Anonymous Referee #1

R: The study describes the development of OceanVar2.0, that is an oceanographic data assimilation method provided with open source software written for highly parallel computing environment. The manuscript further compares the method's performance in the Mediterranean Sea with respect to different options for assimilating Sea Level Anomaly (SLA) observations. The performance of assimilating SLA observations is particularly important in the Mediterranean Sea and many other regional seas and the global ocean, because often they represent the largest observational data set available in real time. I think that the study is well written. It provides an important insight into the improvements obtained by sophisticated application of a variational data assimilation scheme in oceanography that can be used as a reference for future applications of OceanVar2.0 and development and testing of other oceanographic data assimilation schemes. In particular, it highlights how applying a dynamical barotropic ocean model to simulate SLA perturbations within the model of the background error covariance matrix may significantly improve the variational data assimilation performance with respect to more commonly used simpler assumptions. I recommend the publication of the manuscript in Geoscientific Model Development after addressing a few minor comments.

A: We thank the Anonymous Referee #1 for their careful reading of our manuscript and for their constructive comments. We greatly appreciate their positive assessment and are pleased that they found our work to be a useful contribution to the scientific community.

We have carefully considered all of their suggestions and have made the necessary changes to improve the manuscript, as detailed in the point-by-point responses below.

R: Minor comments:

1. R: Line 140 and several other lines: I guess that DB08 should be DP08.
   A: Thanks for pointing out this typo. The manuscript will be corrected accordingly.

2. R: Line 270: Can removing the bias due to tides in observations and model forecasts increase the observational error? The two models have different bathymetries and use different computational methods for simulating the impact of tides. Is the atmospheric pressure forcing removed from SLA observations with the barotropic model?
   A: We thank the reviewer for this important comment. We agree that the accuracy of the observational error is a critical component of any successful data assimilation scheme. We did not modify the SLA observational error to account for the removal of tides for few key reasons. Tides in the Mediterranean Sea are minimal and constitute only a small fraction of the overall SLA signal. Second, our method avoids assimilating SLA data in regions shallower than 100m. This exclusion is particularly important because it prevents the assimilation of data from the few shallow areas where tidal signals might be amplified. As shown in Figure 2 of the submitted manuscript, we provide an example of SLA with and without tides. The bottom panel of this figure demonstrates the model's ability to reproduce tides. Along-track tidal gradients are well reproduced in the model simulation, and the difference between the modelled and observed tidal signal is nearly constant. By removing the mean residual along each satellite track, this bias is effectively eliminated and should be very similar to the misfit computed without tides, which is why we decided not to modify the observational error when including tides. Regarding the atmospheric pressure, we followed the same procedure outlined in Dobricic et al. (2012). Since our numerical model is already forced by atmospheric pressure, we do not remove this effect from the SLA observations. Also in this case, the mean residual is removed along each satellite track in the assimilation scheme. This process effectively removes large-scale oscillations, including those from atmospheric

pressure, as well as the unknown steric height signal. We have also clarified that the barotropic operator in OceanVar2.0 remains forced only by the vertically integrated buoyancy force resulting from temperature and salinity variations.

We believe this setup is physically consistent and produced satisfactory results in our experiments. However, we completely agree with the reviewer that SLA observational error is a crucial factor for the assimilation scheme's success. This is why we have designed the software to provide great flexibility in prescribing this parameter. To improve clarity for the reader, we have revised the manuscript to better explain the observational error, the procedures we used, and the rationale behind our choices. We will add the following sentence in 4.2 Section. "The results presented in Fig. 2 demonstrate that our model accurately reproduces along-track SLA tidal gradients, with the difference between the modelled and observed tidal signal being nearly constant. We effectively remove this along-track bias by subtracting the mean residual along each satellite track. This critical step ensures that the resulting misfit is very similar to the one computed without tides. Based on these considerations, we decided not to modify the SLA observational error, maintaining consistency with previous work (Escudier et al., 2021). Consequently, all the SLA data have an associated error of 3 cm regardless of the satellite and the geographic distribution."

3. R: Lines 380-390: Problems with vertical stratification might be also due to the calculation of vertical EOFs with difficulties to provide correct temperature and salinity increments from SLA assimilation during summer. Are there alternatives for using EOFs calculated from long model simulations in shallow areas?

   A: We thank the reviewer for this insightful comment. We agree that the use of Empirical Orthogonal Functions (EOFs) to model the vertical component of the Background Error Covariance (BEC) matrix can be a limitation, particularly in shallow areas and during the summer, when vertical stratification changes rapidly. The current version of OceanVar2.0 does not yet offer a built-in alternative to EOFs for representing the vertical error covariance. This is why, as a mitigation step, we decided to exclude regions shallower than 100m from the assimilation. While we recognize this remains a limitation of the current system, the modular design of OceanVar2.0 allows for the easy replacement of the EOFs with other, more sophisticated methods. Addressing this constraint is one of our highest priorities for future development, especially for coastal regions. We have added a brief sentence discussing this limitation and our future plans directly where the EOFs are introduced in the manuscript. "However, the persistence of this error maximum suggests a limitation in the current formulation of the Background Error Covariance. Specifically, the static, climatological nature of the EOFs used to model the vertical error component struggles to fully capture the rapidly evolving stratification and strong vertical gradients characteristic of the summer mixed layer. Future development will prioritize replacing these EOFs with more dynamic and stratification-aware operators to address this deficiency."

4. R: Line 500: Does the minimizer converge more slowly, because the barotropic model is slightly non-linear or because it includes more complex dynamical processes requiring a slower convergence?

   A: We thank the referee for this interesting and subtle question regarding the convergence performance. We agree that the observed tendency, though numerically small, is a point worth noting and discussing. Since this figure also drew commentary from another referee, we have slightly restructured the entire discussion section to first acknowledge the finding that the difference in the median values is small when compared to the overall daily variability. However, we believe the physical mechanism behind this consistent, minimal increase is relevant. As we do not have an analytical or definitive answer, we have maintained a cautious, speculative explanation in the revised text. We agree that this observation is worth mentioning, and the text now includes a possible explanation for this small, consistent difference: "The observed increase in the median number of iterations for the barotropic model schemes is likely due to the barotropic model's inherent physical complexity, which results in a more intricate optimization landscape for the minimizer to navigate."

5. R: Lines 503-510: Is the land-see mask used to optimise the domain decomposition? According to Fig. 1, with a higher granularity many subdomains may become completely over land and this can reduce the scalability of the software. Are some subdomains computationally more demanding than the others, for example, due to extra computations near the coast? Eventually, the evaluation of the parallel performance could show the scalability of different parts of the software (e.g. barotropic model, geostrophic adjustment etc.).

   A: We appreciate the detailed questions regarding our domain decomposition and the computational load across subdomains. The land-sea mask was not used to optimize the domain decomposition in these experiments. The decomposition was based purely on the model grid geometry, and we recognize that optimizing based on the land-sea mask is a valuable suggestion for future development to improve load balancing. Our analysis confirms that subdomains completely covered by land still required execution time comparable to the other domains. Regarding the computational load across subdomains, we found that there were no significant differences in computational demands based on geometric peculiarities, such as proximity to the coast. Instead, the computational load of a subdomain, and thus the total execution time, was primarily determined by the overall number of assimilated observations. This finding is consistent with the global nature of the variational solver and the number of iterations required for convergence. To provide a more complete assessment of scalability, we have added a new panel to Figure 12 showing the scalability of the most computationally significant routines. This analysis reveals that the limited overall speedup is not due to the domain decomposition being over land but is entirely due to the diffusive filter routine. Addressing this computational bottleneck is a priority for the future releases.