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.

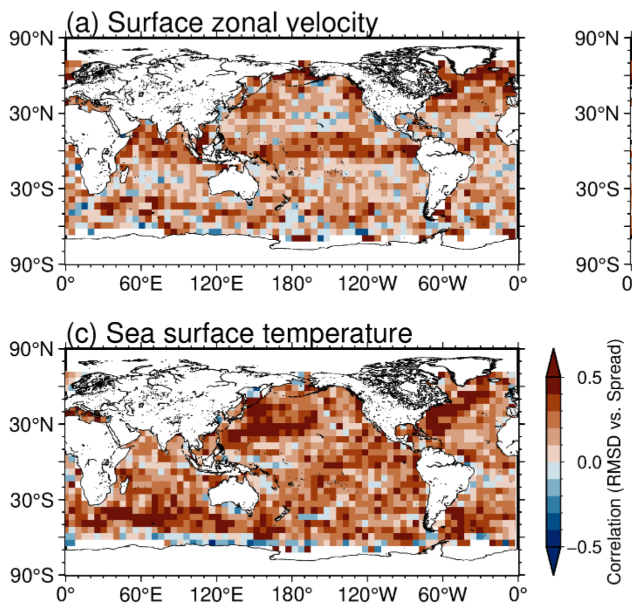

Figure 19: Temporal correlation coefficients between monthly analysis ensemble spread and analysis RMSDs of LORA-QG relative to drifter buoys in 5° longitude × 5° latitude bins for the surface (a) zonal and (b) meridional velocities and (c) SST fields.

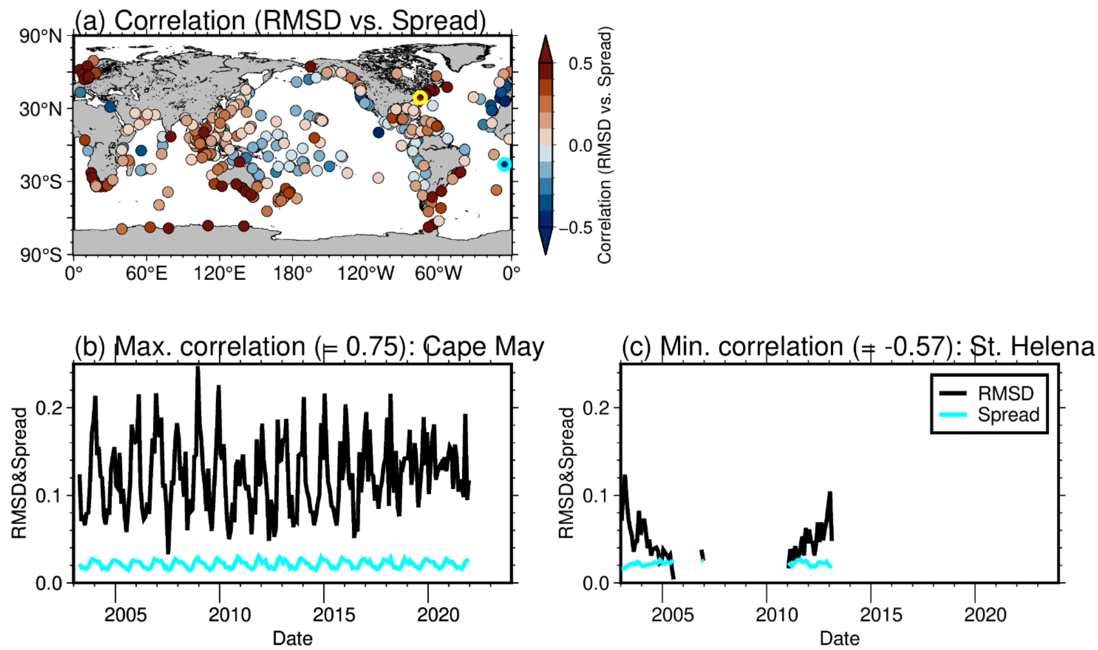

Figure 20: (a) Temporal correlation coefficients between monthly analysis ensemble

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 a new 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 results demonstrate that LORA-QG achieves accuracy comparable to existing global ocean reanalysis datasets. In addition to providing realistic ocean states, 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.