Reviewer 1

The authors investigated the potential role of Data Assimilation in improving the accuracy of barotropic processes induced variant scale/mode sea level anomaly in the Mediterranean Sea. The study is based on the state-of-the-art simulation kernel in SHYFEM. The authors comprehensively investigated the improvement of the astronomical tide, surge and seiches implemented by DA, and promoted the adaptability of SHYFEM with inclusion of EnKF. The manuscript is well written and organized with a sensible logic. However, given I still have these several following major concerns, I cannot recommend an acceptance at its present form.

We thank the Reviewer for the helpful comments, which will improve the quality of the paper. For the answer, we refer also to the first answer that we provided online.

- *Although it is still a nowadays great challenge to DA to treat/improve the hindcast and forecast of sea level anomaly in the region where the SLA oscillation is significant, I'm still wondering why the authors conduct this simulation in a two-dimensional or barotropic configuration? Will the inclusion of, e.g. dynamic height associated with the baroclinic processes be really negligible in the region? If it is not, why the heat fluxes, evaporation and precipitation, as well as riverine discharges are excluded? The larger scale circulation, at least those in the synoptic scale, is another issue related to this concern. Could the authors include some discussion related to the unimportance of these processes? Or, the authors may want to state that they are treating those larger-scaled motions as reference levels already, although I don't think that is a straightforward statement.*

  - Due to the comments of the reviewer we changed the introduction and the definition of the barotropic variations of the sea level (tide, surges and seiches), which are the focus of this paper. We hope that in this way our methodology becomes clearer (rows 15-22, 58-60). With this model configuration we cannot and we won't reproduce SLA due to the baroclinic part, which however varies with a timescale of weeks.

  - In the new section describing the altimeter data that we used (section 2.2.2) we discuss the difficulty to use altimeter data for storm surge applications (actually we used them in a past paper);

  - In section 2.1, in the description of the model, at the end we cite several works, using a similar model configuration, which proves the suitability of this formulation for tides, surges and seiches reproduction. We provided both papers using the SHYFEM model and papers using other models by many different experienced researchers (rows 112-124).

- *I still have concerns about how did the simulation treat the open boundary condition, although the manuscript did clarify that the authors treated the boundary condition with great effort. If sea level is kind of prescribed at the western boundary, how could the circulation (including their impacts in SLA and currents) be connected with that to the further west of the open boundary, which I think is provided by, for example, the CMEMS reanalyses. I may also suggest the authors include a paragraph to elaborate the way the open boundary condition is implemented or explicitly show the algorithm of the open boundary condition.*

As suggested, we added a section with the description of the forcing and boundary condition (section 2.1.1). Now, in this section, we provide a better description and we have also corrected the reference to the CMEMS model used for the boundary conditions (rows 125-142).

- *Why the satellite altimetry data is not used as observed data in this research? Are they at least usable for the astronomical tide correction and forecast? If gridded data is problematic, how about the along-track data? There are dataset of harmonic constants extracted from the along-track data by using this operation, and the authors mainly used much higher resolution records at the surrounding tidal gauge. I mean, there are more observations with much higher spatial coverage may help further improved the DA.*

We thank the Reviewer for this suggestion. We downloaded the Aviso Xtrack tidal data and we used them for a validation of the tide reanalysis. The results are good and provide a validation in the open sea, not only near the coasts. See the new sections 2.2.2 and 3.2.1.

- *In the perturbation runs, why the drag coefficient Cd in the quadratic formulation is not perturbed? Dissipation of energy with the scales smaller than tides through the bottom friction could also be an important process that determines the characteristics of tidal currents, and in this sense, although the authors stated that the current research is focusing on SLA variations, in the current configuration, accuracy in flows will also be an important aspect.*

This part has been corrected, the perturbation of Cd and of the loading tide parameter (that we forgot to expose) are now described in section 2.3.1, row 215.

Did the authors analyze whether the current design could also improve flows or not?

We noted a change in water transports compared to the simulation without DA but we did not compare them to any measures. However, if the cross-correlation between levels and currents is correct (the size of the ensemble - 81 members - should be sufficient), then currents should improve as well. On a smaller scale, we had seen improvements in the current by assimilating sea-level data in the inlets of the Lagoon of Venice (Ferrarin et al., 2021).

- *In my opinion, it is still important to rely on DA to improve the parameterization in the simulation, since it is not that feasible for operational users to generate a large number of perturbation runs to have that short-term forecast improved.*

We added a sentence on parameter estimation as possible improvement in the conclusions (rows 542-547). We also added a part in the discussion with the computational speed of the current system in a daily simulation (about 30 minutes), with a detailed explanation (rows 494-500).

- *It is really hard to intensify the meshes in Figure 1. Could you zoom in to some critically locations to show the spatial variability of resolution?*

We changed the figure, adding a zoom in the northern Adriatic.

Reviewer 2

The manuscript presents the predictive capability of a 2D barotropic model of the Mediterranean Sea sea level with and without the assimilation of the observations obtained from coastal tide gauges stations. The hydrodynamical model setup and ensemble Kalman filter based data assimilation system is described along with the perturbation schemes applied for ensemble generation. The results are presented for the total sea level as well as different contributions from the astronomical tides, surge and seiche for the hindcast/analysis and forecast periods for the December 2019 seiche occurrence following the November 2019 extreme event in the Adriatic Sea.

The manuscript requires a substantial revision before publication. Below are major comments and minor suggestions.

We thank the Reviewer for the dedicated time and detailed review, which helps to increase the quality of this work. Below we provide the answer to the individual points.

Major comments

To start with, for the readability of the manuscript, I suggest including a table of experiments to make it easier to follow, especially the results section. A flow chart for the production cycle would also help since it is difficult to understand where the hindcast/analysis ends and where the forecast starts. This may also help for future works since this system is proposed as a candidate for operational forecasting.

As suggested, we added a Table with all the simulations and their characteristics (Tab. 1). We also set identification labels to call in the paper. Then, we added also a flow diagram with the forecast cycle (reanalysis cycle is simple), similar to the ones used in the CMEMS manuals of the models (Fig. 2).

Moreover, the terminology used can be improved. There are terms used interchangeably such as analysis, reanalysis, hindcast simulation with data assimilation. I suggest homogenising them for an easier read and paying attention throughout the text to use the terminology that is already established, such as using analysis ensemble mean instead of average analysis state.

We checked all the paper, to use a coherent terminology. Now hindcast simulations are the two-month simulations without DA, reanalysis are the two-month simulations with DA, forecast simulations are the forecast simulations, starting or from the background state (no DA) or from the analysis state (DA). See the introduction (rows 53-58) and the section 2.4 (rows 240-260). We also used the definition *analysis ensemble mean* as suggested.

Secondly, I understand that the manuscript targets seiche in December 2019 however, it would be nice to see the evolution of the error in the sea level over a longer period given that the current version of the model is quite cheap as stated by the authors. I expected at least to see some analysis and the skill of the model in the November 2019 high tide event in the northern Adriatic Sea which resulted in the flooding of the city of Venice.

We are planning to run reanalysis simulations for several years and we wrote this in the conclusions. Moreover, as suggested, we added the description and the reproduction of the November 2019's storm surge event (which was one of the most extremes). This event is described in section 3.3.1, with some plots. The results show that the wind forcing was bad one day before, but it was good the same day of the event. As expected, DA does not improve the forecast, since that storm surge event was mainly due to the wind and pressure forcing. However, the storm surge of the day after (wich was still extreme) was forcasted much better with DA, as there was a component of seiche (Fig. 10).

On the other hand, SHYFEM is shown to be a skillful model in various previous studies. It is hard to understand why a simplified version is used in a development that is a candidate for an operational forecasting system. I think that in the cases where the errors and bias are large there is missing the steric steric part from the thermohaline contribution to sea level variability. This should be clarified and justified.

We specified better in the introduction that the focus of this paper is on the barotropic components (tides, surges and seiches), as specified in the title (rows 58-60).

Although the use of a barotropic model is a consequence of this, we added also a part, in the model description, with several works on tides, surges and seiches where we used a similar configuration. We also cited many works of other authors that use similar configurations with different models (rows 112-124).

Finally, it is not easy to completely grasp the improvements brought by the data assimilation of observations from tide gauges since they are limited in space coverage. Satellite observations could be used at least for validation to see the impact, if not assimilated. The results should be supported by maps of, for example, mean dynamic topography, increments. I think there may be other resources for the coastal sea level data for assimilation such as Copernicus Marine, SeaDataNet or EMODNet to better cover the eastern basin.

Based on this comment and on a comment of the first Reviewer, we looked for altimeter data to use for the validation. We downloaded the AVISO X-track data of amplitudes and phases of the tidal constants along the satellite tracks. From these constants, we computed the tidal sea level and we compared it with the model one. The average CRMSE decreases from 11.6cm to 4.3cm. See the new sections 2.2.2 and 3.2.1 and the new figure 5.

Minor suggestions

Title: Mediterranean -> Mediterranean Sea

Ok.

L27 “easily predictable” -> please refer to the sources of uncertainty in the estimates of tides e.g. bathymetry

Ok (rows 43-46).

L92 Please be more precise about the mesh resolution and give a measure of change from the open ocean to the coastal seas. Danilov (2022) may help. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022MS003177

Ok (row 109-112 and Fig.1).

L101 as done with the atmospheric forcing product, please cite the Copernicus Marine multi-year product explicitly in the references, not only with DOI. It should be clarified *why the authors used the multi-year product for the lateral open boundary conditions in the Atlantic Ocean while a NRT analysis/forecast product as in the atmospheric forcing is available in Copernicus Marine catalogs with tides for the experiment period. This is also one of the parameters that defines the type of experiment performed: an analysis, a reanalysis etc…*

The DOI in the paper was wrong, actually we used the analysis/forecast product that the Reviewer cites. Now we cite the right product, with the paper and the right DOI (rows 135-142).

L102 please explain how you de-tide the sea level.

As we wrote in the previous point, we put the wrong product. Actually, the de-tided sea level is provided in the model's variables (rows 138-139).

L124 missing citation in the parentheses. Please add it.

Done.

L128 Please add the mean sea level map and compare with the MDT products such as MDT-CMEMS_2020_MED in Copernicus Marine Catalog

This sentence was badly written. Actually, the MDT products are different from the MSL of the model. Now we removed the word "MDT" referred to the model. Now the sentence should be clearer, we used a similar approach as Byrne 2021 (rows 159-163).

L145 please justify 2 cm of observational error, is it only the instrumental error considered? How do the increments with such a small observational error look like? A map of increments may help to see whether there is an overfitting.

The stations' sensors have an accuracy much lower than 1cm (radar sensors, see e.g.: https://www.mareografico.it/?session=0S1476768288B907168WO8287&syslng=ita&sysmen=-1&sysind=-1&syssub=-1&sysfnt=0&code=SENS&idse=C). Some have pressure sensors with an error of about 1cm.

However, in order to obtain the best results we tested several values (1,2,3 cm). We did not write this before, now we explain better this and other DA settings (rows 272-290).

L153 grid -> node

Ok.

L153 "A_a^* is that of the analysis states not corrected". What do you mean? The definition of analysis implies a corrected background. Do you mean background?

This part was badly explained. Now it is written better (row 189-199).

L156 "levels" -> of what?

The sea level in the model equations ζ in equations 1 (row 201).

L162 "average analysis state" -> analysis ensemble mean

Ok, we changed it in the whole paper.

L169 Please justify 400 km. For example, Sakov et al. 2012 chose 250 km in a north Atlantic - Arctic Ocean system using the same methodology.

As in Sakov et al. (2012), our justification is mainly empirical (we tried also different values). We saw that 400km gives good results. Anyway, 400km SLP perturbations produce sub-synoptic systems that are of the same length scale as the typical ones in the Mediterranean Sea. We added a better explanation (and we corrected the pressure std value which was wrong) (rows 222-225).

L192 This is the definition of analysis ensemble mean. Please use it.

Ok, we checked all the paper to use the right definitions.

L 195 Not clear what the discussion here is.

This concept is now explained better and moved in the introduction (rows 60-64).

L 201 Why brevity? Why not robustness?

This sentence was not clear, now the text has changed.

L202-206 a production cycle flow chart may help.

We made it as suggested (Fig.2).

L222 What are the parameters of DA? Inflation and localization?

This is now explained better. In the section 2.3 the general description and in 3.1 the final values of the DA parameters.

L223 Local analysis is only one way of localization.

The text was wrong it is removed.

L234 Looks like too big error (9.3 cm) even for a free model and with a 2 cm of observation error reduces to only 3.6 cm. Is it because the barotropic model is missing the steric contribution? Please compare with altimeter products.

Actually, as the Reviewer noted, such error could be partly due to the fact that the model without DA is missing the steric contribution and partly to the fact that November-December 2019 was a period in which also the barotropic part of the mean sea level was high. The DA corrects both.

About the altimeter products, now we used them for the tide validation (sections 2.2.2 and 3.2.1).

L347 "is not present in our observations"? Do you mean in the period of observations used?

Yes, we wrote better this sentence (row 439). We also added an explanation for the peak at 5.2h visible in some Adriatic stations (Fig.?), and we found a recent paper (Sepic, 2022) citing it (rows 441-443, 454).

L352 There are other sources of error in DA besides model and representativeness error. Please correct.

This sentence is off topic here and we removed it.