Review of "LETKF-based Ocean Research Analysis version 2.0 for a quasi-global domain (LORA-QG): Validation and intercomparison with eddy-permitting global ocean reanalysis datasets" by Ohishi et al.

This manuscript presents the development of the LETKF-based Ocean Research Analysis version 2.0 (LORA-QG), a quasi-global ocean reanalysis covering the period from June 2002 to January 2024. The product is validated against existing ocean reanalyses, including C-GLORSv7, GLORYS2V4, and ORAS5, and evaluated using drifter observations of sea surface temperature (SST) and surface currents, as well as temperature and salinity observations from the KEO and Papa mooring stations.

The development of a new ocean reanalysis is an important contribution to the ocean modelling and data assimilation community. However, in its current form, the manuscript does not clearly demonstrate the scientific advances or added value of LORA-QG relative to existing state-of-the-art ocean reanalysis products. The validation is limited, the diagnostics are not sufficiently comprehensive, and the discussion does not adequately explain several notable differences between LORA-QG and the other products. Consequently, the manuscript does not yet provide sufficient evidence of the advantages or unique capabilities of the proposed reanalysis. I therefore do not recommend publication in its present form.

The following comments are intended to help improve the manuscript.

We thank the reviewer for the constructive comments on the overall scope and contribution of the manuscript. We address the reviewer's specific points below.

Major Comments

1. Ocean reanalysis datasets (Table 2)
Table 2 is incomplete and not up to date. Several widely used operational and research ocean reanalysis products are missing, including:

GloSea5 (UK Met Office)
NCEP Global Ocean Data Assimilation System (NCEP-GODAS, NOAA/NCEP, USA)
IGORA v1 (INCOIS Global Ocean Reanalysis Version 1, India)
Other recent operational ocean reanalyses listed at: https://reanalyses.org/ocean/overview-current-reanalyses

Therefore, the statement that the table represents the "best of the authors' knowledge" is not justified and should be revised.

We thank the reviewer for pointing out these additional datasets. We have carefully reviewed each of the suggested products.

NCEP-GODAS and IGORA v1 have horizontal resolutions of 1.0° longitude × 0.33° latitude and 0.5° longitude × 0.25–0.5° latitude, respectively, which are coarser than the eddy-permitting resolution adopted as the criterion for Table 2. Regarding GloSea5, we note that it is primarily a coupled seasonal forecast system, in which the ocean data assimilation component serves to initialize seasonal forecasts, rather than to produce a standalone ocean reanalysis dataset; therefore, we consider it to be outside the scope of Table 2. We also confirmed the list on the website the reviewer suggested and found two additional eddy-permitting global products, UR025.3 and UR025.4, produced by the University of Reading under the MyOcean project (1989–2010). We have not included these two products in Table 2, as the MyOcean project has transitioned into the Copernicus Marine Service (CMEMS), whose current quasi-global and global eddy-permitting reanalysis products are already represented in Table 2 (GLORYS2V4, ORAS5, and C-GLORSv7). Therefore, we have added a clarification of the dataset selection criteria to the first paragraph of Section 1, and retained the original descriptions and tables for eddy-resolving and eddy-permitting systems otherwise.

2. Ensemble size

The rationale for selecting 128 ensemble members is not explained. Was a sensitivity analysis performed? Why was this number chosen instead of 64, 96, or 256 members? Please justify this choice and discuss its impact on analysis quality and computational cost.

As reported by previous studies (e.g., Miyoshi et al., 2014; doi:10.1002/2014GL060863), a larger ensemble size generally improves the accuracy of the ensemble-based analysis by better representing the forecast error covariance and reducing sampling error. In this study, the maximum ensemble size is constrained by the maximum number of concurrent background jobs (128) allowed per job submission on the Fugaku and JSS3 supercomputers, and we therefore set the ensemble size to 128. This choice is also consistent with the ensemble size used in our previous regional systems (LORA-WNP and LORA-MC; Ohishi et al., 2022a, b, 2023, 2024a, b). Since larger ensemble sizes are expected to improve, rather than degrade, analysis accuracy, we did not conduct additional sensitivity experiments with smaller ensemble sizes (e.g., 64 or 96 members), as such experiments would be expected to yield lower accuracy without providing additional benefit for this study. We have added the following description to the end of Sect. 2.1.1:

Revised:

The maximum ensemble size in this study is constrained by the maximum number of concurrent background jobs allowed per job submission on the Fugaku supercomputer at RIKEN and the Japan Aerospace Exploration Agency (JAXA) Supercomputer System Generation 3 (JSS3). We therefore set the ensemble size to 128, consistent with our previous regional systems (Ohishi et al., 2022a, b, 2023, 2024a, b).

3. Mean Dynamic Ocean Topography

The manuscript states that the mean dynamic ocean topography (MDOT) is estimated from simulated sea surface height averaged over 1997–2001. Please discuss the reliability of this approach. Why was a model-derived MDOT used instead of an observation-based product such as CNES-CLS MDT?

We have confirmed that using an observation-based MDOT (e.g., CNES-CLS MDT) results in substantial biases and tends to violate the no-bias assumption of the EnKF. We therefore estimated the MDOT from the simulated SSH averaged over 1997–2001, during which full-depth temperature and salinity nudging toward the WOA18 monthly climatology (30-day timescale) was applied to keep the simulation close to the observational climatology. We have added the following descriptions of the spin-up to the end of Sect. 2.1.1.

Revised:

The ensemble simulation (i.e., spin-up) was conducted from an initial state of no motion for the period from January 1996 to May 2002, using the perturbed atmospheric and lateral boundary conditions and applying the full-depth temperature and salinity nudging with a 30-day timescale toward the monthly climatological dataset from the World Ocean Atlas 2018 (WOA18; Locarnini et al., 2018; Zweng et al., 2018).

4. Surface restoring

The manuscript clearly states that sea surface salinity (SSS) is relaxed toward observations. However, it is unclear whether SST is also restored or relaxed during the model integration. Please clarify the surface boundary conditions applied to SST.

We revised the text in the third paragraph in subsection 2.1.2 as follows:

To suppress SSS drift, we applied salinity nudging toward the WOA18 monthly climatology with a 30-day timescale in the mixed layer, whereas no nudging was applied to temperature.

5. Large subsurface bias in the equatorial region

Figure 4 shows considerably larger temperature and salinity biases in the equatorial ocean compared to the other reanalysis products. The explanation provided (Lines 223–225) appears insufficient.

Please discuss whether these larger biases could also result from:

the relatively short model spin-up period,

deficiencies in the data assimilation configuration,

model physics,

vertical mixing parameterization, or

sparse subsurface observations.

A more comprehensive discussion is needed.

We first clarify the extent of the biases referred to by the reviewer. In the equatorial region, LORA-QG exhibits larger SST biases than the other three datasets (Fig. 10a), but its ocean interior temperature biases are of comparable magnitude to the other datasets (Fig. 4b). In contrast, LORA-QG has the smallest SSS biases (Fig. 5b), while its ocean interior salinity biases are substantial (Fig. 6b). Thus, the notably large biases in LORA-QG in the equatorial region are limited to SST and ocean interior salinity. We have added further possible factors to the discussions of the SST and ocean interior salinity biases in the fifth paragraph of subsection 3.2 and the second paragraph of subsection 3.1, respectively.

Revised (subsection 3.2):

The bias differences of LORA-QG with GLORYS2V4 and ORAS5 may reflect differences in both the ocean model and assimilation configurations (Table 2), such as vertical mixing parameterizations and data assimilation schemes. In particular, the warm biases of LORA-QG may be related to limitations in microwave SST observations. Microwave SST retrievals are generally considered to represent the subskin temperature and can differ from the skin temperature observed by infrared sensors. In addition, microwave SST retrievals are affected by precipitation and can introduce systematic errors under heavy rainfall conditions. The spatial distribution of the warm biases in LORA-QG is broadly consistent with that of the rainfall climatology (cf. https://gpm.nasa.gov/data/imerg/precipitation-climatology), suggesting that precipitation-related retrieval errors may contribute to the warm biases. In contrast, the other three datasets assimilate optimally interpolated SST products that combine infrared and microwave satellite observations, which may help mitigate such precipitation-related errors.

Revised (subsection 3.1):

These temperature and salinity biases in the ocean interior are likely associated with the limited availability of observations, data assimilation schemes, and model-related factors, including the vertical localization in LETKF, the specification of lateral boundary conditions from SODA3, limitations in representing high-latitude processes such as Antarctic Bottom Water formation, and differences in vertical mixing schemes: the Nakanishi and Niino (2009) level 2.5 scheme used in LORA-QG and the turbulence kinetic energy (TKE)-based level 1.5 closure scheme used in the other datasets.

6. Inconsistency between text and figure

The discussion in Lines 257–259 (Page 18) does not appear to agree with the results shown in the corresponding figure. Please revise either the figure interpretation or the text.

We re-examined Fig. 8b–d and confirmed that the text of the second paragraph of subsection 3.2 is consistent with the figure: positive biases in the subtropics weaken the westward currents, and negative biases along the mid-latitude western boundary currents weaken the eastward currents. We therefore retained the original text as below.

Original:

GLORYS2V4, ORAS5, and C-GLORSv7 exhibit substantial biases that tend to weaken the large-scale general circulation, with positive and negative biases weakening the westward and eastward currents in subtropical and mid-latitude regions, respectively (Fig. 8b–d), whereas LORA-QG does not show such systematic bias patterns (Fig. 8a).

7. Regional differences in RMSD

The manuscript states:

"The RMSDs of LORA-QG are significantly larger than those of the other three datasets in the tropics, especially in the western tropical Pacific, while they are significantly smaller around the western boundary currents and the Antarctic Circumpolar Current."

This is an important result but is not adequately explained. Please discuss the physical or methodological reasons behind these contrasting regional performances.

We have added the discussion at the end of the fourth paragraph in subsection 3.2 as follows:

Revised:

Flow-dependent forecast error covariance in LORA-QG and GLORYS2V4 might capture the mesoscale variations better and result in the smaller RMSDs around the western boundary currents and ACC compared with the static error covariance in ORAS5 and C-GLORSv7.

8. Accuracy of drifter observations

Please provide the accuracy and uncertainty of the drifter observations used for validating SST, zonal current, and meridional current. This information is necessary to assess the significance of the reported differences.

As far as we could determine, the accuracy and uncertainty of the surface zonal and meridional velocities are not explicitly reported in Elipot et al. (2016), although the uncertainty in the drifter locations is specified. In Elipot et al. (2022), the SST uncertainty is reported as 0.031°C and 0.036°C for drogued and undrogued drifters, respectively. However, we have not added these values to the manuscript because the observation errors are not directly relevant to the statistical significance of RMSD differences, as explained below.

Assuming that the observation and forecast/analysis errors are unbiased, the squared analysis RMSD can be written as

$$(RMSD^a)^2 = Var(\boldsymbol{\epsilon}^o) + Var(\mathbf{H}\boldsymbol{\epsilon}^a) - 2Cov(\boldsymbol{\epsilon}^o, \mathbf{H}\boldsymbol{\epsilon}^a),$$

where $\boldsymbol{\epsilon}^o$ and $\boldsymbol{\epsilon}^a$ are the observation and analysis errors, respectively, and $\mathbf{H}$ is the linear observation operator. Since drifter buoy velocities are independent of the four analysis datasets, the last term on the right-hand side vanishes. Consequently, when comparing the RMSDs among the analysis datasets, the contribution of the observation error variance is common to all datasets and therefore cancels out. Thus, the RMSD differences primarily reflect analysis accuracy differences. However, this independence assumption does not hold for the comparison of SST RMSDs with GLORYS2V4, because GLORYS2V4 assimilates in situ SST observations from CORA, which includes drifter buoy observations.

9. Sea Level Anomaly (SLA)

The methodology for computing SLA is not adequately described.

Please clarify:

How was SLA calculated for LORA-QG?

Were identical reference periods used for all reanalysis products?

Were all datasets processed consistently before comparison?

In addition, spatial maps alone are insufficient. Time-series comparisons at representative locations should be included together with statistical metrics such as:

Mean

Standard deviation

Bias

RMSE

Correlation coefficient

To clarify that the same procedure is applied to the observations and analysis datasets, we revised the description of the SLA calculation at the end of the first paragraph of Sect. 3.3 as follows.

We have already shown temporal series of sea level anomalies for the three representative stations in the original manuscript (Fig. 13). As is clear from the above procedure, the mean and bias are zero for all datasets by construction, and we do not report these separately. The squared analysis RMSD can be written in another form:

$$(RMSD^a)^2 = Var(\boldsymbol{y}) + Var(\mathbf{H}\boldsymbol{x}^a) - 2Cov(\boldsymbol{y}, \mathbf{H}\boldsymbol{x}^a),$$

where y and $\boldsymbol{x}^a$ are the observation and analysis, respectively, so that the analysis RMSD inherently includes the contributions from the standard deviations of the observations and analyses and their cross-correlation coefficient. Therefore, the analysis RMSD provides a comprehensive summary of the overall agreement between the analyses and observations, and we consider additional decomposition into these individual components unnecessary for the scope of this validation.

Revised:

Here, at each station, the analysis SSH is defined as the model grid value closest to the station within a 100 km radius, and the reference level is set to 0 m for all observation and analysis datasets by subtracting each own temporal SSH average over the period when all observation and analysis datasets exist.

10. KEO and Papa buoy validation

Very little information is provided regarding the KEO and Papa buoy observations.

Please include:

observation period,

sampling frequency,

number of available profiles,

quality control procedures.

The validation should also include comparisons of:

mean temperature profiles,

mean salinity profiles,

standard deviation,

correlation coefficient,

We have added the observation periods and the sampling intervals to the first paragraphs of subsection 3.4, and the number of available profiles can be readily estimated using the observation frequency (Fig. 14b–i). We have also added the quality control procedures to the first paragraph of subsection 2.2.

Regarding the mean temperature and salinity profiles, the analysis biases shown in Figs. 15b, 16b, 17b, and 18b represent the mean profile differences between the analyses and observations; we therefore consider the existing bias profiles to convey the same information as a direct comparison of mean profiles. As described in our response to comment 9, we also consider that the analysis RMSDs reflect the standard deviations and correlation coefficient.

Revised (observation period and sampling intervals in subsection 3.4):

Based on daily-mean observations of ocean interior temperature, salinity, zonal velocity, and meridional velocity observed from the independent KEO and Papa buoys (Fig. 14a), we calculated temporally averaged analysis RMSDs and biases (panels a and b of Figs. 15–18, respectively), temporally and depth-averaged analysis RMSDs (triangles at the bottom of panel a in Figs. 15–18), and the corresponding RMSD ratios and absolute bias differences (panels c and d of Figs. 15–18). The observations at the KEO buoy are available from 16 June 2004 for temperature and salinity and from 30 May 2005 for horizontal velocity, and those at the Papa buoy are available from 8 June 2007 for temperature, salinity, and velocity. Data points with less than 20% availability are excluded (Fig. 14b–i).

Revised (quality control procedures in subsection 2.2):

When quality-control flags are available, observations flagged as bad quality are excluded from the validation.

11. Evaluation over climatically important regions

The manuscript would be considerably strengthened by evaluating the product over major climate-sensitive regions.

Examples include:
Niño-3.4 region (Pacific Ocean),
Thermocline Ridge/Dome region (Indian Ocean),
western tropical Pacific warm pool.

Time-series comparisons with drifter observations or independent datasets should be presented for SST and surface currents, together with statistical measures (mean, standard deviation, bias, RMSE, and correlation coefficient).

In addition, it would be valuable to demonstrate how well LORA-QG reproduces important climate indices such as the Niño-3.4 SST index.

We agree that regional, process-oriented diagnostics such as ENSO indices and thermocline ridge/dome dynamics might be valuable applications of LORA-QG. However, the scope of the present study is to validate the overall accuracy of LORA-QG through spatially and temporally averaged biases and RMSDs relative to independent observations, in comparison with existing global/quasi-global reanalysis products. It is beyond the scope of this study to assess the skill of LORA-QG in reproducing specific climate phenomena or indices. We note that the western tropical Pacific warm pool region is already discussed in the manuscript in relation to the pronounced warm SST biases in LORA-QG.

12. Subsurface validation using EN4

The manuscript evaluates subsurface temperature and salinity only against observations from two mooring sites.
A more comprehensive assessment should also include comparisons with widely used objective analysis products such as EN4 over climatically important regions. This would provide a more robust evaluation of the basin-scale performance of LORA-QG and facilitate comparison with existing ocean reanalysis products.

Since the EN4 includes in-situ observations assimilated in all analysis datasets, it cannot provide an independent validation. Consequently, we used the KEO and Papa buoy observations to ensure independence, although this limits the spatial coverage of the validation.

Overall Recommendation

The manuscript describes the development of a new ocean reanalysis product with a long analysis period, which is potentially valuable for the oceanographic community. However, the current validation is insufficient to demonstrate a clear improvement over existing operational reanalysis products. More comprehensive validation, improved diagnostics, deeper scientific interpretation, and clearer justification of the methodology are necessary before the manuscript can be considered for publication.

In response to the specific comments above, we have clarified and revised the methodology (ensemble size, MDOT, temperature and salinity nudging, SLA calculation), expanded the discussion of regional bias and RMSD patterns, and added observation details for the KEO/Papa buoys. We consider EN4-based validation and specific climate-index diagnostics (e.g., Niño-3.4) to be outside the scope of this study, as explained above. We believe these revisions address the reviewer's concerns.