Response to Reviewer 1

We would like to thank reviewer 1 for suggesting minor corrections which will improve the overall quality of the paper. Responses to the reviewer's comments are highlighted in blue.

Beside minor corrections, I can support the manuscript for publication however I am concerned about the fact GMD strongly recommends the dissemination of the code. In the present case, the main improvements come from the DA code that is not free. Ocean Science can be a good alternative

We have previously submitted this paper to Ocean Science (OS). However, the OS editor believed that this paper would be a much better fit to GMD, due to all the technical aspects of the system that are presented in the paper. Therefore, we decided to withdraw our OS submission after the OS editor feedback and resubmit the paper to GMD.

MINOR COMMENTS

1) **GENERAL**

There is a sort of confusing nomenclature throughout the manuscript that I encourage the Authors to clarify for readability purposes. To my understanding, GO6/GOIS9 are generally used to point previous and new system while ORCA12 /ORCA025 are used to identify the resolution.

Yes, you are correct. GO6 and GOSI9 correspond to different versions of the global ocean and sea ice models used by the different Met Office forecasting systems, including FOAM. They are used in the manuscript to distinguish the old and new FOAM systems. ORCA12 and ORCA025 are used to identify the grid resolution of the sea ice and ocean models.

- In the Figures GO6 /GOI9 are used to differentiate the experiments, while the resolution (ORCA12 or ORCA025) is written in the caption. This is however confusing since ORCA12 parametrisation is different between GO6 and GOI9. I can suggest expanding the table 2 by introducing an different name for each experiment, something like GOIS9e12, GOIS9e025 etc.

We understand the reviewer's point here, however if we use the notation of "e12 and e025" to distinguish the different model resolutions, experiment names can be too long, for example GOSI9e025-NoTSProf. Furthermore, the two model resolutions are not directly compared in the paper. Section 4.2 only evaluates results from ORCA12 (which is the model resolution used to run FOAM forecasts operationally) and Section 4.3 only evaluates results from ORCA025 (which is the model resolution used to run FOAM reanalyses).

To make this point clearer in the paper, we included the model resolutions in the title of each result section. Section 4.2 has been changed to "Impacts on the global FOAM system **at 1/12º** with all observation types assimilated", whereas Section 4.3 has been changed to "Potential impacts on the pre-Argo reanalysis with the global FOAM system at **1/4º**".

- The use of ORCA12 DA is to be avoid since DA is performed at ¼° to my understanding. If two different DA set up at 1/4 ° are used, a description can be added to table 2.

We agree with reviewer and the use of ORCA12 DA has been removed from the paper.

- "FOAM" is used an attribute to both the system or the resolution, i.e. FOAM GO6 or FOAM ORCA12, etc., my impression is that you can safely remove "FOAM" everywhere.

We agree with the reviewer that this point can be improved. The term FOAM has been removed when it is associated to the model resolution. However, we will keep the term FOAM when it is associated to GO6 and GOSI9. Both GO6 and GOSI9 correspond to global ocean and sea ice model versions that are also used by other Met Office forecasting systems, such as the coupled NWP and seasonal forecasting system. Therefore, we believe it is important to reinforce throughout the paper that GO6 and GOSI9 results are associated with the old and new versions of the FOAM system.

2) SPECIFIC

Line 25: “In the absence of profile DA [..]”. The sentence is difficult to follow and “profile DA” is not clear. Please rephrase it you can start with something like “Limited to the assimilation of surface data only [..]”

Done.

Line 96-101: I do not understand if those improvements are applied only to 1/12 ° or also to 1/4°

GOSI9 model changes are always applied to both ORCA12 and ORCA025 configurations. To make this point clear, the following statement has been added at the beginning of Section 2:

“GOSI9 model changes presented below are applied to both ORCA025 and ORCA12 configurations”.

Figure F3 is blurry please replace it.

We believe you meant Figure S3 from the supplementary material. A new Figure S3 has been added.

Line 193 and Figure 2: Those results seem not conclusive, is there a reason why statistics are calculated only between Jan-May 2019 without covering the full year? Do you expect the results to be the same by including more statistics?

In the process of testing each individual GOSI9 change to the FOAM system in Section 3.1, we ended up running shorter experiments (varying from 1 to 6 months) which were enough to demonstrate the main impact of each change. Therefore, we could efficiently evaluate the impact of each GOSI9 change on FOAM and justify our choices before running the new FOAM system for 1 year with all changes together.

In our opinion, there is already a clear positive impact on the OmB temperature statistics at depth in the South Pacific and Southern Ocean from removing the long length-scale of temperature corrections, even though the runs only correspond to January-May 2019. This is even more evident in the OmB temperature statistics of 1-year GOSI9 runs with all changes

in Section 4.2 and 4.3, which validate our decision to remove the long length-scale temperature corrections from DA (as shown by Fig. 2).

Line 220: Is the minimization in GOIS9 always converging within 120 iterations? If not, can the Authors comment whether it is a noise problem, caused by the use of a first-order minimizer, or else.

In NEMOVAR, convergence depends on some convergence criteria. The one currently specified is not always met in FOAM with 120 iterations. However, we have run FOAM with NEMOVAR using different number of iterations: 40, 80, 120, 160 and 240. After 120 iterations, the OmB statistics are statistically indistinguishable between the runs. The use of a maximum number of iterations, rather than just waiting for the minimisation to converge, is done to ensure that the computational time of running NEMOVAR operationally is similar each day, so that we know we can produce the analysis and forecasts in time for our users.

Fig4: Similarly to my previous doubt, I am not sure the results in Fig4 are conclusive. There is a small difference among experiments and the statistics consider only 4 months. These results can be related to the specific position of few Argo that do not fully span the depths and the area. Can the Authors add the number of observations, extend the statistics or provide some study on the significativity?

We found a mistake with the land-ocean mask used to select the profiles and calculate the OmB temperature statistics for the Labrador Sea in Fig. 4d. We were not considering all the ocean area of the Labrador Sea but instead a much smaller region. This has been corrected and the positive impacts of using Brunt-Vaisala checks in the Labrador Sea are now larger at depth and convincing when compared to profile observations there.

We have plotted in Fig. 4a-c all the lat/long positions of the profile observations used to calculate the OmB temperature statistics between January and April 2019. We also added the number of profile observations in the caption of Fig. 4. The OmB statistics were calculated against 157 profiles which are well spread in the Labrador Sea.

Line 288: How many cores per node? Which is the frequency of the cpu?

We have added those details in the paper:

“Therefore, although FOAM GOSI9 has tripled the number of iterations, the DA at ORCA025 resolution takes 13 minutes in both FOAM GO6 and GOSI9 running on 15 computational nodes, each of which has 36 CPU cores on two 18 core Intel Broadwell Xeon 2.1 GHz processors.”

Lines 310-315: Can the Authors specify whether there is some preprocessing on the drifter velocities? are the drifters without drogues, used?

Before being disseminated by the Copernicus Marine Service, the drifters are subjected to several quality control tests (Notarstefano et al., 2010; Elipot et al., 2016). The quality tests consist of applying drogue-loss detection algorithms and applying the wind stress and 10-m wind components from ECMWF weather forecasts at each drifter location. This determines consistency from model winds and calculated currents using spectral analysis and a regression analysis. The references above have been added to the paper, so the readers can further look at the details of the drifter velocity pre-processing methodology in case they would like to.

Drifters that are deemed to have lost their drogues are excluded from our assessment. Additionally, we have not accounted for the Stroke drift effects on the model currents. These details have also been added to the paper in Section 4.1.

References:

Elipot, S., Lumpkin, R., Perez, R. C., Lilly, J. M., Early, J. J., and Sykulski, A. M.: A global surface drifter data set at hourly resolution, J. Geophys. Res.-Oceans, 121, 2937–2966, https://doi.org/10.1002/2016JC011716, 2016.

Notarstefano, G., Gerin, R., Bussani, A., Bolzon, G., and Poulain, P. M.: Real Time Quality Control and Validation of Current Measurements Inferred from Drifter Data Within Copernicus in Situ TAC, CMEMS-INS-DRIFTER-RTQC, http://dx.doi.org/10.13155/74299, 2010.

325 "4.2 Impacts on the […] with all observation types assimilated": which is the full set of observation assimilated? It is written that it is similar to the operational system, but the latter may vary in time. For example it is not clearly whether the satellite SST is assimilated or not together with the drifter SST. Is the new rejection algorithm impacting the number of insitu profile assimilated significantly?

The full set of assimilated observations in this paper is the same as described in Tab. 1 of Barbosa-Aguiar et al. (2024). This is a very extensive table listing all the satellite and in situ products that we assimilate operationally. This is the reason why we refer the readers to this table in the first paragraph of Section 4.1, so we do not need to have the same long table in the paper. In this table we assimilate both satellite and in situ SST observations, although OmB SST statistics in the paper are only calculated against in situ (drifters) SSTs. We made this clear in Section 4.1:

"Additionally, even though OmB SST statistics are calculated only with respect to in situ drifters, swath SSTs are also assimilated from a variety of satellite sources (see Tab. 1 in Barbosa-Aguiar et al., 2024)."

The rejection of T/S increments is determined during the process of applying them to the model temperature and salinity, so that we end up rejecting increments at the model grid points (rather than specific observations). Since the rejection algorithm targets regions of water mass convection, only less than 2% of the model grid points have their vertical profile of T/S increments rejected on each assimilation cycle. This has been added to the first paragraph of Section 3.2.

References:

Barbosa Aguiar, A., Waters, J., Price, M., Inverarity, G., Pequignet, C., Maksymczuk, J., Smout-Day, K., Martin, M., Bell, M., King, R., While, J., and Siddorn, J.: The new Met Office global ocean forecast system at 1/12th degree resolution, Q.J. Royal Met. Soc., 1–26, https://doi.org/10.1002/qj.4798, 2024.

418: "Additionally, the same SI3 settings from […]": if I understood correctly the sea-ice DA is not changed passing from the old to the new system. Is it so? Is the same parametrization used for ORCA12 and ORCA025 in SI3?. What about CICE (old system)? Was there a different parametrization for ORCA12 and ORCA025?

The sea ice DA and how the SIC increments are applied to the model are the same in GO6 and GOSI9. To make this clear, we have changed the text in the last paragraph of Section 3.1:

"Although the sea ice model changed from CICE to SI^3 in GOSI9, the DA and how the total SIC increments are added to the sea ice model remain the same".

The sea ice model settings are the same between ORCA025 and ORCA12 in both CICE and SI^3. To make this clear, we have changed the text in the last paragraph of Section 4.2:

"Additionally, the same sea ice model settings from ORCA025 are applied to ORCA12 in both CICE and SI^3…"

421 "Potential impacts […] pre-Argo reanalysis": I have not understood whether satellite sst are assimilated in the pre-argo era configuration.

This has been made clear at the beginning of Section 4.3:

"Along-track SLA, satellite and in situ SST, and SIC data are still all assimilated in those experiments, since there are data available of those observation types during the pre-Argo period."

422-426: The Authors missed to comment the importance of a good initial condition in pre-argo era. Especially in the southern hemisphere, the model state can be far from true state and the assimilation of few SLA measurements can exacerbate any hidden biases. Initial condition in 2019 are instead well spin up and the absence of drastic changes in OHC can partially come from that. Please add a comment on this.

We have made this point clear in the paper when presenting the heat content results:

"It is also important to highlight that the initial condition of our experiments is well spun up, and one would expect the model to be much less constrained by observations in the pre-Argo period, particularly in the Southern Hemisphere. Although the heat content results are promising, future work will involve running longer FOAM GOSI9 reanalyses to look at the heat content drifts in the pre-Argo period."

430: "GOSI9-NoTSProf run […] SLA RMSDs in comparison to its original […]" I was expecting GOSI9-NoTSProf to be closer to SLA obs since the system is assimilating mainly SLA data. Why the RSME is similar to the one that ingest insitu data too?

The experiment results show that the assimilation of in situ and SLA data, when the horizontal temperature length-scales are updated, is done in a physically consistent way (through the multivariate balance relationships in NEMOVAR), so that the assimilation of in situ data does not degrade the performance of the model's SLA fields.

493-497: Which is the method used to estimate the AMOC? Are the Authors using the RAPID decomposition? Probably this is the cause of the difference. I noticed that the assimilation seems to degrade the correlation with RAPID timeseries, do you have any reason why?

We are using the model velocities across the whole section to calculate the AMOC maximum as opposed to following the RAPID methodology. There have been a few papers, particularly Danabasoglu et al. (2021), showing that the differences in the AMOC calculated from these methods are small. Therefore, we can reasonably say that the differences seen at 26ºN are real, especially between GO6-NoTsProf and the other experiments, rather than dependent on the method. We have added to the paper how we calculate the AMOC for the RAPID array and for the OSNAP array, which has been added to the comparison in Fig. 14.

We also found a mistake which explains the low correlation between the model runs and the RAPID array. We were erroneously considering the RAPID transports from **February 2019** to **January 2020** and this should have been from January 2019 to December 2019. After fixing this, the model transports are much better correlated when compared to the RAPID array.

References:

Danabasoglu, G., Castruccio, F. S., Small, R. J., Tomas, R., Frajka-Williams, E., and Lankhorst, M.: Revisiting AMOC transport estimates from observations and models, Geophysical Research Letters, 48, https://doi.org/10.1029/2021GL093045, 2021.