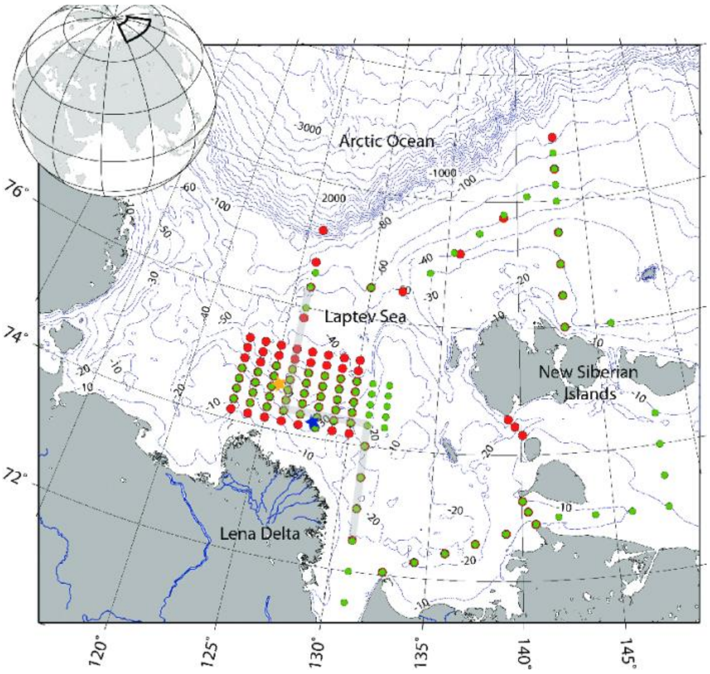
REVIEWER'S COMMENTS and ANSWERS

Reviewer #2

General comments

C1. My primary comment relates to the simulated trajectories used to support the proposed origin of the freshwater anomaly. It is well known that different ocean models can produce substantially different spreading pathways for freshwater and water masses in the Arctic. As I am not a modelling specialist, I cannot confidently assess the reliability of the specific simulations presented here. However, the authors themselves note (Line #57) that numerical models often struggle to accurately reproduce freshwater redistribution. A key limitation of the trajectory figure is the absence of information about flow pathways at different depths. The observed freshening signal is clearly detectable at least within the upper ~150 m. Therefore, it would be highly informative to show simulated trajectories at several representative depths (e.g., 50, 100, and 150 m), rather than only at the surface (0–5 m). In particular, do the modeled pathways at these depths indicate transport through Vilkitsky Strait? If not, and if the simulations instead suggest that water at these levels primarily arrives from the northwest along the continental slope, this would weaken the argument that the entire freshening signal originates from Kara Sea river discharge. In that case, the key conclusion stated in Line #19 of the Abstract may require qualification. At present, the anomaly is traced in the surface 0–5 m layer only, whereas the origin of the anomalies observed at greater depths may have been different.

A: Extensive mooring and shipborne observations and reanalyses data used in our study suggest that salinity anomalies are surface-intensified and originate from Siberian River mouths. As a result, anomalies associated with anomalous riverine discharge initially form as a thin surface layer near the river mouths. They are then advected toward the mooring locations as a shallow freshwater layer, where they are subsequently mixed into the ocean interior. As the attached map (sourced from Google, ignore color dots, see depths near Vilkitsky Strait) illustrates, the shallow bathymetry limits vertical mixing to greater depths along most of the pathway, such that significant deepening of the signal is unlikely until it approaches the continental slope and the mooring sites. Therefore, in the original analysis, we performed Lagrangian tracer simulations for the surface layer (0–5 m).



However, following the reviewer's suggestion, we conducted additional tracer simulations for layers deeper than 50 m. These simulations (not shown in the revised text) suggest that tracers are predominantly advected along the continental slope, from the areas west of the eastern Eurasian Basin with climatologically much higher salinities, highlighting the essential role of waters of the Atlantic origin in shaping water structure at the intermediate layers. Thus, we think that our conclusion regarding the origin of the observed freshening is robust.

C2. I also share the concerns raised by the other reviewer regarding the analysis intended to demonstrate a potential cross-slope shift of the Atlantic Water core. In its current form, this approach is not sufficiently convincing. First, the method does not account for the vertical nonlinearity of either salinity or current velocity, both of which may significantly affect the resulting estimates. To obtain physically meaningful values, the quantity $V\Delta S$ should be calculated over smaller vertical intervals, and only thereafter integrated to derive a depth-averaged effect. Averaging first and integrating afterward may obscure important vertical structure and lead to misleading conclusions. Second, I believe there is an issue in the interpretation of the calculated values presented in Table S2. The table provides estimates of $V\Delta S$ between adjacent mooring pairs. However, these values represent the lateral salt flux between the moorings, not the divergence or convergence of salt flux at a given mooring location. For example, to determine whether advection-driven freshening or salinification occurs at mooring M12, one would need to compute the difference between the fluxes across M11–M12 and M12–M13 (−0.32 and +0.25, respectively). This difference would indicate the sign of the expected salinity tendency. However, such an analysis is not feasible for the boundary moorings (M11 and M15), as there is no information on lateral salt flux outside the array. Consequently, convergence or divergence cannot be estimated at the array edges. In my view, this part of the analysis currently introduces more ambiguity than robust insight. I would therefore strongly recommend either substantially revising this section with a physically consistent framework or considering its removal. In its present form, the approach does not allow one to assess whether lateral shifts occurred upstream at different depths, nor how such potential shifts might have contributed to the observed salinity anomalies.

A: We have revised this section to address the reviewer's concern. In the revised text, we first provide a holistic description of our approach, followed by a more formalized formulation. In response to the concern that the method neglects the vertical structure of currents, salinity, and fluxes, we note that the analysis is intentionally highly simplified. This is motivated by our objective, which is not to quantify the exact contribution of individual terms to potential cross-slope shifts in the AW jet, but rather to assess whether this mechanism could plausibly explain the observed widespread freshening in the eastern Eurasian Basin in 2015–2016. We therefore consider the chosen approach appropriate for this purpose. At the same time, we agree that robust estimates require considering flux divergence rather than fluxes alone. We have accordingly updated Table S2 and revised the text to base the analysis on flux divergence. Thank you for this helpful suggestion.

C3. Figure 11 is arguably one of the strongest figures in the manuscript. However, the reported significance level of the correlation coefficients ($p < 0.001$) appears unexpectedly high. I encourage the authors to carefully verify these calculations. Such a strong level of statistical confidence warrants explicit clarification. In addition, panel (b) does not seem essential to this figure. The rectangle indicating the Kara Sea region could instead be incorporated into Figure 1, which would improve the overall clarity and reduce redundancy.

A: We have identified an error in the estimation of the p-value for the top panel ($M1_1$ mooring) and corrected it. For the given length of the monthly time series and the reported correlation values, all other p-values are correct. We have made changes adding the rectangle in Fig. 1.

C4. I also find the proposed impact of oceanic heat flux on sea ice concentration (SIC) to be insufficiently substantiated. The correlations shown in Figure 14b and 14c appear to primarily reflect the seasonal cycle, rather than demonstrating a robust physical relationship between SIC and velocity shear. I refer the authors to my detailed comment on Line #403 for further elaboration.

A: See our answer to the Reviewer's specific comment (l. 403).

Despite the criticisms outlined above, I view this as a high-quality and potentially impactful contribution that is suitable for publication in Ocean Science. The manuscript is generally well organized, presents a unique and valuable observational dataset, and develops several convincing lines of evidence. Nevertheless, there remain a number of issues that require clarification or revision to strengthen the robustness and interpretation of the results. Below, I provide a detailed list of specific comments and suggestions.

A: We thank the reviewer for his time and effort in evaluating our manuscript. We hope that the revised version addresses most (if not all) of the reviewer's concerns.

Specific comments:

Q1. Line #68:
All river-driven freshenings are seasonally episodic by nature. It would therefore be more appropriate to refer here to an anomalous freshening event rather than implying episodicity as a distinguishing characteristic.

A: Valid point. We have corrected this to "anomalous."

Q2. Line #87 (Table 1):
There are several formatting inconsistencies and potential errors in Table 1 that should be carefully checked.

- Some dates contain odd empty spaces (e.g., "26 .08. 2013").
- M12a lists "70–754" (MMP?), but according to Figure 2 these data appear to be unavailable.
- M11a and M11b are shown with identical SBE37 depths, whereas Figure 2 suggests that the shallowest instrument at M11a was deployed deeper.
- The coordinates of M15a and M15b are identical to the last digit, which may indicate a mistake.

In addition, the MMP on M15 is reported to have profiled between 88–754 m and 172–806 m. However, Figure 2 shows data availability beginning at approximately 50 m and 45 m. Based on Figure S3, it appears that during the first deployment the MMP started from at least ~77 m (not 88 m as indicated in the table), and during the second deployment from approximately ~185 m (rather than 172 m). Figure S3 also suggests that a MicroCAT may have been deployed at ~77 m during the second deployment.
I strongly recommend carefully verifying all information presented in this table and ensuring full consistency with the data shown in the figures.

A: We have corrected the table: all spacing issues in the dates have been removed, several instruments that were missing in the original version have been added, and deployment depths have been corrected. In particular, the $M1_5$ mooring information previously omitted two SBE37 instruments, which explains the reviewer's question.

Q3. Line #90-93:
This statement may be valid for the MMP profiles specifically, but it does not appear to hold for the discrete sensor measurements at other moorings. Based on Figure S3, I have significant doubts that the vertical sampling configuration at moorings such as M11 and M14 would result in only a ~1% error. The vertical distance between the shallowest two loggers at these sites is substantial and could lead to considerable under- or overestimation of freshwater content if linearly interpolated. The authors should clarify how that small potential error could be transferred to the moorings with measurements at discrete and sparse levels.
A: We have removed this statement from the text, as it could not be independently verified.

Q4. Line #94:
It would be useful to include information on ADCP instrument depths and measurement ranges in Table 1 to improve clarity and allow reproducibility.

A: We added this information.

Q5. Line #96: "*MMPs profiled from ~40*"
This statement is inconsistent with the data presented in Table 1 and should be corrected.

A: We corrected this sentence.

Q6. Lines #149-153:
Using wavelet analysis to extract and remove the seasonal signal does not appear to work consistently, as seen in Figure S1. In many locations, the seasonal signal appears artificial and potentially misleading. Conceptually, the seasonal component should correspond to a 12-month period. While this periodicity is reasonably clear at moorings M11, M12, M13, and M3, it is not evident at the other two moorings.

I suggest the authors reconsider the terminology used to describe these anomalies. Rather than labeling them strictly as "seasonal-adjusted" or "anomalies," it may be more accurate to describe them as the high-pass component of salinity, with the low-frequency (periods ≥12 months) signal removed via wavelet decomposition. My concern is primarily terminological rather than methodological. Additionally, I note that the red line in the upper panel of Figure S1 is absent before 2015.

A: We tend to disagree with the assertion that wavelets identify an "artificial" seasonal signal. Rather, we interpret the seasonal signal as inherently variable in time and space (as the reviewer also noted). In our experience, wavelet analysis is a powerful tool for decomposing time series into their constituent components. We also disagree that the method applied here constitutes low-pass filtering; instead, it is more appropriately described as a band-pass filtering approach. However, we agree that the original sentence (ll. 151–153) was not sufficiently accurate and have revised it accordingly. We also added a red line to highlight the early part of the M11 record.

Q7. Line #179:
In the integral, dz represents the differential element of depth, not the layer thickness. It would be clearer to simply state that z is the vertical coordinate.

A: We have removed this statement regarding the definition of dz, as it was not necessary.

Q8. Line #187: "*Here, Q can be defined as the relative heat content, which measures the amount of heat that must be removed to create ice crystals at a given salinity and pressure.*"
It would be more accurate to define Q as the heat that must be removed to lower the water temperature in the entire layer to its freezing point at a given salinity and pressure. The phrase "to create ice crystals" implies removing additional latent heat beyond this amount. For consistency with the freshwater content (FWC) formula, consider adding a subscript *z* to *T* (and possibly ρ) in the ocean heat content calculation, and replace the subscript *i* with *z* in the APE formula.

A: The sentence is edited.

Q9. Line #210:
It would be helpful to clarify in the text that the Kara Sea data are derived from the model, while the Laptev Sea data come from the mooring array, to avoid potential confusion.

A: Added.

Q10. Line #228 and Figure 2: "*... and the freshening extended to depths up to 175 m ...*"
The statement that the freshening extended to depths of up to 175 m requires clarification. During the early-2016 surface freshening, MMP measurements indicate that the anomaly was mainly confined to the upper 60 m at M15 and 70–80 m at M13. On other moorings, the apparent deeper extension (~110 m at M11, ~120 m at M12, ~100 m at M14) results from the discrete depth levels of the sensors (53 & 140 m, 67 & 138 m, 38 & 107 m, respectively) and the linear interpolation between them.
Figure S3 illustrates this effect clearly: interpolation can artificially extend the freshening signal to greater depths, particularly at M11, M14, M15 (where MMP started at 188 m), and M3. This is not a critical problem, but the text should at least mention these potential artifacts. In particular, it should be noted that the calculated freshwater content (FWC) carries additional uncertainty at moorings with sparsely spaced MicroCATs. Therefore, the previously stated maximum error of 1% may not apply to these sites.

A: We agree and revised this paragraph.

Q11. Line #229:
It would be clearer to write "*evolution of the spatial (or cross-slope) pattern of salinity and temperature anomalies*" to better convey the intended meaning.

A: We edited the sentence.

Q12. Line #237: "*strengthening the vertical density gradient.*"
It is preferable to use either "*strengthening the vertical density stratification*" or "*increasing the vertical density gradient*".

A: We edited the sentence.

Q13. Line #238: "*This enhanced stratification was reflected in the pronounced increase in estimated available potential energy across all moorings*"
There appears to be a conceptual issue here. Based on the formal definition, APE is the portion of the total potential energy of a stratified fluid that can be converted into kinetic energy. It represents the energy associated with deviations of density surfaces from a state of minimum potential energy. Critically, the correct calculation of APE requires a reference density profile corresponding to the minimum potential energy, which is obtained by adiabatically sorting all vertical layers by density. In this framework, APE is zero when there are no vertical density inversions. It definitely does not use a constant reference density in the formula!
From the description in the manuscript, it appears that the formula currently used may not represent true APE. Instead, it seems to calculate the energy required to mix a given water column into a fully uniform state with some mean (reference) density. This distinction is important, as the two quantities represent fundamentally different physical processes.
The authors may wish to clarify this point and, if necessary, revise the text and formula to accurately reflect the concept of APE.

A: We thank the reviewer for this important clarification. We agree that the formal definition of available potential energy (APE) requires the construction of a reference state obtained by adiabatic sorting of the density field, as originally formulated by Lorenz (1955). In the present study, we do not compute this full APE. Instead, we use a linearized diagnostic metric based on density anomalies to characterize changes in stratification strength. This metric is proportional to the small-amplitude approximation of APE under the Boussinesq assumption and constant background stratification, and therefore should be interpreted as a proxy for stratification energy rather than a rigorous estimate of Lorenz APE. We have clarified this distinction in the revised manuscript.

Q14. Line #259 (Figure 3):
I have a concern regarding the placement of M3 in this plot. Although M3 is slightly closer to M15, it may be more intuitive to position it on the left side of the figure. From both a topographic perspective and considering the general flow path along the slope, it would visually align more naturally with M11. Doing so would likely reduce the apparent contrast in temperature and salinity between M3 and M11, which currently appears stronger than the contrast between M3 and M15.

A: Our rationale was to (1) show all cross-slope moorings $M1_1$–$M1_5$ without interruption, (2) include the M3 records, and (3) place them as close as possible to the most comparable cross-slope moorings ($M1_3$ and $M1_4$, given their similar depths). It is evident that the M3 signal shares many similarities with those observed at these two moorings. Therefore, we prefer to keep the figure unchanged.

Q15. Line #273 (Figure 6):
The caption does not specify the depth range over which the mean salinity was calculated, and this information is also absent from the main text (Lines 245–247). Although it can be inferred from the winter panels, it would improve clarity to explicitly state the depth range in the text. Additionally, it is unclear why the summer and winter data are presented separately. Would it not be clearer to show them as bars (with the same altering seasonal colors) on a single panel for each mooring, allowing direct seasonal comparison?

A: The depth ranges were presented in the original figure. The requested changes have been implemented.

Q16. Line #278:
"The (negative?) salinity anomalies..."

A: Yes, thank you – we have added this word.

Q17. Line #288: "The potential underlying mechanisms for these contrasting signals are discussed in Section 6."
There is no Section 6 in the manuscript. Additionally, it is unclear how ocean heat content was applied or interpreted in the context of this study. Is it worth removing OHC part at all?

A: We edited this sentence.

Q18. Line #309: "*Thus, the 2015-2017 freshening cannot be explained by the advective cross-slope shift of the salty jet's core.*"
See my general comment at the beginning regarding the validity of the approach used to assess cross-slope shifts.

A: See our answer and edits.

Q19. Line #324: "*3.2.2 Kara Sea as a driver of salinity change*"
A geographical entity such as a sea cannot act as a "driver" of change. It would be more precise to refer to the Kara Sea riverine discharge as the driver of salinity variations.

A: We revised the title.

Q20. Line #325: "*we indeed found that the Yenisey and Ob rivers showed an anomalously high runoff peaking between 2014 and 2015 relative to 2013*"
It is difficult to draw robust conclusions based solely on the plotted daily series. The seasonal peaks in the figure may represent single-day maxima rather than the high total (cumulative) discharge for the season. I recommend adding the mean seasonal discharges for each summer (as a text) above the timeseries in Figure 9 to provide a clearer quantitative context.

A: We agree that it is difficult to fully discern the entire story from this plot alone. Therefore, we have revised the text to make the interpretation as clear as possible and to provide sufficient detail for understanding.

Q21. Line #327: "*when spread over the area covered by mooring observations (roughly 77-82°N and 110-140°E; 5 x 105 km2)*"
This appears to be a very strong assumption. The authors should justify why this specific region

accurately represents the domain over which Kara Sea riverine water becomes eventually distributed. The actual affected area could easily be larger or smaller, and the estimated freshening in meters would strongly depend on this choice.

A: To clarify this point, we have revised the text to explicitly state that the region is only loosely defined and is used solely as a representative area for estimating freshwater distribution. This clarification has been added in the revised manuscript.

Q22. Line #327-330:
There appears to be a minor arithmetic or rounding error. The sum of 0.19, 0.16, 0.14, and 0.26 cannot equal 0.78 under any reasonable rounding scheme. Even allowing for maximum trimmed decimals of 0.0049(9) for each member, the total would not exceed approximately +0.02 (that would give 0.77), not +0.03 (for having 0.78). I know this is not a big issue, but it caught my eye.

A: We edited this paragraph.

Q23. Line #349 (Figure 10):
It is unclear whether the presented anomalies are temporal or spatial. I attempted to follow the tracks shown in Figure 13 to trace the modeled spread of freshwater anomalies, but this proved difficult, likely due to the different color scales used in each panel. While varying scales can help highlight distributions within individual panels, they make it challenging to track the evolution and movement of anomalies across panels. Additionally, the zero value corresponds to different colors in each panel, further complicating careful interpretation. Overall, I am not certain that these modeled anomalies add substantial insight to the narrative as currently presented.

A: Following the advice of Reviewer #1, we have removed all but one panel from this figure and modified the text accordingly.

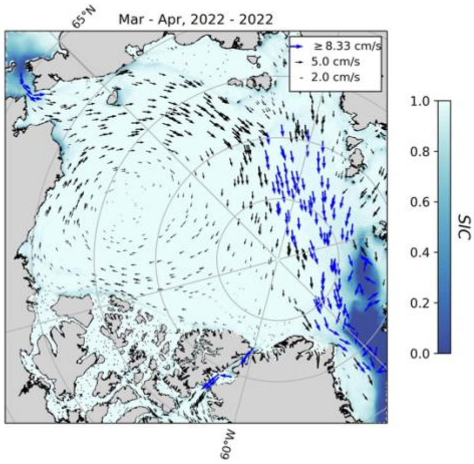
Q24. Line #360 (Figure 11):
For M11, the maximum correlation was reported at a 7-month lag, whereas the graph appears to show a maximum around 22 months, similar to M13–M15 and M3. The authors should clarify why the 7-month lag was chosen. For M12, there is a relatively flat plateau between 9 and 23 months, indicating some uncertainty. I recommend adding confidence intervals (e.g., shaded areas) to each correlation panel to better convey this uncertainty. Overall, it would strengthen the figure and interpretation to acknowledge that a lag of ~22–23 months likely applies across all mooring records.
A minor point: there is a small inconsistency between the text (Line #354, "*all statistically significant at 95% level*") and the figure (significance level p<0.001). If all correlations indeed meet the p < 0.001 threshold, it is acceptable to state 99.9% confidence level in the text.

A: We

Q25. Line #368: "*These changes can alter the direction of Ekman transport, shifting the accumulation of freshwater along the slope rather than driving it into the basin and contributing to the observed freshening buildup.*"
I suggest adding a comma after "*basin*" for improved readability.

A: We have edited this sentence.

Q26. Line #374 (Figure 12):
Please check the ordering and layout of all labels, as some text is currently obscured by overlapping graphics. While not critical, I also recommend improving the figure style. Instead of plotting vectors on a regular spherical grid, consider weight-randomizing their positions with weighting each vector according to the actual area of the corresponding 0.25 × 0.25° ERA5 cell. This approach is relatively straightforward to implement and can produce much cleaner and more visually informative plots (e.g. like an example below)



A: We have corrected labels in this figure. The reviewer’s suggestion may be important for regions close to the pole; however, we find that the presentation of vectors in our domain using a spherical grid remains clear and informative and is not affected by the convergence of meridians. We therefore prefer to retain the original presentation. Note that the figure has also been revised following the advice of Reviewer #1.

Q27. Line #388: “*Instead, the large Yenisey and Ob River discharge between 2014 and 2015 (Fig. 9) was probably the dominant source of the freshening observed in the eastern Eurasian Basin.*”
See my general comment at the beginning regarding the limitations of the approach used to fully attribute the freshening to Kara Sea river discharge.

A: We thank the reviewer for this comment. We have revised the text to clarify that the Kara Sea river discharge is a plausible and likely important contributor to the observed freshening. We hope that in the revised version we made our arguments stronger.

Q28. Line #403: “*It shows that during the winters of 2015 and 2016, around the onset of the freshening event, the divergent heat flux across the halocline decreased from 20 W/m2 to 3 W/m2. This weakening of oceanic heat flux was reflected in the sea ice concentration observed at all moorings during the following summers of 2016 and 2017 (Fig. 15).*”
I am not convinced that the reduction in sea ice concentration can be directly attributed to the decreased oceanic heat flux. The correlation between normalized SIC and velocity shear timeseries is largely driven by the strong seasonal cycle in both variables. The apparent relationship could also result from a third factor, such as seasonal wind forcing, which affects both ice drift and surface currents. Filtering out the seasonal signal would likely reduce the correlation considerably.
If Polyakov (2020b) has already discussed the effect of reduced turbulent heat flux associated with the 2015–2016 freshening on ice cover, then Section 3.3 — at least its latter part — may be redundant.

A: Yes, Polyakov et al. (2020b) discussed this effect in detail. In the present study, we argue that the observed sea ice changes can be attributed specifically to the impacts of ocean stratification and heat release during the freshening event. Therefore, this discussion is highly relevant, and we have retained it in the text.

Q29. Line #404:
The term "*divergent heat flux*" is unclear in this context and should be explicitly defined. It is important to clarify whether it refers to the net upward heat flux across the halocline, a horizontal (vertical) divergence, or another component.

A: We clarified its meaning.

Q30. Line #417 (Figure 13)
I am skeptical that the tracks shown in this figure represent 23 linear monthly segments. The description of the calculation and the reference to Polyakov et al. (2023) do not clarify this. I assume the simulated velocities were interpolated in both space and time with much smaller time steps than one month, which would produce such smooth trajectories.
Regarding the "Pre" and "Post" trajectories: rather than keeping these, it may be more informative to include panels showing trajectories at depths other than the 0–5 m range. The observed freshening in the mooring data occurs over a thicker layer, so it would be valuable to illustrate where the signal originates at, for example, 50, 100, and 150 m depths.

A: The Reviewer is correct that the monthly current velocities were interpolated in time to obtain a much shorter time step. We have added this information to the Methods section. Regarding the inclusion of additional panels at deeper levels, we refer to our response to the Reviewer's general comments, where we explain that the deeper layers are not directly relevant to the main narrative of this study.

Q31. Line #429: "*are statistically significant at 0.05% level*"
I assume the authors intended to indicate significance at p = 0.05 rather than 0.05%.

A: We corrected this sentence.

Q32. Line #439: "*nearly doubling the available potential energy*"
I again have concerns related to the definition of APE and the assumption that freshening and stratification automatically increase APE, as discussed in my comment for Line #238. The authors should clarify whether the quantity they compute is truly APE or rather the energy required to mix the water column.

A: We have clarified this in the Methods section.

Summary:

While I have raised critical concerns regarding individual pieces of evidence, the overall body of observations — correlation analyses, trajectory simulations, and the consistency of anomalies across multiple moorings — strongly supports the main conclusions of the paper. I fully agree with the authors' statement that: "Despite these limitations, the core findings, including the anomalous freshening of the eastern Eurasian Basin and northern Laptev Sea driven by Yenisey and Ob river discharge, and its impacts on sea ice and upper-ocean currents, are robust and well supported by extensive observations." With some targeted clarifications and minor revisions addressing methodological details, figure presentation, and terminology, this manuscript can be considered suitable for publication in Ocean Science. The study

presents a unique and valuable dataset and offers important insights into Arctic freshwater dynamics.
Sincerely, Sergei Kirillov

A: We thank the reviewer for his careful and constructive review, which has significantly helped us improve the manuscript.