Our responses are marked in red; the line numbers correspond to the track change version of update manuscript.

Review 1 comments

In this work, some data assimilation methods of the PDAF framework are tested, together with the ESMF framework (for model coupling). Two cases are presented, a theoretical one (lock-exchange) and an application for a real typhoon near Taiwan. The paper is easy to read, the results are quite good and the use of the two frameworks together can be interesting and useful for other works. However, there are some issues in the methodology that in my opinion need to be resolved before the final publication. Comments below.

General comments:

- The paper aims to demonstrate the use of the PDAF assimilation framework together with the ESMF framework. However, the system is tested with only an oceanographic model, stating that the extension to two models is "trivial". I don't think that's the case. Anyway, a test case, using two coupled models, should be presented. If this takes too much work, you should at least better explain the steps for using a coupled system
  Thank you for your suggestions and constructive comments. In light of the comments from the other reviewer, we have now added a new test that demonstrates the new DA capability on a simple coupled ocean-atmospheric model (Section 3.1.6), and described the extra codes to achieve this. As you can see in this section and Fig. 6, the mechanics of extending the DA system to cover the coupled models are indeed trivial under ESMF (which was the original design goal for ESMF). Doing joint DA for realistic coupled systems would follow the same procedure. However, validating realistic coupled models would take some effort and is out of the scope of this paper.

- In both test cases, too few ensemble members are used. Only 8 in the first case and 16 in the second, numbers definitely too low to avoid problems of spurious correlations, even with localization. I recommend consulting the literature about it. However, in my experience, you should never go below 40 members. The fact that the assimilation of SLA does not have a great impact perhaps depends on this and on the fact that the cross-correlations are not very correct;
  We have now added results with 40 members for the two cases (See Figs. 5,13

etc) and you can see that the errors are indeed reduced. The possibility to use small ensembles in PDAF came from the use of Pham's 2nd-order exact sampling to initialize the ensemble because this ensures that the major covariance structure is represented already in a small ensemble. This is important for operational forecast as it requires much less resources.

- It is necessary to better describe the perturbation of the initial state and of the boundary conditions (lateral and/or superficial). In the second test case, it is not clear whether the boundary conditions (for example the wind) have been perturbed. A correct perturbation method greatly affects the results.
  For the initial condition, we first extract from HYCOM as the central state, then use Pham's method to give each member a different initial state. We didn't perturb lateral boundary conditions or atmospheric forcing because the influence of those perturbations is relatively minor in the interior of the domain for *short duration* validation cases shown in the paper. For operational applications however, we do plan to use ensemble atmospheric forcings by using the same Pham's method. We have mentioned this on pg. 19 (Line 311).

Specific comments:

p2r56: See general comment. If the use of a coupling tool (ESMF) is one of the purposes, you cannot say that this is trivial;

We have revised this.

p3r73-76: not very clear, rephrase;

We have slightly rephrased this part.

p3r80: Provide references for the two methods developed;

We've added ETKF/ESTKF references in the text.

p4r96: Do you need to change the SCHISM code? This affects the portability of the method;

For 'fully parallel mode', no changes are necessary in SCHISM code. The changes are only necessary to implement the 'flexible mode' of PDAF (we remark that the latter is a major novelty of PDAF); in particular, we added "other_hot_init" to allow rewinding of clock in schism_init. Other than that, both SCHISM & PDAF are independent and simply

linked as libraries. In short, portability of the method is easy under the ‘fully parallel mode’ but some code changes are needed under the ‘flexible mode’. We have slightly revised the sentences there.

Fig. 2: the panel on the bottom-right is not explained in the caption;

We have added a description in the caption.

Tab.1: the "status" is not related to the development in PDAF I suppose. Clarify;

Yes, we have modified this as “Implementation status” (in our tool).

p8r164: 8 members are few, I recommend at least 40;

Added in this and some other figures.

p9r166: 'forecast' -> 'background';

Done.

Fig. 3: The labels are too small and overall the figure is unclear. I recommend removing it;

We have overhauled this figure.

p11r183-197: I do not much agree with the interpretation of the results. Even if the MAE is lower with localization, the simulations without localization seem better, since they are able to reproduce the vertical structure. Localization methods are perhaps too limited vertically. However, increasing the number of ensemble members should improve all the DA methods;

We’ve revised the discussion to avoid the blanket statement. Localization can indeed have unintended effects because it can decouple different parts of the domain. Increasing ensemble size also helps.

p13r232: I suggest at least 40 members and to take care of the perturbations of the initial state and of the boundary conditions (surface in particular);

We have added 40-member results.

p16r258: By increasing the number of members the cross-correlation should improve if the perturbations are done correctly;

Increasing the number of members does help (Fig. 13).

Fig. 8: the differences are minimal, I would remove this figure, adding a short sentence describing it;

We have replaced this plot with the difference plots to more clearly show the differences.

Fig. 10 and 12: Also these figures can be removed. You can just give a couple of numbers in the text.

We have added results from the 40-member ensemble in these figures (new Fig. 13 &15). We believe they are of interest to readers.