to the sea ice contours shown on Fig 2.

A reference to Figure 2 has been included here.

Line 290: "bottom waters", wrong word. Bottom waters are a specific water mass, found in the Arctic deeper than 2500 m. Here you mean "fresher water, below the surface", or "in the subsurface".

This has been changed to "fresher subsurface waters."

Line 313: "Vilkitsky Strait"; mark this location on Fig 1.

The location of the Vilkiysky Strait has been added to Figure 1.

Figure 7 and line 380: Same comment as on the previous version, even though you said you changed this in the response. With the uncertainty, 2012 could be in either direction. What is the impact on the results of not including that year?

The JPL composite does not include 2012 and gives similar results (but does include 2021). If both these years (2012 and 2021) are excluded from the eastward composite, the SSS pattern remains very similar. The composite of just 2016 and 2017 is almost identical in the Laptev Sea (to the 3-year composites) but the low salinity signal that extends around the coast in the East Siberian Sea is stronger in the 2016-2017 composite, as it is particularly strong in 2016.

As per the response to both reviews "Note: In Figure 7 (previously Figure 4), the grey overlay is the maximum and minimum wind stress over the 4 month period (not the uncertainty)."

Figure 7: The AOI standardised by the standard deviation of the surface stress is a strange measure. Why not have a second y-axis to the right, or even using the same axis as the runoff? The values should be in the same range (-4 to 4)

A second y-axis has been added to this figure for the AOI.

Lines 419-421: Same comment as on the previous version. These area values are meaningless without more clearly defining the regions, and are anyway not useful for the analysis. Remove this sentence.

These area values have now been removed.

Line 473: You should show the full runoff timeseries somewhere, so that the reader can see whether the satellite period is anomalous.

The satellite period is not particularly anomalous (Figure 1). This has now been stated in text in the results section, where the shorter spring runoff timeseries is presented.

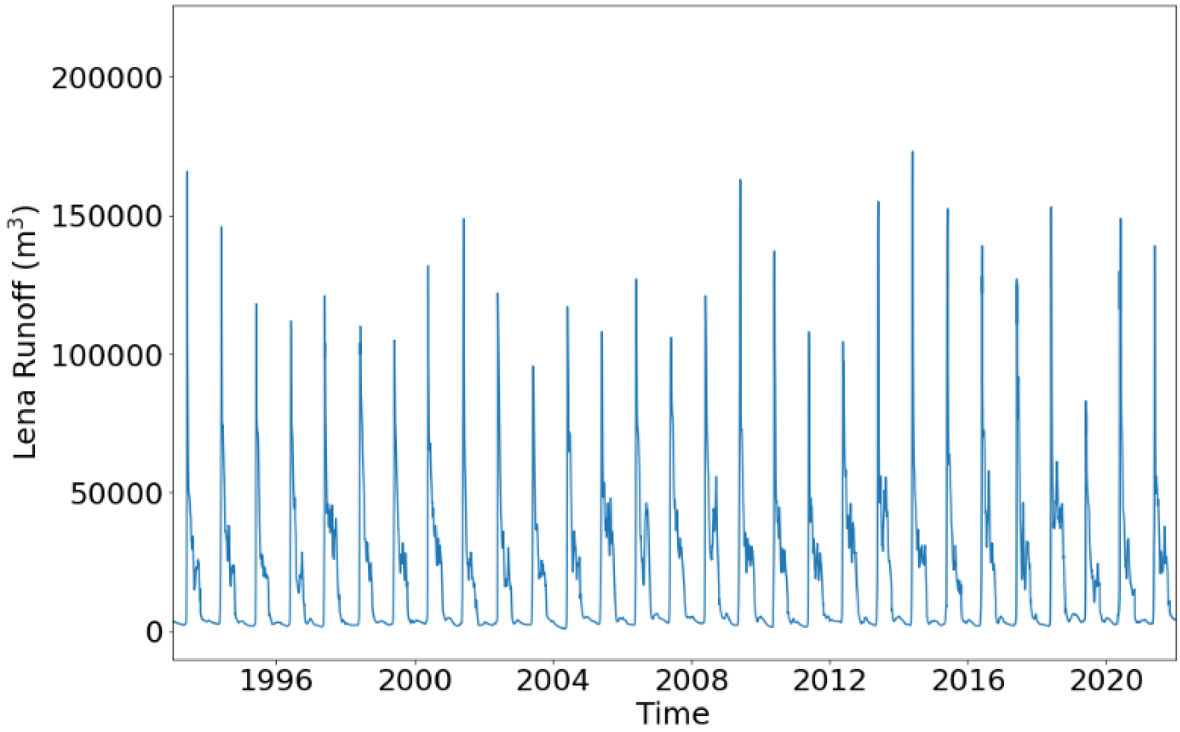

Figure 1: Timeseries of Lena river runoff over 1993-2022

Whilst we chose to not show the full runoff timeseries in text, to help not lengthen the paper further, GLORYS12V1 correlation coefficients are all calculated over 1993-2020. In addition, we now include correlation coefficients (between spring runoff and eastward turbulent surface stress and the AOI) over a longer 1993-2022 timeseries, as well as over the short 2010-2022 satellite timeseries. The relative consistency of correlation coefficients over these two time periods, helps to show that spring runoff in the satellite period is relatively consistent with earlier variability in spring runoff.

Lines 493-495: Same comment as on the previous version. You can calculate the angle between the wind and the coast. You choose not to, for some reason. Ok, but in this case, you cannot write such a sentence.

This sentence has been altered to not discuss the angle of plume transport from the wind.

Lines 530-532: What do you mean? You have 30-years worth of data, you could verify this. Either don't write such a sentence, or say why you would not do the test.

In the following paragraph, we explain that this hypothesis is not tested because "the constant well-mixed plume nearshore suggests GLORYS12V1 is not capable of fully representing plume stratification dynamics in this complex environment." In order to test this hypothesis, you would need a model that appears capable of accurately representing interannual variability in the mixed layer.

Line 536: TOPAZ is the only one built on HYCOM; the other four are NEMO-based. It is not surprising that they suffer from the same bias when 4/5 are basically variations on the same model.

The aim of this paper was not to do a reanalysis inter-comparison, so an extensive product comparison was not done and other products were not considered for use. It would be interesting to see how a wider array of models perform in this region (in particularly if any manage to capture both the surface interannual variability and the variability in stratification) but this is beyond the scope of this study.

Line 541: Not fishing for references here, but yes, that is a known issue with models: The plume needs to move horizontally and vertically with this type of vertical grid (z-level), so even if it had perfectly accurate properties at the beginning (which it does not because of other biases), every time it gets to another grid cell it is strongly mixed with the ambient water, so you lose the signal quickly. Likewise, the issue of the too-well mixed shelf in z-level models has been reported on repeatedly. I would actually expect TOPAZ to perform better, since it is based on a sigma-level model, whereas GLORYS is z-level.

This paragraph has now been altered to reflect that the z-level vertical grids in the models used here are a likely cause of the overmixing issue, based on inclusion of a relevant paper that was previously missed.

The initial plan was to use TOPAZ for this work but, at the time, the version available had only 28 levels, the shallowest of which was 5m (which was decided to be too deep for comparison with satellite SSS). Both this version and even in the more recent TOPAZ version with 40 levels, TOPAZ does not capture interannual variability of the surface plume visible in satellite or SSS data, and appears to also have an overmixing issue, so we chose to use GLORYS12V1.

Line 588: Same comment as on the previous version, you are not showing the "initial" plume propagation. This would require that you use higher temporal resolution data.

This word "initial" has been removed.

Lines 590-592: Same comment as on the previous version, you cannot say who is a "dominant control" on what if you do a same-month correlation. Besides, you yourself right after seem to hint at the possibility that the correlation goes the other way round (less ice = more ocean exposed = warmer ocean).

The beginning of this paragraph has been altered to clarify that although there is clearly strong correspondence between SST and SIC, neither can be assumed to be the dominant driver of the other given work shown here, and that it is likely that they both feedback on each other.

Line 651: Which "increase in correlation strength over the more recent time period"? You either need a reference, or to remind the reader where you showed that (but you did not, right?).

The increase in correlation strength was previously only discussed in the discussion section 4.2. These correlation coefficients have now been included in the results section, alongside the correlation coefficients for 2010-2022 (and the methods section has been updated to reflect this).