Reply to Reviewer Comment 1 (RC1)

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

April 16, 2019

The authors thank the reviewer for their insightful comments. To adequately address the concerns raised by Reviewer 1 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 introduction section and highlighted the innovation of our work.
3. The discussion of the parallel efficiency has been re-written.

We have also proofread our manuscript according to the reviewer's comments. Some paragraphs are rewritten to address the concerns raised by the reviewer. We have also attached an annotated manuscript to highlight the revisions.

Detailed replies to specific comments by the reviewer are presented below:

Comment 1: *This is a review of "A regional coupled ocean–atmosphere modeling framework (MITgcm–WRF) using ESMF/NUOPC: description and preliminary results for the Red Sea", by Sun et al.*

This is an article describing the development of a geophysical model, a regional coupled ocean-atmosphere model, which fits perfectly with the journal. Having worked closely with atmosphere models for years and dabbled with coupled modeling systems, this coupled system represents an enormous effort. The timing numbers and physical results indicate that the coupled system appears to be behaving correctly and is ready for further testing. My concerns are largely with the presentation of this new tool to the community. The actual development and engineering aspects of the coupled *system, which should be the priority, tend to take a back seat to a lengthy description of the physical results from a series of test cases, and the process of identifying those test cases could have been more discriminating. A better showcase of modeled physical phenomenon exists for a coupled ocean - atmosphere system than choosing a data sparse region with no obvious ocean-atmosphere feedback mechanism to model. Finally, some of the discussion about parallel issues is misleading regarding statement of fact, leading to flawed assumptions concerning implications.*

Reply 1: The authors thank the reviewer for acknowledging our efforts in building the coupled modeling system. We also thank the reviewer for their comments, which have helped us improve the quality of the manuscript.

Comment 2: *There are a number of existing coupled regional ocean-atmosphere systems available to the community. With "framework" in the title, I was assuming that this article was more about a technique or related to some new tools that would assist or improve the infrastructure of model coupling.*

Reply 2: We thank the reviewer for pointing out this. We have edited the paper to focus it as a technical introduction to the model. We intend this paper to be technical documentation explaining the new modeling capability, with a demonstration that has scientific value as well.

Comment 3: *The first sentence of the abstract is that a new regional coupled model is developed. The authors then proceed to present a justification which they do not back up.*

1. This will be a "new coupled regional ocean-atmosphere model with 'state of the art' physics and using modern framework". Coupled regional models within the past ten years or so include: FROALS, SCOAR, CROAM, COAWST, COSMO. Some of these models also include data assimilation, chemistry, waves, sediment transport. Various modern toolkits are used for coupling. There is plenty of state-of-the-art in the existing systems.

Reply 3: We agree with the reviewer that many regional coupled models have been developed using modern model toolkits. We also agree that some of these models include data assimilation, waves, sediment transport, and chemistry packages. The innovations in our work are: (1) we used

ESMF/NUOPC, which is a community supported computationally efficient coupling software for earth system models; (2) we used MITgcm and WRF, and the coupled model is being developed as a coupled forecasting tool for coupled data assimilation and S2S forecasting. By coupling of WRF and MITgcm for the first time with ESMF, we provide a regional coupled model resource to a wider community of users. These atmosphere and ocean model components have an active and well-supported user-base.

The introduction now includes a brief review of the regional coupled models developed and used over the past ten years. We have also added a discussion of our purpose in developing this new regional coupled model. We thank the reviewer for highlighting some of the vague descriptors (state-of-the-art, high res, etc.) and we have replaced these terms throughout the manuscript using more quantitative or informative words.

Comment 4: *2. In the comparisons of 2-m temperatures for several episodic events, the month-long diurnal 2-m temperatures, the month-long plot of deviation and RMSE, snap shots of SST over the Red Sea, and the deviations of the SST vs HYCOM and GHRSST the authors state that the regional coupled model behaved similarly to a standalone model with "dynamic" SST. So, the authors indicate that the existing (and much simpler) stand alone models work as well as the coupled system.*

Reply 4: We compared the coupled model with the stand-alone models with "dynamic" SST. By doing this, we aim to show that we successfully coupled the two models. In the forecast, one would not have evolving SST. We have made this point clear in the text.

In this case the goal of the simulation is to accurately estimate the ocean's effect on the atmosphere, and having the coupled SST forecast match the reanalysis is a positive outcome. We did not describe this well in the text and we have revised accordingly.

Comment 5: *3. As the title says, these are preliminary results. However, there are existing global models running at 9 km, so a study of "small-scale processes" for a regional coupled model (specifically set up with an 8 km resolution for this case study) does not seem to be the best possible demonstration of the available capabilities. The selected verification data sets are fairly coarse resolution and the* *verification techniques are those traditionally employed for large-scale fields: bias and RMSE. The only indication of high resolution is the oft repeated "high-resolution" phrase.*

Reply 5: We thank the reviewer for pointing this out. Our aim is towards high-resolution process studies and experimental forecasting. These models have both been shown effective at simulating kilometer-scale processes in the ocean and atmosphere. We agree that we have overused the phrase high-resolution and have edited the text. We now make it clear that our example is a 0.08 ° grid spaced model. We have emphasized our motivation in the introduction that these models do give the capability to run (down-scaled) detailed process study simulations.

Comment 6: *4. For a demonstration of the benefits of a coupled ocean-atmosphere system, one would expect some sort of traditional ocean-atmosphere feedback mechanism to be on display: tides, storm surge, post-hurricane cold wake, inundation, sea breeze, etc.. A heatwave event in a desert region does not seem to identify and highlight the new model's coupled capabilities.*

Reply 6: We investigate the heat wave events because they are extreme events and have societal relevance. We have added the discussion on why we selected the heat wave events in Section 3.

Comment 7: *As both the atmosphere and the ocean model are widely used in the public, specific details of the changes in those codes would appropriate. For the atmosphere model component, it would be nice to know details of how the ocean model's data is hooked into the WRF model's surface layer scheme. Are the surface layer tendencies constant during the intervals between coupling? Was any sensitivity seen in this coupling frequency?*

Reply 7: We did not modify the WRF model's default surface layer scheme (version 3.9.1.1). We use the coupler to provide the ocean's SST and surface velocity as the bottom boundary condition. The WRF model's surface layer schemes can read the updated boundary condition at each coupling interval. The SST and surface ocean velocity are considered constant during each coupling interval.

We select the coupling frequency to resolve the diurnal cycle according to the reference (Seo et al., 2014). We have added this in our revised manuscript. Actually, we tried a few different coupling intervals (2 min, 20 min, and 180 min), but the diurnal cycle is not sensitive to the coupling frequency in this case.

Comment 8: *Quite a number of examples point to poor atmospheric surface comparisons after the Red Sea SST is kept constant for a month. This is physically unrealistic. While this constant SST test case may serve as a data point, a month-long constant SST experiment should not be the primary comparison to display the skill of the new coupled modeling system.*

Reply 8: We agree with the reviewer that the ATM.STA run is a physically unrealistic scenario. In our manuscript, we use the results obtained in ATM.STA as the baseline case to show the difference between coupled and uncoupled runs. We have emphasized that the constant SST test case is unrealistic in our manuscript and detailed our motivation in running the constant SST test. Please refer to Section 3.

Comment 9: *Buried towards the end of the paper is a mention of the importance of the resultant size of the decomposed domain with strong scaling. That a reasonably well designed atmosphere or ocean model scales to 128 processors is germaine only so far as we know the number of computational cells within that MPI rank.*

Reply 9: We thank the reviewer for pointing out this. We have added this to the discussion on the parallel efficiency and the number of CPUs. Although we only used up to 256 CPUs, the parallel efficiency is satisfactory when we only have about 20,000 grid cells in the coupled model. Our results are also consistent with other parallel efficiency test using similar number of processors (Zhang et al., 2013).

Comment 10: *There are several statements that would be easy to verify, and likely that the authors' stated reason is not among the top contenders.*

1. "This may be attributed to the fluctuation of the CPU time when solving the systems of linear equations. When using different numbers of processors, the decomposition of the domain leads to different linear equation systems requiring different CPU load and accordingly different convergence time." The atmosphere model accounts for 75- 90% of the elapsed time, and the WRF model does not solve linear systems with convergence criteria. This assertion is not defensible.

Reply 10: We thank the reviewer for pointing out this. We have removed this sentence and revised our discussion according to the literature (Christidis, 2015). The oscillation of the CPU time might be because of the increase of communication time, load imbalance, and I/O (read and write) operation per processor.

Comment 11: *2. "This is likely because the simulations on T2 suffer from the mismatches between the model terrain and the actual terrain, especially over complex mountains". Smooth the model data to the resolution of the validating analysis to check this out. The domain is mostly a desert, and atmosphere models tend to underestimate and bias the amplitude of the diurnal surface air temperature. Atmosphere models tend to do poorly with a diurnal amplitude when the observation site is on a coast when the sea breeze effect is not well captured. Any number of quick tests are available to find why the T2 behavior is not as expected.*

Reply 11: We thank the reviewer for pointing out this. The test case in this manuscript aims to demonstrate the ocean and atmosphere components are successfully coupled. We have rewritten the motivation to have greater emphasis on the technical details. The discussion of the bias from the mismatch of the terrain is removed from the revised manuscript.

Yes, the domain is mostly desert and the atmospheric model tends to underestimate the diurnal surface air temperature. Our simulation captures much of the diurnal temperature cycle in Fig. 6, but the bias is obvious. We have revised our manuscript in Section 4.

Comment 12: *3. "However, when using 256 processors, the proportion of this cost increases to 10% because of the increase of inter-processor communication with more processors." The per-MPI task cost of communications is approximately constant, but the relative cost of the communication compared to the computation becomes important as the amount of work is reduced as the number of grid cells is reduced during strong scaling tests.*

Reply 12: We thank the reviewer for the comments. Yes, our results show the per-MPI task cost of communications is approximately constant in Table 3. We also find the relative cost of the communication compared to the computation increases as the amount of computation work per-MPI is reduced. We have added this in Section 5. We have also revised our discussions on the parallel efficiency.

Comment 13: *Referring to 128 processors as a "large number of processors" is inaccurate for either a state of the art atmosphere or ocean model. The atmosphere model domain is 256x256 grid cells. The coupled system using this atmosphere domain is not really "simulating large" problems.*

Reply 13: We thank the reviewer for pointing this out. We agree that 128 processors is not "a large number of processors" in the context of earth system model computations, and that our domain using 256x256 grid cells is not "simulating large problems" compared to global weather and climate models or regional cloud resolving models. We have re-written this discussion in the revised manuscript.

Comment 14: *Stating that scaling to a large number of processors makes a model applicable to "high-resolution" studies does not logically follow. The scaling test was for strong scaling, the problem size remained identical. The problem was the same "high-resolution" with 8 MPI ranks as with 256 MPI ranks. Scaling, as used in this study, implies that the same problem size gets done faster.*

Reply 14: Thanks. We agree with the reviewer on this and we have revised our manuscript. We changed 'scaling' to 'speed-up' in the revision.

Comment 15: *Once the atmosphere model is decomposed onto 256 MPI ranks, the resulting computational area is 16x16 grid cells. To indicate that scaling performance tails off, the relative cost of computation and communication needs to be brought up. "The boundary tiles in each processor are 25% of the total, and the parallel communication cost increases significantly." For the atmosphere model, depending on the communication stencil, between approximately 20 and 60% of the computational area could be communicated. But until we know the relative cost of computations and communications, we are left with "cost increases significantly". There is nothing actionable for a user in that statement. Worse, users are left with the impression that after 256 MPI ranks, the communication costs increase significantly for all model configurations.*

Reply 15: The aim of the parallel efficiency test is to demonstrate that the ESMF coupler interface does not slow down the simulations. We agree with the reviewer that we did not show the relative cost of computation and communication in our manuscript. We have revised to focus on that aspect of the code.

Comment 16: *When assigning relative costs between the atmosphere and ocean model, the most important factor of the ratio of the number of computational cells between the two models is ignored and "more complex physics parameterization packages" is offered. A clear and accurate representation is the relative cost of a single-column of the atmosphere model compared to a single column of the ocean model, for a single time step. This permits a user to assign MPI ranks to different model components.*

Reply 16: We thank the reviewer for pointing out the importance of the number of grid columns between the ocean and atmosphere models. We now discuss this. Please refer to the changes in Section 5 and Table 3.

Comment 17: *A number of external sources (books and online) define a "mean deviation" to be the same as mean absolute deviation. Perhaps bias is a better term and seems to be what the authors are interested in (warmer, cooler, etc).*

Reply 17: We thank the reviewer for pointing out this. We have replaced 'mean deviation' using bias in the discussion of the results.

Comment 18: *Care has to be taken when using words that have a typical meaning in a field, but that meaning is not intended. 1. significant: "but the increase is not significant after that". 2. To an atmosphere model user, the term "micro-physics" refers to the bulk or bin parameterization schemes that deal with resolved scale moist processes. "WRF micro-physics models (e.g., land surface model, the PBL model)"*

Reply 18: Thanks. We have replaced the improper words in the manuscript. We have also gone through the entire manuscript to revise other improper words.

Comment 19: *Given that 30 days x 24 hr/day x 60 min/hr x 2 model time steps / min = 86k atmosphere model time steps (which each have an elapsed time reported individually), there should* *be a set of either error bars or standard deviation on all of the reported timing values. Similar statistical information is missing from the differences of the physical fields.*

Reply 19: Thanks. We have added the comparison of standard deviation in the T2 results of three major cities in Fig. 7. In three major cities, the observed T2 values are available to generate the mean value and standard deviation.

Comment 20: *In the description of the atmosphere model, the option for the resolved-scale moist physics is not mentioned. The atmosphere model lid for the vertical coordinate is not provided. A single, deterministic simulation for a month should probably use spectral nudging in the WRF model to keep the large-scale atmospheric flow in check. An atmospheric modeler would find the model setup section incomplete.*

Reply 20: We thank the reviewer for pointing out that we did not provide adequate details regarding the atmospheric model setup. We have added the description of the moist physics scheme (Morrison 2-moment scheme) and the top atmosphere boundary condition ($P_{top} = 50$ hPa). In our work, we don't use spectral nudging as we also want to test the model in a forecasting framework.

Comment 21: *The discussion of the selection of the ocean model should have included the benefits and applicability near coastal areas, how the horizontal and vertical resolution are modified in shallow coastal areas, the impact of the broad shallow portions of the Red Sea on the vertical levels, spin-up time from initial conditions, the sensitivity of correctly choosing boundary locations, and coupling frequency.*

Reply 21: MITgcm uses a finite volume grid and the vertical resolution is not modified in shallow coastal areas. HYCOM reanalysis data is prescribed as the initial condition and we did not apply a spin-up.

We tried different coupling frequency (2 min, 20 min, 180 min), but we did not see much difference in SST and T2. We selected 20 minutes because because it is adequate to capture the diurnal SST variation according to the reference (Seo et al., 2014).

Comment 22: *It is not a fair comparison to make when you keep the SST constant for a month: "Improvements of the coupled model over the stand-alone simulation with static SST forcing are observed in capturing the T2, heat fluxes, evaporation, and wind speed." Also, it was stated that the momentum fields were not impacted in your study: "This suggests that the ocean–atmosphere coupling does not significantly influence the wind field in the Red Sea region during the heat wave events."*

Reply 22: We thank the reviewer for pointing out this. We agree that it is not proper to simply say the coupling does not influence the wind field in a technical paper. We have removed this and rewritten the discussion of the surface wind.

Comment 23: *It is not conventional to have a statement as this in the conclusions: "On the other hand, the difference between coupled simulation and stand-alone simulations with updated forcings is also discussed."*

Reply 23: We thank the reviewer for pointing out this. We have removed this sentence. We have rewritten the conclusion section to focus more on the technical aspects of the work.

Comment 24: *When verifying an 8-km domain with 30-km gridded results, briefly describe the process.*

Reply 24: We converted the validation data on the lower resolution grid to the 0.08 $^\circ$ model domain using 2D spline interpolation. We have added this to Section 4.

Comment 25: *Without some sort of statistical assistance, we do not know if -1.55 is mostly the same as or pretty different from -1.66. "The mean T2 differences over the sea are -1.55 (CPL), -1.66 (ATM.STA), and -1.7 (ATM.DYN) after 36 hours, and -0.99 (CPL), -1.10 (ATM.STA), and -1.12 (ATM.DYN) after 48 hours."*

Reply 25: We thank the reviewer for pointing out this. We have added the mean temperature and standard deviation of T2 from the ECMWF data. We use the mean and the standard deviation of T2 to show the difference between simulations is very small after 36 and 48 hours.

Comment 26: *Figure 1b has lots of arrows. Are they one-way only? If the parent talks to the child components directly, why is there a child coupler component?*

Reply 26: Yes. They are one-way only. The arrows are showing how the main function calls the parent component and then calls the child components. The coupler component handles the grid interpolation and data transfer between different models. We have added this in the manuscript.

Comment 27: *Most people would take Figure 1c to be an indication that modeling system is running concurrently.*

Reply 27: We thank the reviewer for pointing out this. We have re-plotted Figs. 1b and 1c in Fig. 2 to show the system components are running subsequently.

Comment 28: *Are there computational trade-offs for selecting a sequential rather than concurrent coupling mode? Does your implementation preclude selecting sequential vs concurrent as a build- or run-time option?*

Reply 28: 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.

ESMF usually makes the sequential or concurrent mode as a build-time option. Our case is built only in the sequential mode.

Comment 29: *If a purely marine region was selected, is there an expectation that the cost of the atmosphere and ocean models would more equal?*

Reply 29: Yes, the cost of atmosphere and ocean models can be more equal if a purely marine region is selected in an ideal case. In the realistic Red Sea case, the ocean only covers 16% of the entire region. We thank the reviewer for pointing out this and we have revised the discussion on the ratio between ocean and atmosphere models:

The atmospheric model is much more time-consuming because it solves the entire computational domain, while the ocean model only solves the Red Sea (16% of the domain). The atmospheric model also uses a smaller time step (30 s) than that of the ocean model (120 s) and has more complex physics parameterization packages. If a purely marine region is selected in an ideal case, the cost of ocean and atmosphere models would be more equal. [R1]

Comment 30: *Do the atmosphere and ocean models run on the same processor set? If so, are the parallel tests hampered with fewer ocean points as the number of processors are increased?*

Reply 30: Yes, the atmosphere and ocean models run on the same processor set. We also agree that the parallel tests are hampered with very few ocean points when using 256 processors. However, the parallel efficiency of the coupled code is still good compared with that in the literature (Christidis, 2015; Zhang et al., 2013). We have revised our manuscript.

Comment 31: *After 1 day of simulation, why is the modeled SST so much colder than observed in Figure 7?*

Reply 31: The coupled simulations are all initialized using HYCOM data. The initial HYCOM SST is about 1 degree cooler than GHRSST observation data.

Comment 32: *There are several instances of trying to read too much into differences of the fields: "On the other hand, for the heat wave event on June 24th, CPL and ATM.DYN runs exhibit more latent heat fluxes coming out of the ocean (157 and 131 W/m2) than that in ATM.STA run (115 W/m2)." OK, yes, but if you look at Figure 9a, VIII vs IX vs X, IX and X are more similar than VIII and X.*

Reply 32: The mean difference in Fig. 10(IX) is -9.8 w/m2 and 5.9 w/m2 in Fig. 10(X), while the mean difference in Fig. 10(VIII) is 31.8 w/m2. In the coupled run, the sea surface is warmer and the latent heat flux is higher. We have added it to our revised manuscript.

Comment 33: *Technical Corrections*

Figure 1 is pretty busy. 1a mentions "using bulk formulas" instead of listing the variables passed.

Reply 33: We thank the reviewer for pointing out this. In the present work, we use COARE 3.0 bulk algorithm to calculate the turbulent heat fluxes (Fairall et al., 2003). We have added this in Section 3. We have also added the list of variables passed in the test case in Section 3:

In the coupling process, the ocean model sends SST and ocean surface velocity to the coupler. They are used directly as the boundary conditions in the atmosphere model. The atmosphere model are sending the surface fields to the coupler: (1) net surface shortwave/longwave radiation, (2) latent/sensible heat; (3) 10-m wind speed, (4) net precipitation, (5) evaporation. The ocean model uses the atmosphere surface fields to compute the surface forcing: (1) total net surface heat flux, (2) surface wind stress, (3) freshwater flux. The total net surface heat flux is computed by adding latent heat flux, sensible heat flux, and net surface shortwave/longwave radiation fluxes. The surface wind stress is computed by using the 10-m wind speed (Large and Yeager, 2004). The freshwater flux is the difference between precipitation and evaporation. The latent sensible heat fluxes are computed by using COARE 3.0 bulk algorithm in WRF (Fairall et al., 2003). In the coupled code, different bulk formulae in WRF or MITgcm can also be used. [R1]

Comment 34: *A number of figures would benefit from smaller color bar ranges. For example 10a has a range from about -500 W/m2 to 500 w/m2. The text says "However, a small improvement in the CPL (2.19 W/m2) and ATM.DYN (1.27 W/m2) runs can be observed in the longwave radiation on June 24th".*

Reply 34: Thanks. We have re-plotted the color bar ranges to highlight the values in the figure. We agree that improvement is too small to be observed in the figure. The text on the difference between simulations is revised:

However, compared with ATM.STA run, there is a small improvement in the CPL (2.19 W/m^2) and ATM.DYN (1.27 W/m^2) runs on June 24th.[R1]

Comment 35: *Figure 6 has diffs, diffs of diffs, rmse of diffs, and diffs of rmse of diffs. The y-axis labeling and the in-plot descriptions should be more precise.*

Reply 35: Thanks. We have revised the descriptions in Fig. 6 to make the label and descriptions more precise.

Comment 36: *Figure 5 would benefit from having some highlight that indicated the four heatwave periods.*

In several places, "access" and "assess" are swapped.

Table 1 has ATM.STA twice. The second should be ATM.DYN.

Both are used "Arakawa-C grid" and "Arakawa C-grid".

There are some clumsy wordings "This run allows to access the WRF model", "which means that the SST in CPL run is tending to be similar to the realistic."

Figure 3 has a gray bar that covers the table bar. Figure 6 misspells deviation

Page 15, line 2, ATM.STA should be ATM.DYN?

Reply 36: The authors thank the reviewer for pointing out these technical issues. We have corrected them in our manuscript. We have also gone through the manuscript and revised some other technical issues.

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.

Fairall, C., Bradley, E. F., Hare, J., Grachev, A., Edson, J., 2003. Bulk parameterization of air–sea fluxes: Updates and verification for the COARE algorithm. Journal of climate 16 (4), 571–591.

Large, W. G., Yeager, S. G., 2004. Diurnal to decadal global forcing for ocean and sea-ice models:

the data sets and flux climatologies. Tech. rep., NCAR Technical Note: NCAR/TN-460+STR. CGD Division of the National Center for Atmospheric Research.

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.