'Comment on essd-2026-261', Anonymous Referee #1, 02 May 2026

Li et al. 2026 presented a temperature-salinity dataset with a 4D-MGA method using WOD in weekly 0.25-deg resolutions in the global oceans from 2005 to 2023, compared with datasets from ARMOR3D and GLORYS. The analyzed temperature and salinity distributions shows reasonable magnitude of geostrophic currents at western boundaries. The datasets should be helpful to the community and could be published in ESSD. However, I am unable to accept the manuscript in the present form due to numerous questions when I go through the manuscript.

We are immensely grateful for the reviewer's encouragement and insightful suggestions, which proved invaluable in enhancing the quality of our manuscript. The detailed point-to-point response is provided below in blue.

Major concerns

1. OISST dataset is an observation-based analysis filled to the global ocean, which is not suitable to ingest into the 4D-MGA as observations

We would like to thank the reviewer for highlighting this critical issue. We agree with you. The quantification analysis showed that the influence of the OISST is negligible (Figure R1). The ingesting the OISST was found to result in almost identical analysis increments (Figure R1a–1d), with a minimal improvement in temperature (<0.06°C, Figure R1e).

Consequently, in accordance with the recommendation of the reviewer, a new 4D-MGA product has been developed that does not ingest the OISST dataset. All related figures and contents in the manuscript have been updated.

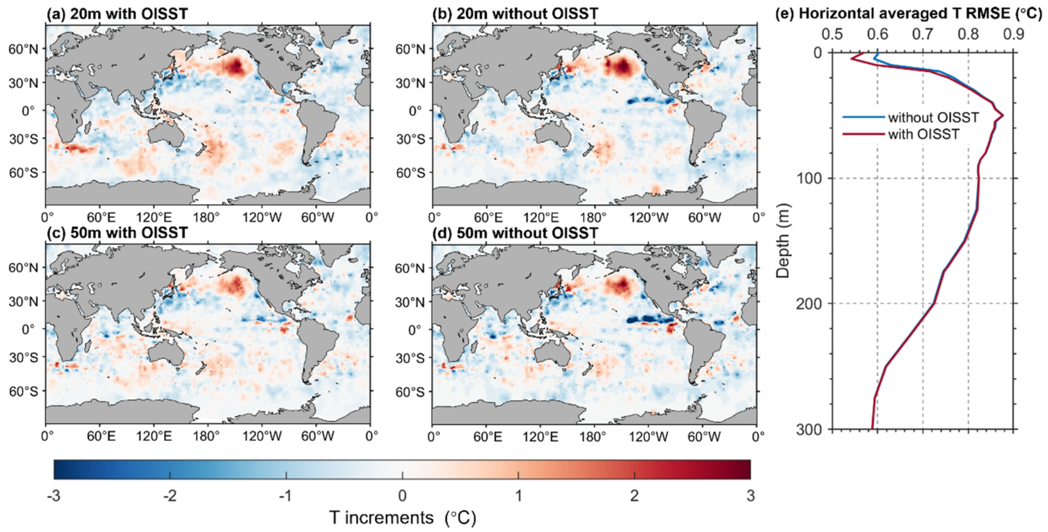

Figure R1. (a)–(d) Temperature analysis increments at 20 m and 50 m on 4 January 2014 for the 4D-MGA product, with and without OISST. (e) Vertical distribution of horizontally averaged root-mean-square error (RMSE) of temperature for the year 2014, relative to independent observations. In this test, 60% of the in-situ observations were used.

2. AVISO SLA is ingested into 4D-MGA while is compared with as independent data

We would like to express our gratitude to the reviewer for highlighting this significant issue. We fully agree with you. To maintain independent, the global daily sea level estimated by UHSLC tide gauges (Caldwell et al., 2001) was used as the independent data for the evaluation.

Here, a total of 129 tide gauges covering 2005–2021 were selected for analysis (Fig. 4b in the manuscript). For all selected stations, 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 gaussian filter is then applied to the tide gauge data to suppress high-frequency signals. As shown in Figure R2, the small root mean square error (RMSE) of the three products show that they are closely comparable 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 demonstrating the reliability of our product. The associated text has been updated (Lines 227–229), and Figure R2 has been incorporated into the text (Figure 4h).

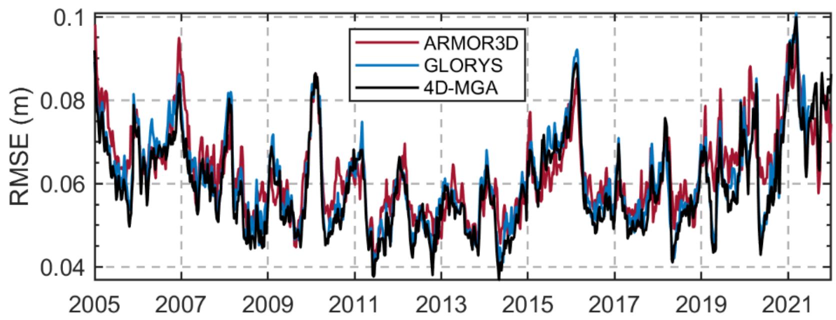

Figure R2 Timeseries of the RMSE of the DHA relative to sea level from UHSLC tide gauges.

Besides, the title "Validation" in Section 3 has been revised to "Results". Nevertheless, the evaluation against AVISO SLA analysis (Figure 4a–g) is still retained in the manuscript. Because both ARMOR3D and GLORYS have also ingested AVISO SLA data. Therefore, it can be hypothesized that the inter-comparison of the three products against the AVISO SLA remains informative. The error levels relative to AVISO SLA analysis can reflect how well each product coordinates these multivariate (temperature, salinity, and SLA) physical constraints. A comparative analysis was therefore conducted, the results of which demonstrate the robust performance of the 4D-MGA product.

Reference

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.

3. Comparisons with ARMOR3D and GLORYS may also be problematic since it is not

mentioned whether they also ingest AVISO and OISST datasets

We're grateful to the reviewer for the reminder! We apologize for the unclear statement in the previous manuscript. Both ARMOR3D and GLORYS products have used SST and SLA data, albeit in different forms. ARMOR3D used OSTIA L4 SST products and AVISO L4 SLA for multivariate linear regression fitting, and GLORYS assimilated along-track AVHRR SST (the primary input to OISST) and AVISO SLA data. Detailed information has been added to Section 2.1.5 of the revised manuscript.

As two excellent representations of eddy-permitted/resolved products, the ARMOR3D and GLORYS have been widely used in the literature of the mesoscale signals analysis (e.g., Zu et al., 2019; Fu et al., 2023; Chaichitehrani and He, 2024; Zhou et al., 2026). Therefore, the selection of the two datasets was made for the purpose of comparing the 4D-MGA products to ascertain whether analogous mesoscale signals can be effectively captured by 4D-MGA.

References

Chaichitehrani, N. and He, R.: Investigation of ocean environmental variables and their variations associated with major Loop Current eddy-shedding events in the Gulf of Mexico, Deep Sea Research Part II: Topical Studies in Oceanography, 213, 105354, https://doi.org/10.1016/j.dsr2.2023.105354, 2024.

Fu, G., Yang, Y., Liang, X. S., and Zhao, Y.: Characteristics and Dynamics of the Interannual Eddy Kinetic Energy Variation in the Central Pacific Sector of the Southern Ocean, JGR Oceans, 128, e2022JC019618, https://doi.org/10.1029/2022JC019618, 2023.

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.

Zhou, Y., Cheng, X., Song, X., and Duan, W.: Springtime Intensification of Mesoscale Eddies in the Southeastern Tropical Indian Ocean, JGR Oceans, 131, e2025JC023723, https://doi.org/10.1029/2025JC023723, 2026.

Zu, Y., Sun, S., Zhao, W., Li, P., Liu, B., Fang, Y., and Samah, A. A.: Seasonal characteristics and formation mechanism of the thermohaline structure of mesoscale eddy in the South China Sea, Acta Oceanol. Sin., 38, 29–38, https://doi.org/10.1007/s13131-018-1222-4, 2019.

4. Comparison against WOA23 appears independent, but actually WOA23 and WOD23 are not independent

We thank the reviewer for raising this concern. We fully agree that WOA23 and WOD23 are not independent, since WOA23 is a global ocean climatological field derived from the historical data of WOD23 through multi-year averaging. The present 4D-MGA product was constructed based on WOD23.

As suggested, all instances of WOA23-related data have been eliminated from the manuscript, including Figures 3 and 9. The text was updated accordingly.

5. Comparisons with selected months and years may misleading, and should be done over the entire period of 2005-2023

We thank the reviewer for this constructive suggestion. Following your suggestion, the independent test has been expanded to encompass the entire 2005–2023 period (Figure R3). In these independent tests, the ingested data included only in-situ *T*/*S* profiles and the SLA product.

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 R3e–f).

All relevant figures and text have been revised (Section 5.1).

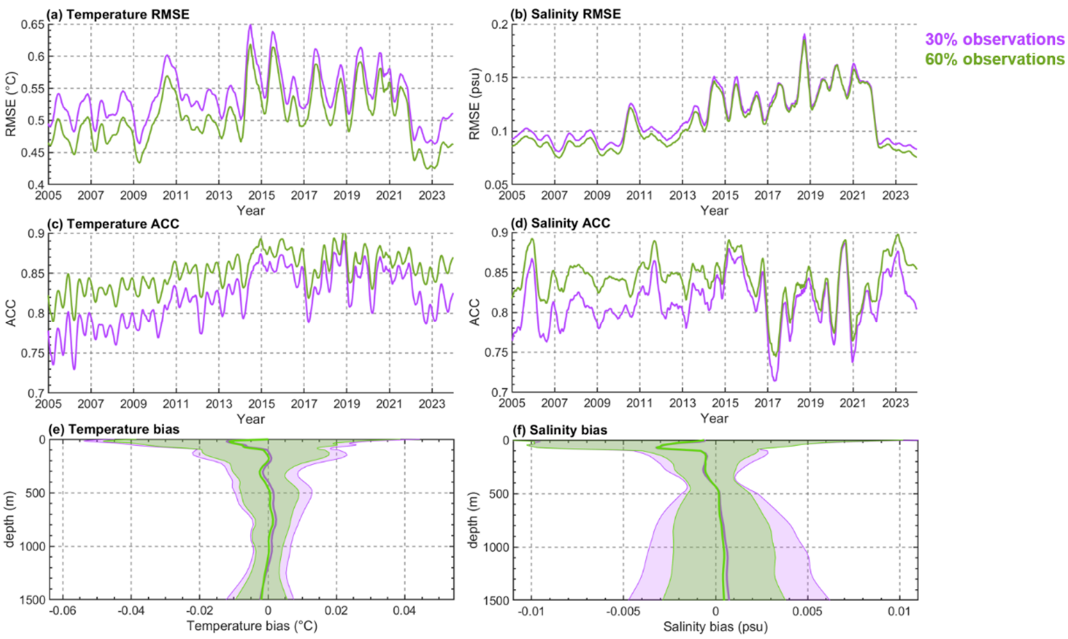

Figure R3. (a)–(d) Time series of RMSE and ACC relative to independent observations. (e, f) Vertical profiles of *T* & *S* bias from the independent tests, solid line indicates the mean bias, and the shaded area represents ±1 STD.

Minor concerns

1. Abstract: It should be stated that 4D-MGA uses WOD24.

Thanks. 4D-MGA has used the temperature and salinity profiles in the latest version WOD23. It has been updated (Lines 9–10).

2. L14, delete “http”

Thanks. Corrected.

3. L14-15, add descriptions of the linear trend of WBCs

Thanks. We have added relevant description of the linear trends of WBCs (Lines 17–18):

“The Kuroshio and Agulhas Current transports exhibit significant decreasing trends across the three products, with rates of 0.34±0.07/0.26±0.06, 0.23±0.05/0.10±0.06, and 0.30±0.06/0.12±0.05 Sv yr^{-1} in 4D-MGA, ARMOR3D, and GLORYS, respectively.”

4. L15-18, it is better to be removed in “Abstract”

Thanks. Removed.

5. In introduction section, should the authors describe the 4D-Var that is popular in data assimilation?

Thanks. As suggested, we have added a description of 4D-Var in data assimilation to the Introduction (Lines 56–59):

“On the other side of the spectrum, as a popular approach in data assimilation, four-dimensional variational (4D-Var; Le Dimet and Talagrand, 1986) uses an adjoint model to directly enforce temporal correlations and deterministic physical constraints. It has been proved that 4D-Var can generate dynamically consistent, mesoscale-resolving ocean estimates (e.g., Forget et al., 2015; Wang et al., 2022). A similar philosophy can be extended to objective analysis.”

6. L86-90, it is problematic to include DOISSTv2.1 and AVISO as “observations” while they are actually observation-based analyses, not real observations.

Thanks. We have moved the description of DOISSTv2.1 and AVISO to a separate subsection titled “2.1.3 Blended satellite-derived products” and renamed section 2.1.2 to “In-situ observations”. Accordingly, all related terms in the manuscript have been adjusted.

7. L112, what is “current grid layer”? Is it the vertical layers of the 4D-MGA system? How many vertical layers has the 4D-MGA, and what is the thickness for each layer?

Thank you for raising this issue. We apologize for the unclear statement. In the 4D-MGA system, the “current grid layer” is not a vertical layer. It refers to “the multigrid layer that is currently being processed”. The key feature of the 4D-MGA system is its multi-grid projection scheme. Once the analysis at the current grid is completed, the results are projected onto the next finer analysis grid and optimization continues based on that resolution. We have revised the relevant description and removed the term “current grid layer” from the revised manuscript accordingly (Lines 167–169):

“The finest vertical layer is divided into the following 33 vertical levels (0.0, 2.5, 7.5, 12.5, 17.5, 22.5, 27.5, 32.5, 37.5, 42.5, 47.5, 52.5, 57.5, 62.5, 67.5, 72.5, 77.5, 82.5, 87.5, 92.5, 97.5, 112.5, 137.5, 187.5, 212.5, 237.5, 262.5, 287.5, 387.5, 487.5, 675.0, 1075.0, and 1500.0 m). The thickness of each layer ranges from 2.5 m at the sea surface to 425m at the deepest layer.”

Thanks again to the reviewer for helping improve the clarity of the manuscript!

8. Figure 1, what is exactly the “adjustment for SST” at different layers?

The term “adjustment for SST” is used to denote the total adjustment relative to the background field (WOA23) after the application of different analysis levels in 4D-MGA. In other words, upon reaching a specific layer, designated as “layer n”, the analysis value represents the cumulative increment from all previous layers. The relevant description to the caption of Figure 1 has been updated.

9. L156, what is the “reference observation”, and where has it been described? Note ARMOR3D and GLORYS in section 2.1.3 can only be used for intercomparison, not as reference observation.

Thank the reviewer for pointing out this issue. The reference data can be in-situ temperature and salinity profiles in WOD23 (now updated to EN4), the AVISO SLA L4 product and the tide-gauge observed SLA. The two datasets (ARMOR3D and GLORYS) are used for intercomparison with reference data. To avoid ambiguity, we have revised Lines 180–181 in the revised manuscript as follows:

“Where X_i denotes the gridded dataset and Y_i is the reference data (including in-situ and satellite derived datasets)”.

10. L166, “WOD23 observations as the reference” cannot be used for Validation since it has been used in 4d-MGA as described in L78.

Thank the reviewer for pointing out this insightful comment. We fully agree with you. We have instead used in-situ T & S from EN4 (Good et al., 2013) for the main comparison, as none of the three products have used EN4. A description of EN4 has been added to Section 2.1.2. The results are consistent. The related figures (Figures 2 & 3) and content have been updated.

Figure R4 shows the RMSE time series for the three products when compared with EN4. It can be seen that the 4D-MGA product performs better.

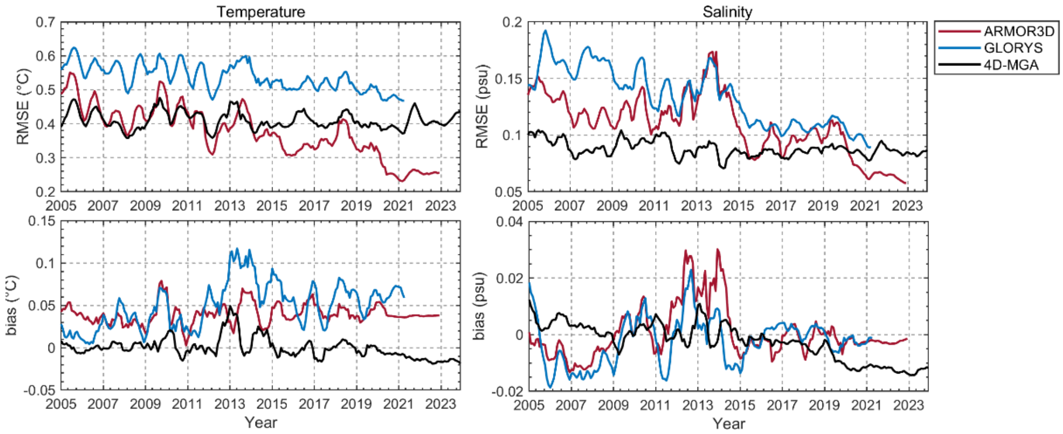

Figure R4. Timeseries of temperature and salinity RMSE and bias compared with EN4 in-situ profiles.

Reference

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.

11. Figure 2. Are the RMSEs averaged from the surface to the bottom of the ocean?

Thank you. You are right. We have added the corresponding clarification in the figure caption.

12. L168-175, the low RMSEs in 4D-MGA should directly result from using the WOD23 since WOD23 is not independent from 4D-MGA, while ARMOR3D and GLORYS may not use the same WOD23 data or with different QC procedures or with different analysis resolutions horizontally and vertically.

Thanks. We agree with you that WOD23 is not independent from 4D-MGA. We have changed the reference dataset to the independent EN4 (Good et al., 2013). EN4 includes observations from different sources and applies its own independent QC procedures. Notably, it was not ingested by any of the three products (4D-MGA, ARMOR3D, or GLORYS). However, it should be acknowledged that most of the in-situ temperature and salinity profiles overlap between different datasets (e.g., EN4 and WOD23). The results still show the low RMSEs in 4D-MGA (Figure R3).

Reference

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.

13. L180, why WOA23 is compared in vertical in Figure 3 but not in Figure 2? Why the RMSEs and biases in WOA23 are much larger than 4D-MGA? “WOA” in Figure legend should be

WOA23.

We thank the reviewer for the questions and suggestions.

As a monthly mean climatology, the errors of WOA23 are much larger than those of the other three datasets (Figure R5). Therefore, WOA23 was not incorporated into Figure 2. The larger errors in WOA23 are attributed to its status as a multi-year average monthly climatology, which lacks the inclusion of time- or mesoscale variability signals. In contrast, 4D-MGA is a weekly-averaged analysis field that can capture real-time ocean variability. Consequently, the RMSE and bias of 4D-MGA relative to independent observations are considerably smaller than those of the climatological field. The WOA depicted in the figure legend is designated WOA23.

As the reviewer suggested, WOA23 is not independent of 4D-MGA. For this reason, the WOA23 comparison has been removed from the figures, and the incorrect reference to WOA in the figure captions has also been deleted.

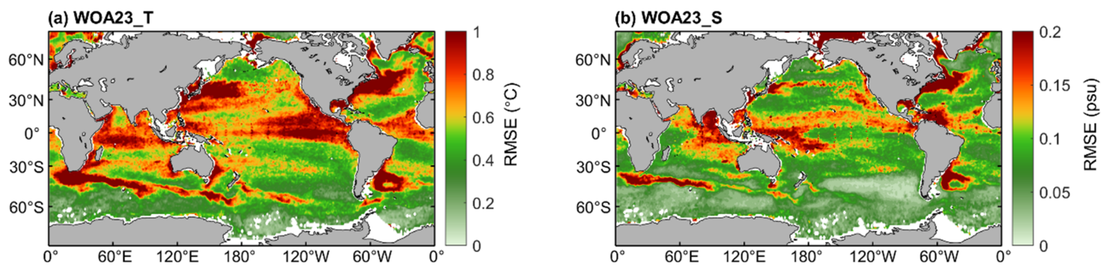

Figure R5 Spatial distribution of the RMSE averaged over 1°×1° bins, the full water column (0–1500 m), and the entire study period for WOA23 temperature (a) and salinity (b).

14. Figure 4, the “SLA observations” is not a correct term, the AVISO SLA analysis or observation-based AVISO SLA analysis should be used.

We thank the reviewer for the correction. “SLA observations” has been updated to “AVISO SLA analysis”.

15. Figure 4, Could the low RMSEs and higher ACCs in 4D-MGA could directly result from the ingestion of DOISST product, while others are not?

We thank the reviewer for the question. DOISST can indeed reduce the analysis error, while the improvement is minor (e.g., Figures R1 & R6).

After excluding the OISST, the new 4D-MGA also shows low RMSEs and high ACCs (Figure R6). Notably, ARMOR3D adopts OSTIA L4 SST and AVISO L4 SLA; GLORYS assimilates along-track L3 AVHRR SST and AVISO SLA as well. It is thus hypothesized that the superior SLA performance of 4D-MGA mainly stems from the direct fitting of AVISO SLA via dynamic height integration and the improved deep-layer density field enabled by ingesting in situ *T* & *S* profiles. The role of DOISST is secondary.

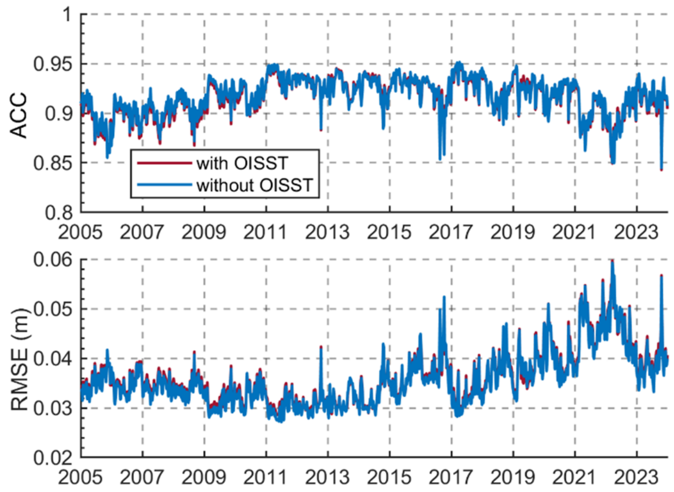

Figure R6. Comparison of DHA derived with and without OISST against AVISO SLA analysis.

16. L227-229, what are the references for these ACCs and RMSEs?

We apologize for the confusion. The reference data is the surface geostrophic current field derived from the AVISO SLA product, which has been clarified in the caption of Figure 4.

17. Figure 5, I doubt it is proper to compare with AVISO, which has been ingested into the 4D-MGA.

Thanks. The independent, observe-based Ocean Surface Current Analyses Real-time (OSCAR; ESR, 2009) was used to compare the three products (Figures R7 & R8). The results are consistent with each other. All related terms in the manuscript have been adjusted (Section 3.3).

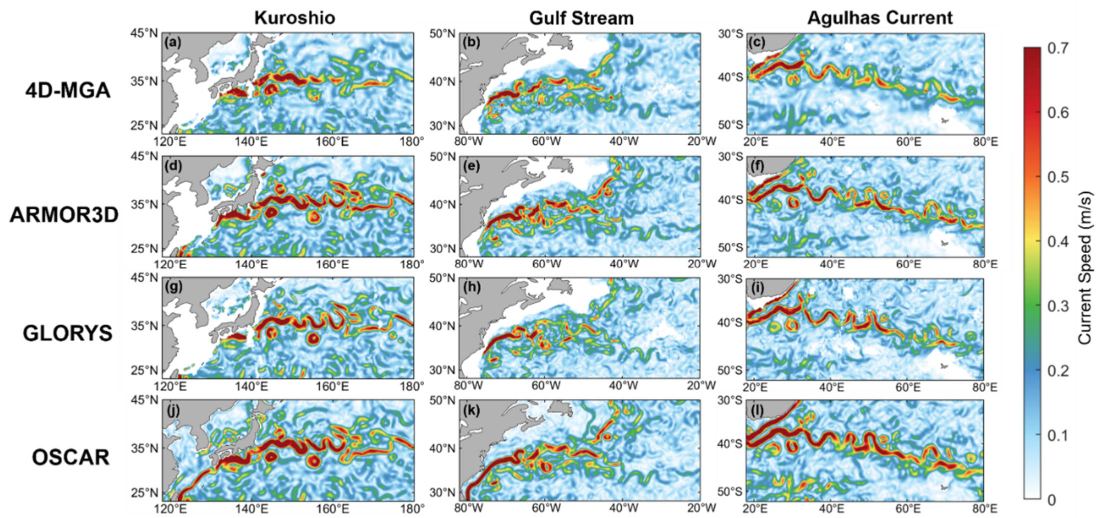

Figure R7. Spatial distribution of geostrophic current speed in selected WBC regions in August 2013. The weekly geostrophic currents from (a–c) 4D-MGA, (d–f) ARMOR3D, (g–i) GLORYS and (j–l) OSCAR are each weighted and averaged to form the monthly mean fields.

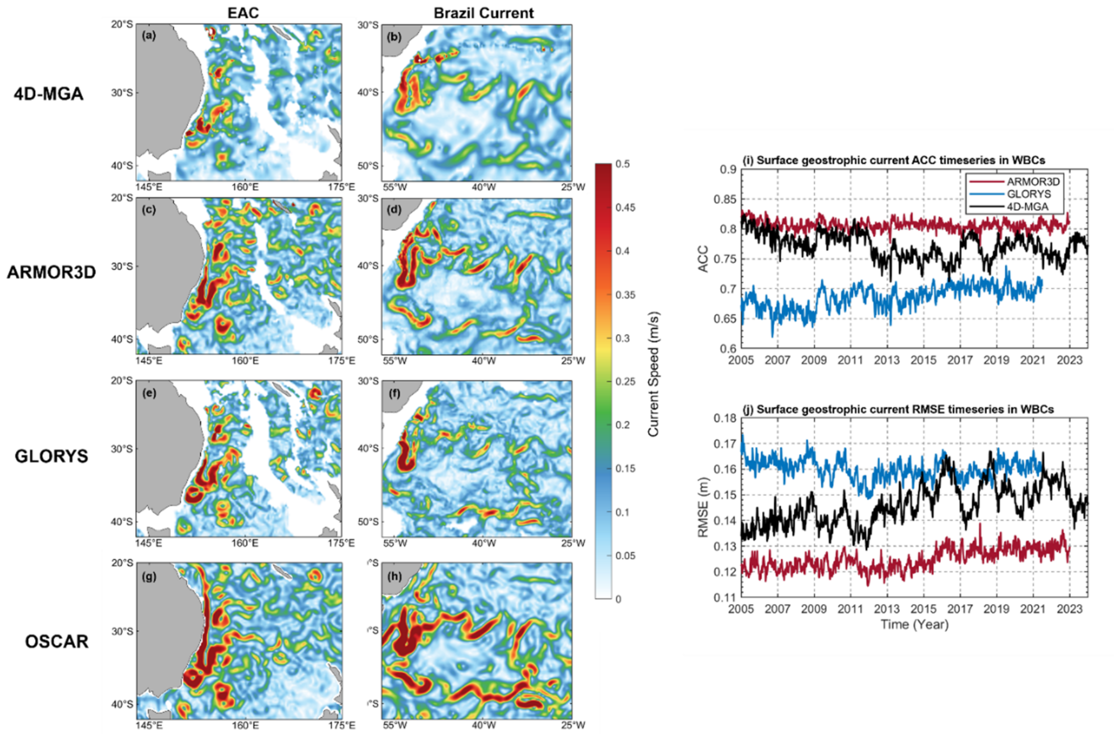

Figure R8 (a–h) Same as Figure R7, but for the EAC and BC. (i, j) Timeseries of spatially averaged ACC (i) and RMSE (j) for surface geostrophic currents in the WBCs against OSCAR.

Reference

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

18. Figure 6 (i,j), What does "WBCs" represent, EAC and BMC? Or includes those in Figure 5? Is "BMC" the Brazil Current, and what does "M" mean?

Thanks. We apologize for the careless statement. WBCs stand for western boundary currents. These include the East Australia Current (EAC) and the Brazil Current (BC). The abbreviation BMC originally stood for the Brazil/Malvinas Current. To avoid confusion, we have changed "BMC" to "BC" (for Brazil Current) in the revised manuscript. The abbreviations for the individual western boundary currents are provided for the first time in the caption of Figure 4.

19. L230, "respectively" in Kuroshio, Gulf Stream, AC, EAC, and BC. It would be better if these currents can be compared with those derived from other products ARMOR3D and GLORYS.

We thank the reviewer for the constructive suggestion. Following the comment, we have added the climatological geographic velocity sections in WBC sections from ARMOR3D and GLORYS to the Supplementary Material as Figures S4–S5. In addition, the following description has been included in the revised manuscript (Lines 273–276):

"Compared with ARMOR3D and GLORYS, the 4D-MGA product shows similar velocity distribution patterns across the five WBC sections, albeit with slightly weaker intensities (Fig. S4 and Fig. S5). The timeseries of globally averaged geostrophic currents in the vertical are further investigated, with all products in good agreement with each other (Fig. S6)."

20. L280, why these month and year are selected?

We thank the reviewer for the question. We comprehensively selected January, April, July, and October of 2005, 2014, and 2023 for the independent validation. The number of available observational profiles differs substantially between these three years (see Figure S2). The selection of these variables enables the assessment of the method's performance under varying observational densities.

In the revised manuscript, we have extended the independent validation to the entire 2005–2023 period.

21. L289, "randomly sampling 30% and 60%", how about the OISST used in 4D-MGA? Should it also be reduced by 30% and 60%?

We thank the reviewer for the question. In this independent test, the OISST was not used. The core objective of this experiment was to test the generalization capability of 4D-MGA when facing sparse and unevenly distributed in situ *T & S* observations. By randomly selecting 30% and 60% of the in-situ profiles, we simulated real-world scenarios with insufficient observations to examine whether the method could reconstruct reasonable *T/S* fields using only a limited number of observations. The OISST has been entirely excluded in the revised manuscript.

22. Figure 8, please clarify the statement "the linear correlation coefficient between the linear trend and the mean anomaly". The abbreviations of KOE, ARC, and BMC should be revised throughout the manuscript as KE, AC, and BC.

We thank the reviewer for suggestion. We apologize for any ambiguity caused. "The linear correlation coefficient between the linear trend and the mean anomaly" refers to the correlation coefficient between the linear trend (dashed lines) and the monthly mean transport (solid lines) shown in the figure. The caption of Figure 8 has been revised.

The abbreviations of KOE, ARC, and BMC have been revised to KE, AC, and BC.

23. Figure 9, (1) quality/resolution is bad, (2) what is "background", (3) what is the reference? Are 4D-MGA compared with independent non-sampled/ingested data? (4) It should be very helpful if the same comparisons are done for ARMOR3D and GLORYS. (5) "shaded area represents ±1 standard deviation" of 30%, 60%, or background? I guess there are three different shadings but are hard to identify. (6) Circles, triangles, and squares are not exactly on the time axis of Jan, Apr, Jul, and Oct, is this intentional designed? (7) What do the error bars represent

for 2005, 2014, or 2023?

We would like to thank the reviewer for highlighting these issues, which will help to improve the quality of the manuscript. The response is below.

(1) Figure 9 has been regenerated to ensure all elements are clear and legible.

(2) "Background" refers to the background field described in Section 2.2.1, i.e., the WOA23 climatology. According to major concern 4, we have removed the WOA23-related figures.

(3) The reference data are independent temperature and salinity profiles that were not integrated into 4D-MGA (i.e., 30% observations were used, and the other 70% observations were non-ingested data).

(4) We agree that it should be very helpful if the same comparisons are done for ARMOR3D and GLORYS. The challenge is that we cannot know which observation profiles were used in ARMOR3D or GLORYS, nor whether they have a similar "reserved independent set" mechanism.

(5) The shaded areas represent ±1 standard deviation for the 30% observation group, the 60% observation group, and the background (WOA23), respectively. The differences between the 30% and 60% groups were very small and the shadings overlapped.

(6) To avoid the circles, triangles, and squares (and their error bars) overlapping entirely at the same month positions, we applied a slight horizontal offset. The actual months corresponding to all data points remain January, April, July, and October.

(7) The error bars represent the standard deviation of the RMSE of all independent validation points within that year and month (i.e., the spread of RMSE values).

We have changed Figures 2 and 3 to show comparisons with EN4. Since none of the three datasets (4D-MGA, ARMOR3D, and GLORYS) were constructed using EN4, EN4 can serve as an independent reference for validation. In addition, we have extended the validation to the full period (2005–2023) and updated the figure accordingly.

24. L299-301, very interesting. Generally speaking, an analysis system can filter out noises and random errors in observations, and the comparison of analysis against those noisy observations may not be problematic. A better way of evaluation is to filter out noises in observations such that the STDs in filtered observations are equivalent to that in analysis (Huang, B., X. Yin, T. Boyer, C. Liu, M. Menne, Y. D. Rao, T. Smith, R. Vose, H.-M. Zhang, 2025: Extended Reconstructed Sea Surface Temperature Version 6 (ERSSTv6): Part II. Upgrades on Quality Control and Large-scale Filter. *J. Climate*, **38**, 1123-1136. https://doi.org/10.1175/JCLI-D-24-0185.1).

Thank the reviewer for the constructive suggestion! We reperformed the independent test using the observation filtering method of Huang et al. (2025). The LS200 filtering method is applied to match the STD of observations with that of the 4D-MGA product. Specifically, the STD (*T*/*S*) values are: observations = 1.22 °C/0.24 psu, product projection = 1.03 °C /0.20 psu, and filtered observations = 0.88 °C /0.16 psu. The STD equivalence criterion is met.

The updated independent validation results are presented in Figure R3. Text has been revised in Section 5.1, and Huang et al (2025) has been cited in the manuscript.

25. L310, what are “temporal grid layers”? are they same as “temporal layers” in section 2.3?

We thank the reviewer for the question. Yes, they are the same. We apologize for the confusion caused by the inconsistent terminology. This has now been unified to “temporal layers” throughout the revised manuscript.

'Comment on essd-2026-261', Referee #2, Giuseppe M.R. Manzella, 04 Jun 2026

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 (Lines 336–342).

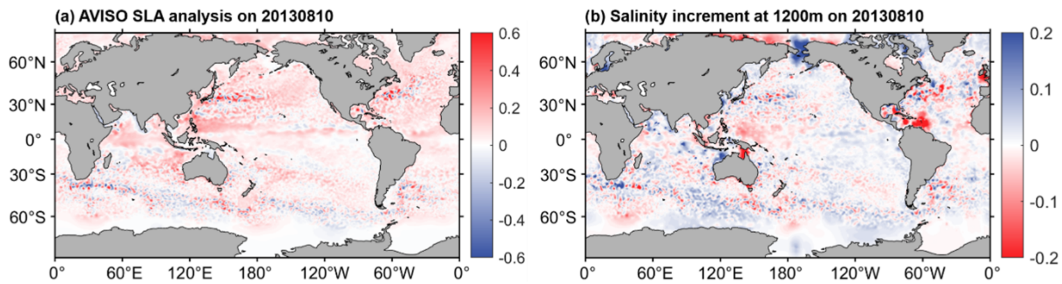

Figure R1. (a) AVISO SLA analysis (m) and (b) salinity increment at 1200 m relative to WOA23, both on 10 August 2013.

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 (Lines 130–132).

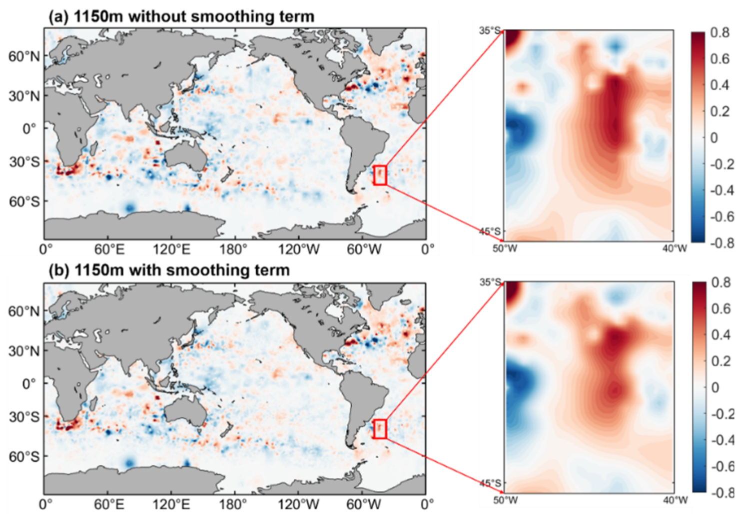

Figure R2. Temperature analysis increments (°C) at 1150 m on 4 January 2014, without (a) and with (b) the smoothing term. The red boxes indicate the enlarged example locations.

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.

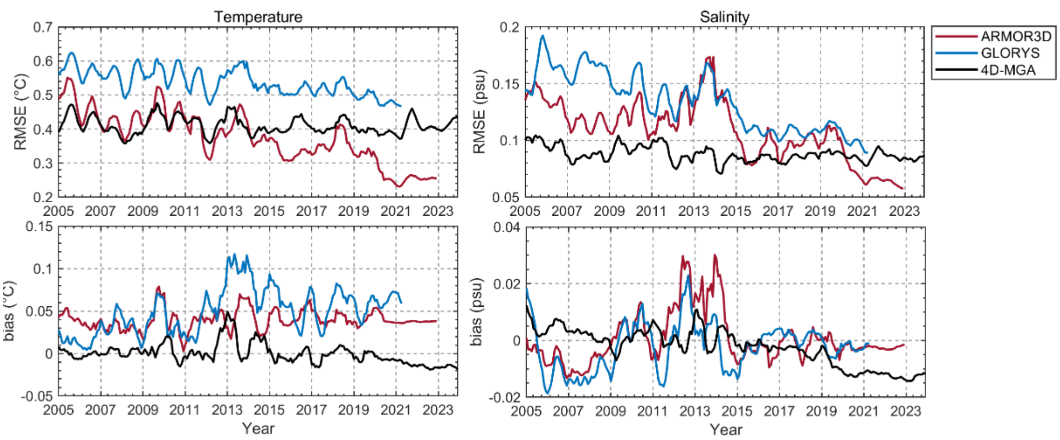

Figure R3. Timeseries of temperature and salinity RMSE and bias compared with EN4 in-situ profiles.

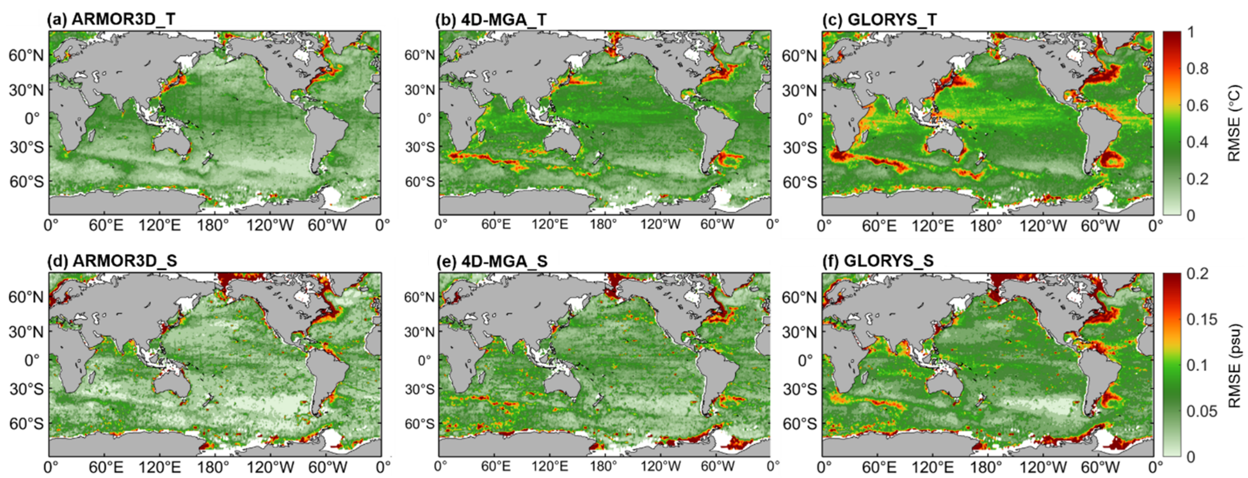

Figure R4. Spatial distribution of the RMSE averaged over 1°×1° bins, the full water column (0–1500 m), and the entire period for ARMOR3D temperature (a), 4D-MGA temperature (b), and GLORYS temperature (c). (d)–(f) are the same as (a–c), but for salinity.

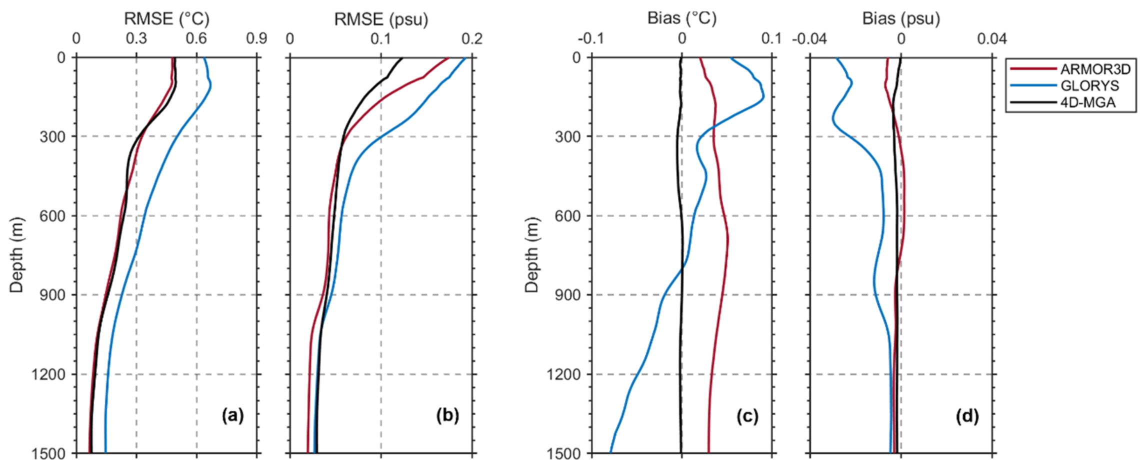

Figure R5. Full-period horizontal averages (smoothed using a moving average) of RMSE (a, b) and bias (c, d) for temperature (a, c) and salinity (b, d) from ARMOR3D, GLORYS, and 4D-MGA.

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.

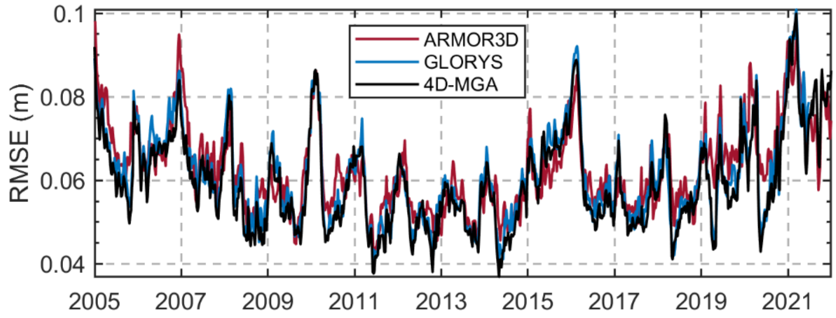

Figure R6 Timeseries of the RMSE of the DHA relative to sea level from UHSLC tide gauges.

(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.

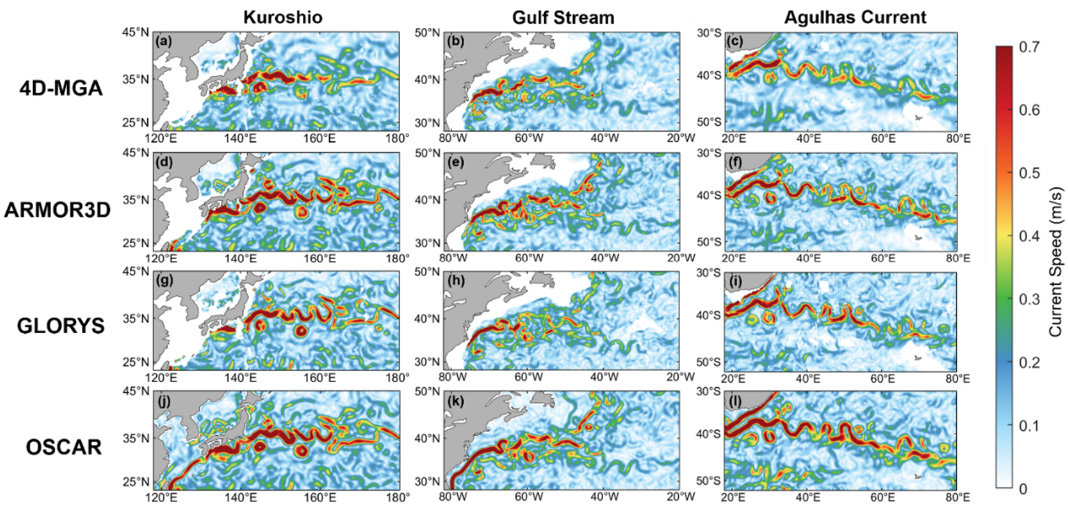

Figure R7. Spatial distribution of geostrophic current speed in selected WBC regions in August 2013. The weekly geostrophic currents from (a–c) 4D-MGA, (d–f) ARMOR3D, (g–i) GLORYS and (j–l) OSCAR are each weighted and averaged to form the monthly mean fields.

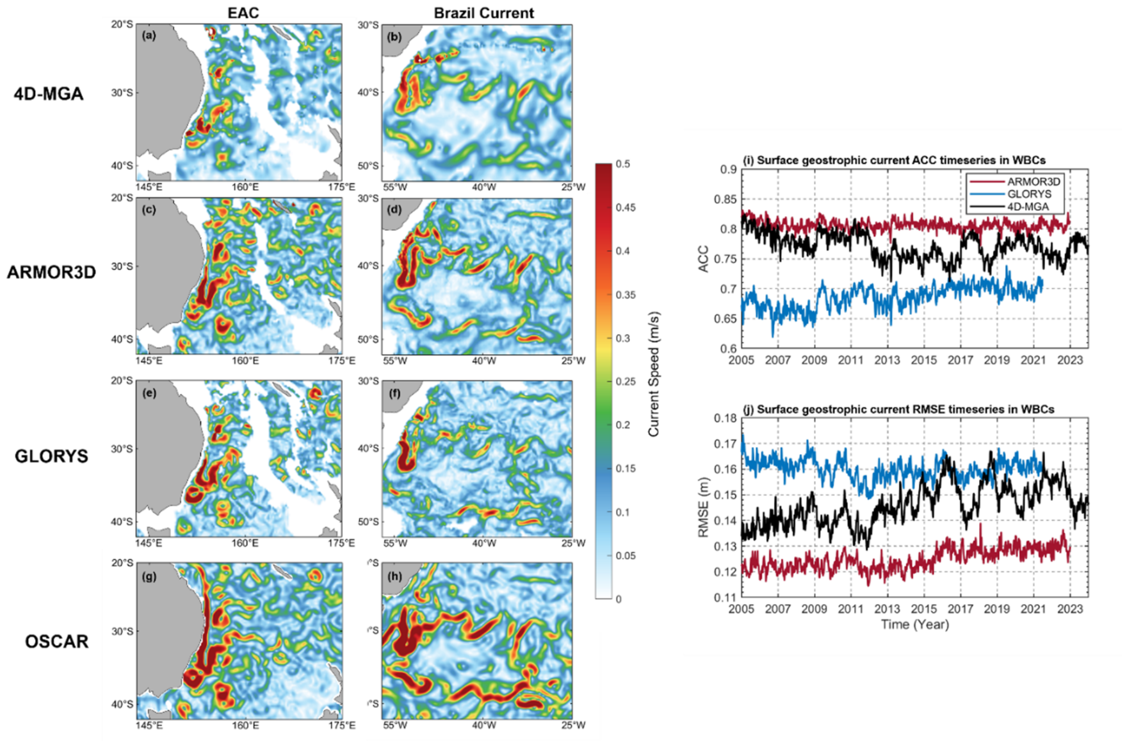

Figure R8 (a–h) Same as Fig. 5, but for the EAC and BC. (i, j) Timeseries of spatially averaged ACC (i) and RMSE (j) for surface geostrophic currents in the WBCs against OSCAR.

(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.

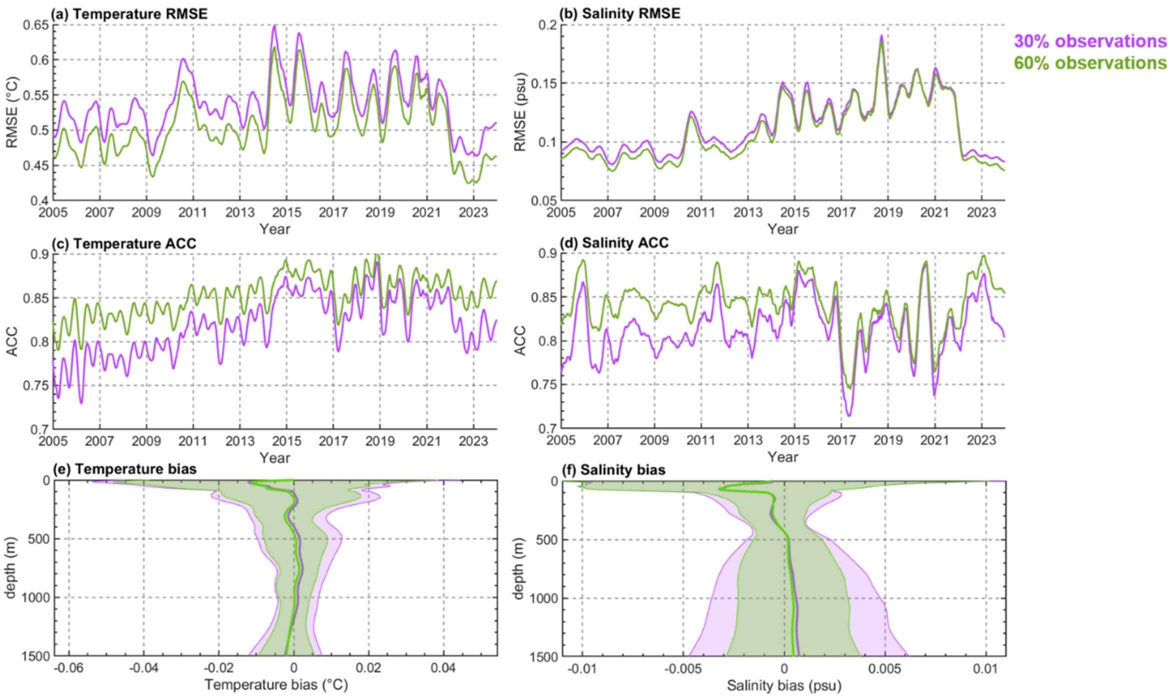

Figure R9. (a)–(d) Time series of RMSE and ACC relative to independent observations. (e, f) Vertical profiles of *T* & *S* bias from the independent tests, solid line indicates the mean bias, and the shaded area represents ±1 STD.

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.