First of all, we would like to thank you for all your constructive comments (shown in black), which greatly helped us improve the manuscript. Below, we provide our detailed point-by-point responses (in blue). We have revised the manuscript accordingly. In addition, we carefully reviewed the entire manuscript for typographical errors and other minor inconsistencies and have corrected all such issues in the revised version.

Reviewer #1

This revised manuscript is improved compared with the previous version. The topic is important, the large-scale patterns are interesting, and the authors have clearly made a serious effort to respond to the earlier comments. In particular, the additions on significance testing, uncertainty, and the discussion of frontal migration make the paper stronger than before.

That said, I still do not think all of the key concerns are fully resolved. My overall view is that the manuscript is getting close, but I still have several substantive concerns about reproducibility, uncertainty treatment, and the strength of some of the interpretations.

The first point is that I could not find an explicit statement of which exact GOBAI-O2 version was used. The manuscript cites the dataset and describes the product generally, but I did not see a clear release/version number in the Methods or Data Availability sections. I think this is important for reproducibility, especially because different GOBAI-O2 versions can give noticeably different regional trend patterns (some versions (pre-2.3) even disagree on the sign of trend in oxygen change!). Since the conclusions here depend on relatively subtle spatial oxygen trends, the exact dataset version should be stated explicitly. Ideally the authors should also comment on whether the main results are sensitive to dataset version.

The version, which we used in this analysis, is 4.4. The version number has been added in Section 2.1 of the current manuscript.

"We use GOBAI-O_2 (Version 4.4), a four-dimensional, monthly gridded product of dissolved oxygen (O_2) in the ocean interior, generated using machine learning (ML) algorithms trained on both Argo float oxygen measurements and ship-based discrete observations [Sharp et al., 2023]."

Second, the uncertainty treatment is improved, but I still think the wording is stronger than what is actually demonstrated. The new robustness metric comparing 20-year trend magnitude with uncertainty is useful, and it definitely helps. However, I do not think this is the same thing as formally propagating uncertainty into the trend estimates. As written, the manuscript sometimes gives the impression that this has been done more rigorously than it actually has. I think the current approach is acceptable as a heuristic or screening metric, but the language should be revised so that this is presented more carefully. In particular, the paper should avoid moving too quickly from "statistically significant" to "physically robust," since those are not equivalent in a reconstructed ML-based product.

Third, some of the conclusions still read as more definitive than the evidence allows. The descriptive patterns themselves are interesting and likely robust at basin scale, but several of the mechanistic interpretations still seem to rely mainly on consistency arguments using OFES, GODAS, and previous literature, rather than direct quantitative attribution. I think that is fine, but the paper should distinguish more clearly between observed trend patterns, interpretations supported by consistency with other datasets, and mechanisms that remain hypotheses. A more cautious tone in some parts of the discussion would improve the paper.

We thank the reviewer for this constructive comment. We have carefully revised the manuscript to ensure that the wording more accurately reflects the level of evidence provided. In particular, we have clarified that the robustness metric is used as a heuristic measure rather than a formal uncertainty propagation, and we have avoided using overly strong expressions such as "physically robust."

In addition, we have revised the discussion to adopt a more cautious tone, clearly distinguishing between observed trend patterns, interpretations supported by consistency with other datasets, and mechanisms that remain hypothetical. The main changes are summarized below. A marked-up version of the manuscript, with all revisions highlighted in red, has also been provided for reference.

Section 3.1

L264: "(Note：This diagnostic provides a heuristic measure of the relative strength of the trend compared to the local uncertainty, rather than a formal quantification of uncertainty propagation.)"

L270: "Based on this metric, larger oxygen trend magnitudes correspond to higher R values, more clearly distinguishable from the background uncertainty."

L271: "Thus, in the upper ocean (2.5–100 m), trends are relatively robust in terms of the R metric, mainly in the northern North Pacific."

L286: "The positive O_2 trends coincide with regions of relatively low uncertainty values (Fig. 1p–s and 1w–z), suggesting that they represent relatively robust features that are better constrained by the high observation density of Argo profiling floats"

L296: "Such large uncertainties would likely arise from sparse observations and high background variability"

L291: "Consequently, these signals may reflect possible regional reoxygenation superimposed on the basin-scale deoxygenation trend."

L302: "Overall, despite the uncertainties associated with the various factors discussed above, the GOBAI-O_2 dataset provides an improved framework for diagnosing basin-scale oxygen variability and its physical drivers."

Section4

L561: "While this filtering procedure likely mitigates a portion of the air-calibration bias, the extent to which residual bias remains in the reconstructed fields is not well quantified."

L551: "This suggests that regional reoxygenation signals can coexist with large-scale deoxygenation, and highlights the importance of sustained BGC-Argo observations for detecting emerging changes in ocean biogeochemistry."

L572: "Accordingly, the interpretation of the diagnosed O_2 trends should be made with caution, particularly in regions where float-based observations dominate."

L578: "In particular, although the large-scale spatial patterns are broadly consistent across datasets, both the magnitude of trends and finer-scale spatial features may still be affected by unresolved observational and reconstruction uncertainties."

Related to this, the new frontal-migration analysis is helpful and makes that part of the argument more convincing than before. However, the agreement across datasets is not completely uniform, and I do not think the results support an equally strong statement everywhere. I would encourage the authors to acknowledge this more directly and emphasize that the evidence supports a broad tendency rather than quantitatively consistent behavior at all locations and depths.

Thank you for your constructive suggestion. Regarding the mechanistic interpretations, we carefully reviewed the wording of Section 3 and revised expressions that could be interpreted as implying direct causal attribution, in order to adopt a more cautious tone.

We also clarified that the location and magnitude of the frontal shifts are not spatially uniform and may vary depending on the region and dataset.

In Section 3.3.1, we made the following revisions:

The sentence "These patterns suggest that waters subducted in the frontal region …" was revised to "These patterns are consistent with the possibility that waters subducted in the frontal region …".

The sentence "the observed positive temperature and salinity trends at 26.8σθ likely reflect influences from changes occurring in the overlying layers (Fig. 5d–e and g–h)" was revised to "the observed positive temperature and salinity trends at 26.8σθ may reflect influences from changes occurring in the overlying layers (Fig. 5d–e and g–h), although this remains a hypothesis based on consistency with known processes."

The sentence "Consistent with these studies, the present dataset exhibits clear northward migration of the 25.0σθ and 26.0σθ outcrop lines (Fig. 6a)" was revised to "Consistent with these studies, although the magnitude and spatial coherence vary across regions among the datasets, the present dataset exhibits an overall tendency toward northward migration of the 25.0σθ and 26.0σθ outcrop lines (Fig. 6a)."

The sentence "Such poleward displacements of frontal structures can help explain the negative temperature and salinity trends in the subtropical gyre" was revised to "Such poleward displacements of frontal structures may contribute to explaining the negative temperature and salinity trends in the subtropical gyre."

The sentence "the associated temperature and salinity changes arise mainly from direct surface warming and freshening" was revised to "the associated temperature and salinity changes may be influenced by direct surface warming and freshening."

In Section 3.3.2, we made the following revisions:
The sentence "likely influenced by" was revised to "which may be influenced by".

The discussion of possible sensor bias is also better than before, but it still remains fairly qualitative. I appreciate that the authors now acknowledge this limitation more clearly, and I do not necessarily think a full new sensitivity analysis is required if that is genuinely not feasible. Still, I think the manuscript should remain cautious when interpreting trends that may be influenced by unresolved observational or reconstruction uncertainties.

In the revised manuscript, texts in bold have been added to the relevant sections to emphasize the need for caution and the explicit acknowledgement of the limits of quantification in Section 4. The revised text now reads as follows:

"Recent work by Bushinsky et al. [2025] has reported the presence of a systematic negative bias (approximately -2.7 μmol kg^{-1}) in air-calibrated BGC-Argo oxygen measurements compared with ship-based reference profiles. This bias does not appear to be explicitly corrected in version 4.4 of GOBAI-O_2 and may therefore influence the magnitude of the estimated oxygen trends–potentially enhancing negative trends or suppressing positive ones in regions with dense float sampling. However, as described in Section 2.1, a substantial fraction of these float data is subject to quality control through comparison with climatological fields derived from ship-based discrete observations, and only profiles with appropriate quality flags are retained and incorporated into the dataset development. While this filtering procedure likely mitigates a portion of the air-calibration bias, the extent to which residual bias remains in the reconstructed fields is not well quantified.

If present, such biases could also affect the apparent vertical structure of the oxycline. In the North Pacific, regions with high float density–such as the Kuroshio–Oyashio transition zone, the North American coastal region, and the vicinity of Hawaii–may be particularly affected (see Fig. 1 of Sharp et al., 2023). While a constant offset would not directly alter linear trend estimates, any time–varying bias associated with sensor behavior or sampling depth could introduce spurious trends. A quantitative evaluation is not feasible at present due to the lack of temporally continuous ship-based reference data at the spatial scales. This limitation should therefore be kept in mind when interpreting the O_2 trends reported here. Accordingly, the interpretation of the diagnosed O_2 trends should be made with caution, particularly in regions where float-based observations dominate.

It is also essential to recognize that GOBAI-O_2 is a machine learning reconstruction derived from available temperature, salinity, and oxygen measurements. While this approach significantly enhances spatial coverage, the results should be interpreted cautiously. In particular, although the large-scale spatial patterns are broadly consistent across datasets, both the magnitude of the trends and finer-scale spatial features may still be affected by unresolved observational and reconstruction uncertainties. Nevertheless, future work incorporating improved calibration of Argo oxygen sensors, expanded ship-based reference datasets, independent machine learning reconstructions [e.g., Ito et al., 2024], and comprehensive ocean reanalysis will be necessary to better constrain these uncertainties."

"

Overall, I think the manuscript is substantially improved and may become publishable, but I do not think it is fully there yet. My strongest requests are that the authors explicitly state the exact GOBAI-O2 version used and that they revise the wording around uncertainty propagation and robustness so that it matches more closely what has actually been done. I also think some of the mechanistic interpretations should be softened slightly.

Thank you for your constructive feedback. We have revised the sections on uncertainty propagation and robustness to more accurately reflect the scope of our analysis. Additionally, we have toned down the wording regarding the interpretation of mechanisms to avoid implying direct causal relationships.

We hope these revisions address your concerns and improve the clarity of the manuscript.

Reviewer #2

The manuscript has been improved with quantitative assessment of data uncertainty. However, there are still a few points that need to be revised.

Figures 1v-1bb: Change the color range from 0–1 to 0–2 (or similar) for easier comparison with Figures 3 and 4.

Thank you for the careful suggestion. We have reconsidered the color scale and revised the color range from 0-1 to 0.5-2.5 in the current manuscript, as shown below.

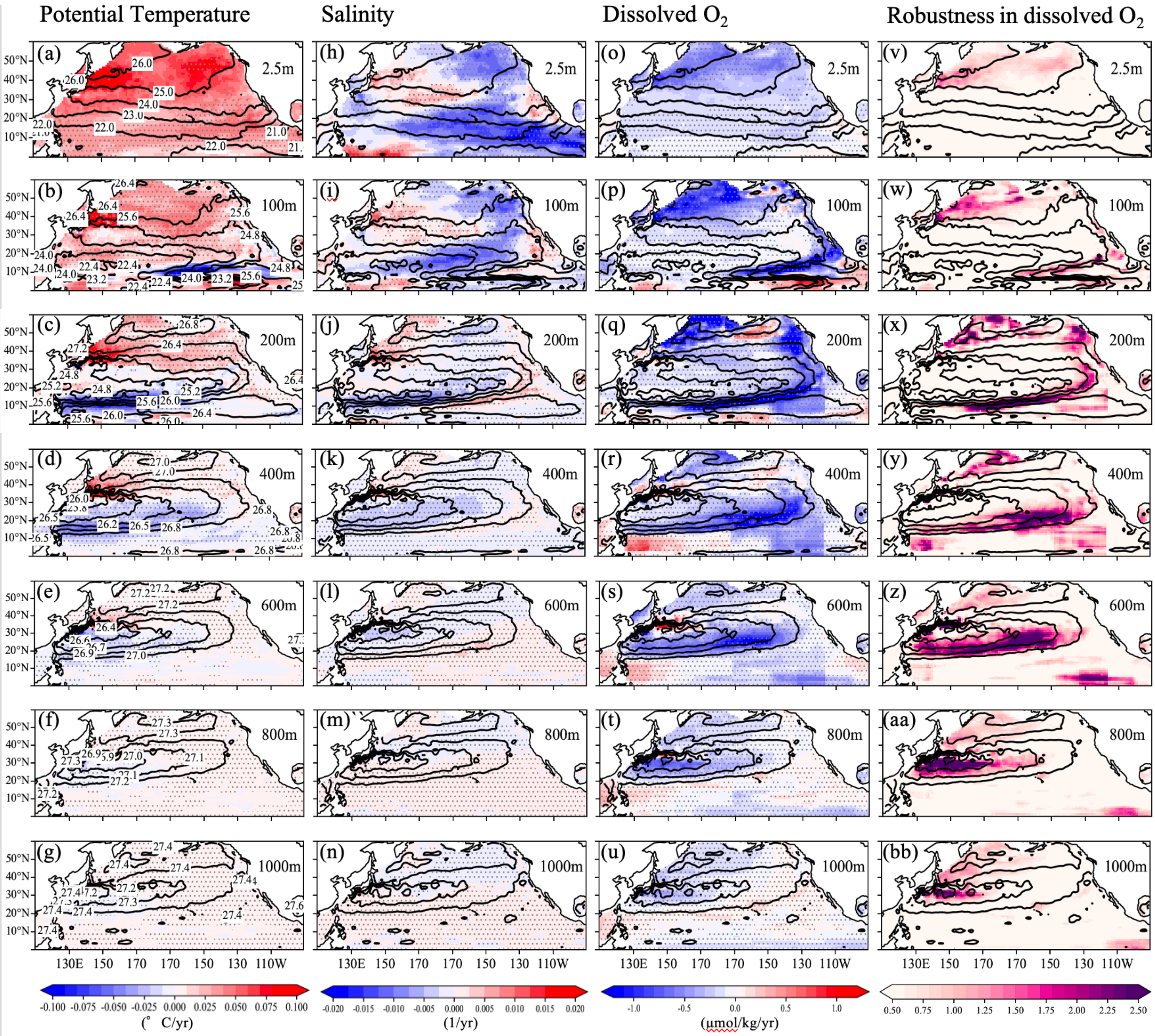

Figure 1 Horizontal distributions of linear trends in (a–g) potential temperature, (h–n) salinity, and (o–u) dissolved oxygen (O_2) during the observational period at depths of 0, 100, 200, 400, 600, 800, and 1000 m, respectively. Hatched areas indicate statistically significant trends at the 95% confidence level based on a Student’s t-test with effective degrees of freedom accounting for temporal autocorrelation. Trend significance

was evaluated using a Student's t-test with effective degrees of freedom accounting for lag-1 autocorrelation. Contours denote potential density at each depth. Labels for the potential density are shown only in the potential temperature sections. Corresponding distributions of the Robustness (R), defined as the ratio of the trend magnitude to the dataset uncertainty in dissolved O_2 are presented in panels (v–bb).

Figures 3d and 3h: Improve contour labeling to ensure values are easily identifiable.

Thank you for your helpful suggestion. We have improved the contour labeling to make the values more clearly identifiable in the revised manuscript, as shown below.

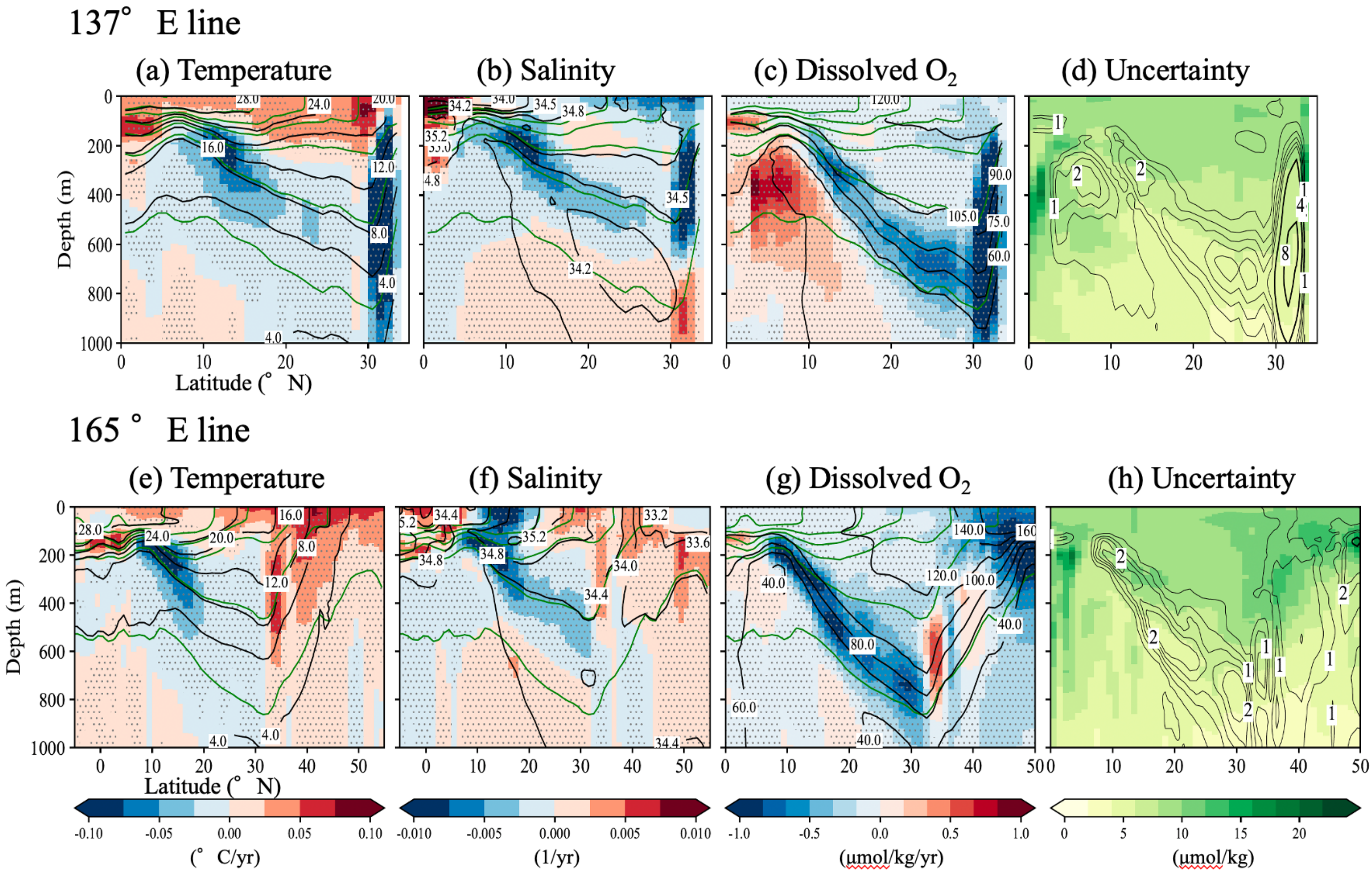

Figure 3 Vertical sections showing linear trends in potential temperature (a, e), salinity (b, f), and dissolved O_2 (c, g) along the 137°E and 165°E meridians, respectively. Black contour lines indicate the mean potential temperature (a, f), salinity (b, g), and dissolved oxygen (c, h) over the period 2004–2023, while green contour lines represent the mean potential density. Labels for the potential density are shown only in the robustness sections. Hatched areas indicate statistically significant trends at the 95% confidence level based on a Student's t-test with effective degrees of freedom accounting for temporal autocorrelation. Trend significance was evaluated using a t-test with effective degrees of freedom accounting for lag-1 autocorrelation. Corresponding vertical sections of the mean uncertainty with the contours of the Robustness (R) in panels (d, h). The contour intervals for thin and thick contours in (d, h) are 0.2 and 4.0, respectively.

Figures 4d and 4h: According to the definition, robustness is a non-negative metric. Therefore, the presence of the 0 contour line is inconsistent and requires clarification.

Thank you for your helpful suggestion. There was an error in the contour labeling. We have corrected this issue and revised the figure accordingly, as shown below.

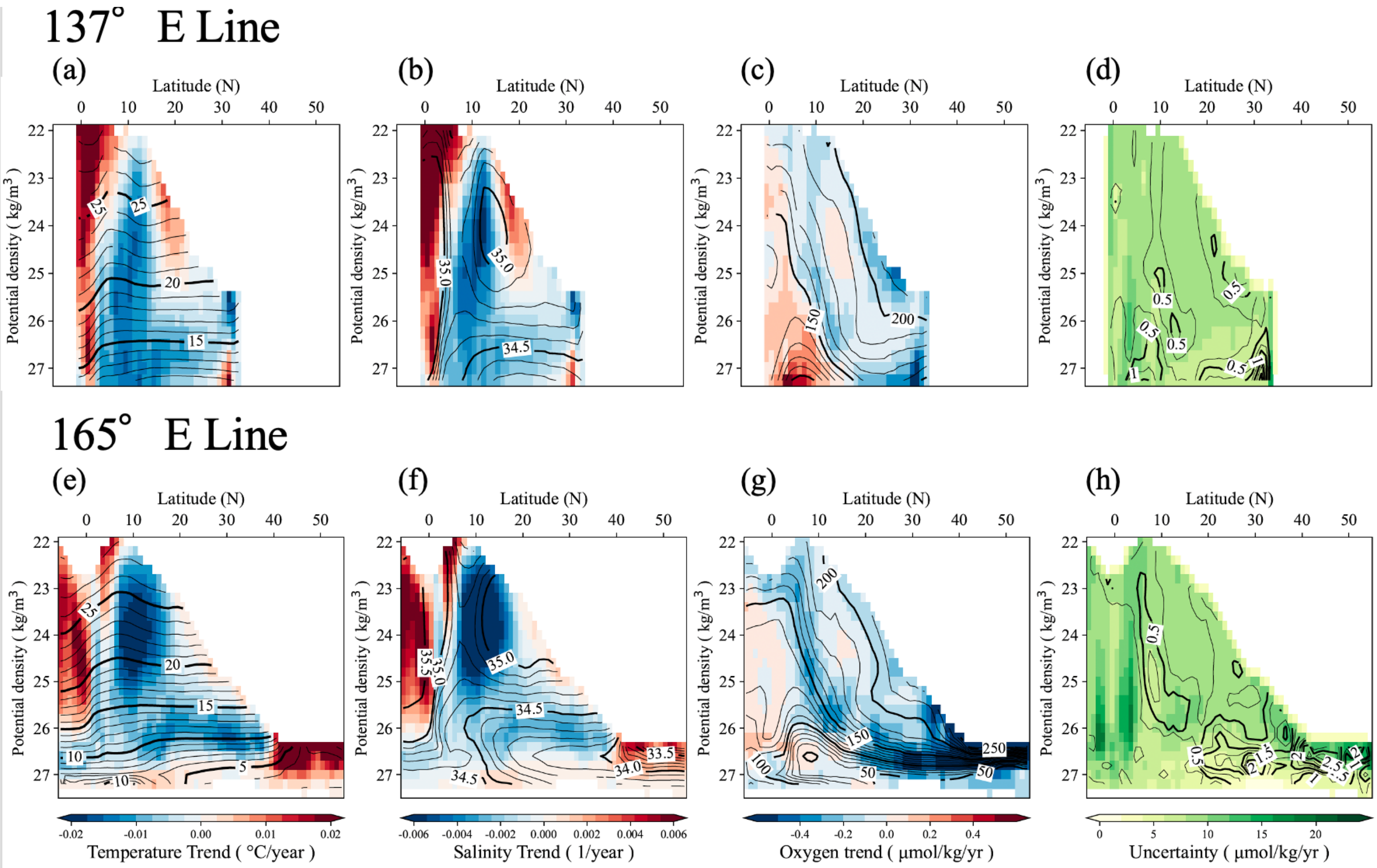

Figure 4 in the revised manuscript: Linear trends in (a, e) potential temperature, (b, f) salinity, and (c, g) dissolved O_2 on each isopycnal surface at intervals of $0.1\sigma_\theta$, calculated at every 1.0° of latitude in 137 °E and 165 °E sections, respectively. Contour lines represent the mean values during the target observation periods, plotted at intervals of $0.1\sigma_\theta$ for each 1° of latitude. Panels (d, h) show the corresponding vertical sections of mean uncertainty, along with contours of robustness (R). The contour intervals for thin and thick contours in (d, h) are 0.25 and 0.5, respectively.

Lines 281 and 545, "climatology-based": Clarify this term. Climatology usually implies a long-term mean, and Ito et al. (2027) analyzed O2 anomalies.

Thank you for your helpful suggestion. We have revised the term "climatology-based" to "O_2 anomaly" in the relevant sections to more accurately reflect the analysis of Ito et al. (2027).