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.

Referee #1

The authors clearly engaged with the reviews and implemented substantial changes (uncertainty discussion, significance tests, novelty clarification, softened claims). However, several of the original major concerns are addressed procedurally rather than substantively, and some still limit robustness.

1-1) The authors have added statistical significance tests for the trends, which is a welcome and necessary improvement. However, this addition does not fully resolve the central concern regarding the robustness of the diagnosed patterns relative to the underlying dataset uncertainty. The t-tests demonstrate where trends differ significantly from zero, but they do not establish whether the magnitudes of those trends are meaningfully distinguishable from the structural uncertainty in GOBAI-O2. In several regions, the trend magnitude remains comparable to the stated uncertainty, and this issue is not directly quantified. The authors also declined to relate trend magnitude to uncertainty on the grounds of unit mismatch, yet it would be straightforward to convert gridded uncertainty into trend uncertainty through propagation or confidence intervals. As a result, it remains unclear where spatial patterns can be interpreted as physically robust versus where they may be indistinguishable from dataset uncertainty. In addition, the uncertainty fields are described but not propagated into the trend confidence intervals or used to define spatial robustness masks, and some discussion sections still treat statistically significant patterns as physically meaningful without clearly distinguishing between statistical significance, robustness relative to dataset uncertainty, and regions that are primarily ML-inferred rather than directly observed. Overall, this core issue is improved but not fully resolved and remains a matter of scientific robustness.

Thank you for clarifying this important point. As you note, evaluating statistical significance alone is insufficient for determining whether the diagnosed trends exceed the inherent uncertainty of the dataset. In response to your suggestion, we have now explicitly quantified the spatial distribution of Robustness (R), defined as: R = | trend over two decades | / uncertainty, by directly comparing the oxygen trends with the mean uncertainty provided by the GOBAI-O_2 product. This analysis allows us to determine where the magnitude of the trend exceeds the observable uncertainty, and to clearly distinguish physically robust regions from those where the trends fall within the range of uncertainty range. Thus, the R-map directly responds to your request to distinguish statistical significance *from* physical robustness.

To further support this interpretation, we added the R-contours to the uncertainty fields in Figures 3(d, h) and 4(d, h). This visualization enables readers to evaluate robust versus non-robust regions directly alongside the associated trend maps, thereby interpretability.

We believe that these revisions adequately address your central concern—namely, whether the detected oxygen trends represent physically meaningful signals rather than artifacts arising from dataset uncertainty.

Corresponding text has been added in Section 3.1, which now reads:

“To assess whether regional trends exceed the dataset uncertainty, we computed the spatial distribution of Robustness (R), defined as $R = |\text{trend over two decades}|/\text{uncertainty}$ (Fig. 1v–bb). This diagnostic identifies regions in which the magnitude of the trend is sufficiently large to be considered physically robust. The results indicate that R exceeds or approaches high values in the eastern and western tropical zones, the Kuroshio Extension region, portions of the subpolar North Pacific, and along the 27.2–27.4 σ_θ density surfaces at 800–1000 m depth. Based on this metric, larger oxygen trend magnitudes correspond to higher R values, indicating greater robustness. Thus, in the upper ocean (2.5–100 m), trends are physically robust mainly in the northern North Pacific. At 200–400 m, robust signals appear both in the northern North Pacific and along the 25.2–26.0 σ_θ surfaces in the southern subtropical region, as well as in the eastern and western tropics. At 600–1000 m, the trends are robust within the subtropical gyre bounded by the 27.0 σ_θ surface.”

Revised figures:

Figure 1: Horizontal distributions of robustness in dissolved O_2 (Fig. 1(v-bb)) added.

Figure 2: Horizontal uncertainty maps moved from Fig. 1 in the previous manuscript.

Figure 3: R-contours added to uncertainty sections.

Figure 4: R-contours added to the horizontal distribution of averaged uncertainty (Fig. 4(d, h)).

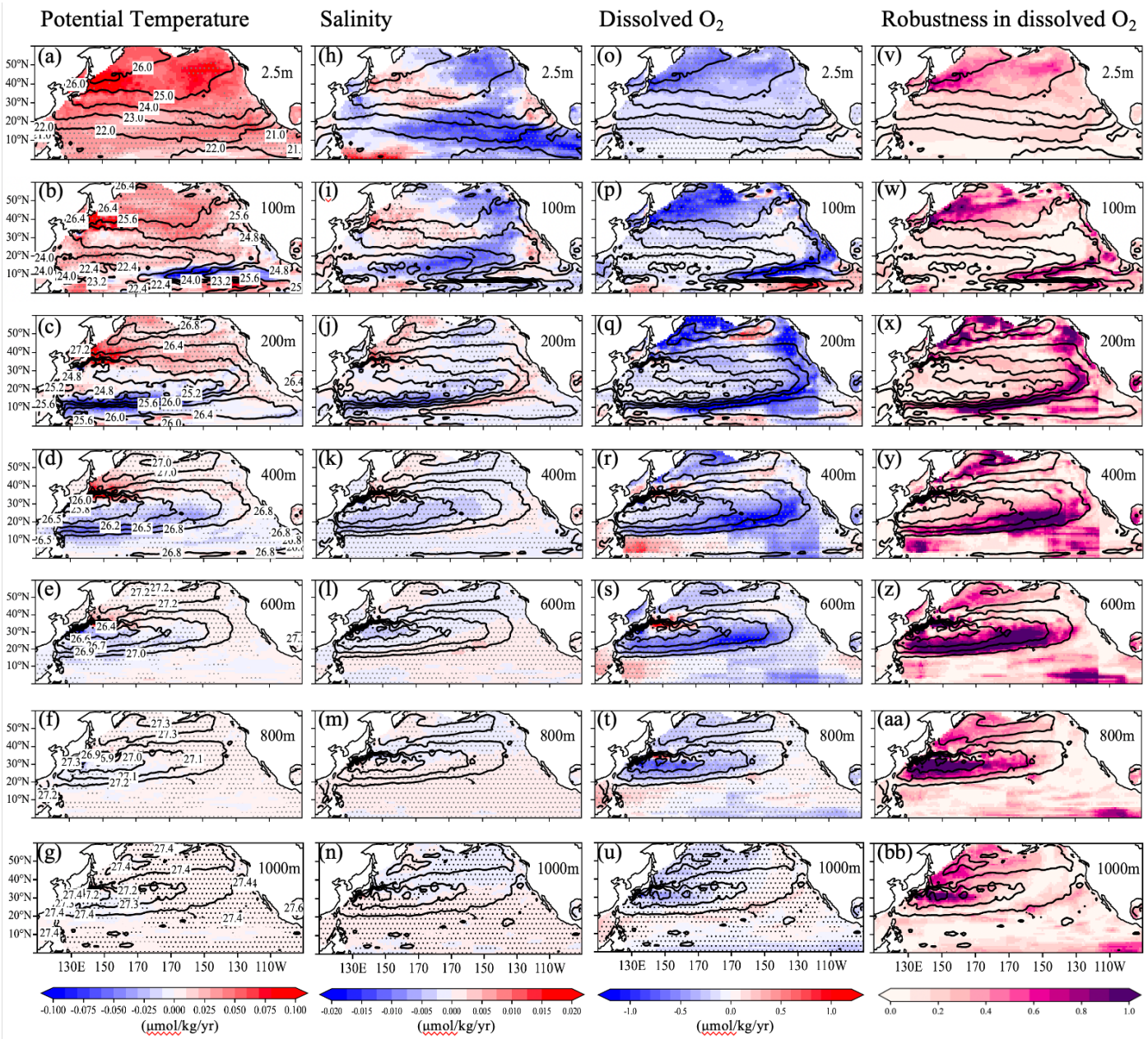

Figure 1 in the revised manuscript: Horizontal distributions of linear trends (μmol/kg/yr) in (a–g) potential temperature, (h–n) salinity, and (o–u) dissolved oxygen (O_2) during the observational period at depths of 0, 100, 200, 400, 600, 800, and 1000 m, respectively. Hatched areas indicate statistically 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).

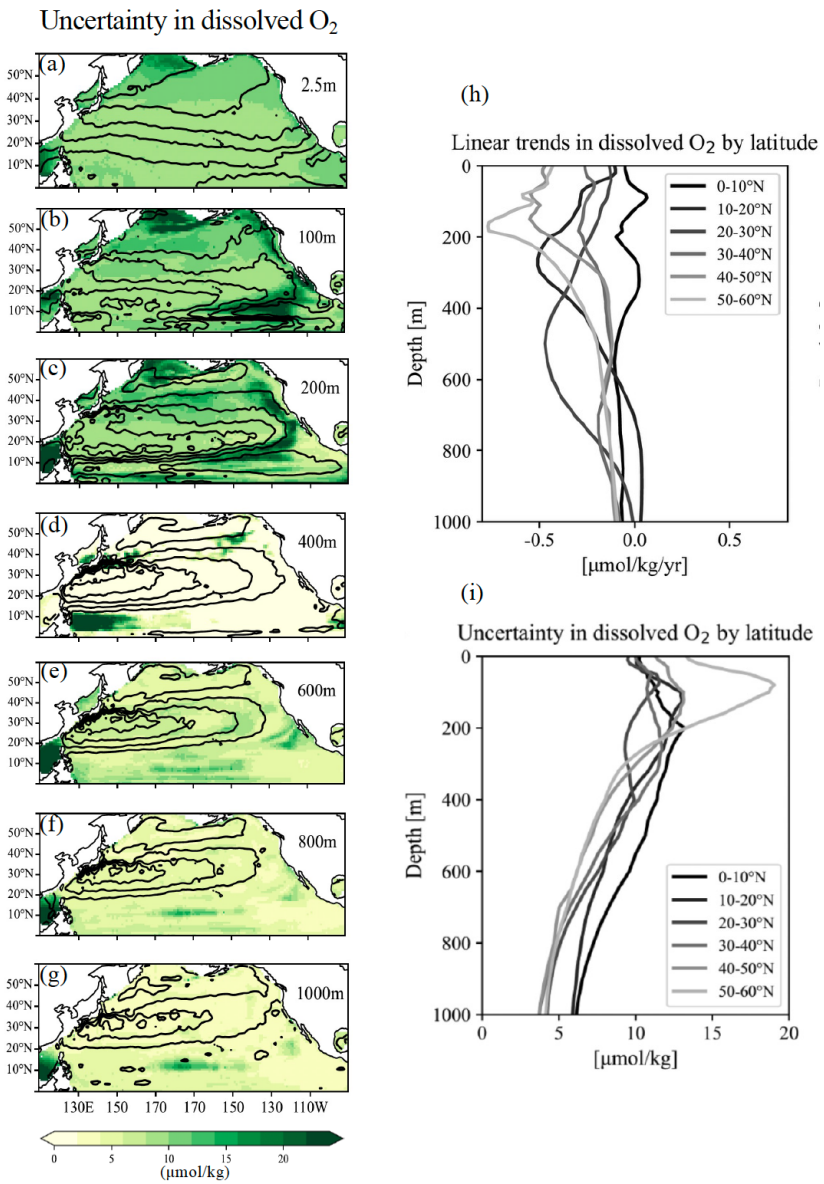

Figure 2 in the revised manuscript: Horizontal distributions of the dataset uncertainty in dissolved O_2 (a–g) and vertical profiles of linear trends and uncertainty in dissolved O_2 by latitude (h–i).

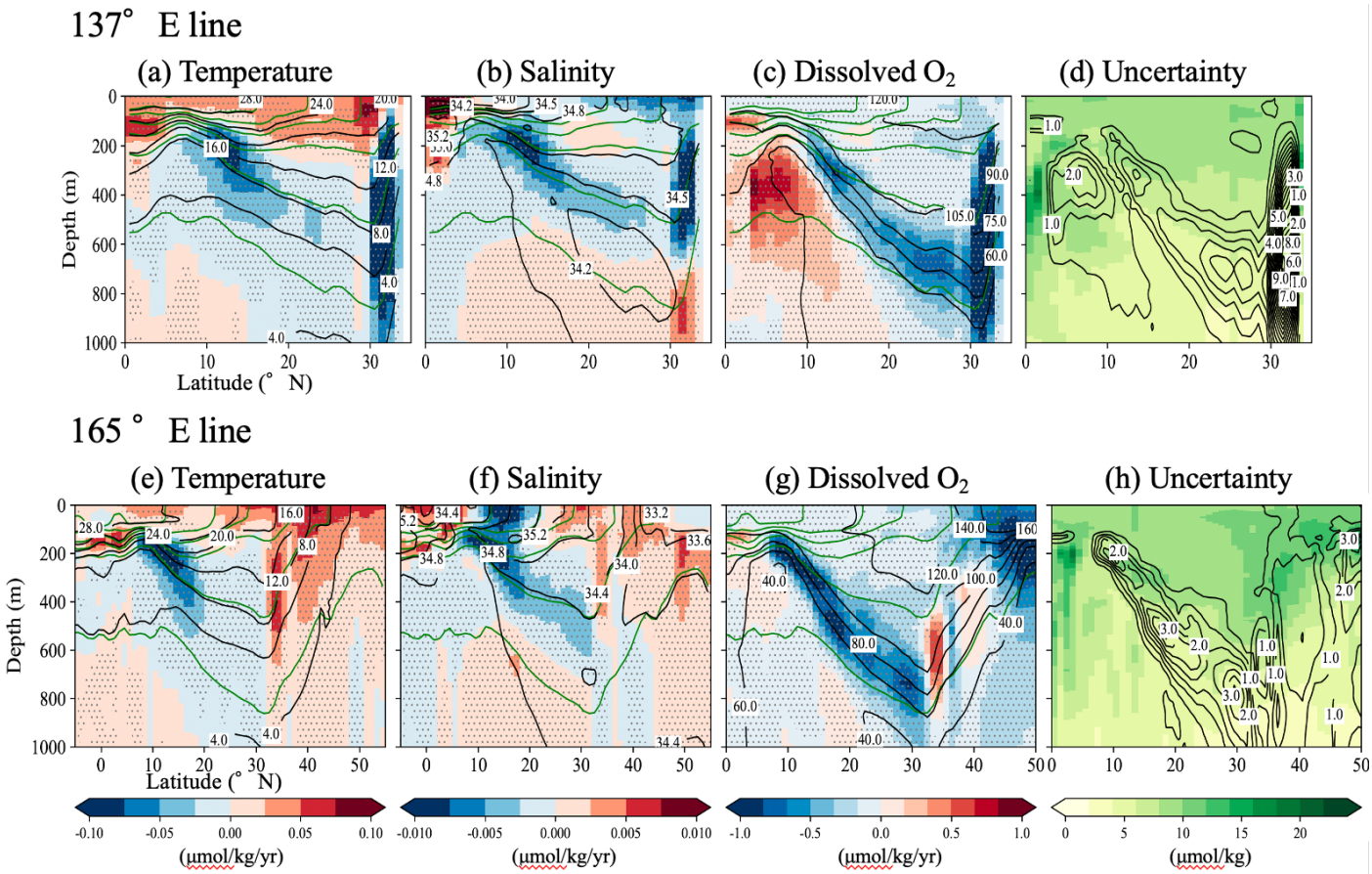

Figure 3 in the revised manuscript: Vertical sections showing linear trends in potential temperature (a, e), salinity (b, f), and dissolved O_2 (c, g) along the 137°E and 165°E meridians, respectively. Black contour lines indicate the mean potential temperature (a, f), salinity (b, g), and dissolved 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).

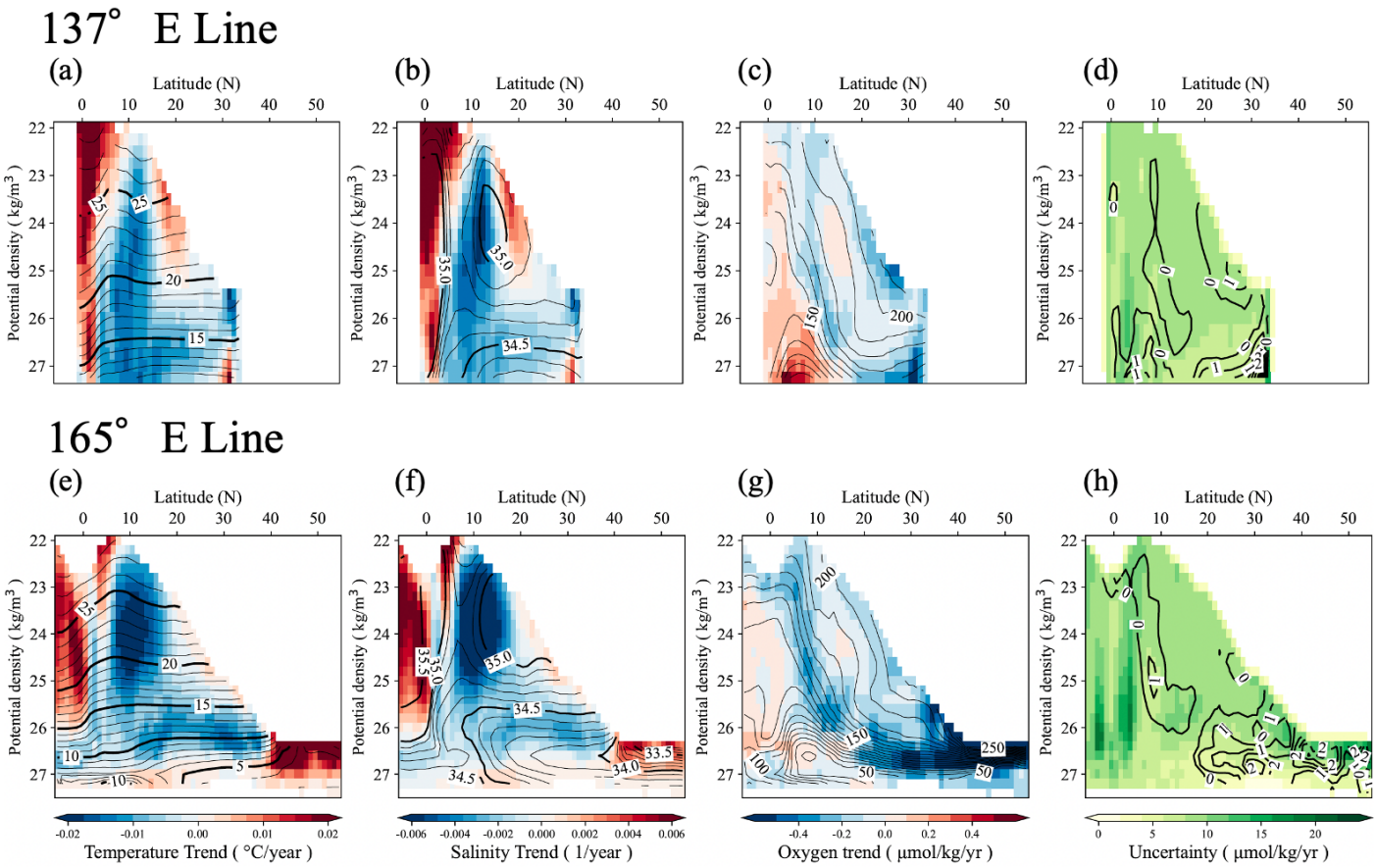

Figure 4 in the revised manuscript: Linear trends in (a, e) potential temperature, (b, f) salinity, (c, g), and dissolved O_2 on each isopycnal surface at intervals of $0.1\sigma_\theta$, calculated at every 1.0 deg of latitude in 137 °E and 165 °E lines, respectively. Contour lines represent the mean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

1-2) The treatment of potential sensor bias has also been improved but remains largely qualitative. The revised manuscript now acknowledges the reported negative bias (~–2.7 μmol kg-1) associated with some air-calibrated BGC-Argo floats and notes that this could influence the magnitude and spatial structure of the diagnosed trends. However, this discussion is not accompanied by any quantitative sensitivity analysis, regional assessment, or estimate of how such biases might affect inferred trends. There is also no substantive comparison with shipboard reference sections beyond a general qualitative statement. The authors argue that a quantitative assessment is not feasible and therefore simply note the limitation. While this may be acceptable in some contexts, the manuscript relies heavily on a derived ML product trained on float data, and the lack of quantitative evaluation leaves the treatment methodologically thin. This concern is therefore only partially addressed and does not yet meet the standard typically expected for resolution of a major revision.

To address this concern, we examined the construction of the GOBAI-O_2 dataset as documented in Sharp et al. [2023]. According to their data-quality control procedures, only float profiles that passed multiple layers of QC—including comparison with ship-based climatological fields, cross-validation with GLODAP, and the exclusion of low-quality profiles—were retained in the dataset. Therefore, any air-calibration bias of approximately −2.7 μmol kg^{-1} reported by Bushinsky et al. [2025] would have been substantially reduced during the dataset construction process.

The revised manuscript now incorporates these details in Section 2.1. We also clarified in Section 4 that, although air-calibration bias may influence oxygen trend estimates, the combination of climatology-based adjustments, float–climatology consistency checks, and the restriction to profiles with appropriate quality flags likely mitigates most of the bias prior to interpolation.

Although a formal quantitative sensitivity analysis is not feasible due to the lack of temporally continuous ship-based reference sections at the required spatial scales, the design of GOBAI-O_2 inherently suppresses large calibration biases. Multiple layers of QC, comparison with ship-based climatological fields, and inter-dataset consistency adjustments are expected to reduce the residual air-calibration bias to levels considerably smaller than the nominal −2.7 μmol kg^{-1} reported for unadjusted floats. This provides methodological justification for why we consider the remaining bias unlikely to dominate the diagnosed trends.

The revised sections (Sections 2.1 and 4) read as follows:

Section 2.1

"According to Sharp et al. [2023], the float data used in GOBAI-O_2 were filtered to retain only delayed-mode adjusted profiles with quality flags of 1 (good), 2 (probably good), or 8 (interpolated/extrapolated) for pressure, temperature, salinity, and dissolved oxygen. Among all available float profiles, 51.4% underwent quality control through comparison with climatological fields from the World Ocean Atlas (WOA) or the Commonwealth Scientific and Industrial Research Organisation Regional Sea Atlas (CARS). An additional 30.3% were evaluated using atmospheric oxygen concentration measurements, and 7.0% were quality controlled through comparison with in-water measurements (WOD, OMS assuming an oxygen zero, or deployment-time CTD profiles). A further 5.3% were adjusted using in-situ optode calibration based on the method of Drucker and Riser [2016], 3.3% were adjusted by other methods, 1.9% were unclassified, and the remaining 0.9% were not adjusted."

Section 4

"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 the 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. This filtering procedure likely mitigates a portion of the air-calibration bias.

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."

1-3) Finally, the physical interpretation has been moderated but still at times extends beyond what is directly demonstrated. The authors have softened their language and now frame mechanistic explanations more cautiously, which is appropriate. Nevertheless, several interpretations continue to rely primarily on consistency arguments rather than independent quantitative evidence. The OFES-based discussion remains largely qualitative, and the evidence for frontal migration is limited to the isopycnal outcrop analysis, which is defensible but not especially strong. These interpretations are likely acceptable if presented strictly as hypotheses consistent with the observations rather than as demonstrated mechanisms, but in their current form they remain somewhat borderline. Overall, this aspect is mostly addressed but not yet fully convincing.

Thank you for your constructive comments regarding the interpretation of the frontal migrations. In response, we have incorporated additional evidence using the original OFES and GODAS datasets to strengthen the support for the northward migration of the fronts. The tendencies derived from these datasets are now presented in Fig. 6b–c and Table 1 of the revised manuscript. As summarized in Table 1, we quantified the northward shifts of the isopycnal outcrop latitudes and compared these values with those reported by Xu et al. (2022), thereby providing an additional benchmark against previous studies. This comparison shows that the magnitude of the outcrop migration detected in our dataset is broadly consistent with earlier estimates.

Xu et al. (2022) examined the northward migration of the North Pacific subtropical gyre and the KE/OE fronts using a comprehensive suite of independent observational datasets, including Argo temperature–salinity profiles, AVISO sea surface height, and NOAA OISSTv2, together with multiple ocean reanalysis products (SODA3.3.2, EN4, IAP, Ishii T–S fields; NCEP and ERA5 10-m winds; JRA-55 surface winds; MERRA-2 wind stress) and eddy-resolving ocean hindcasts (OFES forced with both NCEP and ERA5). Their analysis demonstrated that changes in the mixed layer and outcrop lines are strongly linked to the poleward migration of the North Pacific subtropical gyre and the KE/OE fronts, driven by the poleward expansion of the Hadley circulation under global warming (Yang et al. 2020). They argued that the Kuroshio Extension and Oyashio Extension fronts, mode waters, and subtropical fronts evolve as a coherent system.

By incorporating the findings of Xu et al. (2022) as independent and complementary evidence, we have strengthened the physical interpretation of frontal northward migration in our study and clarified how our revisions adequately address the reviewer's concerns and reinforce the scientific robustness of our interpretation.

The additional references:

Behringer, D. W., & Xue, Y. (2004). Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean. 8th Symposium on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, AMS 84th Annual Meeting, Washington State Convention and Trade Center, Seattle, Washington, 11-15. [Available online at https://ams.confex.com/ams/pdfpapers/70720.pdf]

Behringer, D. W. (2007). The Global Ocean Data Assimilation System (GODAS) at NCEP. 11th Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, San

Antonio, TX, Amer. Meteor. Soc., 3.3. [Available online at https://ams.confex.com/ams/pdfpapers/119541.pdf.]

Yang, H., Lohmann, G., Krebs-Kanzow, U., Ionita, M., Shi, X., Sidorenko, D., et al. (2020). Poleward Shift of the Major Ocean Gyres Detected in a Warming Climate. *Geophysical Research Letters*, *47*(5), e2019GL085868. https://doi.org/10.1029/2019GL085868

Xu, L., Wang, K., & Wu, B. (2022). Weakening and Poleward Shifting of the North Pacific Subtropical Fronts from 1980 to 2018. *Journal of Physical Oceanography*, *52*(3), 399-417. https://doi.org/https://doi.org/10.1175/JPO-D-21-0170.1

The revised parts in Sections 2.5 and 3.3.1 can be read as follows:
Section 2.5

"2.5 GODAS model output

In Section 3.3.2, we also use temperature and salinity fields from the NCEP Global Ocean Data Assimilation System (GODAS) to complement our analysis. GODAS is a global ocean reanalysis system developed at the National Centers for Environmental Prediction (NCEP) and is based on the Modular Ocean Model version 3 [Pacanowski and Griffes, 2000]. The system assimilates surface temperature profiles, XBT data, moored buoy observations, and other in situ measurements using a three-dimensional variational (3DVAR) assimilation scheme [Behringer and Xue, 2004; Behringer, 2007]. The GODAS reanalysis is provided on a 1° x 1° horizontal grid with enhanced meridional resolution (1/3°) near the equator and includes 40 vertical levels. The reanalysis spans from 1980 to the present and is widely used for climate diagnostics and ocean variability studies. In this study, we analyze GODAS density fields over the period 2003-2024 by using temperature and salinity."

Section 3.3.1

"A meridional northward shift of the outcrop line in the North Pacific associated with recent climate change has been documented in OFES analyses [Ogata and Nonaka, 2020] and in other observational, reanalysis, and eddy-resolving ocean hindcasts [Xu et al., 2022]. Consistent with these studies, the present dataset exhibits clear northward migration of the 25.0σ_θ and 26.0σ_θ outcrop lines (Fig. 6a), with the strong shifts occurring in the eastern basin between 150°E and 180°W (Fig. 6 and Table 1). The estimated northward shift rate at 0.004–0.09 °yr^{-1} from 2004 to 2023 is comparable to the value of 0.04 °yr^{-1} reported by Xu et al. [2022] for 1980 to 2018. Xu et al. [2022] further demonstrated that changes in the mixed layer and outcrop lines are tightly coupled with the northward migration of the North Pacific subtropical gyre and KE/OE fronts due to the poleward expansion of the Hadley cell, including the fact that the Kuroshio Extension and Oyashio Extension fronts, mode waters, and subtropical fronts evolve as a coherent system. These changes may also reflect the influence of anthropogenic warming, which has been linked to the poleward expansion of the Hadley circulation and the associated meridional shifts of oceanic fronts [Yang et al., 2020]."

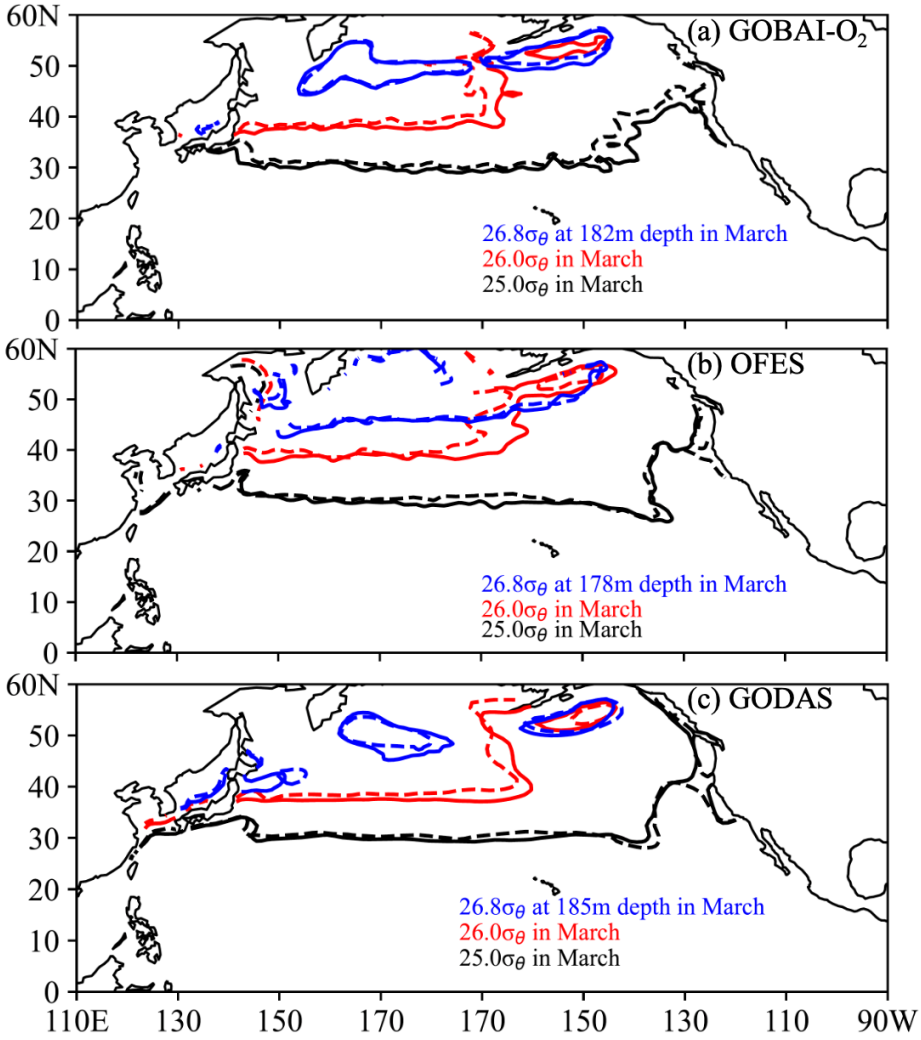

Figure 6 Density contours of 25.0σ_θ (black), 26.0σ_θ (red), and 26.8σ_θ (blue) in each dataset: (a) GOBAI-O_2, (b) OFES, and (c) GODAS. Solid lines indicate the mean March density contours for 2004–2009, while dashed lines represent those for 2019–2023.

Table 1. Northern shifts of the (outcrop) isopycnal latitudes (° yr^{-1}) for 25.0σ_θ (a), 26.0σ_θ (b), and 26.8σ_θ (c) in the GOBAI-O_2, OFES, and GODAS datasets. The estimates are based on data from March of each year. For 26.8 σ_θ, the northern shift is evaluated using the isopycnal depths corresponding to 182, 178m, and 183m in GOBAI-O2, OFES, and GODAS, respectively.

(a) 25.0σ_θ

Longitude	GOBAI-O_2	OFES	GODAS
150 °E	0.0241	0.0157	0.0283
170 °E	0.0444	0.0052	0.0240
170 °W	0.0684	0.0871	0.0481
150 °W	0.0947	0.0353	0.0313
130 °W	0.0420	0.0471	0.0121

(b) 26.0σ_θ

Longitude	GOBAI-O_2	OFES	GODAS
150 °E	0.0368	0.0766	0.0358
170 °E	0.0436	-0.0305	0.0508
170 °W	0.0124	0.1997	0.1412

(c) 26.8σ_θ

Longitude	GOBAI-O_2	OFES	GODAS
150 °E	0.0371	0.1980	0.0046
170 °E	0.0338	0.0217	0.1637
170 °W	0.0728	0.0054	0.0261

Referee #2

2-1) As mentioned previously, the uncertainties should be quantitatively compared with the magnitude of the trends, and then the signals that exceed the uncertainty should be discussed. As the comment from the first reviewer, |trend| > mean uncertainty would be acceptable. Trend/yr multiplied by data period can be compared with mean uncertainty.

Thank you for your constructive comment. We now fully understand the intention behind your earlier suggestion. In this revision, we quantified the spatial distribution of $R = |\text{trend over two decades}|/\text{uncertainty}$, by directly comparing the diagnosed oxygen trends with the mean uncertainty provided by the GOBAI-O_2 product. This analysis enables us to identify regions where the magnitude of the trend exceeds the observational uncertainty and to clearly distinguish physically robust regions from those in which the trends fall within the uncertainty range. In this way, the R-map directly responds to your request to differentiate statistical significance from physical robustness. Furthermore, we have added the R-contours to the uncertainty fields in Figures 3(d, h) and 4(d, h), allowing readers to visually assess regions of physical robustness alongside the corresponding oxygen trend maps. This presentation is intended to make the interpretation of robust versus non-robust regions more transparent. We believe that these revisions adequately address your main concern—namely, whether the detected oxygen trends represent physically meaningful signals rather than artifacts arising from observational uncertainty.

For this improvement, we added some sentences in Section 3. The revised part now reads:

"To assess whether regional trends exceed the dataset uncertainty, we computed the spatial distribution of Robustness (R), defined as $R = |\text{trend over two decades}|/\text{uncertainty}$ (Fig. 1v–bb). This diagnostic identifies regions in which the magnitude of the trend is sufficiently large to be considered physically robust. The results indicate that R exceeds or approaches high values in the eastern and western tropical zones, the Kuroshio Extension region, portions of the subpolar North Pacific, and along the 27.2–27.4 σ_θ density surfaces at 800–1000 m depth. Based on this metric, larger oxygen trend magnitudes correspond to higher R values, indicating greater robustness. Thus, in the upper ocean (2.5–100 m), trends are physically robust mainly in the northern North Pacific. At 200–400 m, robust signals appear both in the northern North Pacific and along the 25.2–26.0 σ_θ surfaces in the southern subtropical region, as well as in the eastern and western tropics. At 600–1000 m, the trends are robust within the subtropical gyre bounded by the 27.0 σ_θ surface."

Revised figures:

Figure 1: Horizontal distributions of robustness in dissolved O_2 (Fig. 1(v-bb)) added.

Figure 2: Horizontal uncertainty maps moved from Fig. 1 in the previous manuscript.

Figure 3: R-contours added to uncertainty sections.

Figure 4: R-contours added to the horizontal distribution of averaged uncertainty (Fig. 4(d, h)).

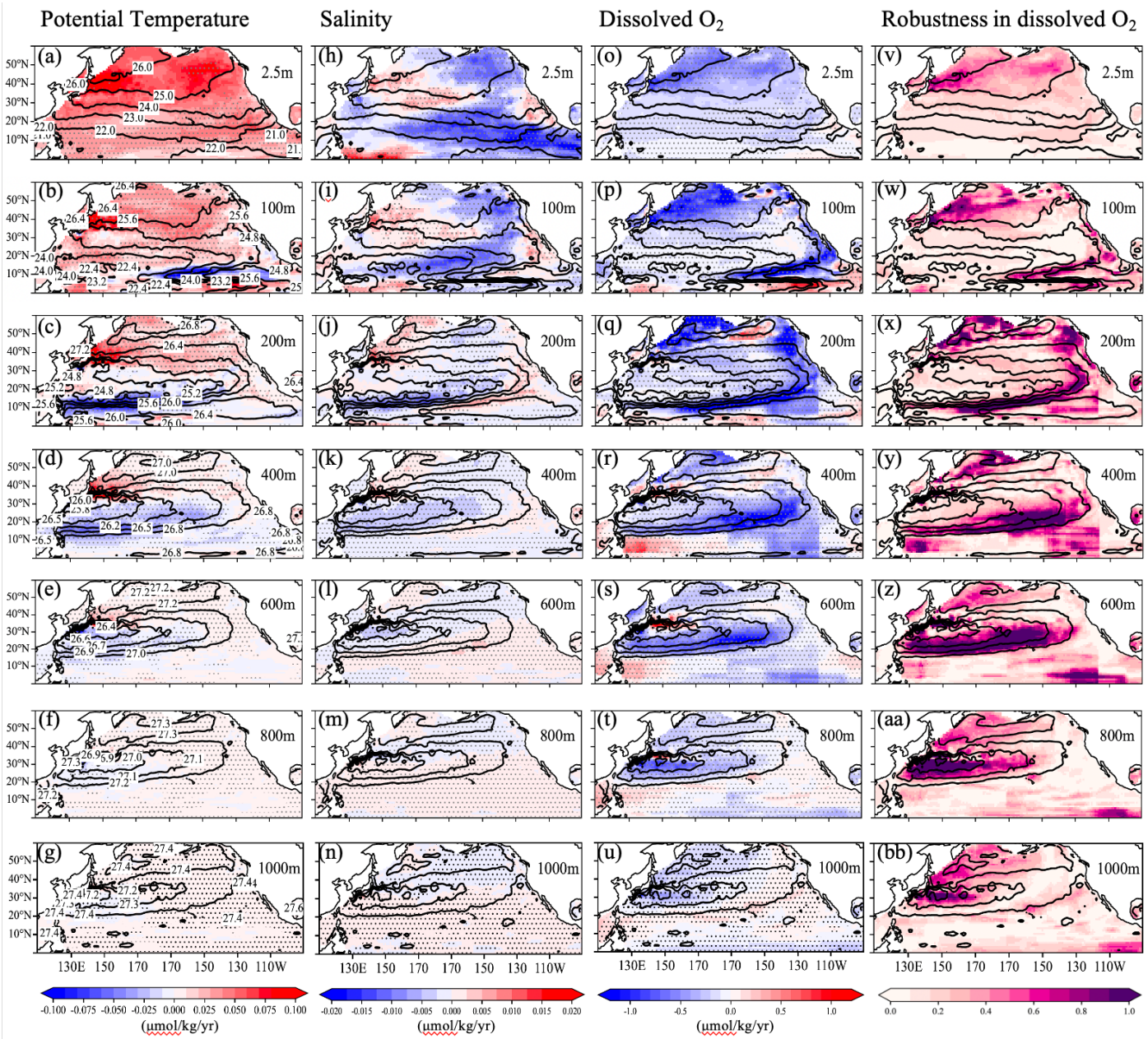

Figure 1 in the revised manuscript: Horizontal distributions of linear trends (μmol/kg/yr) in (a–g) potential temperature, (h–n) salinity, and (o–u) dissolved oxygen (O_2) during the observational period at depths of 0, 100, 200, 400, 600, 800, and 1000 m, respectively. Hatched areas indicate statistically 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).

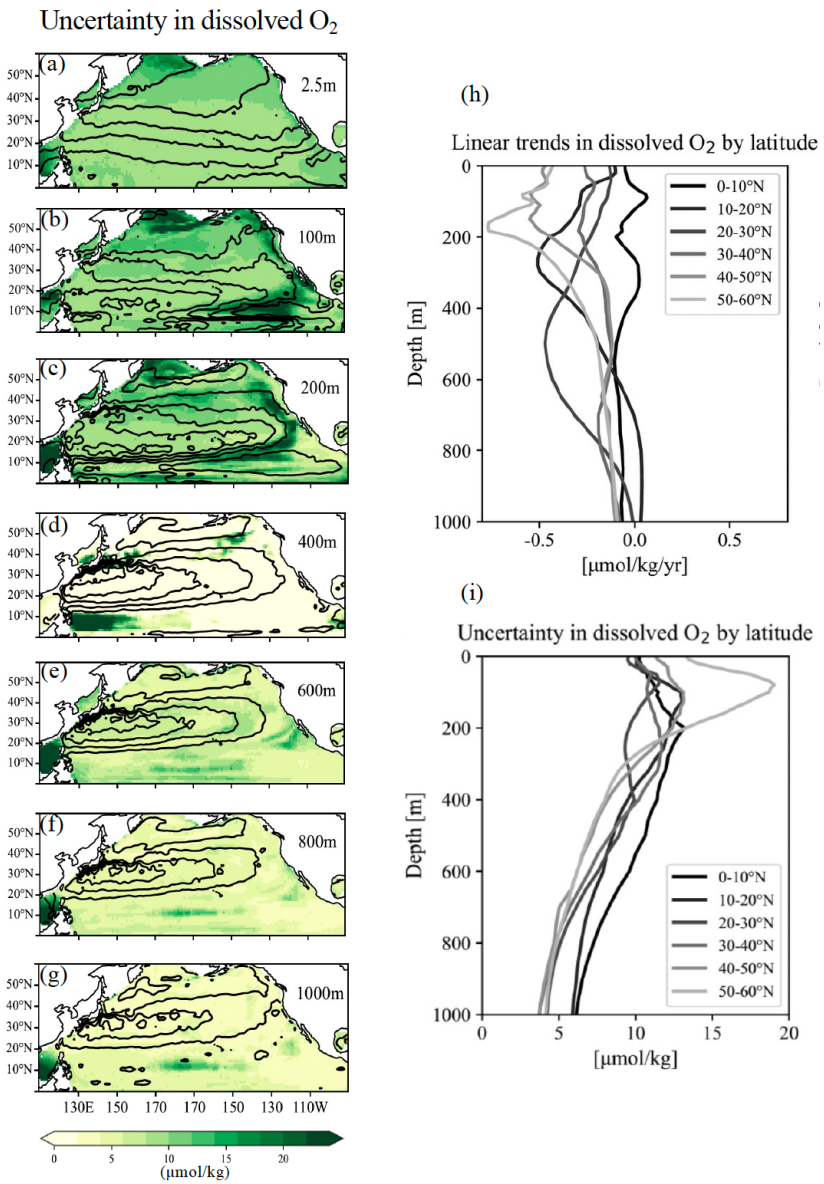

Figure 2 in the revised manuscript: Horizontal distributions of the dataset uncertainty in dissolved O_2 (a–g) and vertical profiles of linear trends and uncertainty in dissolved O_2 by latitude (h–i).

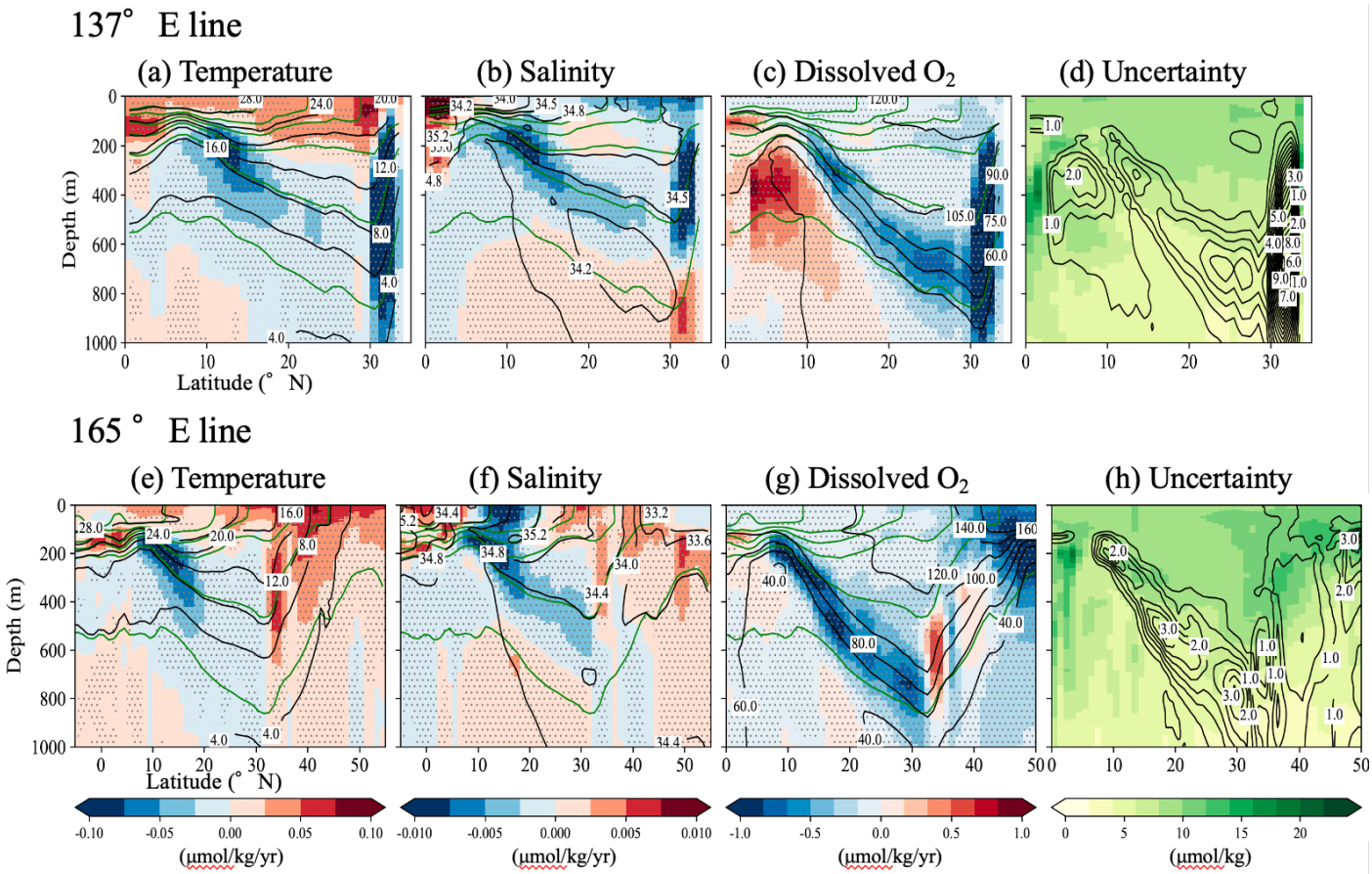

Figure 3 in the revised manuscript: Vertical sections showing linear trends in potential temperature (a, e), salinity (b, f), and dissolved O_2 (c, g) along the 137°E and 165°E meridians, respectively. Black contour lines indicate the mean potential temperature (a, f), salinity (b, g), and dissolved 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).

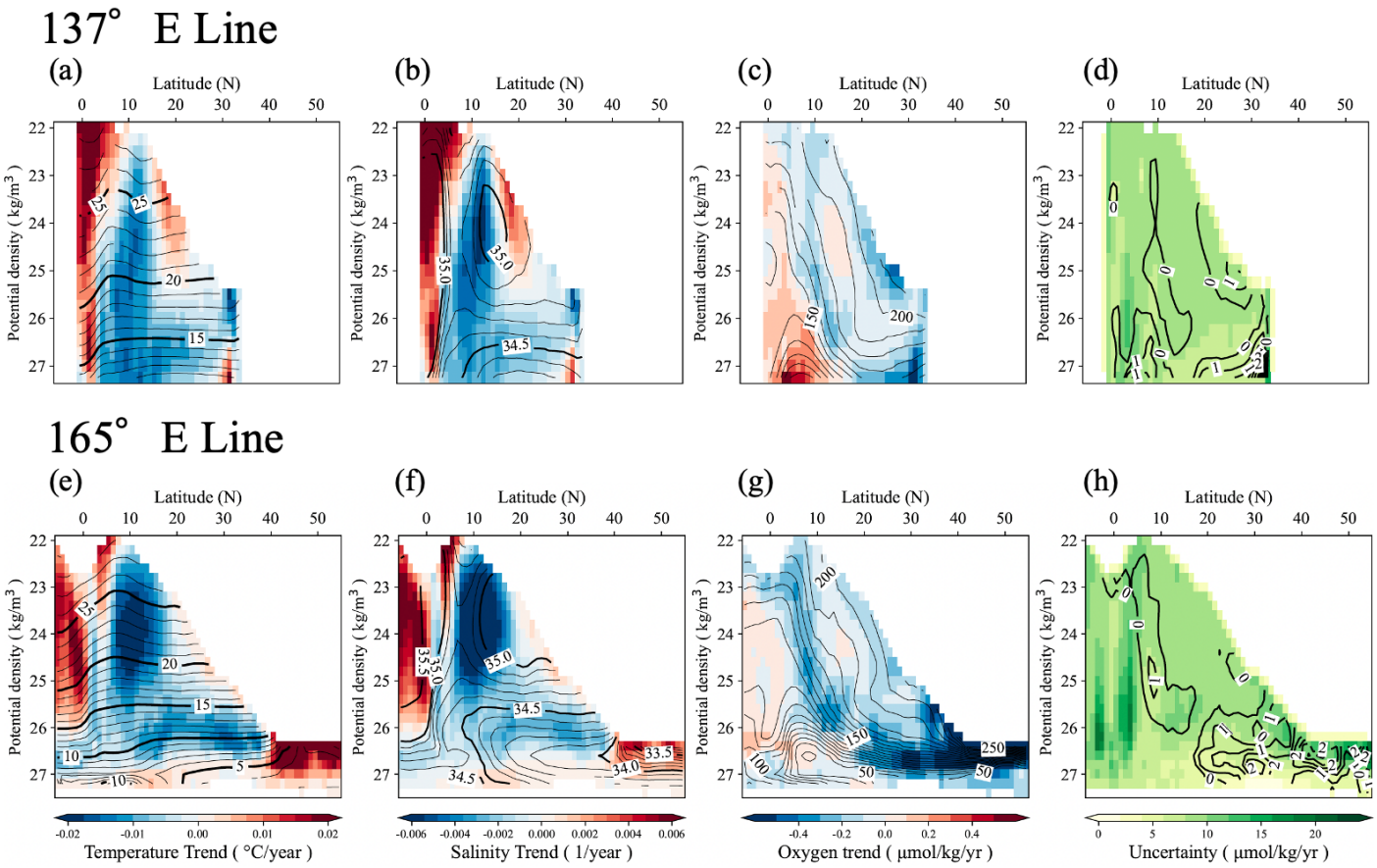

Figure 4 in the revised manuscript: Linear trends in (a, e) potential temperature, (b, f) salinity, (c, g), and dissolved O_2 on each isopycnal surface at intervals of $0.1\sigma_\theta$, calculated at every 1.0 deg of latitude in 137 °E and 165 °E lines, respectively. Contour lines represent the mean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