Ocean Sci. Discuss.,
https://doi.org/10.5194/os-2018-170-AC1, 2019

Interactive comment on "Assessment of ocean analysis and forecast from an atmosphere-ocean coupled data assimilation operational system" *by* Catherine Guiavarc'h et al.

Catherine Guiavarc'h et al.

christopher.harris@metoffice.gov.uk

Received and published: 12 April 2019

We are grateful to Reviewer1 for the time and efforts assessing our work, and for providing comments. These comments are reproduced below, with our responses indicated with asterisks.

General comments: This study presents a new ocean-atmosphere coupled operational system (CPLDA) with weakly coupled data assimilation developed at Met Office. In the first part, the model components are described and compared to previous coupled and uncoupled systems (GloSea and FOAM respectively). The weakly coupled data assimilation method from Lea et al. (2013, 2015) is also detailed. This system is operated

and evaluated during one year (2015) with 6-hour analysis and daily 7-day forecast. The system performances are then compared with the operational ocean-only FOAM (Ryan et al. 2015) and Mercator PSY4 systems and with different observational (SST, MLD, 15 m velocities, SLA) datasets. Atmospheric forcings and turbulent fluxes are also analysed. In summary, the CPLDA system performs as well as ocean-only systems despite its increased complexity and constitute a promising first step toward a fully coupled operational system. The manuscript provides a precise description of the CPLDA system, underlining the benefits and limitations compared to other MetOffice systems and observations. It is globally well written and scientific analyses are seriously presented. However, I recommend the following modifications to improve the manuscript quality and understanding. I have the feeling that some sensitivity tests and analysis are still missing to better cover the system description and validation. This is mainly due to the fact that besides the ocean model similar configurations between CPLDA and FOAM, there a large number of differences between the two systems. Consequently, it is sometimes difficult for the authors to assess precisely the differences between the systems because too much parameters change at the same time. Some additional sensitivity tests allow to better understand those differences (for example, the one month experiment is a different scheduling for the DA). But others sensitivity tests are missing to clearly understand and to disentangle the numerous modifications between the systems. As a consequence, some analysis are not very convincing because of this lack of sensitivity experiments. Depending on the authors will and capacity to run these additional simulations, some corrections will have to be made to the text to confirm or infirm the hypothesis proposed. If no supplementary experiments can be done, some comments must be added to the text to emphasize the limitations of the study. Because atmospheric forcings and surface fluxes are strongly related to SST, upper temperature and MLD, the section 3.4 should be moved before the section 3.3 about velocities (that I would rename "upper or ML velocities"). It is not always clear if the authors are talking about the analysis, the forecast or both in the whole manuscript. It makes the study understanding less clear to follow and more difficult to understand. I think the manuscript can be easily improved by clearly stating this.

**Experiments are currently being undertaken using an ocean-only model to look at the impact of reducing the assimilation window from 24 to 6 hours. These experiments should decompose the effect of the shortened assimilation window from the effect of the coupling and aid in the interpretation of the comparisons to the FOAM system in the manuscript. It is important to note that the scheduling of the operational CPLDA system and the availability of observations will not be investigated in these experiments. It is hoped that the results from these experiments will be available before the final re-submission of the manuscript and the text edited accordingly, but expect this would only be possible if an additional month was provided to allow time for this to be done.

We agree with reviewer1 about the organisation of the paper and we will move the section on surface fluxes before the velocities.**

Specific comments: p.1 - Introduction: The introduction is not really relevant because it is mostly a repeat of the system description in section 2. The introduction should be improved to better describe the context of this study, i.e to describe the main physical and technical arguments in favour of developing a coupled system with weakly coupled assimilation (instead of independent systems and assimilations). A description of what is done in other operational centres (ECMWF, NCEP, . . .) could also help to better understand the framework and the interest of the system presented here.

**To provide more context the Met Office efforts, we will update the introduction and in particular we propose to add:

" Coupled systems are used in wide range of applications (short and medium range forecasts, seasonal forecasts, climate prediction and future scenario projections) and improving the initialization of these systems can play a significant part to reduce the development of errors. Using separate atmosphere and ocean analysis to initialize a coupled system can result in an imbalanced system. The imbalance can cause

an initialisation shock that could potentially increase the development of errors during the forecast. Using a coupled data assimilation approach has been shown to reduce this initialisation shock (Mulholland, 2015). In recent years, numerous research centres have developed operational coupled data assimilation systems. For instance, the European Centre for Medium-Range Weather Forecast (ECMWF) has developed a weakly coupled ocean-atmosphere data assimilation system (Browne, 2019). Their results show that using their coupled data assimilation system reduces forecast errors compared with forecast initialised from uncoupled analysis. The National Centers for Environmental Prediction (NCEP) is using a coupled data assimilation system for seasonal and sub seasonal scale predictions (Saha et al, 2014). Penny et al (2017) proposes an overview of the efforts made on coupled data assimilation in operational centres. It shows the diversity of the approaches available. At JAMSTEC, they developed a low resolution strongly coupled system used for experimental seasonal and decadal prediction while NRL coupled model is initialised by separate analysis but uses a high-resolution ocean component (1/25)." **

p.3 l.10-12: it is not clear if VarBC is activated or not in the present study, please specify this in the text.

**We propose to add: "In the present paper, all the CPLDA experiments use VarBC." **

p.5 l.30: please justify why such changes regarding IAU and observation operator were made in CPLDA compared to FOAM.

3-hour IAU are used so for the 00Z analyses the full increments will have been added to the background. These analyses can subsequently be used by groups running initialised coupled forecasts.

p.6 l.20: scheduling of the operational system ? **it will be corrected**

p.6 l.25: I don't understand why some results should come from the operational version of the system and others from "test" version ? I find this confusing. Perhaps it should be better to talk about the actual operational version only in the conclusion to avoid any confusion.

We propose to include a paragraph on the SLA to the paper. The results from the test experiment 2015 highlighted the degradation of the SLA statistics compared to FOAM cause by the scheduling. The results from the operational system with updated scheduling are presented only to highlight how this problem was solved.

p.6 l.31: please add Lellouche et. al 2018 as a reference for Mercator-Ocean PSY4 system.

This reference will be added : Lellouche, Jean-Michel & Greiner, Eric & Le Galloudec, Olivier & Garric, Gilles & Regnier, Charly & Drevillon, Marie & Mounir, Benkiran & Testut, Charles-Emmanuel & Bourdalle-Badie, Romain & Gasparin, Florent & Hernandez, Olga & Levier, Bruno & Drillet, Yann & Remy, Elisabeth & Traon, Pierre-Yves. (2018). Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time 1/12° high-resolution system. Ocean Science. 14. 1093-1126. 10.5194/os-14-1093-2018.

p7 l.9-11: I think these results are important and should be added to the paper as a distinct section or paragraph with a dedicated figure to illustrate the differences in terms of SLA bias and RMSE between FOAM and the 2 versions of CPLDA with the different schedulings. It is also unclear if the results presented here are those from the current operational system or from the version used in the study (“old scheduling”).

**We agree with the comment and will add a figure to illustrate these results and make clear that the version used for this assessment uses the old scheduling.

We propose to add:

”CPLDA 2015 and FOAM sea surface height are assessed against observations using class4 statistics (Ryan et al., 2015). The observations used are provided by CMEMS and include data from Altika, Cryostat and Jason-2 satellites. Altimeter bias correction

is applied to the observations. For each model comparison against the satellite observations, it is important to use the model own altimeter bias. The altimeter bias contains information from the model mesoscale so correcting observations using the altimeter bias from one model to assess a second model penalised this second model.

Figure 1 shows a timeseries the sea level anomaly (SLA) difference statistics assessed against CMEMS observations. In the 2015 experiment with the old scheduling, CPLDA SLA root-squared-mean error (RMSE) is significantly larger than FOAM SLA RMSE (Figure 1a). The larger RMSE in CPLDA can be attributed to the difference in the number of SLA observations assimilated by both systems. Figure 2 shows the number of observations assimilated by both systems in 2015 the number assimilated by CPLDA is significantly smaller than the number assimilated by FOAM. Differences in scheduling can explain this. In comparison with FOAM, the CPLDA 2015 experiment ”best analysis” runs earlier in the day so fewer observations are available. Following these results, the scheduling of the CPLDA operational system was updated in April 2018 in order to improve CPLDA performance. The best analyses at 0600Z, 1200Z and 1800Z are now delayed allowing more observations to be assimilated. This change along with a change in the Met Office database (MetDB) that now allows a more frequent ingestion of SLA observations, has resulted in a significant reduction of the CPLDA RMSE (see figure 1b).

Figure 1: SLA class4 statistics with respect to CMEMS satellite product (Jason2, Cryosat and Altika) for CPLDA and FOAM. a. For the long CPLDA experiment in 2015 with the old scheduling. b. from the operational systems for 2018, the CPLDA scheduling was updated to allow more observations to be assimilated in April 2018.

Figure 2: Number of SLA observations assimilated in FOAM and CPLDA in the 2015 experiment” **

p.14 section 3.4 should be moved close to the sections 3.1 and 3.2 as it is directly related to changes in SST and MLD. It would improve the readability of the manuscript.

We agree with this comment and will change the order of the sections accordingly.

p.7 l.28-30 This hypothesis should be directly tested by doing some additional sensitivity experiments using a 24h window with CPLDA or a 6h window with FOAM to confirm it or not.

Experiments are currently being undertaken using an ocean-only model to look at the impact of reducing the assimilation window from 24 to 6 hours. The results from that experiment will be added to the manuscript.

p.7 l.30 “in CPLDA” -> “in CPLDA analysis”: it is not clear when the author is talking about the analysis or the forecast. Please specify it everywhere it is necessary in the text.

We will update the text accordingly

p.7 l.31: is there a way to measure this “overfitting” ? It is not clear why FOAM SST forecast performs better than CPLDA and how is it related to this overfitting.

**Overfitting can be measured by withholding a sub-set of observations from the DA, these are then used for validation. This approach is not pragmatic in an operational setting where you want to get the best analysis possible at runtime. Assessing the overfitting will be undertaken in the assimilation window experiments detailed above. These experiments will include investigating scaling the background errors for the 6-hour window (current values calculated from system using 24 hour window). It is thought that in the current CPLDA system background errors are too large (less weight in the DA) thus observations are given too much weight in the assimilation and introducing observation noise.

Similar performance is seen in CPLDA and FOAM forecasts at lead times of 36 hours and beyond.**

p.8 fig.1: the forecast duration is 7 days (168h), please expand the figure axis accordingly

*Addition of an extra day to this figure would require a significant amount of extra data processing which we do not feel would add much

p.8 l.19: it could be interesting to give the effective SST resolution for both system even without showing the figure to get an idea of the scale range resolved.

The effective SST resolution is the same for both system around 50km.

p.9 section 3.2: please add PSY4 in the analysis and Figure 3 if possible to be more coherent with others sections and to get a more complete comparison between products.

**Although this would be a useful addition to the manuscript it would require a substantial amount of data processing. As the scope of the paper is the CPLDA system the comparison to FOAM is valuable as the systems are relatively similar which isn’t the case for PSY4. **

p.9 l.9: please give the reference associated with this product.

**We will add the reference to ARGO product: INSITU_GLO_NRT_OBSERVATIONS_013_030 **

p.9 l.11: please give the range of the RMSE increase to be able to compare it to the SST RMSE. T rmse increased from 0.67 at day1 to 0.73 at day5

p.10 l.15: the warm bias and associated significant RMSE (_1_C) located at _100m is not discussed nor explained in the text. Please add a paragraph about this. If the discussion about King et al. 2018 results explain it, it should appear more explicitly in the text.

**This section has been re-written (see below) and the reviewer comments on RMSE addressed.

"Temperature and Mixed layer depth The temperature of CPLDA is assessed against Argo profile observations provided by CMEMS. The class 4 global temperature statistics for the best analysis are presented in Fig 3a. The results for the forecast (not shown) are similar to the best analysis with the mean bias staying unchanged and the RMSE slightly increasing. CPLDA has a cold bias in the subsurface which is maximum (around 0.1K) at 10 m and present in the thermocline down to approximately 50 m. A smaller warm bias is present at around 100 m, but this is over a greater range of depths so represents a large amount of heat. A subsurface cold bias is present in the ocean-only FOAM system but with a smaller amplitude. For the CPLDA results shown, the number of profile observations being assimilated was smaller than in FOAM due to differences in scheduling of assimilation cycles. Even in tests with the operational scheduling (referred to in section 2.6) the sub-surface bias in CPLDA persists despite the increased number of observations being assimilated. The sub-surface bias can be attributed to the vertical propagation of the surface temperature increments through the water column. King et al (2018) shows that the succession of positive and negative temperature increments has an asymmetric effect on the vertical temperature structure due to the way the temperature increment at the surface is propagated to the bottom of the mixed layer. A negative surface increment weakens the stratification and so deepens the mixed layer, this means that a subsequent positive surface increment is projected deeper. Due to the shorter assimilation cycle in CPLDA (6 hours), relative to FOAM (24 hours), the increments exhibit more temporal noise which leads to the larger subsurface bias observed. In both CPLDA and FOAM the temperature RMSE is largest in the thermocline which highlights its variability, the RMSE is larger in CPLDA due to the over-deepening of the mixed layer leading to a misplacement of the thermocline. CPLDA has slightly larger RMSE than FOAM in the sub-surface due to the propagation of surface increments through the mixed layer described above. Some options for reducing the asymmetric effect of the SST increments in the sub-surface were described in King et al (2018). These may be tested and implemented in future versions of the CPLDA system. Mixed Layer Depth statistics confirm that CPLDA has a deeper MLD than the assimilated profile observations. The mean error against those assimilated observations is -5.2 m while the RMSE is 34.7 m. As expected, the CPLDA

MLD is deeper than FOAM (Fig 4). As for CPLDA, the MLD in FOAM is deeper than in the observations but both the mean error and the RMSE are reduced (-2.1 m for the mean error, 32.6 m for the RMSE). The deeper MLD in CPLDA could be caused by the asymmetric effect of the sub-surface temperature increment on vertical temperature structure but also by differences in wind stress between the two system. Further experiments running FOAM with a 6 hour assimilation time window (consistent with CPLDA) are needed to help to separate the impact of the assimilation time window from the impact of the wind stress. In CPLDA the large negative increment applied at the surface (Fig 3b) is propagated down to approximately 50 m, below this a small warm increment is applied down to approximately 150m. The fact that the negative increment is projected deeper in CPLDA than in FOAM and the dipolar structure in the vertical is consistent with idealised experiments into vertical propagation of temperature increments, (King, pers. comm. 2018). CPLDA has a larger negative increment at the surface than FOAM, this is likely due to the under-estimation of the wind stress in CPLDA, described in section 3.4, which causes a warm bias that the assimilation is trying to correct. Below approximately 200-400 m the magnitude of the average temperature increment is small (Fig 3b) and the increments applied by CPLDA and FOAM are similar." **

p.10 l.17: are you talking about analysis or forecast ? Please detail both aspects regarding MLD.

We will update the text to make it clear that the MLD results presented are from the analysis. During the forecast the rmse and bias persist (not shown).

p.11 l.5: I totally agree with the author: an additional experiment using FOAM with a 6h window or CPLDA with a 12h window is needed to disentangle this effect from other possible factors such as atmospheric forcings, turbulent fluxes schemes, . . . Hence, I strongly suggest to the author to do this simulation if the modifications of the systems are not too heavy and depending on available computing resources. I would greatly improve the manuscript discussion and strengthen the results presented here.

An experiment to test the impact of the time window is ongoing see 1st comment.

p.11 section 3.3: please add a word if the product is also assimilated or other products related to oceanic currents. Please make a better distinction between analysis and forecast biases and RMSE.

The currents are not assimilated. We will update the text to make clear what is analysis what is forecast

p.11 l.20: the sentence is in contradiction with what is stated above at p.11 l.15-16 ("bias and RMSE stable during forecast"). Please correct this or give more explanations.

**The bias and RMSE are stable during the forecast on a global scale but in the tropical pacific where there are large erroneous currents in the analysis the RMSE decreases during the forecast. **

p. 12 l.1-2: is it possible to confirm this statement by comparing directly the number of assimilated observations between both systems please ? Would it be possible to conduct the same kind of tests to address other questions or comments in the manuscript related to the difference in term of assimilation time window or observation number ?

In the new paragraph on SLA we will add a plot highlighting the difference in the numbers of observation between FOAM and CPLDA 2015 experiments. We will address the question of the assimilation time window by running an experiment with FOAM and a 6-hour window. We don't have the time and resources to run more impact on the number of observations.

p.12 l.6-7: again, an additional test with a different time window would strongly clarify these statements and give a real matter for discussion.

Yes as above

p.12 l.10-11: again, an additional 1-month experiment using the new MDT would clarify

this.

**The MDT will not be updated in the current iteration of the CPLDA system so these experiments will not be undertaken. The next generation O-A coupled system which is currently being developed uses an updated MDT from CNES-CLS09 to CNES-CLS13 and initial results look promising but these results are outside the scope of this paper. **

p.12 fig. 5:Figure 5 is not described nor employed in the text. Please supress it or comment it.

We will remove Figure 5

p.13 l.2-3: on the contrary, it has been shown that coupled models have more EKE damping than ocean-only forced models (see Renault et al. 2016 for example). Consequently, this explanation is incomplete or erroneous.

**We agree with the reviewer; the text is not clear. We propose to update it: "Eddy Kinetic Energy (EKE) was also compared in CPLDA, FOAM and PSY4. In an ocean only model, using the ocean velocity to calculate the winds tress has a damping effect on the eddies (Duhaut and Straub, 2006; Dawe and Thompson, 2006; Renault et al., 2016). Coupled system should not suffer from the same eddy damping. Despite a reduced damping effect in coupled system, we did not observe a higher EKE in CPLDA than in FOAM. At a 1/4 resolution, the mesoscale is poorly represented and most of the EKE is injected into the model by the SLA assimilation. Hence any differences in EKE between CPLDA and FOAM are mainly caused by differences in SLA observations assimilated, or impacts of the shorter assimilation window, rather than due to a reduced eddy damping by the wind-stress." **

p.14-15 – section 3.4: There is no numbers in this section text to quantify the differences between atmospheric forcings heat fluxes and stresses. It will make the comparison easier and more "physical" if you add them.

**We calculated the average value for the flux differences plotted in Figure 8. On average, CPLDA loose more heat than FOAM which contributes to the colder SST observed in CPLDA. If the differences in short wave radiation show some strong patterns with the largest difference in the eastern boundary upwelling on average its contribution is limited. On average for the global ocean, the strongest contributions come from the latent heat differences.

Average for total heat flux diff (CPLDA-FOAM) = -3.9 W/m2 Average for latent heat diff (CPLDA – CPLDA with CORE)=+9.61 W/m2 (1) Average for latent heat diff (CPLDA – NWP)=-14.50 W/m2 (2) Average for short wave diff (CPLDA-FOAM) =+1.05 W/m2 (3) (1)+(2)+(3) = -3.84 W/m2 **

p.15 l.1-3: the shortwave (SW) bias pattern suggest an SW overestimation on all eastern boundary upwelling systems. This is usually related to an underestimation of the low-level stratiform clouds which atmospheric models have difficulties to represent. A comment about this atmospheric bias should be added if it is also present in NWP.

The stratocumulus clouds are under estimated in Met office atmosphere model and causes a short-wave overestimation on the eastern boundary upwelling systems. The bias gradually reduces as the resolution increases as the resolution is lower in the atmosphere component of CPLDA (40km) than in the NWP fluxes forcing FOAM (17km) the overestimation is larger in CPLDA.

p.15 l.21: please give a quantification of the contributions of atmospheric state and latent heat to the total net heat flux difference.

Globally, the differences in latent heat due to the different bulk formulae (+9.61 W/m2) and the different atmosphere (-14.50 W/m2) are the main contributions to the difference in heat fluxes.

p. 15 – wind stress analysis: I have some difficulties to correctly understand this part. First, there is no distinction between atmospheric analysis and forecasts. I suppose

analysis are relatively close to MetOp observations as they are probably assimilated, while 7-days forecasts must have quickly increasing surface stress errors. Consequently, it is difficult to guess where the stress differences between CPLDA and MetOp come from. Then, CPLDA wind stress is strongly and globally underestimated (_-0.1 N/m2) compared to MetOp. This bias is far larger than the difference between CPLDA and FOAM (_+0.5 N/m2). I don't think this is related to bulk formulae differences (COARE 3.0 is nearly identical to COARE 3.5 except in very strong wind conditions which statistically almost never occur). It can be partially related to the fact that scaterometers usually provide surface stress in atmospheric neutral stability condition, but it cannot explain the global underestimation observed here. This can also be related to the absolute or relative wind/current coupling in CPLDA. But finally, I don't understand why this global stress underestimation is not associated with warm SST and shallow MLD biases in both FOAM and CPLDA as it should be the case. Please explain clearly why it is not the case.

**All wind stress presented are from the analysis. We are currently investigating why the difference are larger than expected and we will add the results from that investigation to the manuscript.

The negative increment suggests the model is too warm, the MLD is constrained by the assimilation of not only profile observations but also the SST observations as described in section 3.2. Without a free-run it's hard to know what the MLD would look like due to wind-stress differences only without the effect of the surface increments on the MLD. **

p.16 – conclusion: Conclusion should be updated accordingly to the modifications done in the manuscript.

We will update the conclusion

Interactive comment on Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-170, 2019.

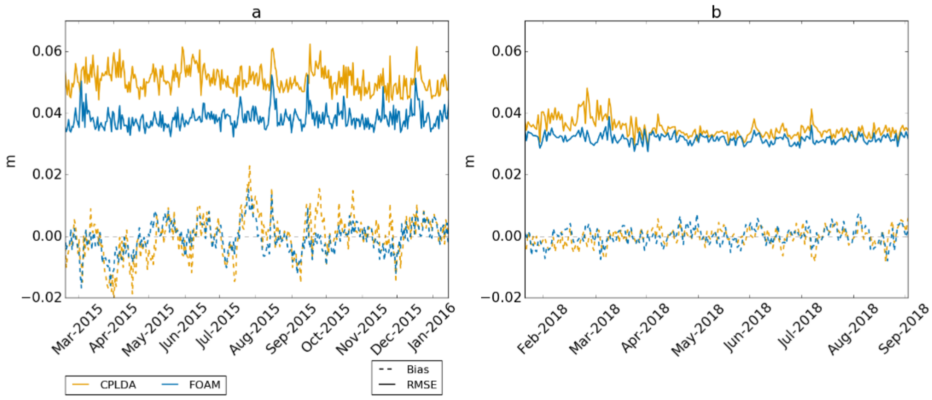

Fig. 1. SLA class4 statistics with respect to CMEMS satellite product (Jason2, Cryosat and Altika) for CPLDA and FOAM. a. For the long CPLDA experiment in 2015 with the old scheduling. b. from the operational

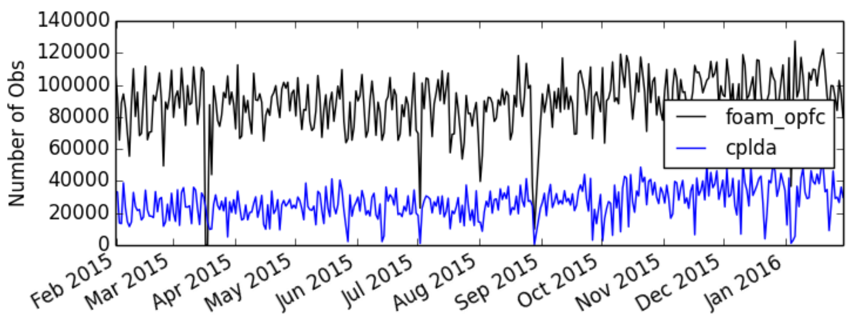

Fig. 2. Number of SLA observations assimilated in FOAM and CPLDA in the 2015 experiment