Interactive comment on "Recent updates on the Copernicus Marine Service global ocean monitoring and forecasting real-time 1/12° high resolution system" by Jean-Michel Lellouche et al.

Anonymous Referee #2
Received and published: 22 April 2018

General comments:

This Discussion paper discusses the main updates of the Mercator Ocean operational forecasting systems at 1/12 resolution, which is the highest resolution deterministic forecast product released by CMEMS. The manuscript is certainly interesting and deserves publication because documents the main changes and quality increase achievements of a state-of-the-art oceanographic analysis system. It can be useful for both developers and users.
We thank anonymous Referee #2 for his careful reading of our manuscript and for this comment.

However, in my opinion there are many scientific issues that are only empirically formulated, lack scientific justification and require a deeper explanation. In general, it is also not clear why some updates are discussed in details in section 3 and some others only mentioned in section 2: this seems quite arbitrary. I therefore recommend revising the manuscript to address the specific points below to help the readership to understand the motivation and justification of such changes, which will really help the usefulness of this work in the oceanographic community and the robustness of the paper.
The issues empirically formulated concern essentially the choice of the "threshold values". For those involved in the quality controls QC1 and QC2, the justification and the criteria for choosing the value of these parameters have been added in the text.
For those involved in section 3.3, we followed the tunings used by Greiner et al., 2008 (internal report) and we checked that these values allow the method to work properly. We give to the Referee the access to this report: https://cloud.mercator-ocean.fr/public.php?service=files&t=2f3c0f2d260b51aac32a4d03da71e2d3.

We also described in details only some of the updates mentioned in section 2. The choice of updates separately illustrated and discussed in section 3 may seem arbitrary. It corresponds in fact to the updates that doesn't result from routine system improvements (bathymetry, runoffs, assimilated databases, Mean Dynamic Topography, etc.). This is mentioned in the manuscript.

I found the article well-written; it is a bit long, and I suggest the authors to consider removing some figures (29 figures are really too many in my opinion).
We agree that. We think also that the number of figures is large. We tried to reduce it before submitting but, on the other hand, we believe that all figures are beneficial to understanding the system. This manuscript describes a complex system with a lot of new ingredients. A solution would be to split the paper in two parts: description of the system and details of the main updates (sections 2 and 3), and scientific assessment (section 4). This has not been our final choice, also wanting to measure, in the same paper, the overall impact of the integration of all updates on the products quality.
We would like therefore to keep all the figures.

Specific comments:

Following constructive comments, we tried to make the manuscript clearer and more detailed. All remarks detailed below by the referee were considered and/or discussed.

Abstract L23: forecast error → background error.

The text has been modified overall the manuscript.

Introduction P1L16: I believe the fact that Mercator is entrusted by EC is not relevant: here it is relevant that Mercator Ocean is in charge of the global analysis and forecast system.

We agree that. Only the relevant part has been kept in the text.

P2L27: "four main areas": I count 6 areas from the manuscript, moreover this number is subjective.

We added numbering of these four main areas (from (i) to (iv)) to better highlight them. This classification, and consequently the number of these main areas, is the one that appears in http://marine.copernicus.eu/markets/use-cases.

P3L5-10: seems a repetition and suggest to merge in P2L14-26.

We agree that. The text has been merged.

P3L27: "three twin ..." the number three appear evident only later in the paper, suggest dropping it.

We would like to maintain this paragraph in the introduction because we think that it is important to precise in the introduction that these three versions of system have been used to quantify the impact of the updates. We added some details about these simulations to clarify the paragraph.

P5L4-6: The point here is not that parameterizations in the version 3.1 of NEMO are still in the version 3.6, but how many new parameterizations and improvements of NEMO are you missing using version 3.1? In my opinion it should be discussed this way: although I perfectly understand that upgrading version is not easy for an operational system, and this is a justification for me, there are many years of ocean model developments not exploited here, which should be honestly mentioned.

We agree that. The text has been modified at the beginning of section 2.1: "The system PSY4V3 uses version 3.1 of the NEMO ocean model (Madec et al., 2008). This NEMO version is available since a few years and has been already used in the previous system PSY4V2. This was the available stable version of the code when we started the development of the system PSY4V3 a few years ago. Note that, using this version of the code, we do not access better algorithms and more sophisticated parameterizations present in the current NEMO 3.6 stable version that is now the standard version of the code."

P6L8: maybe is good to say what are the problems coming from the use of z-coordinate you are referring to? Would it be better then to use sigma-coordinates? Or you mean something else?

The following sentence has been added: "z-coordinates, compared to sigma, isopycnal or hybrid coordinates, induce excessive numerical mixing over overflow sills (Winton et al., 1998). Mediterranean overflow, without any relaxation, would settle at an equilibrium depth of 800 m or so otherwise instead of 1100 m observed. Sigma coordinates could indeed improve the representation of overflow processes but are likely to induce other problems elsewhere due to sigma gradient pressure error over steep topography or excessive diapycnal mixing in the interior (Marchesiello et al., 2009)".

The two following references have been added:
Winton, M., R. Hallberg and A. Gnanadesikan, 1998: Simulation of Density-Driven Frictional Downslope Flow in Z-Coordinate Ocean Models. J. Phys. Oceanogr., 28, 2163-2174.
Marchesiello, P., L. Debreu and X. Coulevard, 2009: Spurious diapycnal mixing in terrain-following coordinate models: The problem and a solution. Ocean Modelling, 26 (3-4), 156-169.

P6L22: it would be interesting to know what you found for 0, 50 and 100% of relative wind and which was the criteria to choose 50%.
We followed the results obtained by "Bidlot J.R., 2012: Use of MERCATOR surface currents in the ECMWF forecasting system: a follow-up study, Research Department Memorandum R60.9/JB/1228, Internal report available on request".
In the conclusion of this report, it is written: "An impact study was performed with the ECMWF forecasting system in which surface currents from MERCATOR OCEAN were incorporated into the analysis as well as the forecast system. The data from MERCATOR were processed in such a way that only the slow varying features were retained. By prescribing surface current as part of the ocean surface boundary condition, it was demonstrated that both the surface stress and the surface wind profile above will adjust such that the effect on surface stress is **only about half** of what would have been intuitively obtained by subtracting the ocean current from the surface wind in which no account was taken of surface current."
We followed also what it was said in the slide 13 (Figure A) of this presentation: http://cersat.ifremer.fr/templates/cersat/resources/meetingTalks/1505_Bidlot.pdf

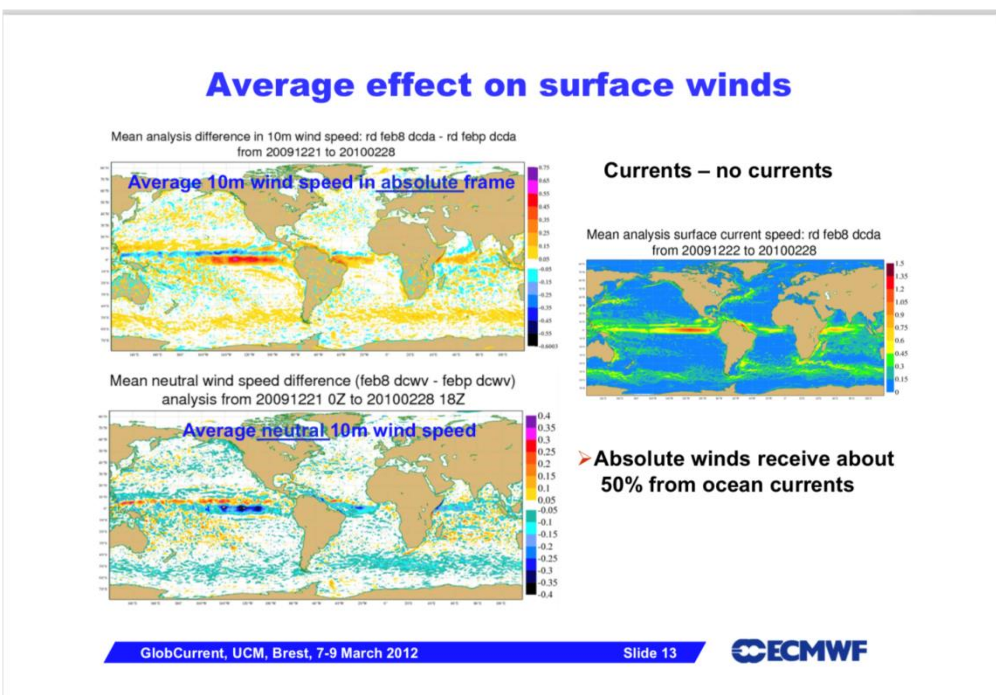

Figure A: Slide 13 of the presentation at Brest of Bidlot in 2012.

P7L1: negative gridded anomalies: maybe better to say negative variations of water masses estimated from GRACE (if I interpret correctly).
We agree that. The text has been modified following your suggestion. Moreover we have clarified the paragraph concerning the building of mean seasonal freshwater fluxes representing Greenland and Antarctica ice sheets and glaciers runoff melting.

P7L11: "...known..." suggest adding a reference.
References have been included.

P7L32: I assume covariances are static (seasonal) but do not vary inter-annually in the real-time system. It is better to state it explicitly.
The background error covariances in SAM rely on a fixed basis. They do not evolve in real-time but they contain the seasonal signal and the inter-annual signal from the 9-year simulation.
The following sentence has been added: "The background error covariances in SAM rely on a fixed basis, seasonally-variable ensemble of anomalies. They also contain the inter-annual signal from the 9-year simulation. This choice implies that, at each analysis step, a sub-set of anomalies is used to improve the dynamic dependency. A significant number of anomalies are kept from one analysis to the other (250 anomalies), thus ensuring error covariance continuity."

P9 paragraphs starting at L15 and L19 seem in contradiction: if the obs errors are adaptive, why do you need a retuning?
Adaptive tuning of errors has been implemented for satellite SLA and SST observations. The method has not been used for temperature and salinity vertical profiles because of the lack of in situ data. Three-dimensional fixed observation errors are then used for the assimilation of in situ temperature and salinity vertical profiles. It is mentioned in the paragraphs starting at L15 and L19 of the original manuscript and at the beginning of section 3.5.

P9L33: this requires a clarification on how you changed the formulation: from the anomaly dataset how do you define the SSH in the old and new system? Wind effect is also barotropic, i.e. is not clear what you actually changed.
You are right. We added some explanations in the text.
In the previous system PSY4V2, the SSH was split in barotropic and baroclinic components, as explained in Benkiran and Greiner, 2008 (page 2060). Moreover, in the system PSY4V2, barotropic height was computed without the wind effect.

The following reference has been added:
Benkiran, M. and Greiner, E.: Impact of the Incremental Analysis Updates on a Real-Time System of the North Atlantic Ocean, J. Atmos. Ocean. Tech., 25, 2055-2073, 2008.

P10L12: suggest adding that the new approach is more consistent with what you actually do (using not a free run but a bias-corrected free run, which better mimics the operational system).
We added a sentence as suggested by the reviewer.

Section 2.3.1 & 2.3.2 It is not clear if these criteria are completely empirical or have some theoretical justification. If empirical as I guess, please discuss the criteria you used to obtain the values for the thresholds.

We agree that. At the beginning of section 2.3, we added some explanations about the criteria we used to obtain the values of the thresholds.

P12L11 it seems weird that there are more suspicious obs in 2012 and 2013. Any idea why?
We agree that. The CORA 4.1 CMEMS in situ database includes the years 2012 and 2013 and we expected a percentage of suspicious profiles relatively stable until 2013. It is almost the case for the temperature profiles but not for salinity. It can not be connected to a strong ENSO event that could explain that more suspicious salinity profiles than usual are detected for instance in the tropical Pacific. We tried to see if it was related to a network effect, but it is not the case. We asked also to the database producers and they have no explanation.

P13L1: how do you define the adjustment, achieved in 3 months?
We consider that an acceptable adjustment is achieved when between 80% and 90% of global energy is reached. To illustrate that, Figure B shows the evolution of the percentage of the three monthly quantities: turbulent kinetic energy (TKE), mean kinetic energy (MKE) and eddy kinetic energy (EKE). For all quantities, 90% of global energy is reached after 6 months. So we changed in the text “3 months” by “6 months”. It does not change the discussion.

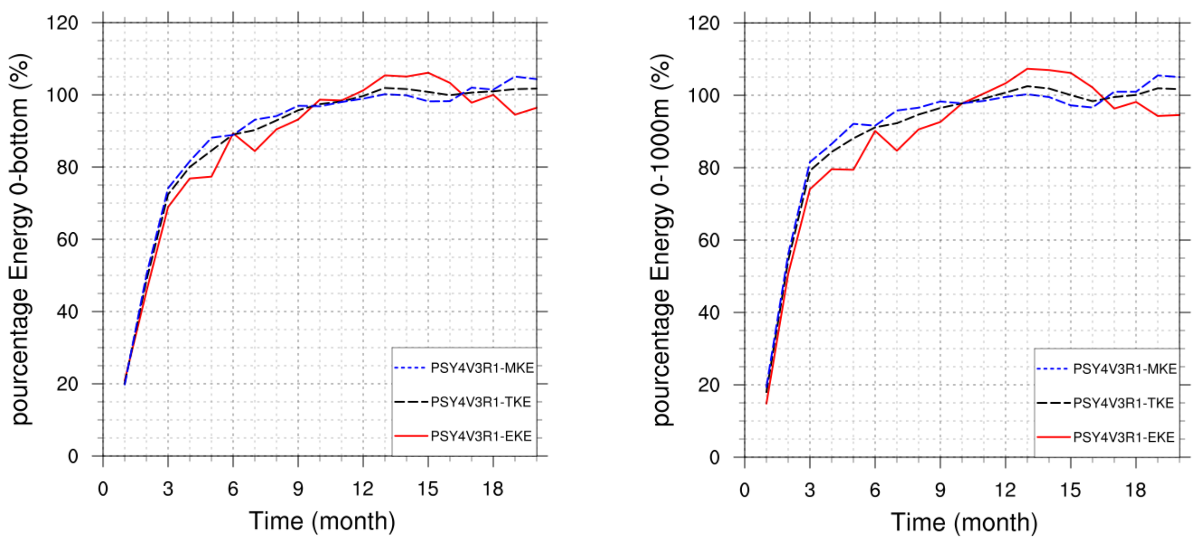

Figure B: Evolution of the percentage of the three monthly quantities: turbulent kinetic energy (TKE), mean kinetic energy (MKE) and eddy kinetic energy EKE. The 100% percentage corresponds to the mean of the months 9 to 20.

Figure 5 and discussion. It looks like the new initialization bears more subsurface biases, so that it is not really convincing that it is better than the old one. I think it deserves a better discussion.
The text has been modified.

Section 3.2: the discussion on the result (Figure 7) will benefit from a quantitative assessment (RMSE and bias reduction of the model vs salinity obs at global scale will be sufficient).
We agree that. The text has been completed.

Section 3.3: A reason for drift might be also inadequate background-error covariances that contain spurious correlations. This should be mentioned at the beginning. Again, the thresholds seem to be empirical and suggest writing the criteria for their adoption.
Referee #1 has also mentioned this point. The text has been modified.
Regarding the value of the thresholds, we followed the tunings used by Greiner et al., 2008 and we checked that these values allow the method to work properly for PSY4V3.

P18 Fig 12: It is weird that without the SEEK you have more variability than the observational product: I wonder whether the two datasets are really comparable, given that the 1/12 model may have a signal at higher resolution than the gridded altimeter product.
We try to make the two datasets comparable by subsampling the 1/12° model (1 point every 3) before doing the comparison with DUACS which is a product on a ¼° regular horizontal grid. This has been clarified in the text.
Also here is an explanation regarding the excess of energy present in the BIAS simulation compared to the observational DUACS product. Figure C shows the mean currents on October-November-December 2013 with superimposed stream lines for the three simulations (FREE, BIAS, OPER) in the zoom (175° W – 125° W / 65° S – 20° S).

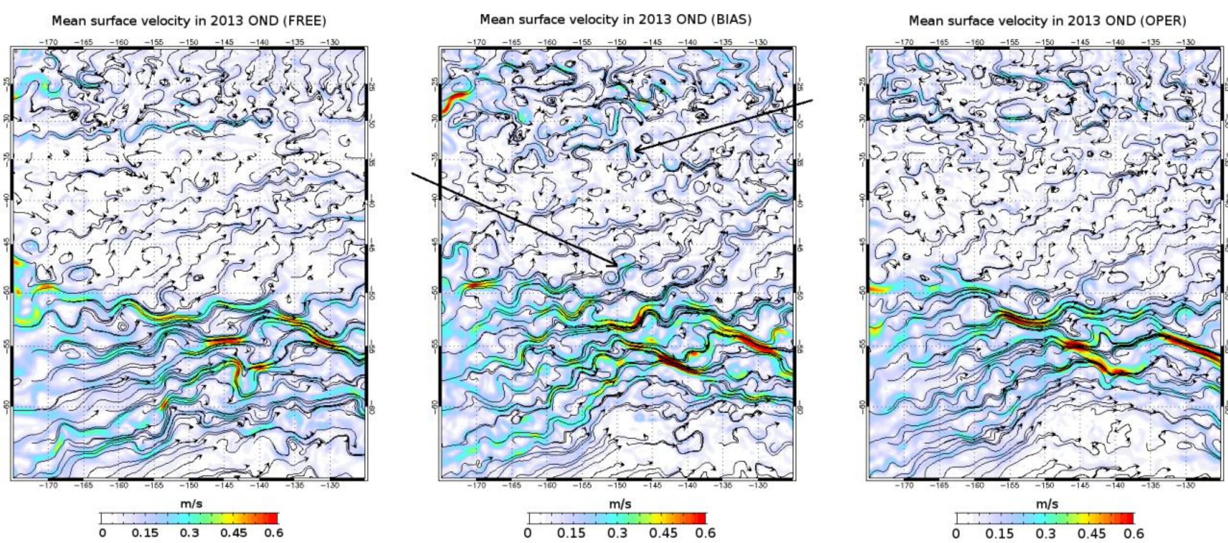

Figure C: Mean currents on October-November-December 2013 with superimposed stream lines for the three simulations (FREE, BIAS, OPER) in the zoom (175° W – 125° W / 65° S – 20° S).

South of 50° S, the Polar Front (PF) is more pronounced in BIAS and even more in OPER. On the other hand, between the East Australian Current (EAC) and the PF, the meridian gradient decreases in OPER. The gradient is not well maintained in BIAS north of the PF. The front leaks to the north in the less energetic zone where we thus find "spurious" meanders and eddies.
The average of the currents shows that the FREE has two very marked veins on the southern edges of the EAC and on the northern edge of the PF. This prevents the export of vorticity (which is advected zonally). In BIAS, the edges are less marked, and there are veins of current towards the less energetic zone. These veins disappear in OPER, especially south of the EAC. It can be noted that the BIAS shows PF connection at 150° W/48° S and EAC connection at 148° W / 35° S (black arrows).
To reduce that in the future, we plan to increase viscosity model coefficient or temperature and salinity in situ observations errors (or both) for the simulation using only temperature and salinity 3D-VAR large-scale biases correction.

Section 3.4.2: Suggest putting it more in the context. I assume that the filtering is applied to the anomaly from the BIAS experiment before covariance computation, and then these differently filtered covariances are used in single-track experiment. However, it is a deduction and recommend to begin the section explaining this.
We have switched the two subsections of section 3.4 to make it clearer. We have also better introduced these two subsections in the introduction of the section 3.4.

Section 3.5: This is the section that I found very hard to justify. I don’t see a reasoning why observation errors are flow-dependent and should change so much with time, except the representativeness error component that might slightly change with season and/or particular events (eg presence of fronts, etc.). But this is less crucial than the background-errors that are certainly modulated by observation availability, large- and small scale processes, forcing, etc. This seems particularly true when looking at an observational dataset with nearly constant sampling (SST, Fig 18/19). I think the results improve not because observational errors really change with time, but because you are changing the ratio between background and observation errors, provided that background errors are of course flow-dependent as mentioned before. Moreover, the Desroziers method implies simultaneous tuning of background and observation errors. If the authors are able to provide a similar complementary retuning of background errors (I mean not with experiments but with diagnostics), it will really improve the robustness of the section. Otherwise a better discussion is needed, probably mentioning that what is actually done is to change the relative weight between background and observation errors, rather than changing observation error themselves.
We agree that. When we say “tuning of observations errors”, we mean the sum of the instrumental and representativeness errors. It’s true that the instrumental error doesn’t change with time. On the contrary, the representativeness error is really flow-dependent.
We tried to apply the “Desroziers method” simultaneous on background and observation errors. But both errors tend to increase or decrease together. This evolution is slow but it is regular and

meaningless regarding the true errors. The ratio between background and observation errors remains constant. Moreover, in the OPER simulation and as mentioned in Lellouche et al. (2013) in the description of the data assimilation system SAM, an adaptive scheme corrects the background variance and gives an optimal background error variance based on a statistical test formulated by Talagrand (1998). This is why we let "Desroziers method" to adjust "instrumental + representativeness" error and "Talagrand method" to adjust the background error.
The text has been modified to make it clearer.

Figure 21: suggest better putting the figure and related discussion in context: the figure shows scores for assimilated vs non-assimilated (NOAA) datasets, so it is not clear the goal of the figure.
The text in section 4.1.1 has been completed.

Section 4.1.2 Title and text: as the SST source is similar between OSTIA and CATSAT (night time measurements from infrared sensors), I don't think the latter is really independent. I would define it "external" or similar.
We changed it.

Section 4.2 L21: 2005-2012 is not a decade; moreover, suggest trying the entire (inter-decadal) climatology to get rid of the weird increase of RMSE after 2012, probably due to the fact that the decadal mean is much too affected by the inter-annual variability therein.
"2005-2012 decade" has been changed to "2005-2012 truncated decade". The five previous decades of WOA13v2 monthly climatology from 1955 and that can be found on the NODC website, properly represent 10-year periods.
It is true that the "2005-2012 truncated decade" contains strong La Niña event (2010-2011) and, as a consequence, is biased to cold. The previous decades (before 2005) are even colder and can no longer be used for recent dates. Moreover, 2005-2012 "truncated decade" doesn't contain the period of transition towards El Niño events and in particular the strong one occurring in 2015. This explains the increase of RMS difference between the WOA13v2 monthly climatology and the in situ observations after 2012. Using the entire (inter-decadal) climatology will globally increase this RMS from 2007 to 2012 but the "weird increase" after 2012 will be still present even if it will be a little reduced.
We clarified the text of section 4.2.

Section 4.3.1 Please clarify how you estimate 2 and 4 cm for instrumental and MDT errors; the MDT one seems in particular arbitrary; also, in the computation of the statistics, do you use any threshold to filter out certain misfits, in order to obtain that global value of RMS, or you use all observations?
The 2 cm is the instrumental prescribed error. It is the error value recommended by data centers.
We prescribe also in the system an a priori MDT error (Figure D), which is equal to 4 cm in average on the regions observed by altimetry.
The text has been modified.

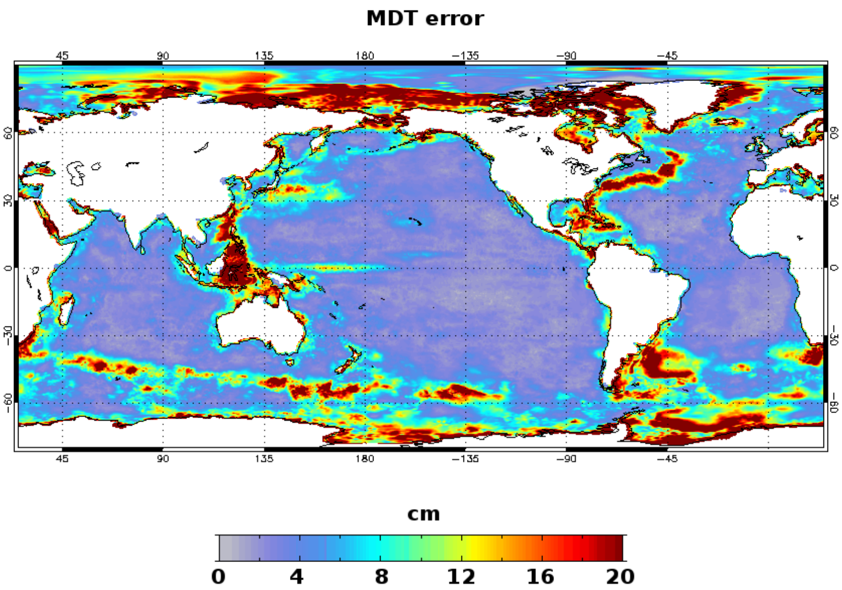

Figure D: MDT error a priori prescribed in the system PSY4V3.

Section 4.3.2. Please provide a reference for BADOMAR and GLOSS/CLIVAR. Also in section 4.4.1 and 4.5 there are products that are referred to only through links: is there any better way to refer to them?

References have been added for BADOMAR product and GLOSS tide gauge stations in section 4.3.2.

The link in section 4.5 has been replaced by a more classical reference. It concerns the Quality Information Document (QuID) for the product in question, which can be accessed via the CMEMS catalogue.

The link in section 4.4.1 has not been replaced because no evident classical reference has been found.