Review

The manuscript provides a nice description and assessment of the new LORA-QG reanalysis in comparison with other state-of-the-science eddy permitting reanalyses, positioning LORA-QG in the envelope of these and providing a summary of strengths and weaknesses. The manuscript is also well written and clear, and I also liked the efforts in summarizing the results (e.g., Figure 19 and the text across the manuscript).

I recommend the manuscript for publication after a few points are better framed and discussed, which likely require some more discussion (and maybe a few additional analyses).

We thank the reviewer for the constructive and insightful comments, which have helped us improve the quality and clarity of the manuscript. We have carefully revised the manuscript in response to all of the reviewer's comments. Detailed responses to each comment are provided below.

Two main points

#1) The argument that the reanalysis provides uncertainty estimates appears in various parts of the manuscript (notably abstract and conclusions), and it is unique to the ensemble DA scheme adopted, but is not really validated, especially because the reanalysis proves under-dispersive. I suggest answering the questions (either through a discussion, and if possible, corroborated by results): is the choice of the (quite large) ensemble size (>100) really required, adding signal to the spread? Even though the reanalysis is under-dispersive, is the temporal evolution of the spread consistent with that of the RMSE (namely, net of the systematic offset), for some diagnostics (e.g., correlation of spread and RMS in Fig. 11)? Can the authors suggest reasons/ways to improve the under-dispersion despite the large ensemble size (i.e., lack of stochastic physics? Suboptimal inflation scheme in the LETKF?)

We thank the reviewer for providing constructive comments on the ensemble spread.

Response to the first question (ensemble size):

Previous studies (e.g., Miyoshi et al. 2014; doi:10.1002/2014GL060863) demonstrated that an ensemble size of about 100 members is still insufficient to fully suppress spurious ensemble correlations in an intermediate atmospheric data assimilation system (SPEEDY-

LETKF). Due to the computational constraints, we set the ensemble size to be 128, but these findings suggest that this ensemble size might still be insufficient to fully eliminate spurious correlations.

Response to the second question (Correlation between analysis RMSD and ensemble spread):
Following the reviewer's comment, we calculated temporal correlation coefficients between monthly analysis ensemble spread and analysis RMSDs of LORA-QG relative to the drifter buoys, tide gauges, and Ocean Climate Stations, and added the following new subsection 3.5 presenting these results. In summary, the ensemble spread is positively correlated over most of the open ocean, particularly around the western boundary currents and the ACC, although it remains substantially under-dispersive.

Revised:
3.5 Temporal correlation between analysis ensemble spread and RMSDs

LORA-QG has the advantage of providing flow-dependent uncertainty estimates, which are expected to be positively correlated with the analysis accuracy. We therefore calculated the temporal correlation coefficients between the monthly analysis ensemble spread and analysis RMSDs of LORA-QG relative to the drifter buoys (Fig. 19), tide gauges (Fig. 20a), and Ocean Climate Stations (figure not shown).

For the surface horizontal velocity and SST fields, positive correlations are distributed over most of the domain and are particularly high around strong current regions, such as the western boundary currents and the ACC (Fig. 19). Similarly, at the tide gauge stations, strong positive SSH correlations are found around these strong current regions, although substantial negative correlations also occur (Fig. 20a); the fraction of stations with positive correlation coefficients (about 60 %) is larger than that with negative correlation coefficients (about 40 %). As shown in Fig. 12e, the analysis ensemble spread is substantially smaller than the analysis RMSD, and its temporal variations are also small (Fig. 20b, c). This suggests that LORA-QG lacks sufficient sources of uncertainty despite the perturbed atmospheric and lateral boundary conditions, and that its horizontal resolution might be insufficient to reproduce fine-scale coastal variations. At the KEO and Papa buoys, the correlations are positive at most observation points, except for subsurface salinity at the KEO buoy and meridional velocity at the Papa buoy (figure not shown); the correlation coefficients between the depth-averaged analysis ensemble spread and RMSDs are 0.25, 0.06, 0.13, 0.08 at the KEO buoy and 0.56, 0.34, 0.25, 0.00 at the Papa buoy for temperature, salinity, zonal velocity, and meridional velocity, respectively. Overall, although the analysis ensemble spread is under-dispersive, it is positively correlated with the RMSDs over broad areas of the open ocean.

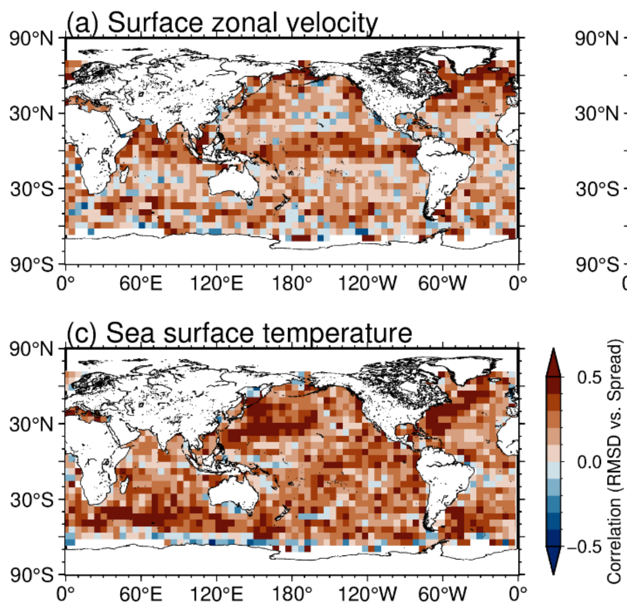

Figure 19: Temporal correlation coefficients between monthly analysis ensemble spread and analysis RMSDs of LORA-QG relative to drifter buoys in 5° longitude × 5° latitude bins for the surface (a) zonal and (b) meridional velocities and (c) SST fields.

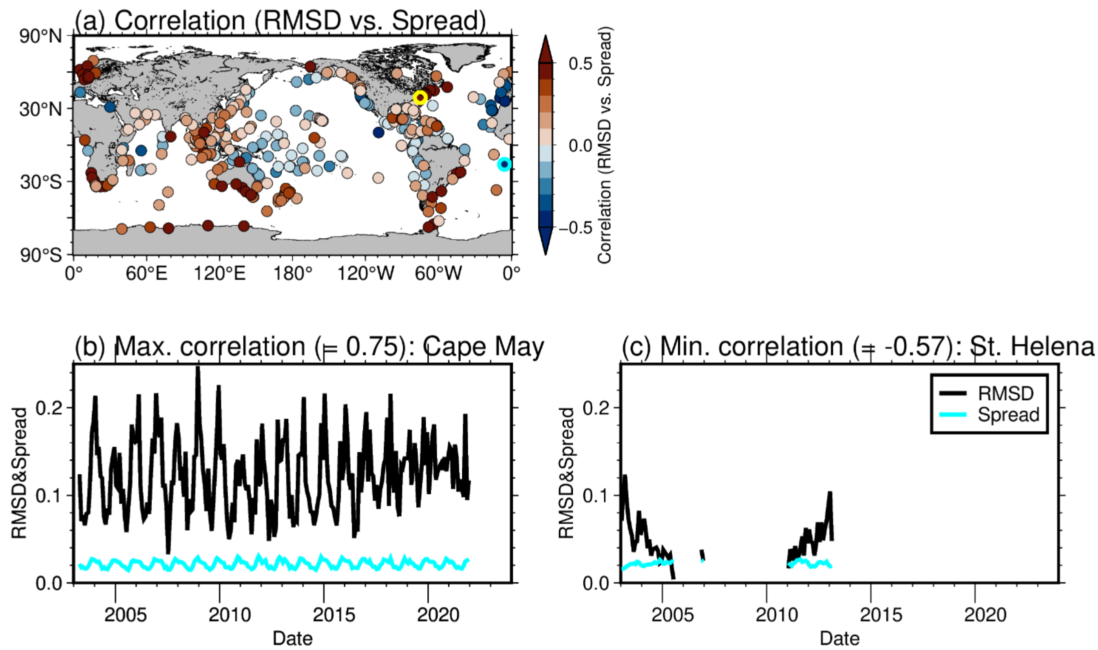

Figure 20: (a) Temporal correlation coefficients between monthly analysis ensemble

spread and analysis RMSDs of LORA-QG relative to sea level anomalies from tide gauge observations. Time series of the monthly analysis ensemble spread (cyan) and analysis RMSDs (black) at (b) Cape May and (c) St. Helena with the maximum and minimum correlation coefficients (yellow and cyan circles in Fig. 20a, respectively).

Response to the third question (underdispersion):
As described in the second paragraphs of subsections 2.1.1 and 2.1.2, we perturbed the atmosphere and boundary conditions and applied RTPP to inflate and maintain the ensemble spread, respectively, but the ensemble spread remains under-dispersive. This is likely due to suboptimal tuning of these perturbations and the lack of stochastic physics. We plan to investigate these in future work. We will add the following discussion to the second paragraph of Section 4.

Revised:
Consequently, some sources of model uncertainty remain insufficiently represented (e.g., uncertainties associated with vertical mixing and dissipation parameterizations), suggesting that additional stochastic perturbations or improved inflation techniques are needed to further increase the ensemble spread.

#2) While bias/RMS metrics may be, in general, difficult to explain in detail, some more efforts could be made to explain the large-scale biases (i.e., in the comparison with climatology), at least at speculation level. For instance, large surface warm biases in the tropics (could this be ascribed to the use of microwave SST, i.e., assimilating subskin SST, while all other reanalyses rely on skin/infrared SST?) or perhaps the warm/salty biases at depth (maybe depend on Antarctic bottom water formation that is limited by the global domain extension and the use of LBCs from SODA?)

We added discussions of the possible causes of the deep-layer temperature and salinity biases and tropical warm SST biases in the last paragraph of subsection 3.1 and the fifth paragraph of subsection 3.2, respectively. The revised texts are as follows:

Revised (subsection 3.1):
These temperature and salinity biases in the ocean interior are likely associated with both the limited availability of observations and model-related factors, including 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.

Revised (subsection 3.2):
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.

Minor points
#3) Abstract L22 and later in the text (L170): I think you should be more precise about the dependence/independence of the validating dataset. I think the reanalyses you included do not assimilate any drifter/tide gauge data; the climate station data dependence may depend on whether these are included in the in-situ dataset assimilated (EN4, CORA), but, if possible, this needs to be clarified.

We carefully examined the observations assimilated in each reanalysis dataset. We confirmed that in situ observations assimilated in LORA-QG (GTSPP and AQC Argo) do not include observations from drifter buoys, tide gauge stations, or the Ocean Climate Stations. We also investigated the dependence of the in situ observations assimilated in GLORYS2V4 (CORA) and in C-GLORSv7 and ORAS5 (EN4). CORA includes temperature observations from surface drifter buoys and temperature and salinity observations from the KEO and Papa buoys, whereas EN4 does not include any observations used for our validation. Therefore, among the four datasets, GLORYS2V4 is not fully independent of validation observations. To clarify this point, we added the dependency information to Table 5 and revised the relevant descriptions in Abstract and the first paragraphs of subsection 2.2 and Section 4 as follows.

Revised (Table 5):

Observation	Variable	Dependency	Reference
Surface drifter buoys	Surface zonal velocity Surface meridional velocity SST	– – GLORYS2V4	NOAA Global Drifter Program[†1] and Elipot et al. (2016, 2022)
Tide gauge stations	SSH anomaly	–	UHSLC[†2]
Ocean Climate Stations (KEO and Papa buoys)	Temperature Salinity Zonal velocity Meridional velocity	GLORYS2V4 GLORYS2V4 – –	NOAA PMEL Ocean Climate Stations[†3], Cronin et al. (2008), and Freeland (2007)

Revised (abstract):

Although these observations are independent of LORA-QG, ORAS5, and C-GLORSv7, they are not entirely independent of GLORYS2V4.

Revised (subsection 2.2):

All of the observations are independent of LORA-QG, C-GLORSv7, and ORAS5. However, SST from the drifter buoys and temperature and salinity from the Ocean Climate Stations are not independent of GLORYS2V4 (Table 5).

Revised (Section 4):

LORA-QG, ORAS5, and C-GLORSv7 are independent of these observations, whereas GLORYS2V4 assimilates SST observations from the surface drifter buoys and temperature and salinity observations from the KEO and Papa buoys (Table 5).

#4) I think the short summary should be self-contained, so I suggest rewriting it to avoid wording like “the second-most” etc., which assumes the reader to already know the work

We modified the short summary as follows.

Revised:

We developed an eddy-permitting local ensemble Kalman filter (LETKF)-based Ocean

Research Analysis (LORA) version 2.0 for a quasi-global domain (LORA-QG) covering the period from June 2002 onward. Validation shows that LORA-QG achieves accuracy comparable to existing global ocean reanalysis datasets. LORA-QG includes ensemble-based uncertainty estimates and individual terms of the heat and salinity budget equations, making it a valuable resource for ocean forecasting and process-based studies.

#5) Tab 1: Please check, I think SODA also released an eddy-resolving version recently

We added an eddy-resolving SODA version 4 to Table 1 and revised relevant descriptions accordingly.

Revised (Table 1):

	BRAN2020	GLORYS12V1	GOFS3.1	JCOPE-FGO	SODA4
Production center	CSIRO (Australia)	Mercator Ocean (France)	Naval Research Laboratory (US)	JAMSTEC (Japan)	University of Maryland (US)
Reference	Chamberlain et al. (2021)	Jean-Michel et al. (2021)	Cummings and Smedstad (2013)	Kido et al. (2022)	Chepurin et al. (2025)
Model	OFAM3[†1]	NEMO[†2]	HYCOM[†3]	POM[†4]	MOM5.1/SIS1[†5]
Domain	Quasi-global	Global	Global	Quasi-global	Global
Horizontal resolution	0.10°	1/12°	0.08°	0.10°	0.10°
Vertical resolution	51 z*-layers	50 z-levels	32 hybrid layers	44 σ-layers	75 z*-layers
Data assimilation	EnOI	SEEK[†6] with 3D-Var bias correction	3D-Var	3D-Var	OI
Assimilation interval	3 days	7 days	1 day	7 days	10 days
Assimilated observation	- Satellite SST - Satellite SSH - In situ temperature - In situ salinity	- Satellite SST - Satellite SSH - In situ temperature - In situ salinity	- Satellite SST - Satellite SSH - In situ temperature - In situ salinity	- Satellite SST - Satellite SSH - In situ temperature - In situ salinity	- Satellite SST - In situ temperature - In situ salinity
Output interval	1 day	1 day	3 hours	1 day	5 days
Available period	1993–2023	1993–2025	January 1994–September 2024	1993–2022	1980–2024

#6) L58 Adding a concise explanation of advantages (all weather) and disadvantages (resolution) compared to infrared SST could be good for non-experts

We added a concise explanation of the respective advantages and disadvantages of infrared and microwave SST observations for non-expert readers in the third paragraph of Section 1. The revised text is as follows:

Revised:

The WindSat polarimetric radiometer aboard the Coriolis satellite observed SST from August 2011 to October 2020 and partially bridged this gap, although its data coverage is more limited than that of AMSR-E and AMSR2 because of its narrower swath. Compared with microwave sensors, infrared sensors provide higher-spatial-resolution SST measurements but cannot observe SST under cloudy conditions, whereas microwave sensors can provide near-all-weather observations at the expense of coarser spatial resolution. Thus, these microwave SST observations have collectively enhanced global SST assimilation capabilities.

#7) L68, maybe say already there what you mean by quasi-global?

We added the system domain (0°–360°E, 70°S–70°N) in the first sentence of the last paragraph of Section 1. The revised text is as follows.

Revised:

Therefore, this study aims to develop an eddy-permitting EnKF-based ocean data assimilation system configured for a quasi-global domain (0°–360°E, 70°S–70°N), to produce an analysis dataset from June 2002, and to evaluate its accuracy in comparison with existing eddy-permitting global reanalysis datasets.

#8) L125 I am skeptical that you improve computational efficiency by converting routines from Fortran 77 to Fortran 90, given that compilers are typically the same. Did you really obtain more efficiency, or did you mean something else?

We agree that the conversion of Fortran 77 code to Fortran 90 does not directly improve computational performance. We therefore revised the description in the first paragraph of subsection 2.1.1 to clarify that the modernization was primarily intended to improve code maintainability, while the improvements in computational efficiency were achieved through the implementation of OpenMP parallelization and Parallel NetCDF I/O. The revised text is as follows.

Revised:

To improve forecast accuracy and computational efficiency, we substantially modified sbPOM. These modifications included modernizing legacy Fortran 77 code to Fortran 90 to improve code maintainability and readability, implementing Open Multi-Processing (OpenMP) parallelization and Parallel NetCDF input/output, and incorporating a scheme that subtracts and adds the horizontally averaged density before and after the pressure gradient calculation (Mellor et al., 1994).

#9) L132 not clear, may you mean “to prevent ensemble filter convergence/collapse”?

We thank the reviewer for pointing out the ambiguity of the term “filter divergence.” We intended to describe the reduction of ensemble spread that can degrade the representation of forecast uncertainty and lead to filter divergence. We therefore revised the sentence in the second paragraph of subsection 2.1.1 to clarify that the perturbations are introduced to prevent ensemble collapse and associated filter divergence.

Revised:
To prevent ensemble collapse (i.e., an excessive reduction of ensemble spread that can lead to filter divergence), the atmospheric and lateral boundary conditions, except for rainfall and river discharge, are perturbed (see Appendix 1 of Ohishi et al., 2023).

#10) L139 is only OpenMP parallel, or the OpenMP tiles MPI parallelization

The LETKF code has been already parallelized using MPI, and we additionally implemented OpenMP parallelization to further improve computational efficiency. We revised the description in the first paragraph of subsection 2.1.2 to clarify that the system employs hybrid MPI/OpenMP parallelization.

Revised:
To improve computational efficiency, OpenMP parallelization was additionally implemented in the LETKF code, which has already been parallelized using MPI. The resulting system employs hybrid MPI/OpenMP parallelization.

#11) The validation description seems to be missing if you use the ensemble mean for your reanalyses, and the temporal frequency (daily, monthly, in some cases?) for all validated reanalyses

We have clarified that the validation of LORA-QG was performed using daily-averaged ensemble-mean fields, whereas the other reanalysis datasets were evaluated using daily-mean fields. We have revised the following sentence at the end of the first paragraph of subsection 2.2.

Revised:
All validations are performed in observation space using daily-averaged ensemble-mean fields from LORA-QG and daily-mean fields from C-GLORSv7, GLORYS2V4, and ORAS5 reanalysis datasets. The validation period covers January 2003–December 2023.

#12) All other sections and analyses are well structured and clear for me (besides the need to try to understand the reason for systematic large-scale biases).

We thank the reviewer for the positive assessment of the manuscript structure and analyses. Following the reviewer's suggestions, we have expanded the discussions on the possible causes of systematic large-scale biases, including the tropical warm SST biases and deep-layer temperature and salinity biases, in the revised manuscript.

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.