Response to Reviewer 1

We would like to thank Reviewer 1 for reviewing the paper for a second time. Responses to the reviewer's comments are highlighted in blue.

The importance of present Figure 2 is still not clear to me. If I understood correctly, the removal of long-range temperature length-scale mitigates the problem with SLA assimilation, moreover it improves the BIAS and RMSE specifically in the Southern Ocean. The behaviour is opposite for salinity, the long-range length-scale can be kept since it does not enter the density calculation, moreover it improves the BIAS and RMSE especially in the Southern Ocean.

You understood correctly and your point above is what Figure 2 is intended to highlight, i.e. that removing the long temperature length-scales will improve the sub-surface temperatures, particularly in the Southern Ocean, by avoiding an inconsistent propagation of the SLA assimilation signal onto the large-scale sub-surface temperatures. Therefore, both RMSE and temperature biases are improved, particularly in the Southern Ocean.

I found that this opposite behaviour represents an interesting point to discuss in some more details since previous papers on FOAM system stated that the use of the sole short-range length-scale goes in the direction of improving the RMSE. The inclusion of long length-scale is meant not to reduce the RMSE but rather to reduce the presence of noisy fields in the case of sparse data, and to constrain the bias of tracer fields when observations are sparse (Blockley et al. 2014, Waters et al, 2013,2014).

The key paper to assess the impact of the NEMOVAR multiple length-scale approach is Mirouze et al. (2016). There is the following statement in the abstract taken from Mirouze et al. (2016):

"The multiple length scale operator has been implemented in NEMOVAR, a variational ocean data assimilation system. A dual length scale formulation was tested in a one-year reanalysis and compared with a single length scale formulation. The results emphasise the importance of estimating with great care the factors used within the combination. They also demonstrate the potential of the dual length scale formulation, in particular through a decrease of the innovation statistics for salinity profiles. The dual length scale formulation is now operational at the Met Office."

There is also this statement in the conclusion of Mirouze et al. (2016):

"When compared with a single length scale formulation, the results show an increase of the RMS error for innovations where a dense observation network is assimilated (SST and SSH)."

In summary, this work found that using a long length-scale for salinity can be beneficial. However, for other variables this may not be true and potentially lead to an increase in RMS errors as well. We believe the findings of our paper are consistent with Mirouze et al. (2016). We have changed the text in our paper to better highlight the aspects of using long length-scales in NEMOVAR:

"These large-scale error covariances are demonstrated to correct near-surface drifts, particularly for salinity in the Southern Ocean, however their impact on other variables, such as SST and SSH, is not clear and might lead to degradations in areas where a dense observation network is assimilated (see Mirouze et al., 2016)."

Reference:

Mirouze, I., Blockley, E. W., Lea, D. J., Martin, M. J., and Bell, M. J.: A multiple length scale correlation operator for ocean data assimilation, Tellus A: Dynamic Meteorology and Oceanography, 68, https://doi.org/10.3402/tellusa.v68.29744, 2016.

Based on the results of Figure 2, are the Authors confident with this opposite behaviour in temperature and salinity, over a longer period says one or few years?

The use of only a short length-scale for temperature and both short and long length-scales for salinity is tested for 1 year as part of the GOSI9 changes in Section 4.2 and Section 4.3. Therefore, we are confident about this change for a timeline of 1 year. Going beyond 1 year, we are currently running a GOSI9 reanalysis to confirm the 1-year results of this paper. We made a statement to reflect this point in the paper:

"Although the positive impacts of GOSI9 changes are shown here only for one-year experiments, we expect that GOSI9 results will lead to more potential for use of Met Office ocean reanalyses in climate studies, particularly for the satellite altimetry era from 1993 onwards. Therefore, future work will involve running a GOSI9 reanalysis for the satellite altimetry era."

If not, I can suggest to remove Figure 2. My impression is that it lacks information that confirm the significativity of such results within 5 months: are such OmB statistics stable in time or do the improvements come only from a short period within the 5 months? Are the number of observations, used to evaluate OmB, similar in the three experiments? A further check on the number of observations can be beneficial, since two mistakes have been already rectified in other figures.

The OmB RMSD for temperature of the 1T-2S experiment improves over the first 5 months, particularly between 500-1500 m, for both the Global and Southern Ocean. As we keep the long salinity length-scale in both 1T-2S and 2T-2S, the impact on the OmB RMSD for salinity is less clear over time, however more consistent improvements are seen towards the end of the first 5 months for both the Global and Southern Ocean. The Hovmöller plots below show the OmB RMSD percentage improvements (degradations) in blue (red) of 1T-2S relative to 2T-2S for temperature and salinity in both the Global and Southern Ocean.

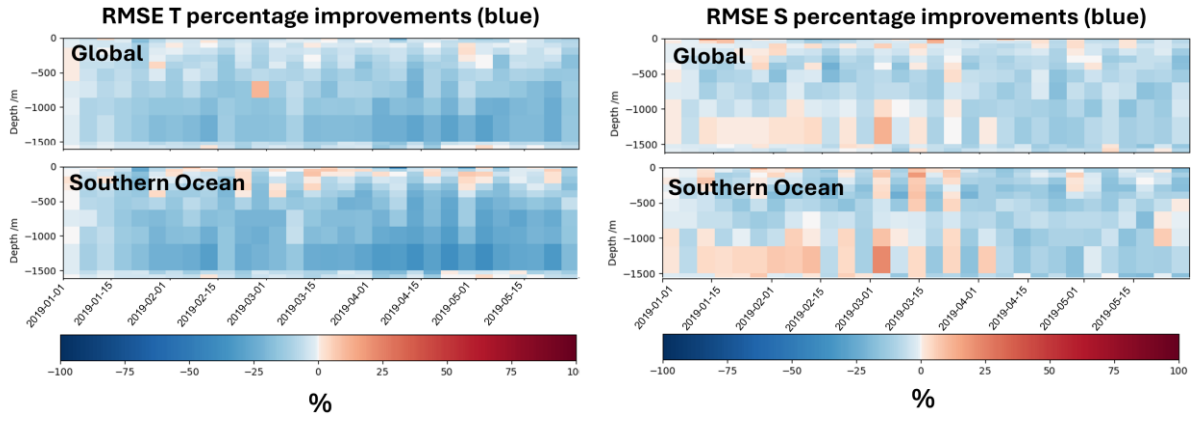

Figure: RMSD percentage improvements (degradations) in blue (red) of experiment 1T-2S relative to experiment 2T-2S between January and May 2019 when both are compared to profile observations.

We also plotted the OmB statistics for temperature and salinity as in Figure 2 of the paper but for specific months over the 5-month period, such as January, March and May 2019. This is consistent with the Hovmöller showing that OmB statistics do improve over time, particularly in southern latitudes.

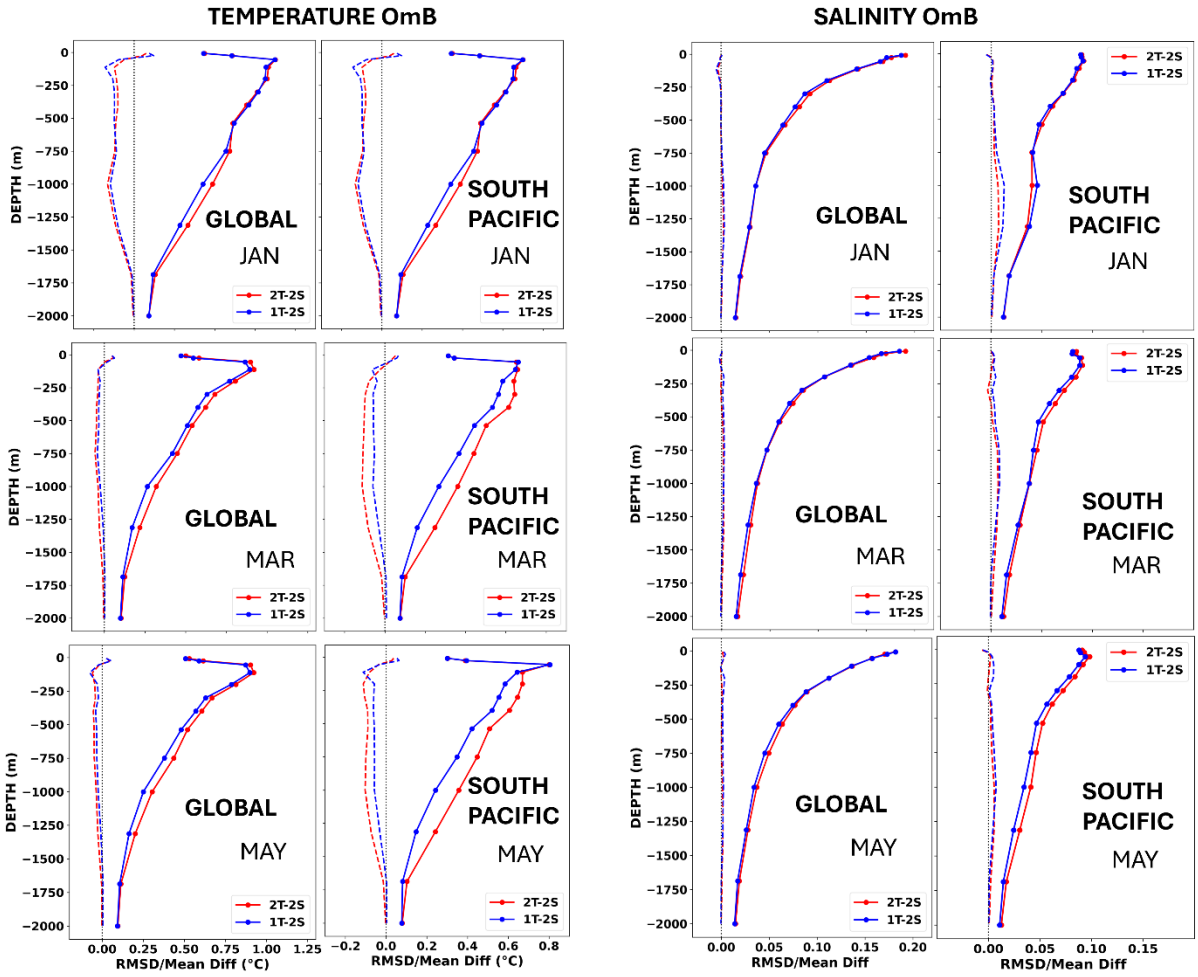

Figure: Temperature (left) and Salinity (right) OmB statistics calculated for three specific months (January, March and May 2019) against profile observations in the Global Ocean and South Pacific for 2T-2S (red) and 1T-2S (blue) experiments.

We confirm that the number of observations used to produce the OmB statistics in Figure 2 of the paper (and in the figures above) is the same between the experiments.

Response to Reviewer 2

We would like to thank K. Andrew Peterson for reviewing the paper for a second time. Responses to the reviewer's comments are highlighted in blue.

L. 320: Validation experiment runs are initialized from a 2017-2018 run of GO6. Exact phrasing "They come from 2017-2018 ORCA12 and ORCA025 runs with the FOAM GO6 configuration (see Barbosa-Aguiar et al., 2024)." This is ambiguous. I believe what you meant is "They are initialized from the end of 2017-2018 ORCA12 and ORCA025 runs with the FOAM GO6 configuration on 1 Jan, 2019 (correct date?).

Done.

L. 325. Presented OmB statistics are calculated from ..., SST from in situ drifters, ... This is an accurate statement, but neglects to specify you are ignoring further OmB statistics from assimilated (but biased) satellite SST observations. Perhaps this is worth stating?

Changed as below:

"In addition to drifters, swath SSTs are also assimilated from a variety of satellite sources (see Tab. 1 in Barbosa-Aguiar et al., 2024). However, satellite SSTs can be biased and therefore are bias-corrected in NEMOVAR (While and Martin, 2019). For this reason, OmB SST statistics are calculated here only with respect to in situ drifters."

All OmB spatial plots: There is no warning/admission that zero OmB difference and no data are both signified by white colour mapping. In general this is not a problem (most observed areas have at least some non-zero shading), or as in the velocity plots, the rms error plots display the observation error holes.

We appreciate the comment, but we believe this is an extremely minor detail and the OmB spatial maps are clear enough to show the positive impacts of GOSI9.

Again OmB spatial plots: Presumably the OmB statistics have been binned into lat-lon bins or onto the model grid? The bin size is never stated, but looks relatively high resolution. Please state if possible in Figure 5 (assuming it remains the same in all figures).

Done. This information has been added to the paper in Figs. 5, 8, 9 and 10.

LL 462-464: However, it is worth highlighting that FOAM SLA statistics in the central and eastern North Atlantic are slightly better in GOSI9-NoTSProf (Fig. 10g) than in GOSI9 (Fig. 10e). This suggests that there could still be minor issues in assimilating SLA and T/S profile data together, even after the substantial SLA improvements caused by GOSI9 DA changes.

This also appears to be true in the equatorial Pacific -- and possibly the whole of the Pacific, just not degradations from GO6 (deeper blue improvements). Were velocity statistics also performed on the no profile runs to further support SLA OmB statistics. Not suggesting to add plots -- just possibly state velocity statistics confirm better enhanced adjustment to SLA innovations without (potentially) conflicting T/S innovations.

The text was changed as below:

“However, it is worth highlighting that FOAM SLA statistics in the central and eastern North Atlantic are slightly better in GOSI9-NoTSProf (Fig. 10g) than in GOSI9 (Fig. 10e), **as well as in the Pacific.**”

We did not calculate the velocity statistics for the no profile runs. We rather focussed on providing the heat content and AMOC diagnostics in Section 4.3 since they are key indicators from a reanalysis perspective.