Comment on egusphere-2022-1337

Anonymous Referee #1

Referee comment on "The Mediterranean forecasting system. Part I: evolution and performance" by Giovanni Coppini et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-1337-RC2, 2023

Comments of the Referee are in black while the authors' reply is in red.

Review general comments

The paper is very interesting and provides very accurate, robust and useful information for those developing operational ocean forecasting systems and for users of operational forecasting products. The plan of the paper is clear : to provide accurate information about main features of state of the art operational system and on scientific quality of the products which are delivered in the framework of the Copernicus Marine Service. Most of the information provided in this paper is also available in Copernicus Marine documentation and it is a very good initiative to publish this information in peer review journal.

I recommend that the authors provide in the introduction more precise information about the Copernicus Marine Service framework and about product quality strategy and how this paper is focusing this strategy.

We thank the reviewer for this comment that allows us to improve the introduction of this paper.

Concerning the information about the Copernicus Marine Service we will integrate the text in Line 42 as follows: "... Katsafados et al. 2016) and in 2015 it became operational in the framework of the Copernicus Marine Service which is the marine component of the Copernicus Programme European Union service for a sustainable use of the Ocean providing free, regular and systematic information on the state of the Blue (physical), White (sea ice) and Green (biogeochemical) ocean on the global and regional scales. The Copernicus Marine service in Europe has shown the strength of a state-of-the-art operational service implemented by hundreds of experts and teams, distributed throughout Europe, coming from public and private sectors, from operational and research organisations, from different countries, from diverse cultures and relations to the ocean (Le Traon et al., 2017; Alvarez Fanjul et al., 2022). "

Concerning the information about the product quality strategy we will add a sentence after Line 53 as follows: "An essential task of the production activities concerns the continuous assessment of the quality of the products (Sotillo et al., 2021; Alvarez Fanjul et al., 2022) which is achieved at two levels: (i) the pre-qualification of the systems before delivering a new release, including an extensive scientific validation of the products, published in the QUality Information Documents (QUIDs) available on the Copernicus Marine Product Catalogue; (ii) the operational evaluation of the skill metrics during operations, made available through the Copernicus Marine Product Quality Dashboard Website (https://pqd.mercator-ocean.fr), as well as through the Mediterranean regional validation websites implemented at the level of the Med-MFC production units (PHY: https://medfs.cmcc.it/, WAV: http://med-mfc-wav.hcmr.gr/, BIO: www.medeaf.ogs.it/NRT-validation). All the delivered variables are thus validated with respect to satellite and in-situ observations using Copernicus Marine observational datasets, as well as additional datasets, climatologies or literature information when needed."

We will add 2 new references

1) Alvarez Fanjul, E., Ciliberti, S., and Bahurel, P.: Implementing Operational Ocean Monitoring and Forecasting Systems. Paris, France, IOC-UNESCO, 376pp. & Annexes, (GOOS-275). DOI: https://doi.org/10.48670/ETOOFS, 2022.
2) Sotillo, M. G., Garcia-Hermosa, I., Drévillon, M., Régnier, C., Szczypta, C., Hernandez, F., Melet, A., and Le Traon, P.Y.: Communicating CMEMS Product Quality: evolution & achievements along Copernicus-1 (2015- 2021), Mercator Ocean Journal #57, https://marine.copernicus.eu/it/node/19306, 2021.

The scientifical analysis of main uncertainties, analysis and/or forecast errors is poorly described in the paper, I suggest to provide more information where possible on the source of uncertainty, on the missing processes in the forecast system, on the errors and uncertainties in the forcing fields … This should at least be mentioned in the section 3 introduction part, to indicate that the paper doesn't provide a strong and detailed analysis of main drivers/stressors of forecast uncertainties based on sensitivity study (or other experimental framework) but only statistics and accuracy numbers based on a reference simulation produced to calibrate an operational forecasting system.

We thank the reviewer for this comment. Concerning the discussion on main uncertainties we propose to integrate the text in Line 353 as follows: "Ocean model uncertainties emerge from sources of errors relevant to the ocean state, including physics, biogeochemistry, and sea ice, as well as errors in the initial state and boundary conditions (i.e. atmospheric forcing and lateral open boundary conditions). Model uncertainties in ocean physics have a significant impact in all other system components as, for example, in biogeochemistry and sea ice (Alvarez Fanjul et al., 2022). Our results describe the quality of the Med-MFC products presenting the statistics and accuracy numbers based on a reference simulation produced to calibrate and validate the operational forecasting systems, whereas the analysis of model uncertainty sources is outlined in the discussion part also referring to previous specific publications."

I recommend the publication of this paper, if the authors provide answers to the following questions and a revised version of the paper taking into account the main suggestions.

Questions and suggestions:

1 Introduction :

I strongly suggest adding in the introduction part information about product quality activities in Copernicus Marine, QUID documentation and why it is important to publish this information in peer review journal.
It could be useful to add two references i) to the ETOOFS guide « Implementing Operational ocean Monitoring and Forecasting Systems Alvarez, Ciliberti and Bahurel 2022 and especially a citation to the chapter 4 containing a section dedicated to ocean forecasting system validation and to ii) Sotillo et al 2021 who describe validation and product quality strategy in copernicus marine. (Sotillo, M. G., Garcia-Hermosa, I., Drévillon, M., Régnier, C., Szczypta, C., Hernandez, F., Melet, A., Le Traon, P.Y. (2021). Communicating CMEMS Product Quality: evolution & achievements along Copernicus-1 (2015- 2021). Mercator Ocean Journal #57. Available at https://marine.copernicus.eu/it/node/19306)

Following the general comments, the information on the QUID will be added after Line 53 (see previous comment) as well as the references to ETOOFS guide (2022) and Sotillo et al. (2021).

This paper is the part I, it could be useful for reader to have information about the partII, which topic and how it will be related to part I.

We propose to add the following after Line 74:

"The Part II of the paper will be showing the capacities of the Med-MFC components in describing the Medicane effects on the ocean. In particular the Med-MFC physics, biogeochemistry and waves components will be used to describe the effects of Medicane Zorbas (27-30 September 2018) on the ocean variables."

In the introduction you provide quantified useful information for the transport at Gibraltar strait and for the mean wave period and significant height, it would be good to add also information on uncertainties and on variability instead of just only a mean value. It's also related to the main objective of the paper to provide quality information on ocean model simulation which should include uncertainties.

Following the reviewer's comment we propose 2 modifications to the text:

1) the statement in Line 54 will be rephrased as follows: "The Mediterranean Sea is a semi-enclosed basin with an anti-estuarine circulation corresponding to a 0.9/0.8 ±0.06 Sv baroclinic inflow/outflow at the Strait of Gibraltar,...."
2) the statement in Line 64 will be rephrased as follows: "The yearly mean wave period, as estimated from available wave buoys over the Mediterranean Sea, amounts to 3.82 s with typical deviations of 0.92 s, while the mean significant wave height is 0.82 m (1.28 m as estimated by satellite observations) with typical deviations of 0.67m (0.77m for satellite data)."

It will be good to provide a clear definition of offline coupling (line 66) where are the feedback between the model components, which variables … There is often confusion between forcing, two way forcing, offline coupling or full coupling.

The feedback between the Med-MFC modelling systems is achieved by means of daily mean physical variables which force the biogeochemical and wave models, so we agree with the reviewer that a proper definition would be "forced systems" rather than "offline coupled systems". Following the reviewer's comment, the statement in Line 66 will be rephrased as follows: "The Med-MFC modelling systems share the same grid resolution (1/24º), bathymetry and use the same atmospheric and river forcing fields. Moreover daily mean fields evaluated by the physical model are used to force the wave component (surface currents) and the transport-biogeochemical model (temperature, salinity, horizontal and vertical velocities, sea level, diffusivity)."

Line 71 : could you explain wheather this standard is also applied for the other MFC in copernicus marine and whether it is good practice for operational oceanography.

The Copernicus Marine Service MFCs approaches can be summarized as follows according to the description of the systems in the corresponding QUIDs:

- GLOBAL-MFC: GLO-PHY forces GLO-WAV,GLO-PHY forces GLO-BIO (daily 3D fields at degraded resolution from 1/12° to 1/4°)
- ARCTIC-MFC: ARC-PHY forces ARC-WAV, ARC-PHY is online coupled with ARC-BIO
- IBI-MFC: IBI-PHY forces IBI-WAV, IBI-WAV forces IBI-PHY IBI-PHY is online coupled with IBI-BIO which has no BIO-DA
- NWS-MFC AMM15: NWS-PHY 2-way coupled with NWS-WAV, NWS-PHY is online coupled with NWS-PHY
- BALTIC-MFC: BAL-PHY is online coupled with BAL-BIO which has no BGC-DA, BAL-PHY forces BAL-WAV, BAL-WAV forces BAL-PHY
- Black Sea-MFC: BLK-PHY forces BLK-WAV, BLK-BIO runs separately NEMO coupled with BAMBY
- MED-MFC: MED-PHY forces MED-WAV, MED-PHY forces MED-BIO (daily 3D fields), MED-PHY runs separately WW3 2-way coupled with NEMO.

Following reviewer's comment, we propose to add the following text after Line 66:

"In the Copernicus Marine Service the approach of forcing waves and biogeochemistry models with information from the hydrodynamic models is used and represents a standard which is also applied for the other MFCs. Several MFCs also foreseen the online coupling between physics and waves models and between physics and biogeochemical models."

2 Description of the Med-MFC core components

2.1.1 Numerical model description

Could you explain in more detail how the exchanges with the Atlantic ocean are implemented and how these exchanges between med and atlantic are better resolved. Are there any references on this development?

Following reviewer's comment, more details will be included in the sentence at Line 122 as follows: "The circulation model's lateral open boundary conditions (LOBC) in the Atlantic Ocean are provided by the Copernicus Global Analysis and Forecast product (Lellouche et al., 2018) at 1/12° horizontal resolution and 50 vertical levels. Daily mean fields are used and the numerical schemes applied at the open boundaries are the Flather (1976) radiation scheme for the barotropic velocity and the Orlanski (1976) radiation condition (normal projection of oblique radiation case) with adaptive nudging (Marchesiello et al., 2001) for the baroclinic velocity and the tracers. The nesting technique is detailed in Oddo et al. (2009), who also show a marked improvement in the salinity characteristics of the Modified Atlantic Water and in the Mediterranean sea level seasonal variability. "

3 additional references will be included:

1. Flather, RA.: A tidal model of the north–west European continental shelf, Mem. Soc. R. Sci. Liege, 10, pp. 141–164, 1976.
2. Orlanski, I.: A simple boundary condition for unbounded hyperbolic flows". Journal of computational physics, 21.3, pp. 251–269, 1976.
3. Marchesiello, P., McWilliams, JC., and Shchepetkin, A.: Open boundary conditions for long-term integration of regional oceanic models, Ocean modelling, 3.1–2, pp. 1–20, 2001.

Can you describe the changes in the bathymetry that have been made in the different critical areas (adriatic, straits atlantic border). What are the reasons for these modifications ?

The bathymetry of several grid points have been manually modified (i) in the Adriatic Sea especially along the Croatian coastline due to the presence of a large number of small islands causing instabilities in the circulation, (ii) in the straits in order to achieve a more realistic value of the transport, (iii) in the Atlantic border in order to be closer to the bathymetry of the Copernicus Global Analysis and Forecast product to avoid large inconsistencies in the nesting.

The sentence in Line 89 will be then modified as follows: "The topography is an interpolation of the GEBCO 30 arc second grid (Weatherall et al., 2015) filtered and specifically modified in critical areas such as: the Eastern Adriatic coastal areas (to avoid instabilities in circulation due to the presence of a large number of small islands), Gibraltar and Messina straits (to better represent the transports), Atlantic external border (to avoid large bathymetric inconsistencies with respect to the Copernicus Global Analysis and Forecast product)."

One additional reference will be included:

Weatherall, P., Marks, K. M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J. E., Rovere, M., Chayes, D., Ferrini, V. and Wigley, R.: A new digital bathymetric model of the world's oceans, Earth and space Science, 2(8), 331-345, 2015.

The barotropic time step is 100 times smaller than the baroclinic time step, is this justified by code stability or other concerns? This seems large compared to other model configurations already published.

This ratio between the baroclinic and barotropic time step is used to preserve stability, and it is a constant value used for both open ocean and coastal areas.

2.1.2

Could you explain if atmospheric forcing is a mixed of analysis and forecast, or only atmospheric analysis during the ocean analysis phase and only atmospheric forecast during the ocean forecast phase ? And what's the higher temporal resolution from year 2020 ? is it 1h ?

The Med-MFC operational systems are forced by ECMWF analysis fields for producing the analysis time series in the past, while atmospheric forecasting atmospheric data is used for producing the forecast. In this context we provide the evaluation of model outputs of systems forced with ECMWF analysis fields at 1/8 deg resolution and 6 hours temporal frequency before 2020, while after 2020 the atmospheric fields resolution is increased to 1/10 deg but the temporal frequency of the analysis fields is still 6 hours (only the forecast temporal resolution has been increased to 1 – 3 – 6 hours according to the forecast leading time). Following the reviewer's comment the sentence in Lines 118-121 will be updated as follows: "The atmospheric forcing fields for both NEMO and WW3 models are from the 1/8° horizontal resolution at 6 hours temporal frequency (3 hours frequency is used to force the first 3 days of forecast) operational analysis fields from the European Centre for Medium-Range Weather Forecast (ECMWF) Integrated Forecasting System (IFS), a higher spatial resolution of 1/10° (with higher forecast temporal frequency of 1-3-6 hours according to the forecast leading time) is used starting from year 2020."

In addition, Lines 189-191 will be revised as follows: "The WAV component is forced with 10 m above sea surface analysis and forecast ECMWF winds at 1/8° dissemination resolution. The temporal resolution is 6 h for the analysis, 3 h for the first 3 days of the forecast and 6 h for the rest of the forecast cycle. From year 2021, a higher spatial (1/10° for both analysis and forecast) and temporal (hourly for forecast days 1-3, 3-hourly for days 4-6 and 6-hourly for days 7-10) resolution dataset is used to force the WAV component."

You are using closed boundaries in Atlantic for WW3 model, this is strange and not consistent with the justification to have a Atlantic model for the boundary condition of the WAV system. Could you comment on this choice and justify why there is different implementation for these two models.

The WW3 modeling implementation is meant to provide the surface drag coefficient to NEMO to improve the surface wind stress representation in the Mediterranean Sea circulation model, which is not affected by the choice of the closed boundaries in the Atlantic side of the domain.

On the other hand the WAM implementation is intended to provide the Copernicus Marine analysis and forecast wave product for the Mediterranean Sea, thus it considers additional modules and a more complex setup including open lateral boundaries in the Atlantic in order to provide a better representation of swell in the Alboran Sea and further east along the north African continental shelf area (see also answer in section 2.2.2).

You provide the salinity of the river discharge in table A4, there are differences depending of the river, I did not find justification of these differences in Delrosso 2020. How do you explain and justify these differences?

The choice of the salinity associated with the river discharge has been evaluated by means of sensitivity experiments that has been used to provide the values presented in Delrosso 2020 PhD thesis. Due to lack of observations of salinity at river mouths, it was not possible to validate the results close to these areas, but an assessment was made at basin scale and in sub-regional domains. The salinity values that have been chosen are the ones providing a higher salinity skill in the overall basin. These values would be improved once an estuary box model (like to the one presented in Verri et al., 2020) will be implemented to retrieve more realistic and time dependent salinity values.

Following the reviewer's comment, the sentence at Line 127 will be integrated as follows: "The river runoff inputs consist of monthly climatological data for 39 major rivers (characterised by an average discharge larger than 50 m3/s) with a prescribed constant salinity at river mouth (Delrosso, 2020) evaluated by means of sensitivity experiments and listed in Table A.4. More realistic and time varying river salinity values (at least for major rivers) would be evaluated in next model evolutions using an estuary box model, such as the one presented in Verri et al. (2020), coupled to the circulation model."

A new reference will be then added:

Verri, G., Pinardi, N., Bryan, F., Tseng, Y., Coppini, G., and Clementi, E.: A box model to represent estuarine dynamics in mesoscale resolution ocean models, Ocean Modelling. http://dx.doi.org/10.1016/j.ocemod.2020.101587, 2020.

2.1.3 The data assimilation component

You describe a method for rejecting an observation based on a quality check. Could you explain how this square departure is computed in the methodology ? For each individual Temperature or salinity profile along the vertical? For each sla track? In a spatial box, temporal window?

The quality check is applied to each individual observation of each Argo vertical profile and each altimeter track. There is no binning in the vertical or horizontal spatial scales.

We will add a sentence in Line 155 as follows: "The quality checks are applied to each individual observation of each Argo vertical profile and for each altimeter track."

Correction to the background is applied once a day, does it mean that it is applied during the last time step of the day, there is no IAU method used to apply analysis increment?

The reviewer is correct, we apply the corrections to the restart files without any IAU in the present system. The use of IAU in the PHY system is under investigation.

The text in Line 157 will be modified as follows: “The misfits are computed at the observation time by applying the FGAT (First Guess at the Appropriate Time) procedure and the corrections to the background are applied once a day to the restart file using observations within a one-day time window.”

2.2.2. Model initialization, external forcing and boundary conditions

Could you explain what is the impact of the lateral forcing with full wave spectrum in the atlantic and why it is not applied in the WW3 configuration?

A nesting approach of this kind (Med-waves nested within a coarse resolution Atlantic wave model) enables us to properly simulate the effect of the remotely generated Atlantic swell into the Mediterranean Sea as it passes through the Strait of Gibraltar. In fact, Cavaleri and Sclavo (2006) pointed out that the narrow Strait of Gibraltar appreciably affects the wave climate in the close-by area of the Alboran Sea and further east along the north African continental shelf and it is often neglected in wave modelling systems of the Mediterranean Sea. At this stage of Med-MFC system development, as it is described in the present work in terms of wave - currents interactions in the general circulation model component, nesting of WW3 wave model with a coarser resolution Atlantic wave model wouldn’t have had altered the estimation of the neutral drag coefficient computed by WW3 and passed to NEMO to compute the surface stress neither in the Alboran Sea nor in the Atlantic area which is a buffer zone for the NEMO model setup.

Cavaleri, L. and Sclavo, M.: The calibration of wind and wave model data in the Mediterranean Sea, Coastal Eng., 53, 613–627, 2006.

2.3.2. Model initialization, external forcing and boundary conditions

Could you provide more information on the initialisation procedure for the BGC model. Does the BGC model initialisation use a constant profile for each area and for each variable for the initialisation? How long the model is integrated to smooth these discontinuities between the areas?

We thank the reviewer for the comments that allow us to better clarify some aspects of the BGC model setup. For a subset of variables (nitrate, ammonia, silicate, phosphate, oxygen, alkalinity and DIC) the initial condition consists of 16 profiles homogeneously applied to all gridpoints of each of the 16 sub-regions of Fig. 3. The profiles are computed from the EMODnet dataset (Buga et al., 2018). The other biogeochemical state variables (phytoplankton, zooplankton and bacteria biomasses) are initialised in the photic layer (0–200 m) according to the standard BFM values (See BFM manual: Vichi et al., 2020). Then, a 5-year hindcast is run using the first year (i.e., 2017) in perpetual mode to smooth discontinuities among sub-areas (e.g., protocol described in Salon et al., 2019).

The sentence at Lines 255-256 will be revised as follows:

"Initial condition of nutrients (nitrate, ammonia, silicate and phosphate), oxygen and carbonate variables (DIC and alkalinity) consists of 16 climatological profiles homogeneously applied in each of the sub-regions represented in Fig. 3. Climatological profiles are computed from the EMODnet dataset (Buga et at., 2018). The other biogeochemical state variables (phytoplankton, zooplankton and bacteria biomasses) are initialised in the photic layer (0–200 m) according to the standard BFM values. A 5-year hindcast is run using the first year (i.e. 2017) in perpetual mode.

Two new references will be included:

1) Salon, S., Cossarini, G., Bolzon, G., Feudale, L., Lazzari, P., Teruzzi, A., et al.: Novel metrics based on Biogeochemical Argo data to improve the model uncertainty evaluation of the CMEMS Mediterranean marine ecosystem forecasts. Ocean Science, 15(4), 997-1022, 2019.
2) Vichi, M., Lovato, T., Butenschön, M., Tedesco, L., Lazzari, P., Cossarini, G., Masina, S., Pinardi, N., Solidoro, C., and Zavatarelli, M.: The Biogeochemical Flux Model (BFM): Equation Description and User Manual. BFM version 5.2. BFM Report series N. 1, Release 1.2, June 2020, Bologna, Italy, http://bfm-community.eu, pp. 104, 2020.

Could you provide more information about the newtonian damping. Where exactly? Only for the Atlantic boundary? The Atlantic part in the bio system is smaller than for the physics if I am right?

In the BIO model, the Atlantic boundary is set at the longitude of 9°W. It is smaller than the Atlantic area of the physical domain (see dashed red line in Figure 2). The Newtonian dumping (D_N) term consists of the following equation:

$D_N = 1/t^N_{lon} (C^N - C^t_{lon})$

where t^N_{lon} is the time scale of the relaxation, which is set to 1/24 [d^{-1}] at Lon=9°W and 90 [d^{-1}] at Lon=6.5°W and it varies linearly between the two limits. C^N are the seasonal climatological profiles and C^t_{lon} are the tracer concentration profiles at time t and Longitude lon.

The sentence at Lines 257-259 will be revised as follows:

"The model has two open lateral conditions: in the Atlantic Ocean and in the Dardanelles Strait. Nutrients, oxygen, DIC and alkalinity in the Atlantic (i.e., boundary at lon=9°W) are provided through seasonally varying climatological profiles derived from Word Ocean Atlas (WOA 2018) and literature (Alvarez et al., 2014) and a Newtonian dumping is applied. The Newtonian dumping is set between the longitudes 9°W and 6.5°W with a time scale relaxation term linearly varying from 1/24 d-1 at 9°W to 90 d-1 at 6.5°W."

Could you provide more information on the atmospheric pCO2 forcing, it is not clear how this is computed. Is it a constant mean value applied for the Med sea? what is extrapolated?

Atmospheric pCO2 forcing is the timeseries of the annual averages measured at the Lampedusa station (Artuso et al., 2009; Trisolino et al., 2020). Data availability covers the period 1992 and 2018 (http://cdiac.ess-dive.lbl.gov/ftp/trends/co2/lampedus.co2). Thus, the values for the years 2019 and 2020 are computed by extrapolating the linear trend. The annual average is homogeneously applied to all grid points of the Mediterranean Sea. Finally, to avoid discontinuity, daily pCO2 values are computed by linear interpolation of the annual values.

The sentence at Lines 270-272 will revised as follows:

"Spatially constant values of atmospheric pCO2 are derived from the 1992-2018 time series of the ENEA Lampedusa station (Trisolino et al., 2021) with the 2019 and 2020 values extrapolated by linear trend."

We will include 2 additional references:

1) Artuso, F., Chamard, P., Piacentino, S., Sferlazzo, D. M., De Silvestri, L., Di Sarra, A., and Monteleone, F.: Influence of transport and trends in atmospheric CO2 at Lampedusa, Atmospheric Environment, 43(19), 3044-3051, 2009.
2) Trisolino, P., di Sarra, A., Sferlazzo, D., Piacentino, S., Monteleone, F., Di Iorio, T., Apadula, F., Heltai, D., Lanza, A., Vocino, A., Caracciolo di Torchiarolo, L., Bonasoni, P., Calzolari, F., Busetto, M., and Cristofanelli, P.: Application of a common methodology to select in situ CO2 observations representative of the atmospheric background to an Italian Collaborative Network, Atmosphere, 12, 246. https://doi.org/10.3390/atmos12020246, 2021.

You didn't provide very precise information on the computation of error covariance for the biogeochemistry. Is it the same characteristic than for the physical assimilation system? Which resolution, length of the simulation to compute EOF? Also 3-year simulation or there are other constraints related to bgc processes and differences in term of observations

In order to provide more details on the BIO variational assimilation scheme, the sentence at Lines 274-286 will revised as follows:

"The BIO component features a variational data assimilation scheme (3DVarBio) which is based on the minimization of the cost function (Eq. 1) (Teruzzi et al., 2014). Minimization is computed iteratively in a reduced space using an efficient parallel PETSc/TAO solver (Teruzzi et al., 2019) and the background error covariance matrix, B, is factored as $B = VV$, where V is a sequence of linear operators: $V = V_B V_H V_V$. The horizontal error covariance operator (V_H) is a gaussian filter and includes non-uniform and direction-dependent length scale correlation radius to account for anisotropic coastal assimilation (Terruzzi et al., 2018) and vertical profile assimilation (Cossarini et al., 2019). The vertical error covariance operator (V_V) is based on a set of 0-200m vertical error profiles obtained using an empirical orthogonal functions (EOFs) decomposition of a 20-year long pre-existing biogeochemical simulation. EOFs are computed monthly for the 16 subregions with the actual vertical resolution and rescaled at each grid-point considering the ratio between observation and model variances (Teruzzi et al., 2018). The biogeochemical error covariance operator (V_B) is designed to preserve the ratios among phytoplankton functional types and their internal carbon to nutrient quotas (Teruzzi et al., 2014) and supports monthly and spatial varying covariances between dissolved inorganic nutrients (Teruzzi et al., 2021)."

Quality assessment

3.1. PHY component skill

You explain that salinity is characterised by a negative bias, this is not what is shown in tab2. Negative bias is only in the first layer, below the bias seems to be positive.

The reviewer is correct, there were errors in Table 2: all the values of bias were erroneously provided as observations minus model. Thus in the revised version all the signs will be corrected in Table 2. We also noticed several minor typos in the text from Line 363 to Line 368, so the text will be modified as follows: "Table 2 summarises the EAN of 3D model temperature and salinity daily mean values compared to in-situ observations, in particular Argo floats and CTD profiles averaged over the three reference years. Model temperature shows small positive and negative biases depending on the depth, with the largest error (maximum value of the period is 0.85°C) in the sub-surface layers between 10 and 60 m, decreasing with depth. Salinity is characterised by an almost general negative small bias, meaning generally lower salinities than measured, along the whole water column except for the first layer. The salinity RMSD mean value is generally lower than 0.2 PSU, the error is larger in the first layers and decreases significantly below 150 m."

Revised Table 2:

Layer (m)	Temperature RMSD (°C)	Temperature bias (°C)	Salinity RMSD (PSU)	Salinity bias (PSU)
0-10	0.54	-0.02	0.19	0.01
10-30	0.82	-0.04	0.20	-0.01
30-60	0.85	0.04	0.19	-0.01
60-100	0.58	0.03	0.16	-0.02
100-150	0.41	-0.01	0.13	-0.01
150-300	0.28	-0.02	0.08	0.00
300-600	0.18	0.00	0.05	-0.01
600-1000	0.09	-0.02	0.03	0.00
1000-2000	0.05	0.01	0.02	0.00

Regarding the temperature bias, the bias seems to be negative in the upper layer and positive below 60m depth. Could you explain your assumption and how overestimation of shortwave flux will produce a warm bias only below the mixed layer?

The reviewer is correct, the temperature bias is almost negative at surface as from Figure 4 and also as shown in Table 2 (corrected) for the first layers between 0 and 30 m depth. These 2 validation analyses were carried out with respect to in-situ data and we did not consider the surface layers. In the submitted version of the paper, the comment referred instead to the SST bias evaluated with respect to satellite data, which was not presented. So following the reviewer's comment, we will include the SST bias in Table 3 and several changes to the text will be added too.

1) The text at Lines 389-392 will be updated as follows: "The temperature misfits (Figure 4c) indicate an overall overestimation of the temperature, except for the subsurface layer, during winter and spring."

2) The text at Lines 403-405 will be updated as follows: "Table 4 presents the RMSD and bias values computed for SST as well as SLA RMSD averaged in the Mediterranean Sea and over the 16 sub-regions (see Figure 3). Considering SST, the RMSD values range between 0.47 °C and 0.69°C (mean Mediterranean Sea error is 0.54 °C) and the bias is generally positive, possibly caused by an overestimation of the downward shortwave radiation flux which is estimated according to Reed (1977) formula, as already discussed in (Byun et al., 2007) and (Pettenuzzo et al., 2010). The SLA error ranges between 2.3 cm and 5.3 cm (mean error is 3.8 cm)."

3) Table 4 will be updated adding SST bias:

Table 4. EAN RMSD and Bias of SST and SLA RMSD averaged in the whole Mediterranean Sea and 16 sub-regions (see Figure 3) for the period 2018-2020.

Region	**Temp. RMSD (°C)**	**Temp. Bias (°C)**	**Sea Level Anomaly RMSD (cm)**
MED SEA	0.54	0.12	3.8
REGION 1	0.69	-0.05	5.3
REGION 2	0.53	0.06	4.3
REGION 3	0.53	-0.01	3.2
REGION 4	0.55	0.15	5.1
REGION 5	0.47	0.13	3.1
REGION 6	0.49	0.15	3.5
REGION 7	0.51	0.22	5.0
REGION 8	0.55	0.16	3.8
REGION 9	0.51	0.14	3.4

REGION 10	0.58	0.20	2.3
REGION 11	0.63	0.08	NA
REGION 12	0.49	-0.01	4.0
REGION 13	0.59	0.14	3.6
REGION 14	0.57	0.16	3.3
REGION 15	0.53	0.13	4.4
REGION 16	0.52	0.24	3.1

4) Text at Line 535 will be changed as follows: “The PHY component has been validated comparing model data with respect to in-situ and satellite observations showing a good accuracy in representing the spatial pattern and the temporal variability ozf the temperature, salinity and sea level in the Mediterranean Sea. In particular the model has a warm surface temperature bias of +0.12°C when compared to satellite SST. The temperature error along the water column has a clear seasonal signal with the largest errors at the depth of the surface mixed layer and the seasonal thermocline. The model error in salinity is higher in the first layers and decreases significantly below 150 m. The SLA presents an average error of 3.8cm on the three-year averaged period for the whole basin.

Concerning the negative salinity bias due to mixing at Gilbraltar (fig 4) is it something verified with statistic in appropriate boxes for example in the Alboran sea or in a western part of the med sea ?

The salinity bias in different subregions is not presented in this paper, but the values which have been evaluated close to Gibraltar in the Alboran Sea (area 01 in Figure 3) is slightly negative (minimum value of -0.06 PSU in the subsurface) and the negative bias is even larger (minimum value of -0.13 in the subsurface) in the South-West Mediterranean area (area 02 in Figure 3). The exchanges at the Gibraltar strait as well as the salinity transport through the strait are a matter of an analysis carried out by means of sensitivity experiments and using a 2 layer box model, which will be soon published in a paper under preparation.

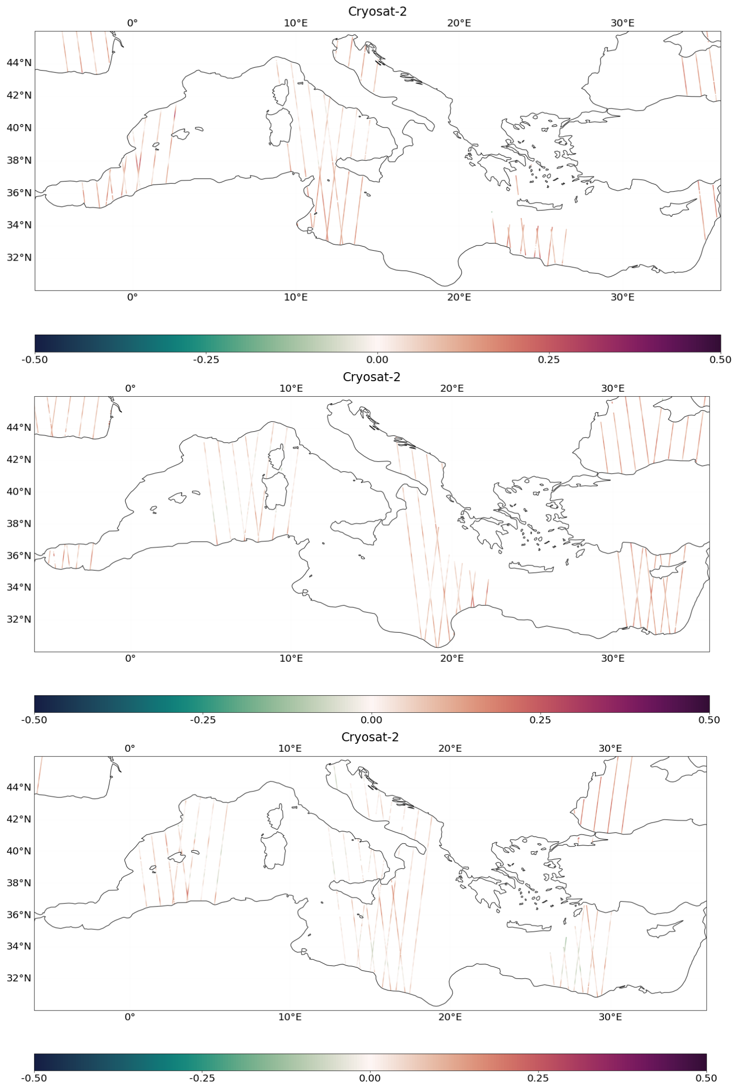
About the spatial variability of the SLA error (line 406), you suggest it could be impacted by the distribution of observations but sea level variability and the eddy kinetic energy should be much larger. Could you comment on this, is it link to your comment on model inaccuracies? Could you identify which components of the model are affected (forcing, assimilation method, numerical scheme, missing processes ...?)

SLA data are assimilated only at depth larger than 1 km and satellite tracks are not homogeneously covering the domain, thus the skill of the SLA can vary in the basin. A different assimilation scheme, as the barotropic model operator, could improve the quality in shelf regions.

We would update the text in Line 406 as follows: "The SLA skill scores vary in different regions, this could be related to the spatial coverage of the observations (not homogeneous in the basin) and on the limit of the 1000m assimilation depth (due to the dynamic height operator which assumes a level-of-no-motion to compute the sea level increments from temperature and salinity increments, see section 2.1.3).”

You have large differences of sla error between satellites. How do you explain these differences between the satellites? Do you use the same measurement errors for all the satellites? Is it due to the satellite coverage?

The measurement error is the same for all the satellites and is 4 cm. It is true that the skill score with respect to different satellites can be a bit different. This is especially true for CRYOSAT-2 which is a satellite designed to observe polar regions. The 3 consecutive 10-days spatial coverage of the CRYOSAT-2 is shown below 01/01/2018-09/01/2018, 11/01/2018-19/01/2018, 21/01/2018-29/01/2018. It is clear the coverage of CRYOSAT-2 is localized compared to other satellites (not shown) which have more evenly distributed coverage.



3.2. WAV component skill

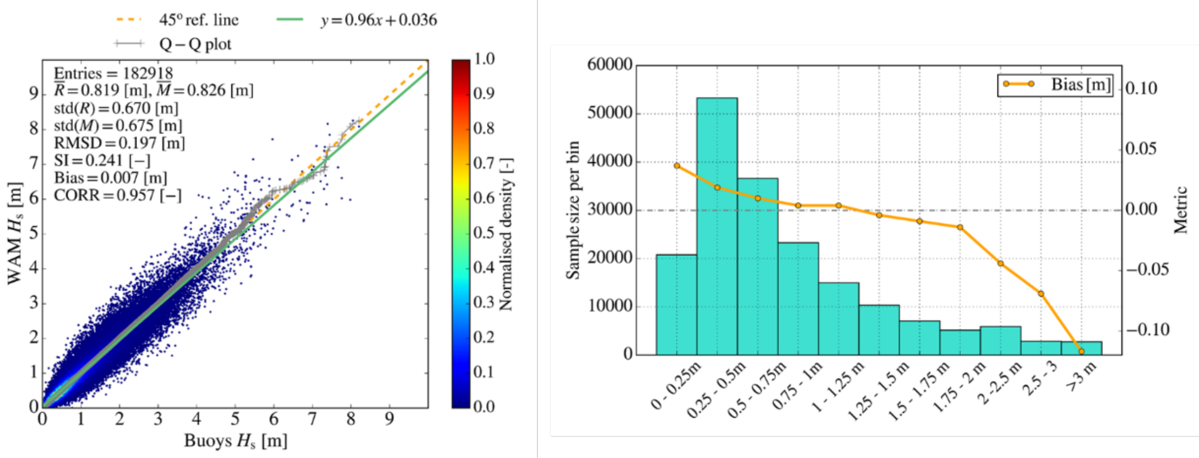
You explain at the beginning of the paper that forecasts are not assessed in this paper, is it different for the waves?

As for the other components of the Med-MFC system, the skill of the wave model is assessed by considering inter-comparisons of the wave solution during the 24-h analysis phase with in-situ (SWH and mean wave period) and remotely sensed (SWH) observations. As the latter are ingested into the model through data assimilation, the first guess model SWH fields (i.e. model background) are used instead of analyses.

It is difficult to see in the figure what is explained in the text, for example we can't see the underestimation for very small wave heights (<0.6m), neither the underestimation for MWP<7s. If I am right, in fig 6, there is an underestimation for the period <5s and overestimation for period > 7s. Do you think the overestimation is

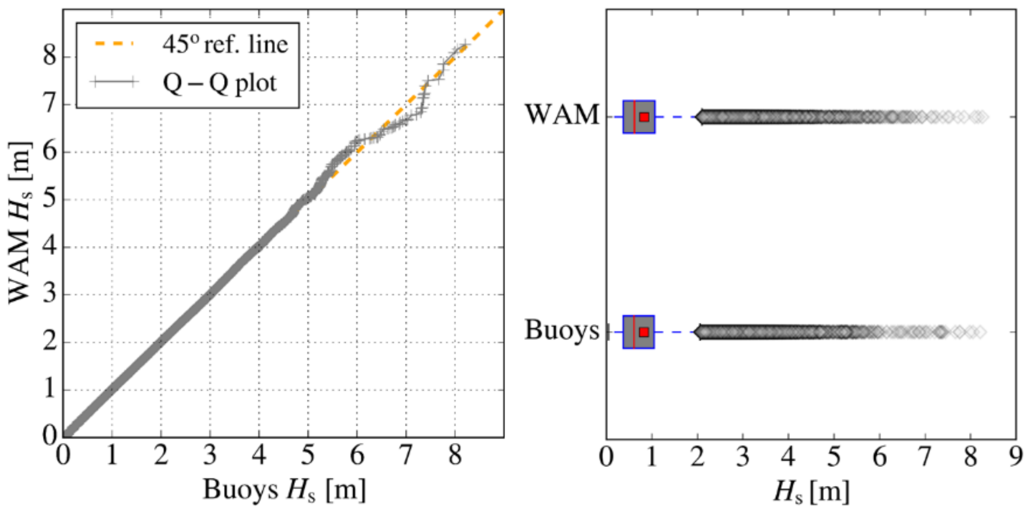
significant for 2m waves? It's difficult to trust this information with figure 2. Could you provide more information and argument to consolidate these conclusions.
Could you explain better the interpretation or the figures.

Generally, the overall tendency of model performance is a positive BIAS (0.007m) for significant wave height, meaning a slight model SWH overestimation. The model and observed data are highly correlated (as the green line of the left-hand side graph indicates). Although for certain buoys with high sample size, the model demonstrates SWH underestimation for small wave heights (<0.6 m), which changes for SWH>0.6 (e.g. for buoy 6101404, BIAS is altered from -0.028 to 0.042 m, and for Barcelona coast buoy 61499 it is altered from -0.033 to 0.078), the reviewer's comment is correct, as this is not the case for all buoys and cannot characterise the model tendency for the total sample size. Even though scatter plots are a practical and valuable way to assess the overall model performance, it is also possible for information to be masked due to the size of the dataset. To avoid any subjectivity potentially linked to scatter plot interpretation, we have undertaken a further investigation to analyse better the model performance of the model for various Hs ranges (0 - 0.25m, 0.25 - 0.5 m, etc.), utilising the average BIAS per cluster (i.e. the average error between modelled and observed data), which allows us to detect the underestimation (or overestimation) of the studied parameter (lower graph, right-hand side panel).



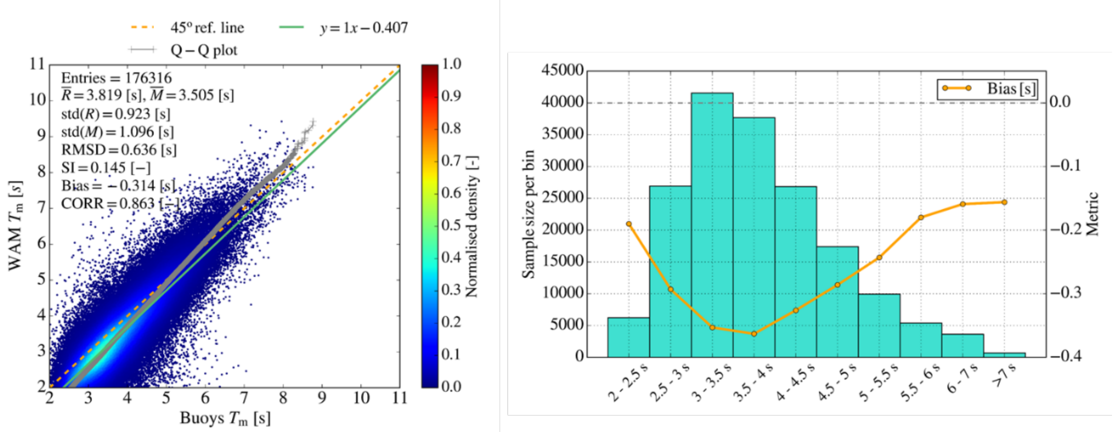
For the immense number of match-up data (within the range (0 – 1.25m), the model overestimates Hs with respect to the buoy measurements. Additionally, the model underestimates Hs during more energetic events (>1.25m), except for the range 5.5-6.2 m, where a positive bias is observed. For large wave heights, model results underestimate Hs compared to the buoys, which agrees with past findings (Ardhuin et al., 2007; Korres et al., 2011) for the Mediterranean Sea. Negative Hs Bias can be attributed to errors in the forcing or inaccurate wave growth and dissipation at high wind speeds (Pineau-Guillou et al., 2018) (note that wind speed was overall overpredicted, compared to satellite data, Bias = 0.173 m/s).

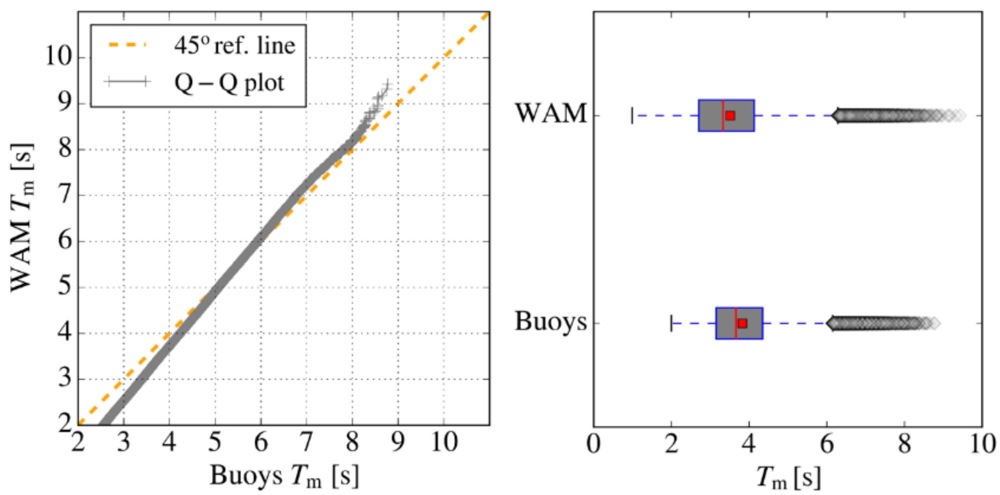
Except for scatter plots, Quantile-Quantile (Q-Q) plots are also employed as they provide the opportunity to check whether the model can reproduce the observed sample wave climate. The dashed orange line (below referred to as the 45° ref. line) stands for the unit gradient line. From the following Q-Q plot (left-hand side panel), we observe that model results follow the dashed orange line very closely, suggesting the model produces well the distribution of Hs observations (this is also evident by the box plots on the right-hand side panel, as the modelled and observed data appear to be distributed in a similar manner). Although for higher waves (> 1.25m) the model tends to underestimate Hs (except for the range 5.5-6.2 m), the high percentile assessment of modelled output suggested that the model overproduces very large wave heights (this is valid for 100th, 99.97th, 99.96th, 99.95th percentiles); hence this is why we observe this deviation from the orange dashed reference line in the Q-Q plot for very high waves.



Quantile	Buoys H_s [m]	WAM H_s [m]	Comment
100th	8.207	8.261	overestimation
99.99th	7.048	6.703	underestimation
99.98th	6.392	6.311	underestimation
99.97th	5.816	6.011	overestimation
99.96th	5.591	5.812	overestimation
99.95th	5.437	5.599	overestimation

Regarding mean wave period, the model tends to systematically underestimate it, as also confirmed by an investigation of BIAS per Tm cluster. A similar analysis as the previous one (focused on high MWP percentiles) revealed that despite the overall modelled MWP underestimation, the system tends to overestimate MWP for high percentiles/very long waves (hence we observe the deviation of the Q-Q plot from the unit gradient line for very high periods).





Quantile	Buoys T_m [m]	WAM T_m [m]	Comment
100th	8.770	9.425	overestimation
99.99th	8.312	8.642	overestimation
99.98th	8.197	8.406	overestimation
99.97th	8.074	8.276	overestimation
99.96th	7.999	8.164	overestimation
99.95th	7.928	8.107	overestimation

During the revision phase, we plan to clarify these aspects within the paper (revising text in lines 419-423 as follows:

"Figure 6 depicts scatter plots of the evaluation of the observed SWH and MWP against measurements obtained from the 28 buoys. For the immense number of match-up data (within the range (0 – 1.25m), the model overestimates SWH with respect to the buoy measurements (left-hand side panel). Additionally, the model underestimates SWH during more energetic events (>1.25m), except for the range 5.5-6.2 m. For large wave heights, model results underestimate SWH compared to the buoys, which agrees with past findings for the Mediterranean Sea (Ardhuin et al., 2007; Korres et al., 2011) . Negative SWH BIAS can be attributed to errors in the forcing or inaccurate wave growth and dissipation at high wind speeds (Pineau-Guillou et al., 2018). The dashed orange line (i.e. the 45o ref. line) in the Quantile-Quantile (QQ) plot stands for the unit gradient line. We observe that model results follow the dashed orange line very closely, meaning the model produces well the distribution of SWH observations. Although for higher waves (> 1.25m) the model tends to underestimate SWH (except for the range 5.5-6.2 m), it overproduces very large wave heights (100th, 99.97th, 99.96th, 99.95th percentiles); hence a deviation from the orange dashed reference line in the Q-Q plot becomes prominent for very high waves. Concerning MWP, the model systematically underestimates it (right-hand side panel). Despite the overall modelled MWP underestimation (BIAS = -0.314 s), the system tends to overestimate MWP for high percentiles/very long waves (hence we observe the deviation of the Q-Q plot from the unit gradient line for very high periods)."

Ardhuin, F., Bertotti, L., Bidlot, J. R., Cavaleri, L., Filipetto, V., Lefevre, J. M., and Wittmann, P.: Comparison of wind and wave measurements and models in the Western Mediterranean Sea, Ocean Eng., 34, 526–541, https://doi.org/10.1016/j.oceaneng.2006.02.008, 2007.

Korres, G., Papadopoulos, A., Katsafados, P., Ballas, D., Perivoliotis, L., and Nittis, K.: A 2-year intercomparison of the WAM-Cycle4 and the WAVEWATCH-III wave models implemented within the Mediterranean Sea, Mediterranean Marine Science, 12(1), 129-152, Mediterranean Marine Science, 12(1), 129–152. https://doi.org/10.12681/mms.57, 2011.

Pineau-Guillou, L., Ardhuin, F., Bouin, M.-N., Redelsperger, J.-L., Chapron, B., Bidlot, J.-R., and Quilfen Y.: Strong winds in a coupled wave-atmosphere model during a North Atlantic storm event: evaluation against observations, Quarterly Journal of the Royal Meteorological Society, 144(711), Part B, 317-332. https://doi.org/10.1002/qj.3205, 2018.

You haven't shown the seasonal results, could you say whether the best results in winter are for the height or for the period or both?

We thank the reviewer for this comment. The following tables present standard quality metrics for significant wave height and mean wave period analyses against buoy data per season. Overall, for both variables, the results suggest that the seasonal variability is adequately captured by the model.

Table 1: Evaluation of the modelled SWH (analysis) against buoy measurements for the entire Mediterranean Sea, for the period 2018-2020, per season.

MED	$\underline{R}$ [m]	$\underline{M}$ [m]	RMSD [m]	SI [-]	Bias [m]	CORR [-]
Whole Year	0.819	0.826	0.197	0.241	0.007	0.957
Winter	0.978	0.980	0.226	0.231	0.001	0.963
Spring	0.886	0.891	0.201	0.227	0.005	0.956
Summer	0.591	0.603	0.154	0.260	0.012	0.932
Autumn	0.849	0.858	0.207	0.244	0.010	0.952

Table 2: Evaluation of the modelled MWP (analysis) against buoy measurements for the entire Mediterranean Sea, for the period 2018-2020, per season.

MED	$\underline{R}$ [s]	$\underline{M}$ [s]	RMSD [s]	SI [-]	Bias [s]	CORR [-]
Whole Year	3.819	3.505	0.636	0.145	-0.314	0.863
Winter	4.035	3.721	0.660	0.144	-0.315	0.878
Spring	3.924	3.612	0.638	0.142	-0.312	0.869

Summer	3.459	3.129	0.610	0.148	-0.330	0.792
Autumn	3.888	3.590	0.638	0.145	-0.298	0.859

Currently, the following sentence exists in the manuscript: "Equivalent seasonal results (not shown) revealed that the performance of the model is better in winter than in summer which agrees with other studies (Cavaleri and Sclavo, 2006; Ravdas et al., 2018)". We intend to refer to this information within the revised manuscript (Line 423) as follows:

"Seasonal results (not shown) for both variables SWH and MWP indicated that the model adequately captures the seasonal variability. For SWH, RMSD values vary from 0.154 m in summer to 0.231 m in winter. Nevertheless, SI is higher in summer (0.26) than during the other seasons. Additionally, the highest Pearson correlation coefficient (CORR) is observed in winter (0.963, while the lower one is equal to 0.932 and it is observed in summer). The metrics reveal that the model follows better the observations in winter than during the other months since the former is associated with more well-defined weather patterns and higher waves. A similar conclusion has been reached also by other studies (e.g. Ardhuin et al., 2007) for the Mediterranean Sea. Summer and autumn are characterised by higher SI values (0.244 and 0.260 respectively), while lower values are obtained for winter and spring (0.231 and 0.227 respectively). Finally, small positive BIAS values are met for all seasons, with the highest values found in summer (0.012 m). Regarding mean wave period, RMSD varies from 0.610 s in summer to 0.66 s in winter and BIAS is negative for all seasons. SI does not present significant seasonal variability, with the highest value encountered in summer. Finally, CORR for MWP is higher than 0.8 in all seasons (values are within the range 0.859 – 0.878, while during summer CORR equals 0.792). These metrics demonstrate that the model wave period (similarly to the wave height) correctly follows the observations in well-defined weather conditions characterised by higher waves and longer periods, agreeing with past studies (Cavaleri and Sclavo, 2006; Ravdas et al., 2018)."

Ardhuin, F., Bertotti, L., Bidlot, J. R., Cavaleri, L., Filipetto, V., Lefevre, J. M., and Wittmann, P.: Comparison of wind and wave measurements and models in the Western Mediterranean Sea, Ocean Eng., 34, 526–541, https://doi.org/10.1016/j.oceaneng.2006.02.008, 2007.

Line 432, could you explain what is the CORR deviation in the figure and the correlation coefficient commented in the text.

In the submitted version, this was not clarified, thank you for bringing this to our attention. CORR deviation in figure 7 stands for the deviation from unity of the correlation coefficient (CORR) commented in Line 432. The latter represents the Pearson correlation coefficient. We decided to maintain both CORR and CORR deviation within the manuscript, as the former is more tangible for the reader (included also in the scatter plots), while the CORR deviation facilitates its depiction along with other metrics (RMSD, Bias, SI) in Figure 7. We intend to clarify this in the text (line 432) as follows: "The Pearson correlation coefficient (CORR) mostly follows the pattern of variation of SI (in this figure we present the correlation coefficient from unity, CORR deviation). CORR ranges from 0.87 at SARON in the Aegean Sea to 0.97 at the deep-water buoy 6100196 offshore Spain, which is well-exposed to the prevailing north-westerly winds in the region."

Line 443 : Is it underestimation instead of overestimation ?

The ECMWF winds are overall overestimated with respect to the satellite observations as it is correctly written in Line 443. Line 442 will be changed to "It is seen that ECMWF forcing overestimates U10 with respect to observations, throughout most of U10 range while some underestimation is observed for high wind speeds (14 – 19 m/s)".

3.3 BIO component skill

L475 : there is no illustration of spatial gradient in fig 10, the figure only show the seasonal cycle

Indeed, Figure 10c reports the RMSDs computed between the maps of satellite and model outputs. The daily values of RMSD are then averaged over the two (winter and summer) periods. Thus, the metric shown in Fig. 10c is a measure of the skill performance of the model to reproduce spatial gradients.

However, we recognize that this point was not clearly presented. Thus, we propose to clarify this aspect by:

1) adding a new panel in Figure 10 with the spatial Pearson correlation between the maps of satellite and model output for the 16 sub-regions (see figure Fig. Rev1d, - new figure 10 -),

2) changing the caption of Figure 10 at Lines 1142-1144 as follows:

"Figure 10. Timeseries of surface chlorophyll for centred composite 7-day satellite (green) and the model forecast (black) in two selected sub-regions (a and b). RMS of differences (c) and Pearson correlation (d) between maps of satellite and model forecast for the day before the assimilation in the 16 sub-regions of Fig.5(c). Metrics are averaged over the winter (from Oct to Apr) and summer (from May to September) periods."

3) adding a new sentence at Line 478 as follows:

"Daily values of RMSD and of Pearson correlation are computed between satellite and model output maps, then averaged over the two periods (Figure 10c and d). The plot of RMDS (Figure 10c) shows that higher errors are registered in the western sub-regions and in winter, when chlorophyll levels and variability are higher. On the other hand, spatial correlation values are moderate and high in all sub-regions (i.e., values always above 0.5 except for a few sub-regions), with summer values better than winter values. Considering the number of grid points in each sub-regions, all values in Figure 10d should be considered significantly non-zero at the 0.05 level."

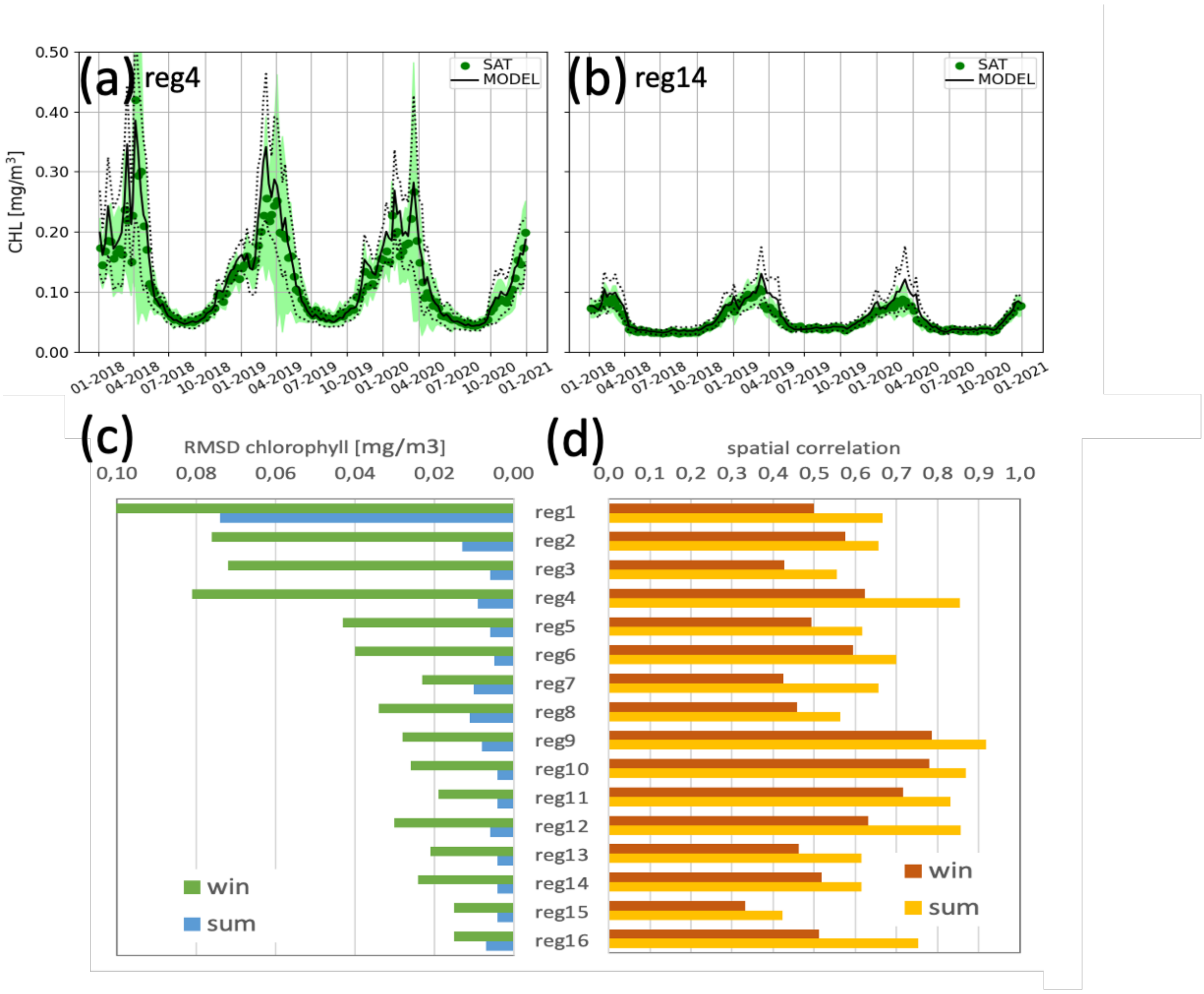


Fig. Rev1 (new Figure 10): Timeseries of surface chlorophyll for centred composite 7-day satellite (green) and the model forecast (black) in two selected sub-regions (a and b). RMS of differences (c) and Pearson correlation (d) between maps of satellite and model forecast for the day before the assimilation in the 16 sub-regions of

Fig.5(c). Daily values of the metrics are averaged over the winter (from Oct to Apr) and summer (from May to September) periods.

L483 : In table 5, the RMSD error is the order of 10 to 40m and not a meter, depending of the domain and the variable. Could you explain how the uncertainty is estimated

Many thanks for spotting this oversight. The sentence at Line 481-483 will be changed as follows:

"The depth of the deep chlorophyll maximum during summer and of the surface bloom during winter, as well as the depth of the nitracline and the depth of the maximum oxygen layer, which results from the interaction of physical and biogeochemical processes, are reproduced with an uncertainty of the $O(10^1)$ meters."

The uncertainty is computed as the RMS of the differences between the depths of the indicators (i.e., DCM, nitracline or maximum oxygen depth) for any matching BGC-Argo and corresponding model profiles.

L490 : Consistency between the observation and the model seems to be good in all the areas, but could you comment on some of the differences, are they significant ? For example, at the surface in the alboran sea, there is not a good agreement for alcalinity or DIC. It's also the case for alcalinity in the Aegean basin. Is it not possible to add the mean profile computed with the observation?

A new version of Figure 11 of Alkalinity with the mean values of EMODnet climatology (grey line and o marker) is shown in Figure Rev2. The figure confirms the good agreement of the model with the climatology. The Alboran (Aegean) sub-region shows an overestimation (underestimation) of about 20-30 μmol/kg in the upper 100m that should be mentioned but it is worth to note that modelled values are well within the range of variability of climatology. To better comment this figure, the sentence at Lines 490-492 will be changed as follows:

"Average maps and profiles of Alkalinity and DIC in selected sub-regions in the zonal directions (coloured lines) are well superimposed to the range of variability of the historical in-situ data (grey shaded areas) demonstrating the capability of the BIO component to reproduce both horizontal basin-wide gradients and vertical profiles in the different areas. A slight overestimation of DIC and alkalinity (underestimation of alkalinity) is simulated in the Alboran (Aegean) sub-region in the upper 0-100 layer."

For sake of readability, we prefer not to include the black solid-and-dotted line for EMODnet climatology in the new version of Figure 11 unless the reviewer suggests for that. However, to increase the readability of the figure, we will introduce thicker coloured dashed lines for the range of variability of the model output (see figure rev3; new figure 11) and we will change the caption consequently.

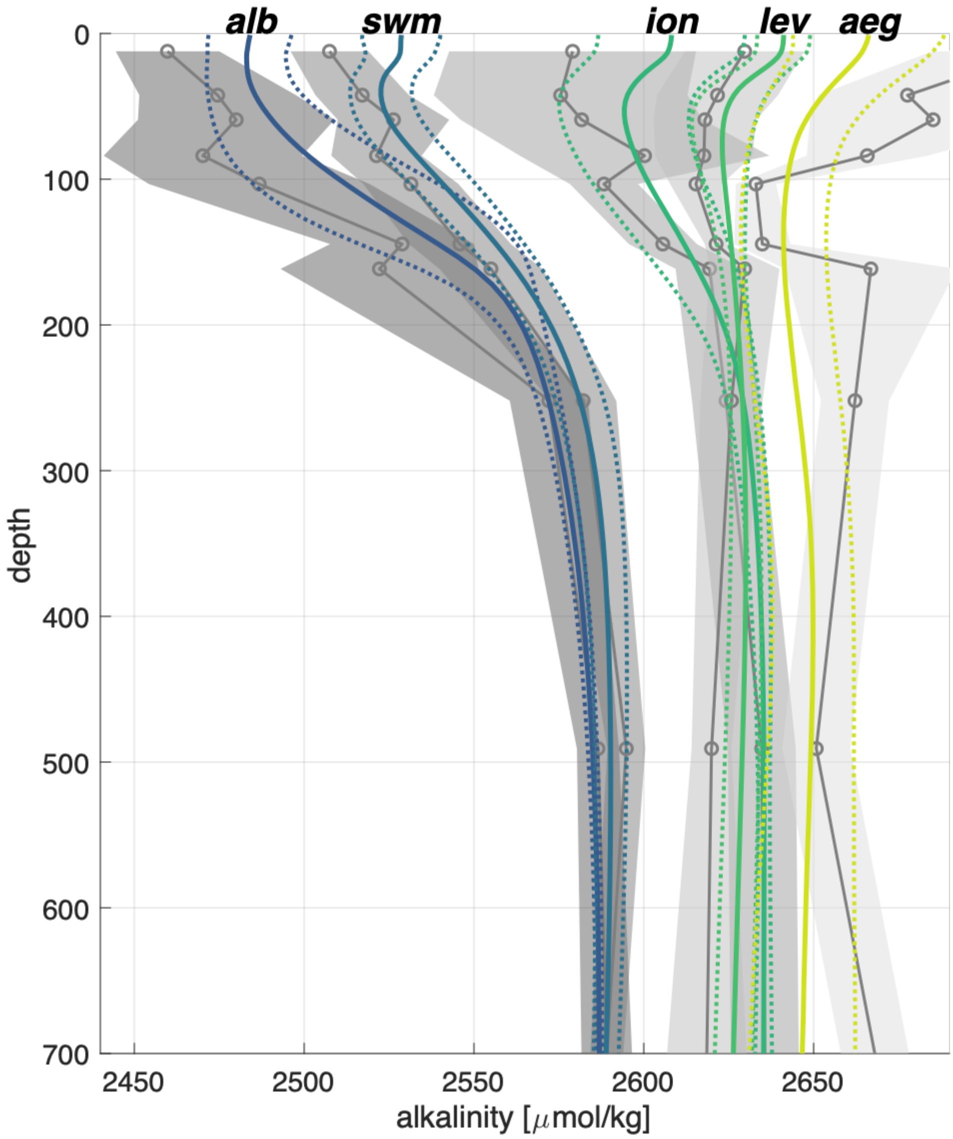


Figure rev2. Alkalinity profiles for model (average and range of variability, solid and dashed coloured lines, respectively) and Emodnet climatology (average and range of variability, black dots and lines and grey shaded areas, respectively) for selected macro areas. Climatological data are computed using historical data (Emodnet, 2018; Bakker 2014). Range of variability is the average ± standard deviation

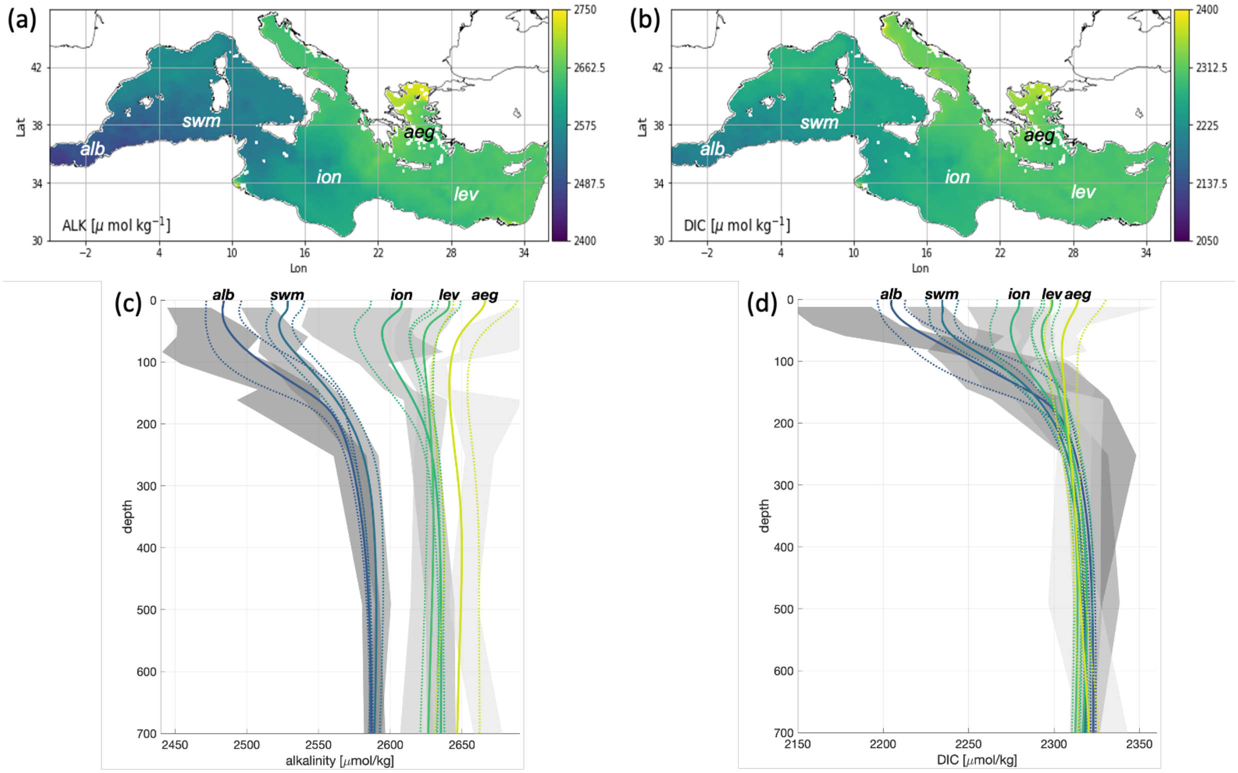


Figure rev3 (new Figure 11). Spatial distribution of modelled DIC (a) and alkalinity (b) and comparison of vertical profiles of DIC (c) and alkalinity (d) for model (average and range of variability, solid and dashed coloured lines, respectively) and Emodnet climatology (average and range of variability, black dots and lines and grey shaded areas, respectively) for selected macro areas. Climatological data are computed using historical data (Emodnet, 2018; Bakker 2014). The range of variability is the average ± standard deviation

Conclusions and Future Perspectives

The analysis of the source of uncertainty is missing. There is some information for the physical part and for the waves but nothing for the biogeochemistry. That should be very useful to synthetise main source of uncertainty and error in the conclusion for the different systems.

It is the first time that the biogeochemical model component is evaluated with such a rigorous multivariate (e.g., more than 10 biogeochemical variables) and multilevel (e.g., GODAE metrics of class1 in Tab. 6 and class4 in Tab. 5) validation framework (as described in Salon et al., 2019). Thus, our proposed results (e.g., accuracy values) represent a benchmark for the future improvements of the BIO system. We are aware that there are several sources of uncertainties for the BIO model: unproofed biogeochemical formulations and uncertainty in model parameterization, uncertainty in boundary conditions, initial conditions and land forcing, and impacts of unresolved physical dynamics. However, targeting any specific uncertainty is not the focus of this general and comprehensive paper. Other details of biogeochemical model uncertainty can be found in Salon et al. (2019) and Cossarini et al. (2021). Since this is an important aspect that we would like to stress, we will revise the sentence at Line 552-554 as follows:

"The BIO system has defined a validation framework (Salon et al., 2019) based on multivariate (e.g., more than 10 variables) and multilevel metrics that include GODAE class 1 and class 4 statistics and process oriented metrics. Particularly interesting, the present validation framework includes also near real time observations (i.e. satellite and BGC-Argo Argo) that show average errors in the 0-200 m layer of 0.04 mg/m3, 0.4 mmol/m3 and 16.8 mmol/m3 for chlorophyll, nitrate and oxygen, respectively. Thus, the validation framework represents a robust benchmark for the future improvements of the Mediterranean BIO model. Indeed, as detailed in Salon et al. (2019) and Cossarini et al. (2021), critical sources of the BIO model errors include unresolved Atlantic boundary conditions as well as land-sea and atmospheric-sea forcing uncertainty in model parameterization and inconsistency of coupled physical-biogeochemical processes."

Additionally, Lines 549-551 will be modified as follows:

"Overall, the quality of the WAV component stems from the ECMWF wind forcing that drives the wave dynamics, data assimilation, forcing from Med-PHY surface currents and improved parameterization of wave wind source and dissipation terms of WAM model. In particular, the WAV component assimilates satellite altimetry data with a well calibrated stand-alone OI scheme and implements regular updates and improved parameterization independently from the other components. Given that wind forcing quality has a substantial influence on the model response, a considerable part of the wave product uncertainty, especially under high winds or extreme conditions, is related to the wind forcing uncertainty and can be substantially improved by undertaking the ensemble approach in wave forecasting. The lower accuracy of the wave product in semi-enclosed regions of the Mediterranean Sea (e.g. Adriatic and Aegean Seas) can be related to the current spatiotemporal resolution of the wind forcing. Near the coast, unresolved topography by the wind and wave models and fetch limitations cause the wave model performance to deteriorate."

L563 : could you explain what is expected assimilating sea level in coastal stations? Complementarity to altimetry sea level? No plan to assimilate altimetry close to the coast and on the shelf?

In the current system, the altimeter observations are assimilated only if the depth below the measurement is greater than or equal to 1000 m. This is due to the dynamic height operator (section 2.1.3) which assumes a level-of-no-motion to compute the sea level increments from temperature and salinity increments. This represents a limitation to improve the forecast accuracy in shelf and coastal regions, which is especially important in the case of extreme events (e.g., storm surge). We are working on a new operator to substitute the dynamic height approximation and avoid its limitations that will allow us to assimilate altimetry also near the coast. Concerning coastal station sea level assimilation, this is not yet provided and the planning for this evolution is at long term, thus we would propose to remove from the paper modifying the sentence in Lines 562-563 as follows: "Another important goal for the future is to assimilate Argo and drifter trajectories (Nelson et al., 2016) and gliders (Dobrici et al., 2009), as well as sea level anomaly in coastal areas."

L580 : could you explain the link between these new model development/improvement and the expected impact on the metric and score computed in the previous section.

We thank the reviewer for this comment and we propose to update the text by adding new specific paragraphs for each of the three components.

Considering the PHY system, the paragraph at Line 557 will be improved as follows:

"Considering the PHY system, the users need finer spatial scales and higher time frequencies of the products especially for improving the representation of the coastal scale and limited area processes in nested models, thus providing a unique opportunity to model the coastal areas at the resolution of few hundred meters using nesting schemes as demonstrated in Federico et al. (2017) and Trotta et al. (2021) among the others. Users also require higher accuracy in storm surge forecasting, which can be achieved by including the explicit representation of the tidal forcing to resolve non-linear interactions between astronomical and internal tides with the baroclinic circulation. An upgrade of lateral open boundary conditions in the Atlantic and the Black Sea would provide better evaluation of the transport at Gibraltar on one side, and improved dynamics in the north Aegean Sea on the other. Higher frequency river runoff data from hydrological models, as well as more accurate salinity values at river mouths, would provide better salinity skill not only along the coastal areas but in the whole basin. Another important goal for the future is to assimilate Argo and drifter trajectories (Nelson et al., 2016) and gliders (Dobricic et al., 2009) as well as sea level anomaly in coastal areas. Finally, the future should consider ensemble forecasting to recast the deterministic forecast within a probabilistic framework assessing the modeling uncertainties (Pinardi et al, 2011, Millif et al, 2009, Thoppil et al, 2021; Barton et al., 2021)."

In order to explain the link between WAV developments/improvements and expected impacts on wave product quality Lines 566-569 will be revised as follows: "The required increased accuracy in wave height and mean periods predictions can be mainly achieved by improving the quality of the wind forcing which is the main driving force of wave models. Bias correction of ECMWF winds and further downscaling of ECMWF forecasts is expected to improve winds and consequently wave product quality especially in semi-enclosed areas (e.g. Adriatic, Aegean) and near the coast. Assimilation upgrades with the ingestion of multimission significant wave heights at 5Hz and in-situ wave heights measurements from HF Radars will improve accuracy in coastal areas of the

Mediterranean Sea while the inclusion of spectral information in the near future (e.g. CFOSAT wave spectrum) will further improve the prediction of the sea state. Finally the development of a WAV ensemble prediction based on ECMWF operational ensemble winds is expected to improve the existing accuracy of the deterministic forecast at lead times beyond 48 hours providing in parallel uncertainty estimates of wave parameters."

Regarding the BIO component, the sentences at Lines 570-580 will be revised to better explain the link between improvements/developments and expected impacts on quality of BIO products: "User requirements for the BIO component developments include improved quality and products tailored for ecosystem and coastal applications. The present results (i.e., the validation framework) have highlighted strengths and weaknesses of the current model system and helped identify biogeochemical model process representations and model parameter estimates that can be improved. These include better representation of vertical dynamics, a greater number of functional phytoplankton functional types and zooplankton compartments to describe the diversity of the plankton community and the different energy and matter pathways in the ecosystem. In addition, the integration of optics and biogeochemistry, including new coupled models and novel hyperspectral and high-resolution radiometric data, will be useful for calibrating parameters of important ecosystem processes (Lazzari et al., 2021). Assimilation of new in-situ profile sensors and variables (e.g., BGC-Argo Float and Glider) will help increase the reliability of BIO products, especially along the water column (Cossarini et al., 2019). Higher quality vertical dynamics can be achieved through better representation of vertical model error covariances by ensemble (Carrassi et al., 2018) or joint physical-biogeochemical data assimilation techniques. Finally, revising nutrient and carbon inputs from rivers (e.g., from monthly climatologies to daily observations or model predictions) will allow better resolution of coastal dynamics and coastal-offshore patterns in critical areas."

Corrections

L 51 : von schuckmann instead of von Schckumann
Powered by TCPDF (www.tcpdf.org)

We thank the reviewer and we will correct in the revised version of the manuscript

L 78 ref is missing

We thank the reviewer and we will include the reference to the OceanVar in the revised version of the manuscript.

L195 : global physical model instead of global wave model

The text will be changed to "global physical model". Thank you.

L417 : right panel instead of lower panel

The text will be changed to "lower panel of Fig. 7". Also changes will be introduced into Fig. 6 legend ("(bottom panel of Fig. 7)" instead of "(bottom panel)")

L442 : ECMWF forcing instead of ECMWF is forcing

Thank you. See our answer to comment (above) regarding Line 443.

Figure 7 : MYKKON buoy is missing?

MYKON buoy location will be added to the lower panel of Figure 7.

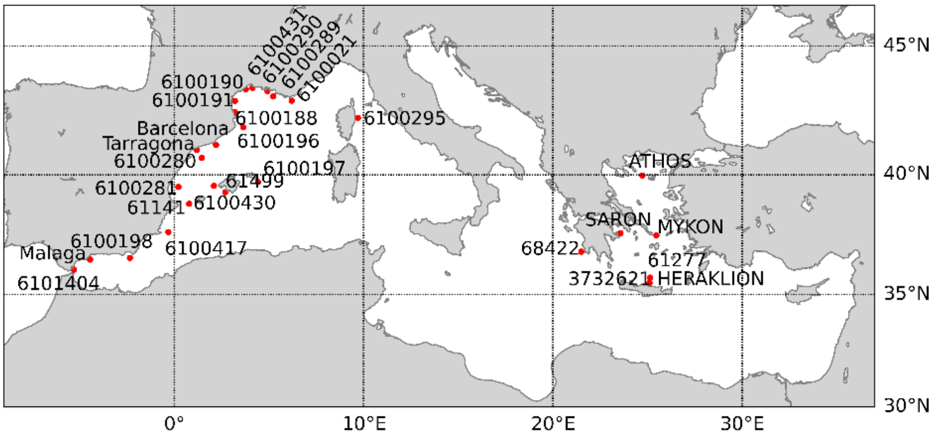


Figure 8 : left panel, the yellow dash line should be in the foreground as on the right panel. There is colorbar and no information about the grey cross and the yellow dashline in the legend.

Thank you for catching these errors. We will accordingly revise figure 8.

In a similar manner, we will also include in the revised manuscript, the corrected version of figure 6, following your latter suggestion (i.e. including the legend and colour bar).

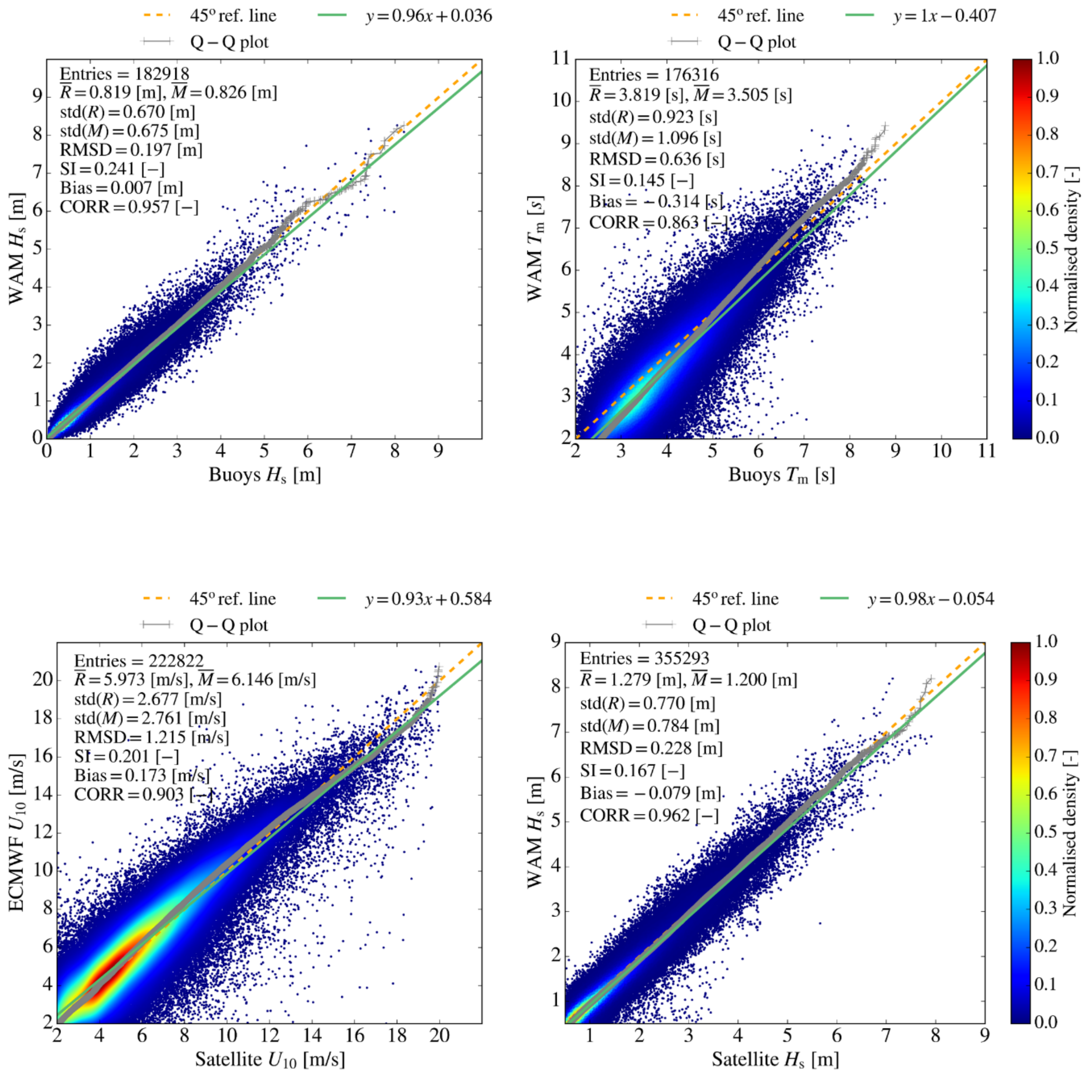


Figure 10 : is it model forecast or model analysis in black ?

Figure 10 shows model analysis. Many thanks for spotting this oversight. The captions will be changed as follows: "Timeseries of surface chlorophyll for centred composite 7-day satellite (green) and the model analysis (black) in two selected sub-regions…."

Table 3 : westward transport should be 0.87 instead of 0.087

The reviewer is correct, there is an error in table 3, the westward transport value will be corrected in the revised version

Gibraltar Transport	Model [2018-2020]	Literature Soto-Navaro et al. (2010) [2004-2009]	Literature Candela (2001) [1994-1996]
Net	0.040±0.017	0.038 ± 0.007	0.04 (max: 0.26, min: 0.11)
Eastward	0.91±0.01	0.81 ± 0.06	1.01 (max: 1.12, min: 0.91)
Westward	0.87±0.06	0.78 ± 0.05	0.97 (max: 0.83, min: 1.11)

Table 5 : mean values are missing for the Chl and Nitrate for the upper layer.

We are sorry, but this point is not clear to us. The missing values in Table 5 are for Nitrate metrics in swm and adr sub-regions. The reason is the lack of available BGC-Argo floats with the nitrogen sensor in these sub-regions.