Response to Referee 2

Sergey Skachko

October 2019

We wish to thank the referee for the constructive review. Here below are our answers to the referee's questions and comments, in bold blue.

This paper describes the weakly coupled approach to data assimilation taken by ECCC for the initialisation of coupled NWP forecasts. The methodology is described in detail, and in particular compared to the uncoupled approach which is used as a control for assessment. Impact is measured by looking at observation statistics (analysis and forecast) and forecast error statistics. The paper is overall of high quality, well written, and of scientific interest. There are no insurmountable problems with this manuscript, but I have one request which may be regarded as major, though I hope this could be addressed quickly.

Major point: The choice of forecast error statistic is one that I have not seen before and I have had long discussions with colleagues about its applicability. Specifically it is the choice of verifying forecasts against the mean analysis of the two experiments. Starting with an example, if forecast 1 matches exactly analysis 1, then it will verify worse than if the forecast 2 drifts towards analysis 1, as it will approach the mean of analysis 1 and 2. Looking in more detail, say at Simmons and Hollingsworth, 2002 (https://doi.org/10.1256/003590002321042135), differencing the equation for forecast errors on pg 668, you see that for this to be a true measure of forecast error differences then you are assuming that the true forecast errors times the correlation between the true forecast error and the mean analysis is equal for both forecasts (f1T cf1a = f2T cf2a). So then you have the additional problem of looking at how forecast 1 correlates with analysis 2, and vice versa. I think this just muddies the waters here! The clean solution to this is to use an independent analysis for verification. Indeed, page 9 line 29 says that you have already produced such plots using ERA5. I would suggest to replace the results you show with those verified against ERA5 to simplify the interpretation of your results.

The verification against ERA5 is shown on Fig. 1. As you can see, the results are similar to those obtained using the mean analysis, except the fact that the comparison against ERA5 doesn't show areas where the difference between two experiments is statistically significant with confidence level above 90%. On the other hand, the ERA5 reanalysis is not performed in a coupled data assimilation framework. Thus, comparing a coupled analysis with an uncoupled analysis with

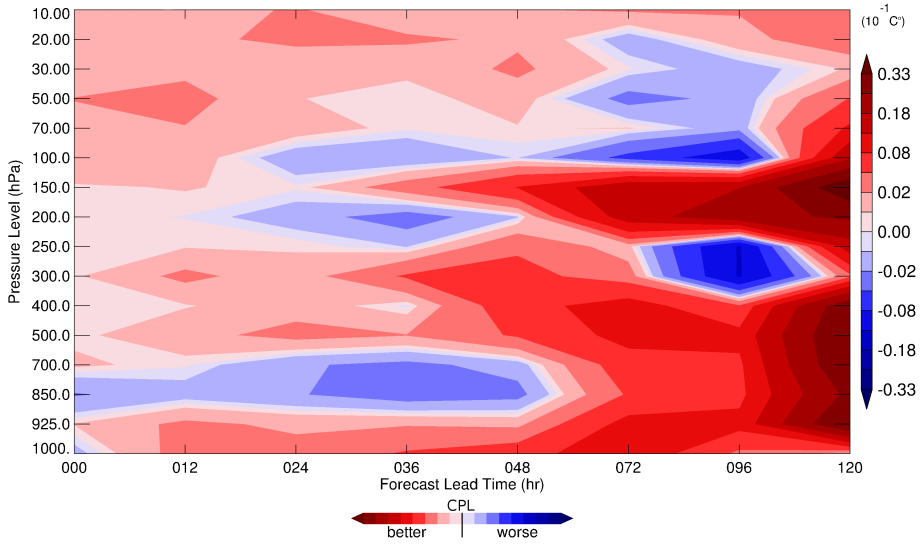

Figure 1: Difference in standard deviation of the air temperature, in degrees C, against the ERA5 reanalysis as a function of forecast lead time. The statistics are computed for CPL and UNCPL in the Northern Extratropics region. Positive values (red) mean that the standard deviation produced by CPL is smaller, whereas negative values (blue) mean the converse.

respect to another uncoupled analysis may lead to preferencing the uncoupled analysis. Besides, the areas where the biggest statistically significant differences in standard deviation computed against the mean analysis are observed near the atmospheric-ocean interface, where the impact of coupling is expected. That is why we would prefer to keep the initial figure in the final manuscript.

Minor points: Page 2, line 4: worth referencing ECMWF here: P. Bauer and D. Richardson. New model cycle 40r1. ECMWF Newsletter No. 138 - Winter 2013/2014, (138):3, 2014. URL https://www.ecmwf.int/node/14581

Done.

Page 4, line 14: are the increments computed on the full 80 levels? please clarify.

The sentence is modified as follows: "The 4D-EnVar analysis increments are computed on a grid with a horizontal grid spacing of 50 km, as in the EnKF system, on all 80 vertical levels."

Page 7. line 26: "The daily ocean SAM2 DA (Sect. 2.2) assimilating only SST data is computed at 0000 UTC." Is this a daily mean SST field, or is it valid at 0000? Please clarify.

The sentence is modified as follows: "The daily ocean SAM2 DA (Sect. 2.2) assimilating only SST daily mean data is computed at 0000 UTC."

Page 8, line 10. Please could you clarify if the ensemble used in the 4D-EnVar uses a coupled or uncoupled model?

To do so, we modified the section 2.1 describing the atmospheric data assimilation: "The ensemble covariances are estimated from the ensemble of 256 uncoupled background states, available hourly within the 6-h assimilation window, obtained from the global ensemble Kalman filter (EnKF) being used operationally at ECCC (Houtekamer et al., 2014) since 2005."

Page 8, line 12:14. "However, by saving the atmospheric fields from the 6-h coupled forecasts and using these to force the ocean model, this is equivalent to the explicit use of the fully coupled atmosphere-ocean-ice model." My understanding of this line is as follows: "However, by saving the atmospheric fields from the 6-h coupled forecasts and using these to force the ocean model, this is equivalent to the explicit use of the fully coupled atmosphere-ocean-ice model with a 6 hour coupling frequency". Is this a correct reading? If so, should it be added for clarity?

The coupling frequency chosen for the computation of the CPL ocean analyses is one hour, as explained in the next sentence: "Preliminary experiments showed that the use of the precomputed atmospheric forcing from the fully coupled model every hour gives results similar to the forcing changing every model time step (the ocean-ice model time step is 15 min in our experiments)."

Page 9, line 2: The test period used here of 2 months is short. Specifically it might be too short to see any major changes in the ocean component. Given the computational cost of the coupled assimilation experiments it would be unreasonable for anyone to ask for an extended period of testing. I think, however, this warrants a comment in the conclusion to reflect that the results should be viewed in this context.

Following the referee's suggestion the following sentences are added to the conclusions: "Besides, the test period of two months used here might be too short to see major changes in the ocean component. Thus, the results presented here should be viewed in this context."

Page 9, line 6:7. "Differences between these two systems are expected for the SST as well as for near-surface layers in both atmosphere and ocean models." This sentence I spent a while trying to understand what may be very obvious to the authors, and in the end I cannot see why SST is expected to be different in the two systems. I thought the SST analysis described in 2.3 was independent of any model, and so should not be different in the analysis of the weakly coupled or uncoupled systems. Perhaps this refers to forecasts of SST? Please can you expand on this to make your point more explicitly.

The sentence is modified as follows: "Differences between these two systems are expected for the forecasts of SST and near-surface layers in both atmosphere and ocean models."

Page 10, line 7:8. "The OmF standard deviations produced by CPL are systematically lower than those produced by UNCPL in all three regions." This may be systematic, but it a is very small difference.

The text is modified as follows: "The OmF standard deviations produced by CPL are systematically lower than those produced by UNCPL in all three regions. Though, the differences are small."

Page 12, line 24: "an integrated software" -¿ "integrated software"

Done.

Figures 6, 7 (top), 9, 10, 11: please state in the caption and in the text that these are plots of errors, not just std etc.

Done.

Figures 13 and 14: The grey colour looks blue which is misleading. Maybe replace the orographic shading with a constant colour and make the grey areas that same colour.

Done.