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 will add a table summarising all the numerical experiments carried out, before showing the results. We will also add a flowchart of the DA procedure and simulations, in the analysis and forecast.

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.

Indeed, there is some confusion in the terminology, especially when referring to simulations in the hindcast period. We will pay more attention to the terminology used, defining it correctly.

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.

Indeed, we showed December's event since DA has an effect mainly on seiches. However, as noted by the Reviewer, it can be interesting also to look at the behaviour in November's storm-surge event. In that event, the 2-day forecast wind totally missed the event, while the 1-day wind was good. DA cannot correct of course the 2-day forecast, while the 1-day forecast did not need a correction. However, the high tide of the following days (due to seiches) was better reproduced by DA. As suggested by the Reviewer, we will add this event to the results.

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 thank the Reviewer for this comment which is similar to the first Reviewer's first comment. Please refer back also to that answer, which has similar points. The version of SHYFEM we used is one of the latest available, and there is no simplification on the numeric or other parts of the code. However, the baroclinic (steric) part was not activated. This is usual in storm-surge and tide modelling since to reproduce the barotropic sea-level components a barotropic model should be used (see the references in the 1st common of 1st rev). As we wrote in the title, this work focuses on the barotropic components (surges, tides, seiches) and not on the baroclinic part of the sea level.

Steric components are usually not considered in storm surge models as their variation is much slower and almost constant in 5-day forecasts (see references and comments on 1st point, 1st rev.).

Finally, we should consider the execution speed and the complexity of the model and of the DA procedure. Usually, 2D barotropic models are over 10 times faster than baroclinic ones. For example, in a simulation with SHYFEM in the Venetian lagoon, the model takes 16 minutes in a 2D barotropic simulation and 176 minutes in a baroclinic simulation with 12 vertical levels (however the Mediterranean would need at least 30 vertical levels). The simulated water levels from the two simulations are almost indistinguishable. Regarding the DA part, a baroclinic model would also require the perturbation of all the other forcings (solar radiation, air temperature, relative humidity, cloudiness, rain) and boundary conditions (T, S), which would take time. Assimilating only sea level would be not advisable, and T, S data should be assimilated as well.

In the revised manuscript we will add a similar explanation in the introduction, as well as some bibliography (see 1st point, 1st rev).

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.

This comment has points in common with the third comment by the first Reviewer, so please read that explanation as well.

Altimeter data could actually be useful in areas with few coastal stations, such as the eastern basin of the Mediterranean Sea. However, using altimetry data to correct high-frequency sea-level signals is problematic, as pointed out also in: https://www.aviso.altimetry.fr/en/data/products/auxiliary-products/dynamic-atmospheric-correction/description-atmospheric-corrections.html. And this is the reason why the storm surge signal is removed from SLA products and is rarely assimilated. Note also

that in order to remove the storm surge and tidal signals, 2D barotropic models are used (Mog2D, FES2014). In the past, we made some attempts to assimilate altimetric data (reintroducing the Mog2D correction) in a project with ESA on storm surges (see: Bajo et al. 2017). The results were modest and slightly positive only in the case of “lucky” tracks (direction along the Adriatic Sea and right time). Anyway, this is rare in the Mediterranean Sea, as the ratio between coastal areas, where the altimetric signal is scarce, and the open sea is very high. There are also issues related to the reference sea level, SLA refers to the Mean Sea Surface, different from model msl and in-situ data msl.

As future progress, we will try again to assimilate altimetric data, considering that they are improving in coastal areas. The imminent launch of the SWOT satellite (https://www.aviso.altimetry.fr/en/missions/future-missions/swot.html) cloud also be of interest, as its swat is large (about 120km). We will add some of this in the paper.

As suggested also by the first Reviewer, the use of altimetric data to determine the tidal harmonic constants is interesting. We will try to use them to validate the mode data (if the quality is good enough in the Mediterranean Sea).

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.

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.

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…

We thank the Reviewer for noting this. Actually, the DOI in the paper is wrong, we used the analysis/forecast product (https://doi.org/10.25423/CMCC/MEDSEA_ANALYSISFORECAST_PHY_006_013_EAS7). The reanalysis product does not even cover the lateral edge of our grid (-7.2W). Furthermore, as also reported in the comments to the first Reviewer, in the paper we wrote that the boundary conditions were imposed in Gibraltar, while the grid ends in the Atlantic Ocean. In the next version, we will correct these errors and expand the description of the boundary conditions. We will cite better the CMEMS product.

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

The sea level without the astronomical component (de-tided) is supplied directly by CMEMS. We will explain this point better in the paper.

L124 missing citation in the parentheses. Please add it.

Ok.

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

In the paper we referred to the MDT of the model, however, this is incorrect since the model simulates only the barotropic part of the sea level. We would therefore like to change the nomenclature to the mean sea level of the model.

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. We used a 2cm error precisely to avoid overfitting. The errors are still low, but the results obtained are good and they don't seem to have overfitting problems (we had also done some tests with 1-3cm).

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?

A_a^* is the original analysis, A_a is the quantity in the formula 2. The text is incorrect, we will modify it.

L156 "levels" -> of what?

The variable sea level (ζ) in equations 1. Will be explained further.

L162 "average analysis state" -> analysis ensemble mean

Ok.

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 saw that 400km gives good results. Anyway, 400km SLP perturbations produce sub-synoptic systems that are of the same length scale as the real ones in the Mediterranean Sea (see e.g., 22 November 2022, 06 UTC – https://www1.wetter3.de/archiv_dwd_dt.html). We will add a sentence to the text.

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

Ok.

L 195 Not clear what the discussion here is.

The concept is badly explained, we will try to write it better. For example, if we wanted the astronomical tide in 2025, we could compute tidal level time series from the harmonic constants in some locations where they are known with accuracy. Later, we could use these observations to run the DA and to make an analysis of the 2025 year.

L 201 Why brevity? Why not robustness?

This sentence is also unclear, we will change it. We meant that we display the first three days of the forecast in the results, not the fourth and fifth, because they would lengthen the section and the advantages of the DA on the fourth and fifth days are reduced.

L202-206 a production cycle flow chart may help.

Ok, we will make a flowchart.

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

We have always used adaptive inflation. The parameters that we varied in the tests are the localisation with different radii (in the end we did not use localisation), the number of ensemble members, the method (EnKF, EnSRF), the observation error and the C-G radius in the formula 2. We will write better in the text.

L223 Local analysis is only one way of localization.

Thank you to note it, we will correct it. We tested Domain Localization and not Covariance Localisation.

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 is partly due to the fact that the barotropic model is missing the steric contribution and partly to the fact that November-December 2019 was a period in which storm surges and seiches were particularly intense.

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

Yes, we will write better this sentence. We didn’t see this peak in our spectral analyses. A reason could be that this mode has not been triggered, it should be not very energetic and we didn’t know any previous work observing it.

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

This sentence does not fit well into the discussion, we will decide whether to remove it or change it also considering the other sources of error.

Bibliography

M. Bajo, L. Zampato, G. Umgiesser, A. Cucco, P. Canestrelli, A finite element operational model for storm surge prediction in Venice, Estuarine, Coastal and Shelf Science, Volume 75, Issues 1–2, 2007, Pages 236-249, https://doi.org/10.1016/j.ecss.2007.02.025.

Marco Bajo, Francesco De Biasio, Georg Umgiesser, Stefano Vignudelli, Stefano Zecchetto, Impact of using scatterometer and altimeter data on storm surge forecasting, Ocean Modelling, Volume 113, 2017, Pages 85-94, https://doi.org/10.1016/j.ocemod.2017.03.014.

Bajo, M, Međugorac, I, Umgiesser, G, Orlić, M. Storm surge and seiche modelling in the Adriatic Sea and the impact of data assimilation. Q J R Meteorol Soc. 2019; 145: 2070–2084. https://doi.org/10.1002/qj.3544

L. Cavaleri, M. Bajo, F. Barbariol, M. Bastianini, A. Benetazzo, L. Bertotti, J. Chiggiato, S. Davolio, C. Ferrarin, L. Magnusson, A. Papa, P. Pezzutto, A. Pomaro, G. Umgiesser, The October 29, 2018 storm in Northern Italy – An exceptional event and its modeling, Progress in Oceanography, Volume 178, 2019, https://doi.org/10.1016/j.pocean.2019.102178.

T. Fernández-Montblanc, M.I. Vousdoukas, P. Ciavola, E. Voukouvalas, L Mentaschi, G. Breyiannis, L. Feyen, P. Salamon, Towards robust pan-European storm surge forecasting, Ocean Modelling, Volume 133, 2019, Pages 129-144, https://doi.org/10.1016/j.ocemod.2018.12.001.

Christian Ferrarin, Debora Bellafiore, Gianmaria Sannino, Marco Bajo, Georg Umgiesser, Tidal dynamics in the inter-connected Mediterranean, Marmara, Black and Azov seas, Progress in Oceanography, Volume 161, 2018, Pages 102-115, https://doi.org/10.1016/j.pocean.2018.02.006.

Ferrarin, C., Bajo, M., and Umgiesser, G.: Model-driven optimization of coastal sea observatories through data assimilation in a finite element hydrodynamic model (SHYFEM v. 7_5_65), Geosci. Model Dev., 14, 645–659, https://doi.org/10.5194/gmd-14-645-2021, 2021.

Christian Ferrarin, Marco Bajo, Alvise Benetazzo, Luigi Cavaleri, Jacopo Chiggiato, Silvio Davison, Silvio Davolio, Piero Lionello, Mirko Orlić, Georg Umgiesser, Local and large-scale controls of the exceptional Venice floods of November 2019, Progress in Oceanography, Volume 197, 2021, https://doi.org/10.1016/j.pocean.2021.102628.

Flowerdew, J., Horsburgh, K., Wilson, C. and Mylne, K. (2010), Development and evaluation of an ensemble forecasting system for coastal storm surges. Q.J.R. Meteorol. Soc., 136: 1444-1456. https://doi.org/10.1002/qj.648

Horsburgh, K., Haigh, I.D., Williams, J. et al. “Grey swan” storm surges pose a greater coastal flood hazard than climate change. Ocean Dynamics 71, 715–730 (2021). https://doi.org/10.1007/s10236-021-01453-0

Ohishi, S., Hihara, T., Aiki, H., Ishizaka, J., Miyazawa, Y., Kachi, M., and Miyoshi, T.: An ensemble Kalman filter system with the Stony Brook Parallel Ocean Model v1.0, Geosci. Model Dev., 15, 8395–8410, https://doi.org/10.5194/gmd-15-8395-2022, 2022

Pugh, D.T. (1996) Tides, surges and mean sea-level (reprinted with corrections) , Chichester, UK. John Wiley & Sons, Ltd., 486pp.

Aron Roland, Andrea Cucco, Christian Ferrarin, Tai-Wen Hsu, Jian-Ming Liau, Shan-Hwei Ou, Georg Umgiesser, Ulrich Zanke, On the development and verification of a 2-D coupled wave-current model on unstructured meshes, Journal of Marine Systems, Volume 78, Supplement, 2009, Pages S244-S254, https://doi.org/10.1016/j.jmarsys.2009.01.026.

Xavier Bertin, Kai Li, Aron Roland, Yinglong J. Zhang, Jean François Breilh, Eric Chaumillon, A modeling-based analysis of the flooding associated with Xynthia, central Bay of Biscay, Coastal Engineering, Volume 94, 2014, Pages 80-89, https://doi.org/10.1016/j.coastaleng.2014.08.013.