1. This is one of many papers that use data from various repositories to provide 'intermediate' products for other products. Its strength lies in its attempt to provide a global data set with weekly temporal resolution. Its weakness lies in its failure to clearly state the product's limitations.

We thank the reviewer for the critical comments! As the reviewer pointed out, the primary strength of the 4D-MGA product is that it can capture comparable mesoscale information to that of GLORYS and ARMOR3D, while exhibiting a low bias and a low root mean square error (RMSE). It is therefore suggested that this product offers a complementary perspective to existing products, particularly in providing an objective, high-resolution and high-certainty estimate of the ocean state.

The present analysis framework has several notable limitations. Firstly, despite the inclusion of additional background constraints, the dynamic height integration process may still lead to distorted adjustments in deep layers. Specifically, the optimization algorithm's propensity for strong gradients has been observed to result in an over-adjusting deep salinity to align with SLA. As shown in Figure R1, the deep salinity adjustments manifest similar mesoscale spatial patterns to the SLA features in specific regions. Secondly, due to computational constraints (Wu et al., 2026b), the temporal grid resolution used in this study is coarse (30 days). The limited number of temporal layers may not only filter out resolvable signals but also introduce additional noise. Thirdly, the approximation of DHA by SLA has the potential to introduce errors owing to the neglect of barotropic signals. However, this assumption is less valid in barotropic-dominated regions, such as the polar oceans and coastal areas. Fourthly, due to the distribution and resolution of in-situ profiles, as well as the inherent smoothing of the analysis method, the effective resolution of the product is lower than its nominal grid resolution.

These limitations have been summarized and incorporated into the "Deficiencies and Future Improvements" section of the revised manuscript.

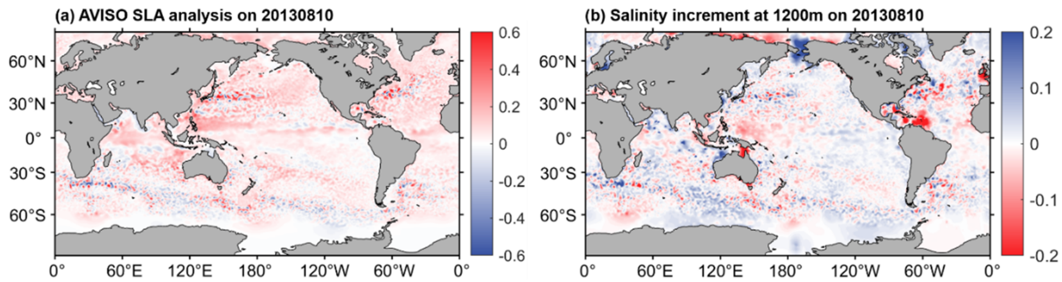

Figure R1. (a) AVISO SLA analysis (m) and (b) salinity increment at 1200 m relative to WOA23, both on 10 August 2013.

2. Based on the authors' presentation, the method appears to work reasonably well in cases where geostrophy plays a predominant role (for example, in the case of along-fluxes - western boundaries currents). The analysis is centered on two equations (1) and (2) which (in my opinion) significantly smooth the data and therefore lose much of the non-geostrophic signal. The results are integrated with dynamic heights which (again) are based on geostrophy. Figures 4 effectively confirm this refutation of mine.

We thank the reviewer for the positive recognition and critical observations. We agree that the smoothing effect inevitably reduces the effective resolution below the nominal grid resolution, and this limitation is now acknowledged in the revised manuscript.

In order to assess the actual impact of the smoothing term in equations (1) and (2), an additional experiment was performed using the 60% observation subset from the independent test described in the revised manuscript, but with the smoothing term removed entirely.

As shown in Figure R2, the incorporation of the smoothing term yields minimal influence. This is due to the fact that the MGA framework delineates the fundamental spatiotemporal correlations. The observation operator **H** links grid points to establish these correlations, and the iterative grid-doubling procedure enables multiscale analysis, extending the correlations to all grid levels. Conversely, the smoothing term ensures the continuity of the increments. In its absence, the bilinear interpolation inherent in **H** generates an anisotropic artifact, merging two warm eddies into a discontinuous, confused structure (Figure R2a). Conversely, adding the smoothing term clearly separates the eddies and restores physical continuity (Figure R2b). Thus, a suitably weighted smoothing term serves to enhance mesoscale features and, thereby promote enhanced physical realism without compromising global-scale consistency.

The description of the smoothing term has been updated in the manuscript.

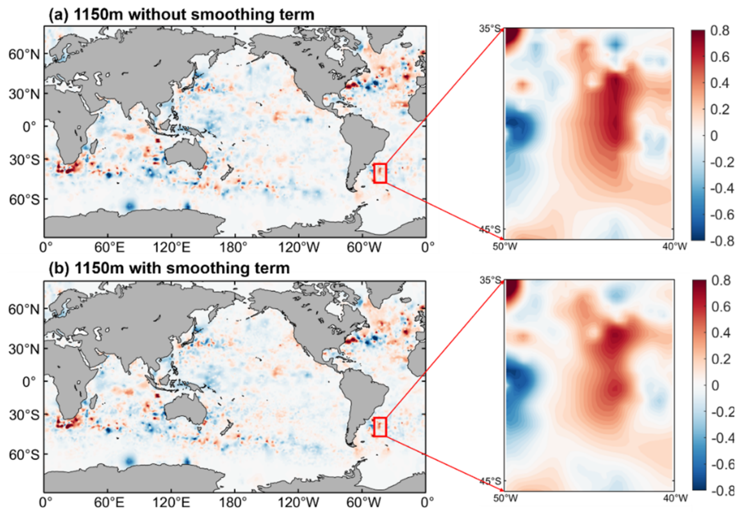

Figure R2. Temperature analysis increments (°C) at 1150 m on 4 January 2014, without (a) and with (b) the smoothing term. The red boxes indicate the enlarged example locations.

3. In any case, the fields obtained with the method at ¼ degree resolution are always smoothed, and this is not stated in the paper.

We thank the reviewer for raising this point. We fully agree that, due to the inherent smoothing of the present objective analysis method, the product fields at 1/4° resolution are indeed smoothed, and this was not explicitly stated in the original manuscript.

As part of our revisions addressing the reviewer's first comment (regarding insufficiently stated limitations), we have already added a clear statement on this issue in Section 5.2. We thank the reviewer again for the valuable suggestion!

4. In verifying the results, the authors use data sets that are not independent, having already been ingested in analyses/models used for comparison (section 3, Validation). This is a very significant weakness and makes the work not yet suitable for publication.

We greatly appreciate the reviewer for pointing out the critical issue of dependent validation datasets, which helps standardize the rigor of our validation framework. It is acknowledged that datasets already ingested into the 4D-MGA product cannot serve as independent references for quantitative validation. The relevant chapter settings and verification schemes have therefore been comprehensively revised to address this concern.

(1) In-situ *T/S* evaluation using independent EN4 (Section 3.1)

In the context of the in-situ *T/S* evaluation, the reference dataset was modified to EN4 (Good et al., 2013), given that none of the 4D-MGA, GLORYS, or ARMOR3D ingest EN4 data. Despite the inevitable overlap of observations, EN4 employs an independent quality control and adjustment system, thereby ensuring a reliable basis for a fair inter-comparison of the three products. As shown in Figures R3–R5, the comparison results using EN4 do not modify our overall conclusions.

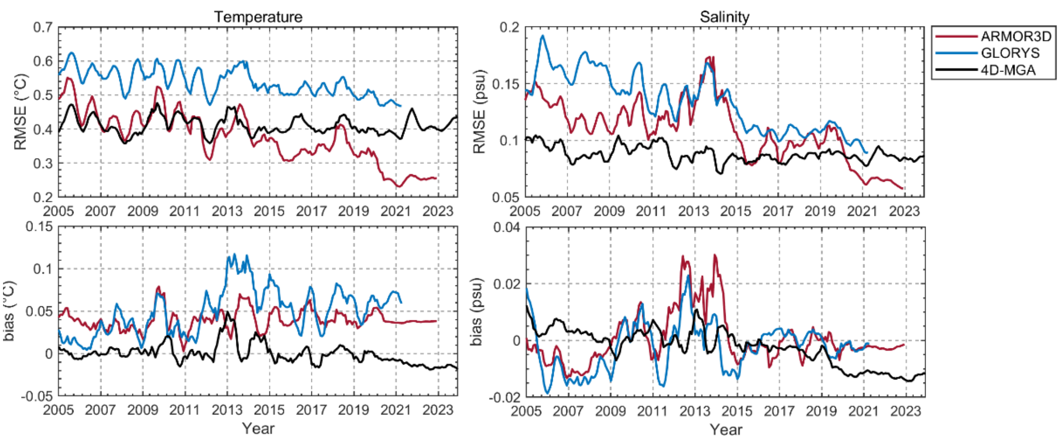

Figure R3. Timeseries of temperature and salinity RMSE and bias compared with EN4 in-situ profiles.

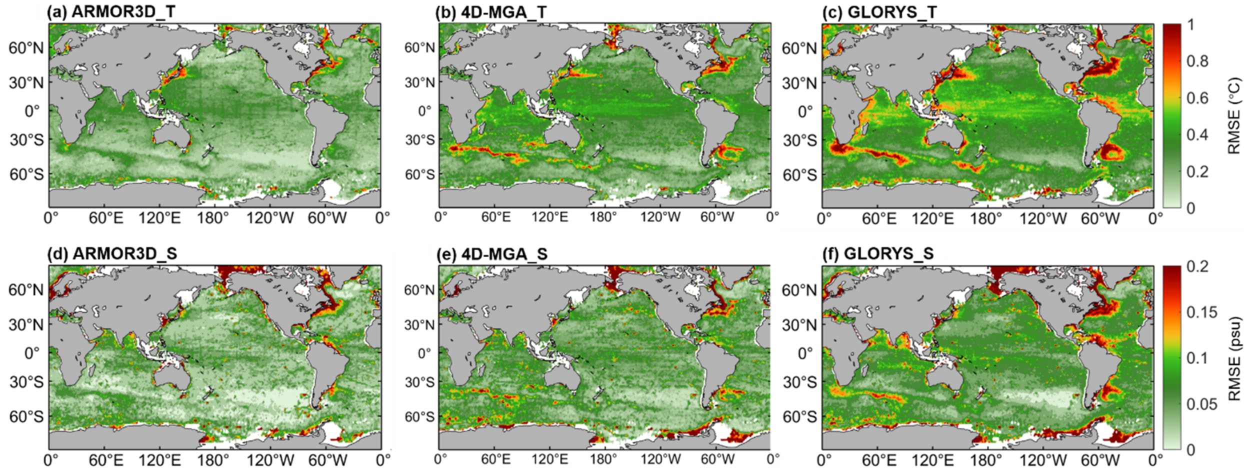

Figure R4. Spatial distribution of the RMSE averaged over 1°×1° bins, the full water column (0–1500 m), and the entire period for ARMOR3D temperature (a), 4D-MGA temperature (b), and GLORYS temperature (c). (d)–(f) are the same as (a–c), but for salinity.

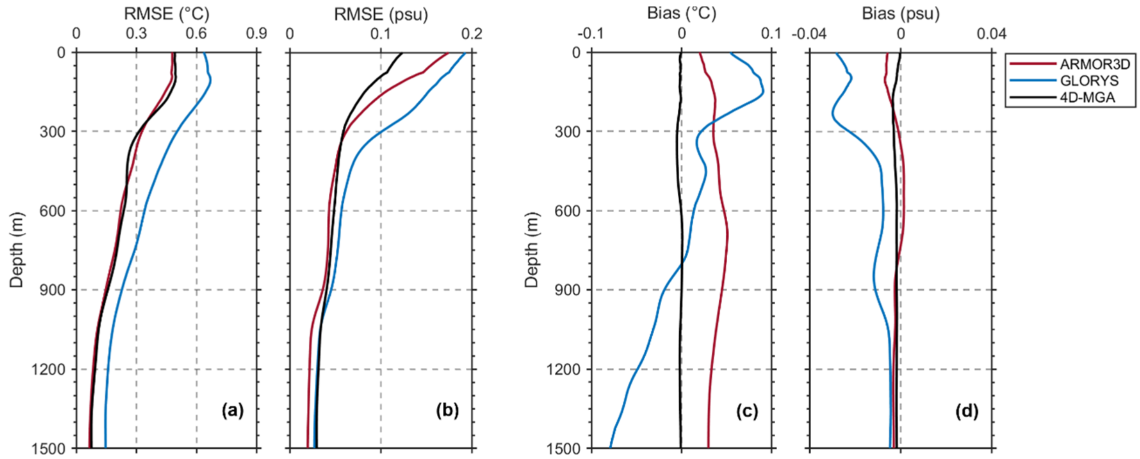

Figure R5. Full-period horizontal averages (smoothed using a moving average) of RMSE (a, b) and bias (c, d) for temperature (a, c) and salinity (b, d) from ARMOR3D, GLORYS, and 4D-MGA.

In the revised manuscript, Figures 2 and 3 (Figures R4 & R5) have been updated to present the EN4-based comparisons, and the corresponding descriptions have been modified accordingly.

(2) The sea surface height anomaly evaluation by independent global tide gauge data (Section 3.2)

For the verification of DHA, we included global tide gauge data as an independent reference and updated Figure 4 accordingly.

The global daily sea level estimated by UHSLC tide gauges (Caldwell et al., 2001) was used as the independent data for evaluation. Here, a total of 129 tide gauges covering 2005–2021 were selected for analysis (Fig. 4b in the revised manuscript). For all stations that were selected, both the tide gauge data and the DHA products were subjected to a process of time-mean value removal, with the objective of aligning them to a common reference level. Subsequently, a 12-

week low-pass filter was then applied to the tide gauge data to suppress high-frequency signals. As shown in Figure R6, the small RMSE of the three products indicates their close comparability to that of the tide gauges, with all achieving a time-averaged RMSE of about 0.06 m. This indicates that the DHA derived from 4D-MGA are consistent with the independent tide gauge observations, thereby substantiating the reliability of our product. The related text has been updated.

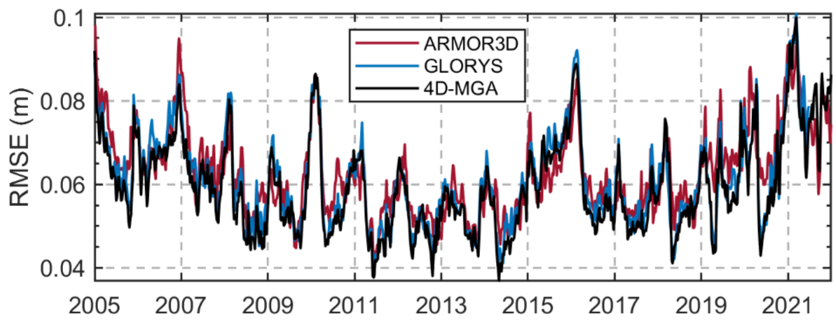

Figure R6 Timeseries of the RMSE of the DHA relative to sea level from UHSLC tide gauges.

(3) The geostrophic currents evaluation by independent OSCAR sea surface currents (Section 3.3)

The Ocean Surface Current Analyses Real-time (OSCAR; ESR, 2009) has been introduced as an independent reference to replace AVISO geostrophic currents in the revised Figures 5 & 6 (Figures R7 & R8). The results show consistency with each other. All related terms in the manuscript have been adjusted.

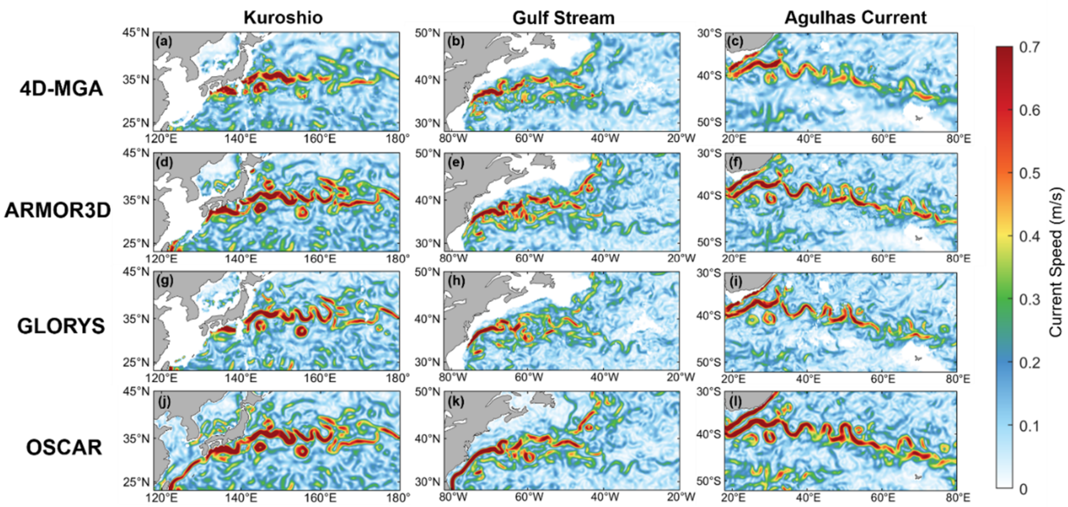

Figure R7. Spatial distribution of geostrophic current speed in selected WBC regions in August 2013. The weekly geostrophic currents from (a–c) 4D-MGA, (d–f) ARMOR3D, (g–i) GLORYS and (j–l) OSCAR are each weighted and averaged to form the monthly mean fields.

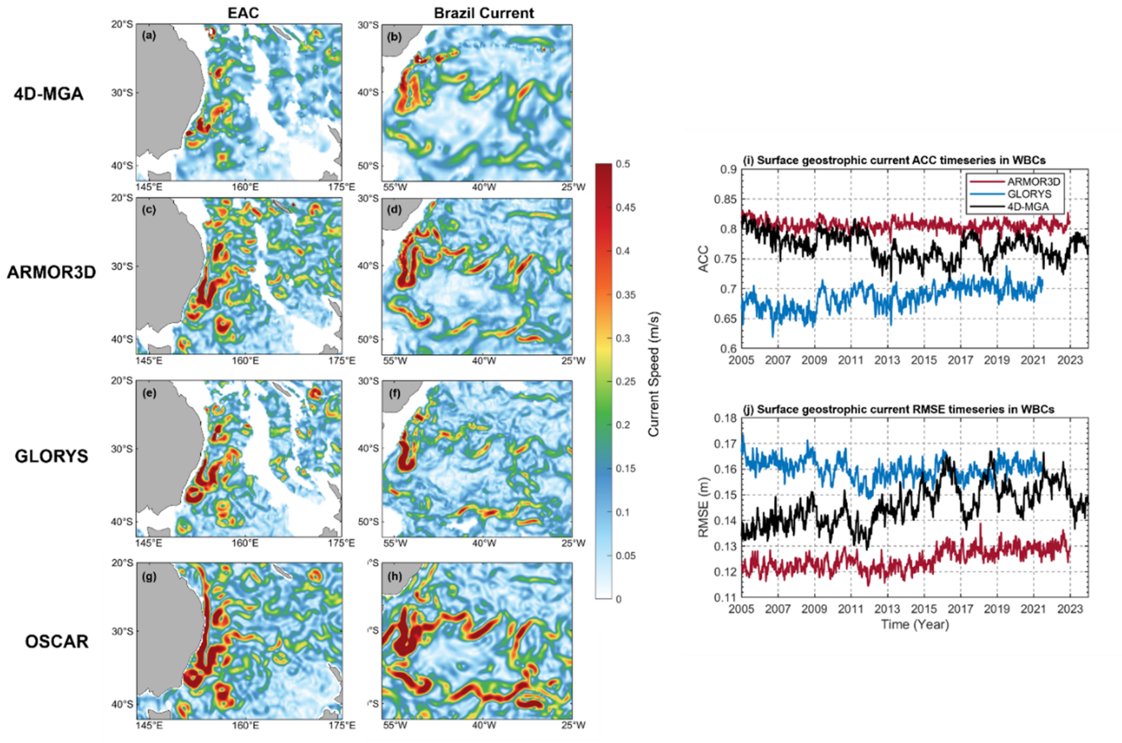

Figure R8 (a–h) Same as Fig. 5, but for the EAC and BC. (i, j) Timeseries of spatially averaged ACC (i) and RMSE (j) for surface geostrophic currents in the WBCs against OSCAR.

(4) Strengthened independent test (Section 5.1)

The independent test in Section 5.1 of the revised manuscript has been extended to the full period 2005–2023 (Figure R9).

The results show robustness. The method can effectively extract core dynamic signals from both small subsets and data-rich environments. Even when only 30% of the in-situ observations are used, the ACC for temperature and salinity (*T* & *S*) still exceeds 0.7, and the averaged RMSEs for *T* & *S* are less than 0.54 °C and 0.11 psu, respectively. The negligibly small mean and standard deviation of the systematic bias further validate its excellent bias-correction capability (Figure R4e–f).

The relevant figures and text have been revised.

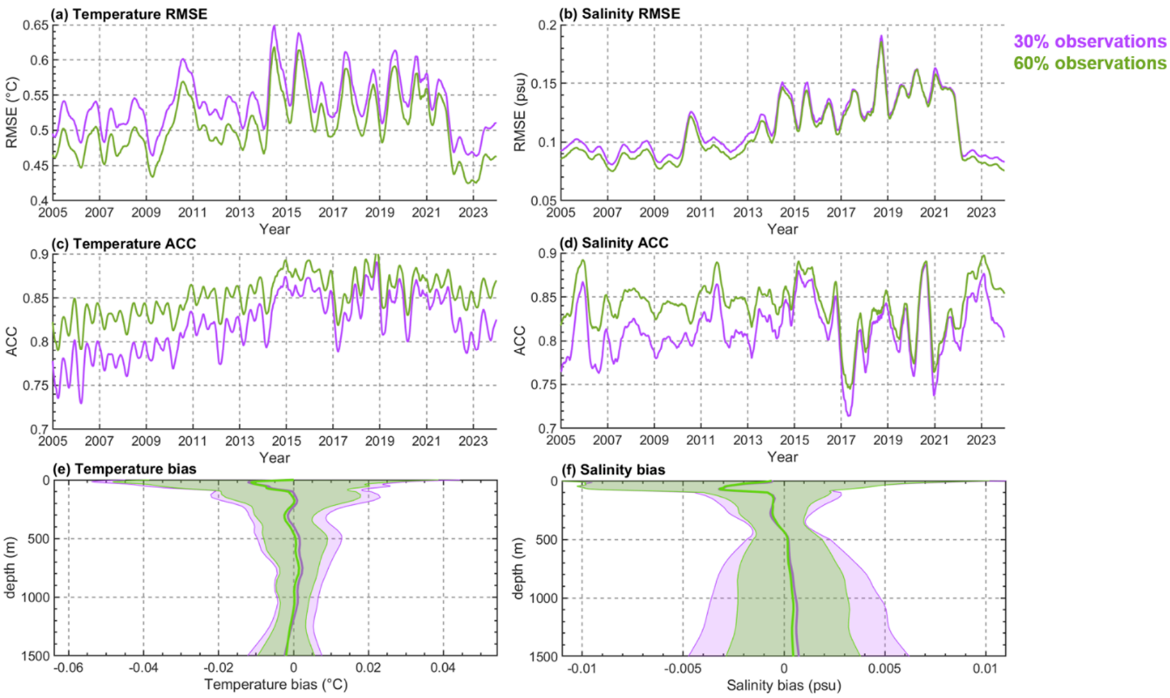

Figure R9. (a)–(d) Time series of RMSE and ACC relative to independent observations. (e, f) Vertical profiles of *T* & *S* bias from the independent tests, solid line indicates the mean bias, and the shaded area represents ±1 STD.

References

Caldwell, P. C., Merrifield, M. A., and Thompson, P. R.: Sea level measured by tide gauges from global oceans as part of the Joint Archive for Sea Level (JASL) since 1846, https://doi.org/10.7289/V5V40S7W, 2001.

ESR: OSCAR third degree resolution ocean surface currents, https://doi.org/10.5067/OSCAR-03D01, 2009.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, JGR Oceans, 118, 6704–6716, https://doi.org/10.1002/2013JC009067, 2013.