Reviewer 2

The authors present a global implementation of a 3DVAR assimilation scheme with a high resolution oceanographic model. I think that the manuscript is well written and the evaluation of the impact of data assimilation within the new global operational system is very detailed and informative. I do not fully agree with some conclusions of the study and I have several small comments that may be addressed by the authors before the final publication.

Comment:
The authors discuss the modest improvement of the data assimilation performance with four times higher model resolution. They suspect that this is mainly due to the inadequate forecast evaluation methodology developed for low resolution models. They also suspect that arbitrary reducing the background horizontal covariance scales in the data assimilation scheme might negatively impact its performance. I think that there may be other equally or even more important reasons for lower than expected system's performance. An important reason might be that this data assimilation implementation produces only temperature and salinity increments that are dynamically unbalanced with respect to sea level and velocity increments. The IUA methodology is used to add increments to the initial state, but this may result in excessively smoothing and eventually losing a significant part of information introduced by data assimilation. This may be a less important issue with low resolution models characterised by simple dynamics, but in high resolution models the dynamical consistency of increments like those in Oddo et al (2026) (doi:10.5194/gmd-19-423-2026) might be a crucial ingredient for maintaining the impact of data assimilation in short term forecasts.

We thank the reviewer for the thorough and constructive assessment of our manuscript. We are pleased that the reviewer found the paper clearly written and the evaluation of the impact of data assimilation within the new global operational system detailed and informative. We have carefully addressed all points raised, and we believe the revised manuscript is substantially strengthened as a result. Our detailed responses are provided below.
Reviewer comments are in light black, our responses in bold black.

We agree that dynamical imbalance of the assimilation increments is a valid and important consideration that we had not sufficiently discussed. The 3DVAR scheme implemented in GOFS16 produces temperature and salinity increments only, without explicit balancing of sea level and velocity fields.
Based on both Reviewers' comments we acknowledge the limitation of the GOFS16 approach and expanded the section 2.2 to discuss the possibility of adding in the future more complex correction schemes that involve balanced velocity fields and barotropic SSH corrections. We would like to note for the reviewer that some co-authors of the present paper were directly involved in the development of OceanVar 2.0 presented in Oddo et al. (2026), and that efforts are already in progress to implement this new assimilation system into a global ocean configuration. This represents a natural and planned next step for GOFS16, and we have included this in the revised conclusions as a priority direction for future development.
About the implementation of IAU methodology, it is a highly debated topic, since at the beginning we used a direct ingestion scheme. The Reviewer is correct in stating that IAU

may result in smoothing that dissipates part of the assimilated information, and we added, in section 4, a discussion on the fact that we moved to IAU to reduce the model instability, which has been the main reason for retaining this choice. Nonetheless the plan is to revise the full strategy following Oddo et al. (2026).

Oddo, P., Adani, M., Carere, F., Cipollone, A., Goglio, A. C., Jansen, E., Aydogdu, A., Mele, F., Epicoco, I., Pistoia, J., Clementi, E., Pinardi, N., and Masina, S. (2026): "Implementation and evaluation of sea level operators in OceanVar2.0: an open-source oceanographic three-dimensional variational data assimilation system", Geosci. Model Dev., 19, 423–445, https://doi.org/10.5194/gmd-19-423-2026.

Minor comments:

1. Line 111: Does "fields" mean "fields of atmospheric variables"

Thank you, we change the wording:

"Atmospheric variables"

2. Line 112: What are "turbulent variables"?

We rephrase the full sentence by specify the variables:

"The turbulent variables (wind components at 10m, Temperature at 2m, specific humidity at 2m) are used at 6 hourly frequencies, while radiative and freshwater fluxes (short/long-wave radiation, precipitation, snow) are read as daily mean fields."

3. Line 113: What are "daily fields"?

The sentence is now changed (see L112 above).

4. Line 116: I guess that it should be "concentration" instead of "value".

Thank you, we modified it accordingly.

5. Line 166: I could not find information on how mean dynamic topography is included in the calculation of SLA misfits.

We include a description of how the SLA model anomaly equivalent is calculated:

"Sea level anomaly observations are related to the corresponding model equivalent via the mean dynamic topography following the formula:

$\xi = \eta\_SSH - \eta\_MDT + \varepsilon$

whereη_SSH is the model SSH, η_MDT the mean dynamic topography and ε identifies both observation and model errors (Storto et al 2011). The mean dynamic topography is defined as the 1992-2012 average SSH calculated from an experiment with insitu

assimilation at ¼°, following the strategy of Vidard et al (2009). This η_MDT is then interpolated on a 1/16° grid and therefore can include local spurious drifts with respect to the circulation at 1/16° which we try to mitigate with the inclusion of a spatial unbias scheme described below. One of the next developments includes the use of an MDT directly coming from a non-assimilative simulation at 1/16° (Iovino et al, 2023). "

Vidard, A., M. Balmaseda, and D. Anderson, (2009): Assimilation of Altimeter Data in the ECMWF Ocean Analysis System 3. Mon. Wea. Rev., 137, 1393–1408, https://doi.org/10.1175/2008MWR2668.1.

6. Lines 170-176: Is there a reference on the debias methodology? On line 171 "d" is defined as innovation, and on line 172 as bias. How are model and observational standard deviations (sigma_n and sigma_o) calculated? Are they prescribed arbitrary or updated dynamically? Are they estimated along tracks, regions or have a high spatial resolution? What is exactly the physical reasoning behind this scheme? I think that this all is important, because "unbiased" SLA observations that might be dynamically inconsistent are used to propagate temperature and salinity increments over deep water columns, and debiasing might significantly influence the forecasting performance.

We agree with the Reviewer and extend the description of the unbias scheme and the reasoning behind. Please see the answer to L170-176 to the first Reviewer.

7. Line 176: It should be "unbiased" instead of "unbias".

We now changed the paragraph and checked the wording, thanks.

8. Line 178: Does "the shallowest" mean "the top"?

Correct, we changed the wording to "first model level".

9. Line 179: What does the "lower bound for the error" mean? Are misfits with magnitude larger than this excluded? This value is much lower than the average RMS for SST misfits shown in Fig. 3a.

We rephrase the sentence to clarify the meaning:

'We impose a lower bound to the SST observation error by artificially increasing all the errors below 0.3°C to be exactly 0.3°C. This is a practical choice that has been implemented to reduce strong warming or cooling generated by an over-confidence in the observations that contain skin/subskin processes not resolved by the model.'

10. Line 185: How is an observation flagged bad? It could be expected that misfits are large at the thermocline and might be flagged as bad. All observations below are excluded? On the other hand, in Fig. 10 the number of assimilated observations seems to increase with depth.

Insitu observations can be flagged bad either by the producers or by our internal background quality check. This conservative check was necessary on a few occasions where the system ingests one profile excluding surface "bad" observations and assimilating "good" observations at depth. This eventually generates large increments causing the model to blow up. We decided to remove all the observations in the profile below the single observation flagged as bad. This limits occasional disruptions of the production at the cost of missing possible few good data (see L185-188).

The number of observations seem to increase with depth since the resolution of vertical levels is not even and degrades with depth. We pass from about 1m in the first 15meters to roughly 250m at 1500m. So, we are comparing the number of observations within 1 meter at the surface against the number of observations within 250m around 1500m. We clarify the point by adding the following sentence:

"At validation level, we adopt the same vertical levels of the model. Vertical spacing is uneven with a higher density of levels at the surface and a lower density at depth. The resolution is about 1 meter in the first 15 meters and roughly 250m around 1500m. "

11. Line 188: I guess that both background and observational error covariances change significantly with four times higher resolution. Should they be recalculated for the new model?

The Reviewer is right, part of the biases and the errors from GOFS16 may come from an under-representation of the model error that was formulated at 1/4°. Recently we have been investigating several improvements of the system, including an upgrade of the model version, ECMWF atmospheric forcings, and the renewal of the Background Error Covariance (BEC). We are investigating the possible use of BEC calculated from a 60-year OMIP run at 1/16° (Iovino et al 2023). This could likely help resolving part of the biases of the present system. We expand the description of the next development:

"In general, each component of the BEC is shaped on experiments at ¼° and then tuned to 1/16°. Recently a new long-term OMIP simulation at 1/16° has been produced by Iovino et al 2023 and can be the basis of a full replacement of the present BEC with the new OMIP-based one at 1/16°. "

12. Line 189: I do not understand why horizontal correlation scales of model errors should linearly scale with the model resolution.

Such scaling is a way to access the small-scale variability which is not present in the experiments at ¼ °. The figure below compares the increments coming from the assimilation of SST obtained using two different grids, at ¼° and 1/16°, respectively. We agree that this strategy is not optimal, and that the horizontal correlation scales should be derived from a simulation at the same resolution. Anyway, the use of such rescaling is beneficial when assimilating dense observations (SLA and SST). Following the discussion in Cipollone et al., (2020), GOFS16 grid has a higher number of degrees of freedom coming from the higher resolution and lower length-scales, thus being better suited for

dense observations. In particular, the overall SST/SLA RMSE using rescaled length-scale is reduced by 15%/9% (Cipollone et la 2020).

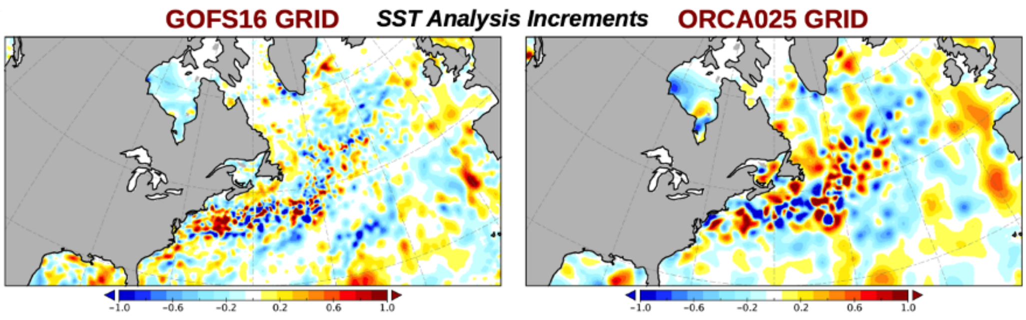

Fig.1 SST increments for one day of analysis with different horizontal lengthscales and grid. Left panel, SST increments for GLOB16 configuration. Right panel, similar but using ORCA025 configuration

We expanded the discussion as follows:

"The latter were estimated from the dataset of anomalies of temperature and salinity (Storto et al. 2014) and subsequently reshaped to capture the small-scale variability. While in the tropics the length-scales have been kept constant, in the other regions, we divided the original length scales by a factor proportional to the final resolution with a smoothing linear transition. Smaller horizontal length-scales are beneficial especially for dense observations such as SST retrievals, as discussed in Cipollone et al 2020."

13. Lines 190-194: Is the SST and salinity nudging in conflict with assimilation of other observations, in particular, SST observations described in one of the previous paragraphs?

About SST, we agree with both Reviewers that a better approach should be considered. We discuss it at line L191 of Reviewer 1. We report here the changes in the text:

"The analysis scheme of GOFS16 includes a nudging scheme to correct the surface heat and freshwater fluxes. DA corrections to surface model temperature does not sufficiently constraint the heat flux in the NEMO model, and we include a nudging scheme to have a realistic air-sea exchange using the gridded sea-surface temperature analyses provided by NOAA that includes infrared retrievals from AVHRR instruments, drifter and buoy in-situ measurements (Reynolds et al., 2007). This is going to be revised in the future, following a more complex SST operator (Broccoli and Cipollone, 2025)."

Similarly, for SSS we discuss the matter in L238 of Reviewer 1. At the end we suggest:

"SSS nudging is not meant to assimilate observations but rather to avoid large drift in the model salinity and reduce possible unrealistic effect on the overturning circulation. Rather than using a climatology, for the month J the following formula is used:

$SSS_J = SSS^{CLIM}_J + (SSS_{J-1} - SSS^{CLIM}_{J-1})$

The reference for month J, is constructed based on the climatology of the month J plus the anomaly of the month J-1."

14. Line 198: The sentence is not clear. What shares the same grid?

Eddy-resolving global forecasting systems usually work with different resolutions at model and DA steps (Lellouche et al 2018a, Barbosa Aguiar et al 2024). GOFS16 shares the same grid between model and DA steps.

We rephrase the sentence as follows:

"The GOFS16 system is designed to share the same 1/16° grid for both the model and the DA scheme, although this choice is computationally intensive and not strictly required (Lellouche et al, 2018a). The use of the same grid for model integration and DA maximizes the amount of information that can be extracted from the datasets, especially for observations that are dense on the grid (Cipollone et al, 2020). "

15. Line 218: It should be "increments" instead of "corrections". Can you provide a reference for IAU?

Thank you, we added the reference:

"with an incremental analysis update (IAU, Bloom et al. 1996) that spreads such increments over the full day."

Bloom, S. C., Takacs, L. L., da Silva, A. M., and Ledvina, D. (1996): Data assimilation using incremental analysis updates, Mon. Weather Rev., 124, 1256–1271.

16. Lines 435-439: Is this discussion related to Fig 10? This result may also indicate the presence of dynamical inconsistency between initial fields of temperature, salinity, sea level and velocity.

The same point has been raised by the first Reviewer. Looking at Fig. 10, it is a matter of fact that the model is not able to sustain the improvements brought by DA in the best estimate, similarly to other old (Ryan et al., 2015) and new systems (Smith et al., 2026). The reason can be the inconsistency of DA increments with the model dynamics, as the Reviewer suggested. Another reason can be that DA increments are too localized and do not sufficiently spread the information around. This is another point to be investigated, we add a comment on this:

"[...] suggesting that the model vertical mixing is not able to maintain an accurate initial state. Reasons can span from an inconsistency of the increments with the model dynamics to increments that are too localized and do not spread the information around."

17. Line 446: I think it should be Fig 10b instead of 8b. Arbitrary filtering out observations with large misfits may influence the system evaluation statistics when comparing with other data assimilation systems.

The Reviewer is correct, exactly the same number of observations should be used to compare the skills in a multi-system comparison. In the manuscript we follow this approach by filtering out observations with large salinity misfits (> 1 psu) either in the model or in the climatology. Moreover, the figure below compares the salinity RMSE profiles without any filtering, for best estimate and climatology in several regions and covering the period 2022-2023. Climatological RMSE is more than 0.1 psu higher than the best estimate down to 300m because of the presence of outliers. In the north Atlantic the climatology error exceeds 1 psu around 100m with GOFS16 best estimate that peaks at 0.5psu maximum. In the manuscript we prefer to remove large salinity misfits from the comparison, thus limiting the influence of outliers and having a meaningful comparison.

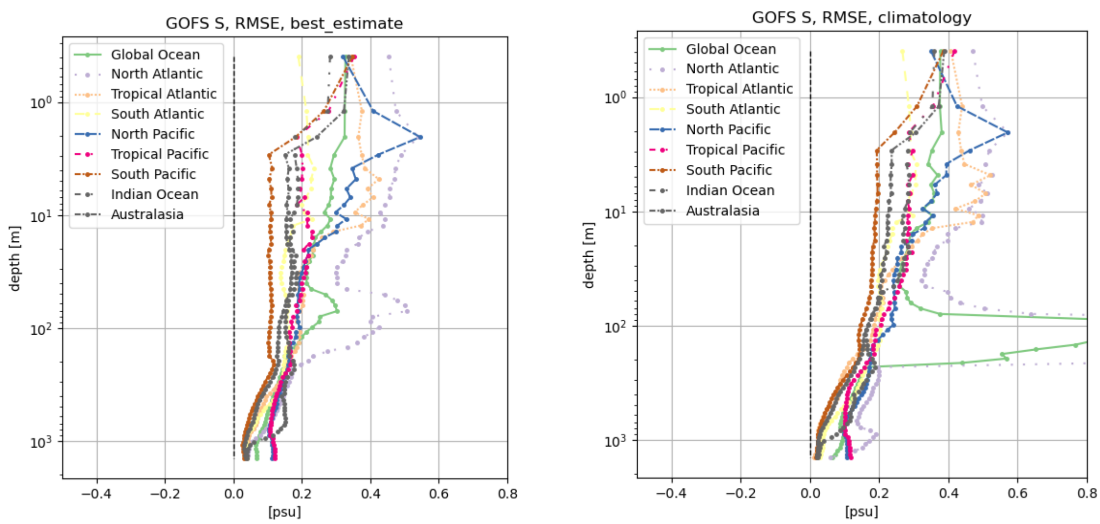

Fig.2. Vertical RMSE profile for salinity without any filtering, averaged over 2022-2023 for different regions. Left panel shows the RMSE for the best estimate, right panel similar for the climatology.

18. Line 670: What does it mean to correct scales? What are “coarse observations”? Is one scale for the spatial bias and the other is for the unbiased errors? How can these coarse and fine scales impact model dynamics?

The sentence was not clear, and it has been rephrased following also the last comment of the other Reviewer. GOFS16 works with two separate stationary datasets for temperature and salinity correlation lengths, that are used to spread information when assimilating SST and SLA as well. One of the possible reasons for the sudden saturation of the errors in T and S can be that increments remain too local and do not sufficiently spread in the surrounding water for sparse observations. Other forecast systems (for example FOAM) consider a different approach with multiple correlation lengths when treating insitu or SLA, for example. The ideal approach would be to have a multiscale DA

with an outer loop assimilating insitu on a coarse grid and an inner loop assimilating dense observations on a finer grid.

"In addition, we propose to start investigating multi-scale analysis technique (Li et al., 2015; Chamberlain et al., 2021) in which two assimilation steps correct the ocean at separate length scales: the first assimilation step is performed on a coarser grid to ingest sparse insitu observations. The second step employs a higher resolution grid and assimilates dense datasets to generate increments at smaller scales. Similar multi-scale analysis can be also performed in one single step (Mirouze et al, 2016b)."

Mirouze I, Blockley EW, Lea DJ, Martin MJ, Bell MJ. (2016b). A multiple length-scale correlation operator for ocean data assimilation. Tellus A 68: 29744.