Reply to Reviewer Comment 2 (RC2)

Rui Sun, Aneesh C. Subramanian, Arthur J. Miller, Matthew R. Mazloff,
Ibrahim Hoteit, Bruce D. Cornuelle

April 16, 2019

The authors would like to thank the reviewer for his/her comments, which have helped improve the quality of the manuscript. To adequately address the concerns raised by Reviewer 2 in the original manuscript, we have made the following changes:

1. We have augmented the text to include more technical details.
2. We have revised the discussion on the scalability in Section 5 according to the reviewer's comments.
3. We have added more discussion regarding the selection of boundary condition, projection, and coupling intervals.

Our detailed replies to specific comments of reviewer 2 are presented below. We have also attached an annotated manuscript to highlight the revisions.

Comment 1: *It is neither a technical nor a science paper. It would be beneficial to re-focus the manuscript on one aspect by clearly stating the problem, hypotheses and discuss the findings. Based on a few snippets of the manuscript it comes across that the authors are vaguely familiar with the foundations of numerical modeling in the atmosphere; they got two open source models, coupled them (no small feat!), and ran a test case. What is missing is a critical look at the approach, results, discussion of why things worked and more importantly, why not. I suggest to omit the whole section on scalability. The experiment design does not support any meaningful conclusions for scaling purposes. I would also recommend proof-reading (not spell checking!) the manuscript.*

Reply 1: The authors thank the reviewer for the general comment and we completely agree with reviewer. We have revised our manuscript to focus on the technical part of our coupled model and removed the scientific discussion in a few paragraphs. We have better motivated the manuscript in the introduction. We have added a paragraph in Section 3 to emphasize that the test case aims to show the ocean and atmosphere models are successfully coupled. We have also re-written Section 5 to emphasize the purpose of the scalability test.

We have proof-read the manuscript carefully.

Comment 2: *Page 6, lines 21-32: Too technical.*

Reply 2: We thank the reviewer for pointing out that the language used in the initial draft is too technical. We have revised the introduction of the ESMF/NUOPC coupler to make it more readable in section 2.4. Please refer to the revised manuscript.

Comment 3: *Page 6, line 32: Why was sequential mode selected?*

Reply 3: The sequential mode is simple when dealing with the data transfer in ESMF, especially when each processor contains the ocean and atmosphere data for the same region (Collins et al., 2005). This makes it a natural starting point and it is chosen in our work. We have added this discussion in Section 2.4.

We have also plotted Fig. 2 to show how the sequential mode is executed in the coupled model.

Comment 4: *Page 7, lines 17-31: Is it a 30-day long run? How frequently are you forcing lateral boundary conditions? What is the lateral boundary condition type? What is the projection? Why is coupling every 20 minutes? Why 8 km grid spacing?*

Reply 4: Yes, this is a 30-day long simulation which allows validation of the coupled model.

The ocean lateral boundary conditions are specified using HYCOM/NCODA global analysis data, and are updated every 24 hours. The atmosphere lateral boundary conditions are specified using ERA5 reanalysis and are updated every 6 hours. They are linearly interpolated between two time steps. We have highlighted the boundary conditions in Section 3 and Table 1. In MITgcm, a sponge layer is applied at the lateral boundaries, with a thickness of 3 grid cells and inner/outer boundary relaxation timescales of 10/0.5 days. In WRF, the lateral boundary values are specified in WRF in the 'specified' zone, and the 'relaxation' zone is used to nudge the solution from the domain toward the boundary condition value. Here we used the default width of one point for the specific zone and four points for the relaxation zone. We have added these details in Section 3.

We used a lat-lon projection in both the ocean and atmosphere models. The grid spacing is 0.08^o and we have replaced 9km by using 'approximately 9km' or '0.08^o' in the manuscript.

The coupling interval is 20 minutes because it was deemed short enough to capture the resolved dynamics. It is 40 atmospheric time steps and 10 ocean time steps. 20 minutes is adequate to resolve the diurnal variation of SST and atmosphere forcing (Seo et al., 2014).

We have revised our manuscript and added the detailed discussions in Section 3.

Comment 5: *Page 8, lines 13, 16: 'accessing', 'accesses' should be 'assessing', 'assesses'*

Page 9, Table 1: The second ATM.STA should probably be ATM.DYN.

Reply 5: The authors thank the reviewer for pointing out this. We have fixed these typos in the manuscript.

Comment 6: *Page 9, line 9: Is MERRA-2 really an independent data compared to ERA5? The forecast model is, but the observations do overlap quite a bit.*

Reply 6: The authors agree that the observation data used in producing MERRA-2 and ERA5 overlap. However the reanalysis of MERRA-2 and ERA5 are performed independently. We choose MERRA-2 because it provides us with the latent heat and sensible heat fields. Hence, we rewrite the sentence as:

The MERRA-2 dataset is selected because it is an independent reanalysis data compared to the initial and boundary conditions used in the simulations. The dataset also provides a $0.625^o \times 0.5^o$ (lon $\times$ lat) resolution reanalysis fields of turbulent heat fluxes.[R2]

Comment 7: *Page 10, line 21-22: Why not use a nest with finer grid spacing to resolve the local* *topography?*

Reply 7: The authors agree with the reviewer that using a finer grid spacing would better resolve the local topography and improve the forecast skill of model in the mountains. However, in our manuscript, we aim to develop a model to capture the ocean-atmosphere coupling in the Rea Sea. Therefore, we did not use a finer grid to resolve the local topography in the mountains. To give our manuscript a more technical focus, we have rewritten the paragraph and removed the discussion of the topography.

Comment 8: *Page 10, lines 10-32: When comparing to ERA5 data, how were the statistics computed? Was the model output interpolated onto the observation points in space and time?*

Reply 8: We interpolated ERA5 to our model as ERA5 data has lower grid resolution (30 km) than our coupled model (approximately 9 km), but omitted this detail in the original submission. We compared the results at the same time so that the results are not interpolated in time. These details have now been included in Section 3.

Comment 9: *Page 11, Figure 3: There was a gray stripe at the bottom, making it impossible to read labels of the color bars.*

Reply 9: Thanks. We have updated the figure and removed the gray stripe of this figure.

Comment 10: *Page 11, line 16: Land surface model and PBL model are not microphysics models.*

Reply 10: The authors thank the reviewer for pointing this out. The land surface model and PBL model are WRF *physics models*. We have fixed our mistakes on page 11 and other places.

Comment 11: *Page 13, Figure 5: Are there missing data points for the observed high and low T2, e.g. Mecca and Yanbu 6/21, Yanbu 6/8, 6/10, 6/14. . .?*

Reply 11: Yes, some T2 points are missing from NOAA NCDC data. We have added this in Fig. 7 of the revised manuscript.

Comment 12: *Page 14, Figure 6: Are model points interpolated to ERA5 points over the Red Sea? Which simulation is ATM.CPL, it has not been introduced in Table 1? How do you explain the drift (blue line)? How can RMSE be negative in the lower right figure?!*

Reply 12: We interpolated ERA5 data (30 km) to our coupled model (approximately 9 km). We have added the discussion on the interpolation method to Section 3.

'ATM.CPL' is a typo. It should be 'ATM.DYN'. Here we are comparing the ATM.DYN results to CPL and ATM.STA runs. We have fixed this in Table 1.

The blue line shows that the error in CPL run is much smaller than that in ATM.STA run (in Fig. 8 of the revised manuscript). This is because a fixed SST is used in ATM.STA run and the ocean response to the atmosphere is not represented.

In the lower right figure, the magnitude is showing the difference in the RSME from all simulations. We have updated the figure and labels to try to make this more clear.

Comment 13: *Page 14, line 4-6: Where there many clouds present during that period?*

Page 16, line 13-14: Any cloud comparison?

Page 14, line 4-6, and line 12-13: First you state the forcing is different due to 'uncertainty in cloud modeling', then you state 'both simulations are driven by realistic atmospheric forcing'. Which one is correct? Please explain.

Reply 13: We thank the author for pointing this out. We focus on validating our coupled model and we assessed the surface variables to demonstrate the ocean-atmosphere coupling. Hence we did not show the cloud data obtained from our model or observational data. We aim to keep a technical focus on the coupling and have removed our discussion on the cloud.

In the OCN.DYN run, the ocean is driven by ERA5 data; in the coupled run, the ocean is driven by the atmospheric fields obtained in the WRF simulation. We revised the sentence as:

Generally, the OCN.DYN and CPL runs have a similar range of error compared to both validation datasets, which shows the skill of the coupled model in simulating the ocean SST.[R2]

Comment 14: *Page 16, Figure 8: Why no time series comparison to MERRA-2 dataset?*

Reply 14: We have added two more figures and Fig. 8 in the initial draft is Fig. 10 now. In Fig. 10(a), we compared our results with HYCOM data. Since our coupled simulations are initialized using HYCOM, this aims to show the increase of the simulation error. In Fig. 10(b), we compared our data to the GHRSST satellite observation data to further validate the simulation results. The MERRA-2 reanalysis data is not used to validate the SST because the GHRSST observational product can be used.

Actually, MERRA-2 is used to validate some simulation results (e.g., latent heat, sensible heat) when the high-resolution observational products are not available. We have added the discussion on validation data to Section 3.

Comment 15: *Page 19, line 6: Which selected micro-physics schemes?*

Reply 15: We have replaced the original sentence using 'selected WRF physics options presented in Section 3'.

Comment 16: *Page 20, line 3: 64 /cm/year should read 64 cm/year?*

Reply 16: Thanks. We have fixed this typo.

Comment 17: *Page 21: line 8-9: What does it mean 'The decrease in parallel efficiency is because when using 256 processors, there are only 16x16 grid points in the horizontal plane'?*

Page 21: line 11-13: Please elaborate: 'This may be attributed to the fluctuation of the CPU time when solving the systems of linear equations. When using different number of processors, the decomposition of the domain leads to different linear equation systems, requiring different CPU load and accordingly different convergence time.

Reply 17: Thanks. We have used different number of processors to investigate the parallel efficiency of the coupled code. When using up to 128 CPUs, the parallel efficiency of the coupled code is close to linear. However, when using 256 processors, the parallel efficiency decreases to 70%.

We have re-written this paragraph:

It can be seen that the parallel efficiency is close to 100% when employing less than 128 processors and is still as high as 70% when using 256 processors. When using 256 processors, there are 20480 cells (16×16×80, 16 lat×16 lon×80 vertical levels) in each processor, but there are 5120 overlap cells (4×16×80, 4 sides×16 tiles per side×80 vertical levels), which is 25% of the total cells. From results reported in previous literature, the parallel efficiency of the coupled model is comparable to other ocean-alone or atmosphere-alone models when having similar number of grid points per tile (Marshall et al., 1997; Zhang et al., 2013). The decrease in parallel efficiency results from the increase of communication time, load imbalance, and I/O (read and write) operation per processor (Christidis, 2015). It is noted in Fig. 16 that the parallel efficiency fluctuates when using 8 to 32 processors. This may be because of the fluctuation of the communication time, load imbalance, and I/O operations. The fluctuation of the CPU time can also be seen in the speed-up curve, but at smaller magnitude.[R2]

Comment 18: *Page 21-23: Did you try weak or strong scaling? What is the communication cost? I/O cost? How many grid points per core are recommended? Are you reporting average times of multiple simulations in Table 3? How does WRF scale, MITgcm scale - do your results fit? Why did the coupling cost increase when using more cores?*

Reply 18: We tried strong scaling in our test. When presenting the scaling test our aim was to demonstrate that our implementation of the coupler does not slow down the individual simulations for varying core count. We have revised our discussion of the parallel performance in Section 5.

Comment 19: *Page 23, Table 3: Please use the experiment names consistently throughout the manuscript.*

Reply 19: Thanks. We have revised the experiment names in Table 3. We have also read through the manuscript ensuring that the experiment names are consistent.

References

Christidis, Z., 2015. Performance and scaling of WRF on three different parallel supercomputers. In: International Conference on High Performance Computing. Springer, pp. 514–528.

Collins, N., Theurich, G., Deluca, C., Suarez, M., Trayanov, A., Balaji, V., Li, P., Yang, W., Hill, C., Da Silva, A., 2005. Design and implementation of components in the Earth System Modeling Framework. The International Journal of High Performance Computing Applications 19 (3), 341–350.

Marshall, J., Adcroft, A., Hill, C., Perelman, L., Heisey, C., 1997. A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. Journal of Geophysical Research: Oceans 102 (C3), 5753–5766.

Seo, H., Subramanian, A. C., Miller, A. J., Cavanaugh, N. R., 2014. Coupled impacts of the diurnal cycle of sea surface temperature on the Madden–Julian oscillation. Journal of Climate 27 (22), 8422–8443.

Zhang, X., Huang, X.-Y., Pan, N., 2013. Development of the upgraded tangent linear and adjoint of the Weather Research and Forecasting (WRF) model. Journal of Atmospheric and Oceanic Technology 30 (6), 1180–1188.