Response to reviewer comment RC1

Reviewer comments are presented first in *blue italics*, then followed by the author's response in normal font. Line numbers are referring to the original manuscript and are denoted as L145 for Line 145. For changes to text, we include the original in red and the changed version or new additions in green.

Dear Editor,

first of all I would like to say that I feel honored by the invitation to be a reviewer of the manuscript gmd-2022-204.

The manuscript entitled "How does 4DVar data assimilation affect the vertical representation of mesoscale eddies? A case study with OSSEs using ROMS v3.9" addresses a relevant and not entirely dominated question related to the quality of subsurface fields of ocean circulation models with observational data assimilation.

In general, the authors report their work in an organized, objective and well structured text, in a way that I could easily understand the problem, the methodological approach used, results obtained and what they indicate. The way the authors conduct the comparison of dynamic modes between solutions is particularly interesting.

I therefore recommend the publication of the paper after minor reviews are addressed.

Some questions, general comments, and suggested corrections are listed below so that the authors might want to address:

General comments:

line 18) wouldn't be "to deliver" instead of "the deliver"?

Changed to "and, they deliver nutrients.." Thank you.

line 90) although details of the DA model setup are reported in the other works cited, I missed some basic information of the 4D-Var implemented in ROMS that could somehow impact the results obtained, for example, the horizontal and vertical decorrelation scales, errors of the observations (table 1) and the number of inner and outer loops of the IS4D-Var.

Following this comment and a comment of RC#2, we have added the following table to the Appendix.

Table 1. Key differences in model configuration are shown between the Free-running ref state and the 4D-Var OSSEs. Further details are given in Gwyther et al. (2022) and references therein.

Configuration	Free run	4D-Var OSSE
Lateral BCs	BRAN2020	BRAN2020
Surface BCs	BARRA-R	ACCESS with bulk flux parameterisation
Mixing schemes	Harmonic horizontal mixing coefficient is 40 $m^2 s^{-1}$ for tracers and 55 $m^2 s^{-1}$ for momentum. Background vertical mixing coefficient is 1×10^{-6} $m^2 s^{-1}$ for tracers and 2×10^{-5} $m^2 s^{-1}$ for momentum.	Harmonic horizontal mixing coefficient is 200 $m^2 s^{-1}$ for tracers and 300 $m^2 s^{-1}$ for momentum. Background vertical mixing coefficient is 1×10^{-6} $m^2 s^{-1}$ for tracers and 1×10^{-5} $m^2 s^{-1}$ for momentum.
DA background error	n/a	Decorrelation length scales are assumed to be homogenous and isotropic. Horizontal length scale is 100 km; Vertical length scale is 10 m.
DA observation error	n/a	SSH error is 0.04 m; SST error is 0.5°C; XBT has a depth-varying error profile with a subsurface max of 0.6°C at 300 m decreasing to 0.12°C at 1100 m.
DA 4D-Var loops	n/a	14 inner loops and 1 outer loop.
More details	See Gwyther et al. (2022) and Li et al. (2021).	See Gwyther et al. (2022) and Kerry et al. (2016).

We have add the following information about the initial perturbation of the OSSEs:

"The OSSE that is simulating the same period as the Ref state is perturbed to introduce error and initiate divergent evolution (see discussion below)."

To

"The OSSE that is simulating the same period as the Ref state is perturbed to introduce error and initiate divergent evolution through the use of different initial conditions. These initial conditions are similar to those used to initialise the Ref state but are extracted from a point 8 days later (the OSSE begins at 2 December 2011 with conditions from 10 December 2011). This offset is chosen so as to fairly test the DA system (see Gwyther et al., 2022 for further information about this choice of perturbation)."

We have also included the following paragraph in the associated Appendix section, which discusses some of the differences between the model configurations.

"Key configuration settings and differences between the ref state and the OSSE model configuration are shown in Table 1. The decorrelation length scales are set following Kerry et al. (2016; section 3.5), and are consistent with estimates used elsewhere (e.g. Zhang et al., 2010; Zavala-Garay et al., 2012; Kerry et al 2018; Siripatana et al 2020; Gwyther et al., 2022)). Observation error covariances (see Table 1) are applied for each observation type. Further discussion of the preparation of the observations, the choices of error, and the minimization scheme is discussed further in Gwyther et al. (2022)."

Has any sensitivity analysis been done in order to verify whether some of these assimilation parameters affect in a significant way the patterns found, i.e. the poor representation of the subsurface structures? Would reducing the errors of the observations bring the models closer to the reference state?

We completely agree that a sensitivity study of parameters is important, in particular the sensitivity to the choice of how background error covariances are estimated. However, we believe this is firmly out of scope of this study. Here, we have focussed on how the oceanography in dynamic regimes is impacted by choices in observation strategies, rather than an exhaustive and technical exploration of DA system choices. The latter would likely require many new experiments in order to be a thorough analysis of the full range of parameter choices and would change the scope of the paper completely. (see also response to Reviewer 2).

With regards to the option of reducing observational error. Reducing the observational error may bring the estimates closer to the ref state/observations, however it may also ‘overfit’ the values leading to increased misfit elsewhere. We selected these values as they are the same or similar values to those used in other real (i.e. non-OSSE) data assimilating models. Furthermore, the synthetic observations are perturbed with errors that are normally distributed with a variance corresponding to the specified observation error, so the specified observations errors are ‘correct’ by definition.

line 117) do the fact that the reference run has a distinct setup with different boundary conditions and vertical mixing schemes, for example, interfere with the ability of the assimilative run to converge to the reference solution with respect to, for example, mixed layer depth? could a distinct setup between the reference and assimilative model lead to biases that could not be corrected through DA? How do the reference simulation and the free integration of DA setup compare?

Thank you for this interesting question. We chose to perform a ‘fraternal twin’-type OSSE, which can be run with different model configurations between the ref state and the OSSE, or with the same model type between the ref state and the OSSE, but with a variety of other configuration differences. This is a deliberate choice so that the growth of errors results from several sources, as opposed to just initialisation error, which would be the case if we used identical model configuration. This is more appropriate for the simulation of a realistic data assimilation system, which includes errors from a variety of sources (e.g. numerical truncation error, initialisation error, errors in resolved and parameterised processes, and errors in boundary conditions). A good overview is given in Halliwell et al., (2014).

Given that we’re also interested in subsurface representation, we don’t want to introduce too much error for the DA system, such that it cannot correct for such large differences. That is why we chose to keep many model configuration parameters and forcings the same, except for the surface forcing and some mixing parameters (see new table for more details).

We acknowledge that we omitted a demonstration of a free-running integration of the data-assimilating configuration. We now show this below. This ‘baseline’ run shows the bias in the integration resulting from the different surface forcing and mixing parameters.

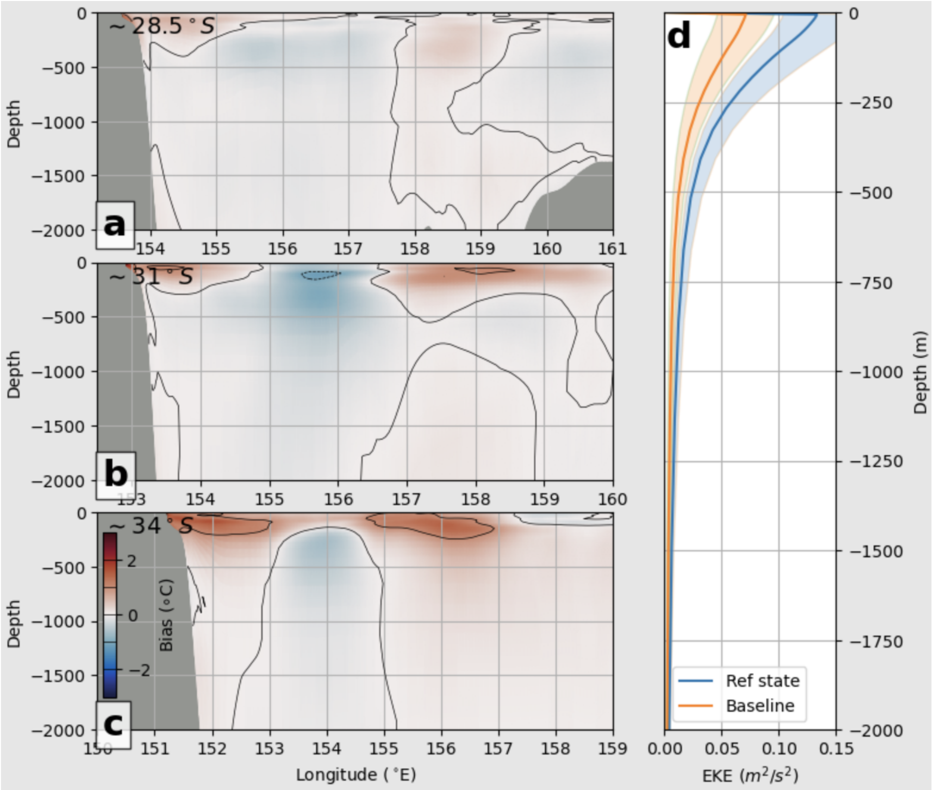

Figure A1. The temperature bias between the baseline and the ref state is shown at three transects, (a) 28.5S, (b) 31S and (c) 34S. In (d), the depth profile of EKE, averaged over the high EKE box (see box in Figure 3a), for the ref state and baseline experiments.

To explain this figure, a new section has been added to the appendix.

Appendix A: The `baseline’: Bias in the OSSE configuration

As described in Section 2.2, we employ a fraternal twin approach, where the ref state and the OSSE are simulated by the same model, but with different configurations. These differences, such as parameterisations and boundary conditions, should produce errors that are similar in nature (i.e. have similar magnitude and properties) to the initialisation error present in a true ocean DA system. However, the errors introduced through differences in configuration should not result in such a large impact, that the long-term representation is no longer realistic. If this occurs, it is difficult to separate out the error resulting from the difference in configuration (the bias), and what is the difference resulting from the DA process itself. Consequently, the free-running and data-assimilating simulations must have different configurations but without a large mean bias.

To quantify this bias, we run a ‘baseline’ experiment, using the free-running model with boundary conditions and parameterisations identical to the OSSEs. The bias is then calculated as the time-mean difference between the ref state and the baseline simulation.

Figure A1 shows the time-mean bias in temperature at three transects: 28, 31, 34 (Figure A1a-c). The surface region displays the greatest bias, of approximately 1.5C in the surface waters at 34S (Figure A1c), while at depth bias is negligible (close to 0C below 500m in all transects Figure A1a-c). The surface bias is very likely to be corrected for by the assimilation of SST observations. The depth profile of EKE for the ref state and baseline have similar shape: surface intensified with a gradual decrease with depth. Compare this to the same profiles for the OSSEs, which display subsurface maxima (Figure 3k).

The lack of strong (subsurface) bias with a consistent sign suggests that the differences in subsurface structure (e.g. Figure 2,4,5), mode structure (Figures 9 and 10), EKE distribution (Figure 3) and energy conversion rates (Figure 8) are principally a product of the DA system; they don't result from any consistent bias in the DA model forcing and configuration.

This appendix has also been introduced in the methods section:

"However, it is also important to ensure that the different configuration of the Ref state and OSSEs (e.g. in this case, surface forcing and some mixing parameters) do not cause such an impact as to introduce a large long-term bias. To assess this, a `baseline' experiment was conducted using the OSSE configuration, but without assimilating any observations. Comparison of this against the Ref state showed a warm bias in the surface waters, which is likely to be corrected by assimilating SST. More importantly, there is no strong bias in the subsurface ocean, which would otherwise be difficult to correct with assimilation (see Fig.A1)."

line 167) how was the density perturbation rho' estimated? was it calculated as the perturbation of a time and area average density rho(z)?

It was calculated as the difference from the time-mean density at each location, not an area average. This information has been added at L167: "are the density perturbation and vertical velocity perturbation from the long-term means calculated at each location, respectively,"

Figure 5) fontsize of vertical axis ticklabels and titles are too small

Thank you for pointing this out. The figure has been adjusted to have a smaller width and bigger font sizes. The same changes were also applied to Figure 4.

Figure 6 and 7) colorbar labels are difficult to read on panels f) and I think a colorbar is missing for panels k) (green to red diverging colormap)

Thank you for pointing this out - we have corrected both of these figures.

line 340) is the word "sim" a typo in "sim350m"?

Yes, we have corrected it to $\sim$ or the ~ symbol. Thank you.

line 391) in this case, could the balance operator implemented in ROMS be favorable?

As described in Kerry et al 2016, in this work, we only prescribe univariate covariance. The dynamics are coupled through the use of the tangent-linear and adjoint models in the assimilation, but not in the statistics of P (that is the matrix is a diagonal matrix and does not include balanced operators). The 4DVar balance operator may be a useful approach, however we have not used it and so could not comment on whether it will produce a large improvement. However we thank the reviewer for this suggestion, and will explore it in future work.

Other changes:

L412: Removed the first full stop in “static covariance estimate.(Bonavita et al., 2011).”

Figure 9: We have modified this plot to show a single mode in each column (instead of all modes in each OSSE per column). We believe this better demonstrates the failure of the OSSEs to represent each mode, especially without nearby observations. We also added the RMS difference from the Ref state for each mode in each OSSE. This new figure is shown below. The caption and figure description will also be adjusted to reflect this change.