Replies to the Comments of Reviewers

R.Sun et al.

April 17, 2019

1 Reply to Reviewer Comment 1 (RC1)

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 ° 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.

2 Reply to Reviewer Comment 2 (RC2)

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 37: *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 37: 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 38: *Page 6, lines 21-32: Too technical.*

Reply 38: 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 39: *Page 6, line 32: Why was sequential mode selected?*

Reply 39: 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 40: *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 40: 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 41: *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 41: The authors thank the reviewer for pointing out this. We have fixed these typos in the manuscript.

Comment 42: *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 42: 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 43: *Page 10, line 21-22: Why not use a nest with finer grid spacing to resolve the local topography?*

Reply 43: 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 44: *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 44: 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 45: *Page 11, Figure 3: There was a gray stripe at the bottom, making it impossible to read labels of the color bars.*

Reply 45: Thanks. We have updated the figure and removed the gray stripe of this figure.

Comment 46: *Page 11, line 16: Land surface model and PBL model are not microphysics models.*

Reply 46: 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 47: *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 47: Yes, some T2 points are missing from NOAA NCDC data. We have added this in Fig. 7 of the revised manuscript.

Comment 48: *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 48: 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 49: *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 49: 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 50: *Page 16, Figure 8: Why no time series comparison to MERRA-2 dataset?*

Reply 50: 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 51: *Page 19, line 6: Which selected micro-physics schemes?*

Reply 51: We have replaced the original sentence using 'selected WRF physics options presented in Section 3'.

Comment 52: *Page 20, line 3: 64 /cm/year should read 64 cm/year?*

Reply 52: Thanks. We have fixed this typo.

Comment 53: *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 53: 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 ($16\times16\times80$, 16 lat$\times$16 lon$\times$80 vertical levels) in each processor, but there are 5120 overlap cells ($4\times16\times80$, 4 sides$\times$16 tiles per side$\times$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 54: *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 54: 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 55: *Page 23, Table 3: Please use the experiment names consistently throughout the manuscript.*

Reply 55: Thanks. We have revised the experiment names in Table 3. We have also read through the manuscript ensuring that the experiment names are consistent.

3 Reply to executive editor (SC1)

Comment 56: *Interactive comment on A regional coupled ocean–atmosphere modeling framework (MITgcm–WRF) using ESMF/NUOPC: description and preliminary results for the Red Sea by Rui Sun et al.*

In my role as Executive editor of GMD, I would like to bring to your attention our Editorial version 1.1. This highlights some requirements of papers published in GMD, which is also available on the GMD website in the 'Manuscript Types' section. In particular, please note that for your paper, the following requirement has not been met in the Discussions paper:

"The main paper must give the model name and version number (or other unique identifier) in the title."

Please provide the version numbers of MITgcm–WRF and ESMF/NUOPC in the title of your revised manuscript.

Reply 56: The authors thank the executive editor for pointing out this issue.

We have revised the title to include the name and version of our model. The versions of the ocean/atmosphere model components and the ESMF coupler are also highlighted in the manuscript.

Now the title of the manuscript is: SCRIPS v1.0: A regional coupled ocean–atmosphere modeling framework (MITgcm–WRF) using ESMF/NUOPC, description and preliminary results for the Red Sea

Comment 57: *GMD is encouraging authors to upload the program code of models (including relevant data sets) as supplement or make the code and data of the exact model version described in the paper accessible through a DOI (digital object identifier). In case your institution does not provide the possibility to make electronic data accessible through a DOI you may consider other providers (eg. zenodo.org of CERN) to create a DOI. Please note that in the code availability section you can still point the reader to how to obtain the newest version. If for some reason the code and/or data cannot be made available in this form the "Code Availability" section need to clearly state the reasons for why access is restricted (e.g. licensing reasons).*

Especially, please note, that it is not enough, that the code will be available in the future. It must be available now and the exact version of the code published in this article needs to be made available.

Reply 57: The authors thank the executive editor for pointing out this in the code availability section. We have uploaded our source code, test cases, and code documentation to a GitHub repository `https://github.com/iurnus/scripps_kaust_model`. This repository is now open to public. We have added the link of the GIT repository to the manuscript.

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.

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.

SKRIPS v1.0: A regional coupled ocean–atmosphere modeling framework (MITgcm–WRF) using ESMF/NUOPC~~:~~ , description and preliminary results for the Red Sea

Rui Sun[1], Aneesh C. Subramanian[1], Arthur J. Miller[1], Matthew R. Mazloff[1], Ibrahim Hoteit[2], and Bruce D. Cornuelle[1]

[1]Scripps Institution of Oceanography, La Jolla, California, USA
[2]Physical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Correspondence: Rui Sun (rus043@ucsd.edu); Aneesh Subramanian (acsubram@ucsd.edu)

Abstract. A new regional coupled ocean–atmosphere model is developed to study air–sea feedbacks. The coupled model is based on two open-source community model components: (1) MITgcm ocean model; (2) Weather Research and Forecasting (WRF) atmosphere model. The coupling between these components is performed using ESMF (Earth System Modeling Framework) and implemented according to National United Operational Prediction Capability (NUOPC) consortium. The ~~regional coupled model allows affordable simulation where~~ coupled model is named the Scripps–KAUST Regional Integrated Prediction System (SKRIPS). The SKRIPS allows affordable regional simulation of oceanic mixed layer heat and momentum ~~interact~~ interactions with atmospheric boundary layer dynamics at mesoscale and higher resolution. This can capture the feedbacks which are ~~otherwise~~ not well-resolved in ~~coarse resolution~~ coarse-resolution global coupled models and are absent in regional uncoupled models. ~~To test the regional coupled model , we focus on a~~ After the model was created and passed a typical suite of consistency checks, we demonstrated it using a real-world example. It simulated a 30-day period including a series of heat wave events that occurred on the eastern shore of the Red Sea region in June ~~2012 using a 30-day simulation.~~ 2012. The results obtained using the coupled model, along with those in forced uncoupled ocean or atmosphere model simulations, are compared with observational ~~and reanalysis data~~data and reanalysis products. All configurations of coupled and uncoupled models have good skill in modeling variables of interest in the region. The coupled model shows improved skill in temperature and circulation evaluation metrics. In addition, a scalability test is performed to investigate the parallelization of the coupled model. The results indicate that the ~~coupled model scales linearly for up to 128 CPUs and sublinearly for more processors. In the coupled simulation, the~~ ESMF/NUOPC interface ~~also~~ scales well and accounts for less than 10% of the total computational resources ~~compared with uncoupled models. Hence this newly developed regional model scales efficiently for a large number of processors and can be applied for high-resolution coupled regional modeling studies.~~ used in the simulation.

1 Introduction

Accurate and efficient forecasting of oceanic and atmospheric ~~circulations~~ circulation is essential for a wide variety of high-impact societal needs, including ~~:~~ extreme weather and climate events (Kharin and Zwiers, 2000; Chen et al., 2007)~~;~~ , environmental protection and coastal management (Warner et al., 2010)~~;~~ , management of fisheries (Roessig et al., 2004), marine conservation (Harley et al., 2006)~~;~~ , water resources (Fowler and Ekström, 2009)~~;~~ , and renewable energy (Barbariol et al., 2013). Effective forecasting relies on high model fidelity and accurate initialization of the models with the observed state of the ~~ocean–atmosphere coupled~~ coupled ocean–atmosphere system. Although global coupled models are now being implemented with increased resolution, ~~high-resolution~~ higher-resolution regional coupled models, if properly driven by the boundary conditions, can ~~contribute additional~~ provide an affordable way to study air–sea ~~feedback information to the study of small-scale~~ feedback for frontal-scale processes.

A number of regional coupled ocean–atmosphere models have been developed for various goals in the past decades. An early example of building a regional coupled model for realistic simulations focused on accurate weather forecasting in the Baltic Sea (Gustafsson et al., 1998; Hagedorn et al., 2000; Doscher et al., 2002), and suggested that the coupled model improved the SST (Sea Surface Temperature) and atmospheric circulation forecast. Enhanced numerical stability in the coupled simulation was also observed. These early attempts were followed by other practitioners in ~~basin-scale~~ ocean-basin-scale climate simulations ~~(e.g. Huang et al., 2004; Aldrian et al., 2005; Xie et al., 2007; Seo et al., 2007; Somot et al., 2008; Fang et al., 2010; Boé et al., 2011~~ ~~For instance~~(e.g. Huang et al., 2004; Aldrian et al., 2005; Xie et al., 2007; Seo et al., 2007; Somot et al., 2008; Fang et al., 2010; Boé et al For example, Huang et al. (2004) implemented a regional coupled model to study three major important patterns contributing to the variability and predictability of the Atlantic climate. The study suggested that these patterns originate from air–sea coupling within the Atlantic Ocean or by the oceanic responses to atmospheric internal forcing. Seo et al. (2007) studied the nature of ocean–atmosphere feedbacks in the presence of oceanic mesoscale eddy fields in the eastern Pacific Ocean sector. The evolving SST fronts were shown to drive an unambiguous response of the atmospheric boundary layer in the coupled model, and lead to model anomalies of wind stress curl, wind stress divergence, surface heat flux, and precipitation that resemble observations. This study helped substantiate the importance of ocean–atmosphere feedbacks involving oceanic mesoscale variability features.

~~On top of~~ In addition to basin-scale climate simulations, regional coupled models are also used to study weather extremes. For example, the COAMPS (Coupled Ocean/Atmosphere Mesoscale Prediction System) was applied to investigate idealized tropical cyclone events (Hodur, 1997). This work was then followed by other realistic extreme weather studies. For example, extreme bora wind events in the Adriatic Sea were investigated using different regional coupled models (Loglisci et al., 2004; Pullen et al., 2006; Ricchi et al., 2016). The coupled simulation results demonstrated improvements in describing the air–sea interaction processes by taking into account ocean surface heat fluxes and ~~wind~~ wind-driven ocean surface wave effects (Loglisci et al., 2004; Ricchi et al., 2016). It was also found in model simulations that SST after bora wind events had a stabilizing effect on the atmosphere, reducing the atmospheric boundary layer mixing and yielding stronger near-surface wind (Pullen et al., 2006). ~~The regional coupled models was also used by Bender and Ginis (2000); Chen et al. (2007); Warner et al. (2010)~~ Regional coupled models were also used for improving the ~~tracking~~ forecasts of the hurricane path and intensity, predicting SST variation, and forecasting wind speeds (Bender and Ginis, 2000; Chen et al., 2007; Warner et al., 2010) .

Regional coupled modeling systems also play important roles in studying the effect of surface variables (e.g., surface evaporation, precipitation, surface roughness) in the coupling ~~process~~ processes of ocean or lakes. One example is the study conducted by Powers and Stoelinga (2000), who developed a coupled model and investigated the ~~frontal passage~~ atmospheric frontal passages over the Lake Erie region. Sensitivity analysis was performed to demonstrate that taking into account lake surface roughness parameterization in the atmosphere model can improve the calculation of wind stress and heat flux. Another example is the investigation by Turuncoglu et al. (2013), who compared a regional coupled model with uncoupled models ~~,~~ and demonstrated the improvement of the coupled model in capturing the response of ~~the Caspian Sea level~~ Caspian Sea levels to climate variability.

~~Despite the existing regional coupled ocean–atmosphere models ,~~ In the past ten years, many regional coupled models have been developed using modern model toolkits (Zou and Zhou, 2012; Turuncoglu et al., 2013; Turuncoglu, 2019) and include waves (Warner et al., 2010; Chen and Curcic, 2016) , sediment transport (Warner et al., 2010) , sea ice (Van Pham et al., 2014) , and chemistry packages (He et al., 2015) . However, it is still desirable and useful to develop a new coupled regional ocean–atmosphere model ~~with 'state-of-the-art' physics and using modern coupling framework~~ implemented using an efficient coupling framework and with state estimation capabilities. The goal of this work is to (1) introduce the design of a newly developed regional coupled ocean–atmosphere modeling system, (2) describe the implementation of the modern coupling framework, (3) present preliminary simulation results in the Red Sea region, and (4) demonstrate and discuss the parallelization of the coupled model. In the coupled system, the ~~ocean~~ oceanic model component is the MIT general circulation model (MITgcm) (Marshall et al., 1997) ~~; the atmosphere~~ and the atmospheric model component is the Weather Research and Forecasting (WRF) model (Skamarock et al., 2005). To couple the model components in the present work, the Earth System Modeling Framework (ESMF) (Hill et al., 2004) is used because of its advantages in conservative re-gridding capability, calendar management, logging and error handling, and parallel communications. The National United Operational Prediction Capability (NUOPC) layer in ESMF is also used (Sitz et al., 2017). The additional NUOPC wrapper layer between coupled model and ESMF simplifies the implementations of component synchronization, execution, and other common tasks in the coupling. ~~To test the coupled model, we focus~~ The innovations in our work are: (1) we use ESMF/NUOPC, which is a community supported computationally efficient coupling software for earth system models, and (2) we used MITgcm together with WRF. The resulting coupled model is being developed as a coupled forecasting tool for coupled data assimilation and subseasonal to seasonal (S2S) forecasting. By coupling WRF and MITgcm for the first time with ESMF, we can provide an alternative regional coupled model resource to a wider community of users. These atmospheric and oceanic model components have an active and well-supported user-base.

After testing of the new coupled model, we demonstrate it on a series of heat wave events that occurred on the eastern shore of the Red Sea region in June 2012. The simulated surface variables of the Red Sea (e.g., sea surface temperature, 2-m temperature, and surface heat fluxes) are examined and validated against available observational ~~and reanalysis data~~data and reanalysis products. To assess the improvements gained from the coupled simulation, the results are compared with those obtained using stand-alone ~~ocean or atmosphere model.~~ oceanic or atmospheric models. This is not a full investigation of the importance of coupling for these extreme events, which is outside of the scope of this paper, which focuses on the technical aspects. In addition, a scalability test of the coupled model is performed to investigate its parallel capability.

The rest of this paper is organized as follows~~: the~~. The description of the individual modeling components and the design of the coupled modeling system are detailed in Section 2. Section 3 introduces the ~~experiment design, observational~~ experimental design, validation data, and analysis methodology. Section 4 discusses the results obtained from the coupled model. Section 5 details the parallelization test of the coupled model. The last section concludes the paper and presents an outlook for future work.

2 Model Description

The newly developed regional coupled modeling system is introduced in this section. The general design of the coupled model, descriptions of individual components, and ESMF/NUOPC coupling framework are presented below.

2.1 General design

The schematic description of the coupled model is shown in Fig. 1(a). The coupled model is comprised of five components: ~~ocean~~ oceanic component MITgcm, ~~atmosphere~~ atmospheric component WRF, MITgcm–ESMF interface, WRF–ESMF interface, and ESMF/NUOPC coupler. They are to be detailed in the following sections.

The coupler component runs in both directions: (1) from WRF to MITgcm, and (2) from MITgcm to WRF. From WRF to MITgcm, the coupler collects the surface atmospheric variables (i.e., ~~surface temperature, pressure, mixing ratio, wind velocitycomponents~~solar radiation, turbulent heat flux, wind velocity, precipitation, ~~longwave and shortwave radiations) from the atmosphere component~~ evaporation) from WRF and updates the surface forcing variables (net heat flux, wind stress, freshwater flux) to drive ~~the ocean component~~MITgcm. From MITgcm to WRF, the coupler collects SST and ocean surface velocity from the ~~ocean component~~ MITgcm and uses them as the surface boundary condition in ~~the atmosphere component~~WRF. Re-gridding the data from either model component will be performed by the coupler, in which various coupling intervals and schemes can be specified by the ESMF (Hill et al., 2004).

2.2 MITgcm Ocean Model

The MITgcm (Marshall et al., 1997) is a 3-D, finite-volume, general circulation model used by a broad community of researchers for a wide range of applications at various spatial and temporal scales. The model code and documentation, which are under continuous development, are available on the MITgcm webpage http://mitgcm.org/. The 'Checkpoint 66h' (June 2017) version of MITgcm is used in the present work.

The MITgcm is designed to run on high-performance computing (HPC) platforms and can run in non-hydrostatic and hydrostatic modes. It integrates the primitive (Navier-Stokes) equations, under the Boussinesq approximation, using finite volume method on a staggered 'Arakawa C-grid'. The MITgcm uses modern physical parameterization schemes for subgrid-scale horizontal and vertical mixing and tracer properties. The code configuration includes build-time C pre-processor (CPP)

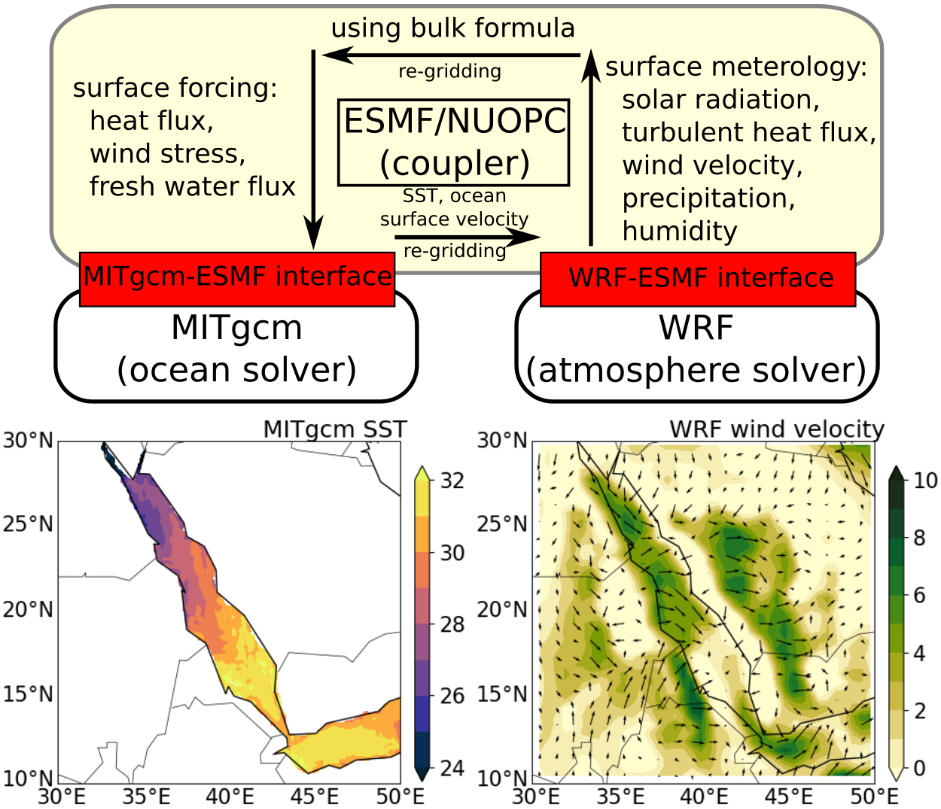

Figure 1. The schematic description ~~, general code structure and run sequence~~ of the coupled ocean–atmosphere model. ~~In panel (a), the~~ The white blocks are the ~~ocean~~ oceanic and ~~atmosphere~~ atmospheric components; the red blocks are the implemented MITgcm–ESMF and WRF–ESMF interfaces; the yellow block is the ESMF/NUOPC coupler. ~~In panel (b), the black block is the *application driver*; the red block is the *parent gridded component* called by the *application driver*; the green/blue blocks are the *child gridded/coupler components* called by the *parent gridded component*. In panel (c), each horizontal arrow indicates the time axis of each component; the ticks on the time axis indicate the time step; the boundary condition fields are updated at each coupling interval in the connector.~~

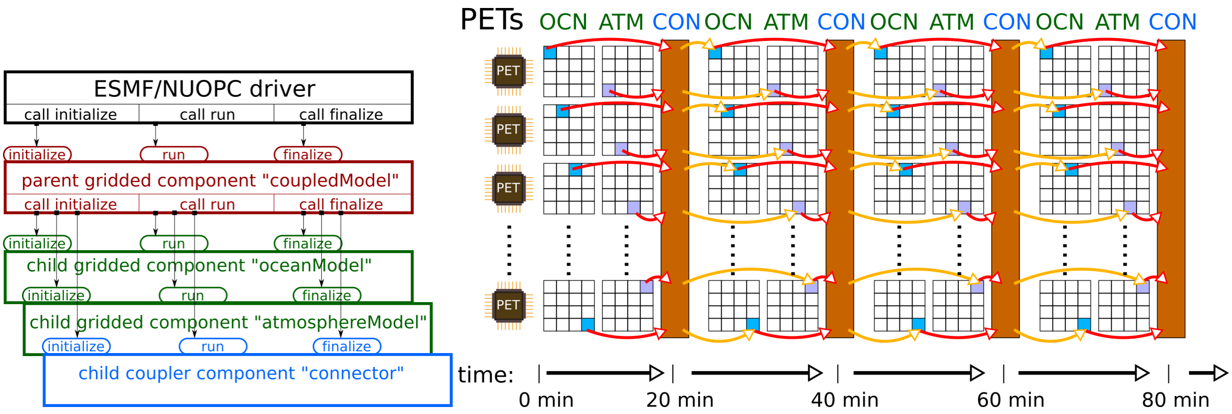

Figure 2. The general code structure and run sequence of the coupled ocean–atmosphere model. In panel (a), the black block is the *application driver*; the red block is the *parent gridded component* called by the *application driver*; the green/blue blocks are the *child gridded/coupler components* called by the *parent gridded component*. In panel (b), OCN, ATM, and CON denote oceanic component, atmospheric component and connector component, respectively. The red arrows indicate the model components are sending data to the connector and the yellow arrows indicate the model components are reading data from the connector. The horizontal black arrows indicate the time axis of each component and the ticks on the time axis indicate the coupling time step.

options and run-time switches, which allow for great computational modularity in MITgcm to study a variety of oceanic phenomena (Evangelinos and Hill, 2007).

To implement the MITgcm–ESMF interface, we separated the MITgcm main program into three subroutines that handle initialization, running, and finalization, shown in Fig. ~~1(b)~~2. These subroutines are used by the ESMF/NUOPC coupler that controls the ~~ocean~~ oceanic component in the coupled run. The surface boundary fields on the ocean surface is exchanged online[1] via the MITgcm–ESMF interface during the simulation. The MITgcm SST and ocean surface velocity are the export boundary fields, and the atmospheric surface forcing variables are the import boundary fields (see Fig. ~~1(b))~~2). These boundary fields are registered in the coupler following NUOPC consortium and timestamps[2] are added to them for the coupling. In addition, MITgcm grid information is also provided for online re-gridding of the exchanged boundary fields. To carry out the high-resolution simulation, the MITgcm–ESMF interface runs in parallel via MPI communications. The implementations of the present MITgcm–ESMF interface ~~is~~ are based on the baseline MITgcm–ESMF coupler (Hill, 2005), but we updated it to couple the modern version ESMF/NUOPC with MITgcm. We also modified the baseline coupler to receive atmosphere surface fluxes and send ocean surface variables (i.e., SST and ocean surface velocity).

[1]In this manuscript, 'online' means the manipulations are performed via subroutine calls during the execution of the simulations; 'offline' means the manipulations are performed when the simulations are not executing.

[2]In ESMF, 'timestamp' is a sequence of number, usually based on the time, to identify the ESMF fields. Only the ESMF fields having the correct timestamp will be transferred in the coupling.

2.3 WRF Atmospheric Model

The Weather Research and Forecasting (WRF) Model (Skamarock et al., 2005) is developed by NCAR/MMM (Mesoscale and Microscale Meteorology Division). It is a 3-D, finite-difference atmospheric model with a variety of physical parameterizations of sub-grid scale processes for predicting a broad spectrum of applications. WRF is used extensively for operational forecasts (http://www.wrf-model.org/plots/wrfrealtime.php) as well as realistic and idealized dynamical studies.

In the present work, the Advanced Research WRF dynamic version (WRF-ARW, version 3.9.1.1) is used. It solves the compressible Euler non-hydrostatic equations, and also includes a run-time hydrostatic option. The WRF-ARW uses a terrain-following hydrostatic pressure coordinate system in the vertical direction and utilizes the ‘~~Arakawa-C grid~~Arakawa C-grid’. WRF incorporates various physical processes including microphysics, cumulus parameterization, planetary boundary layer, surface layer, land surface, and longwave and shortwave radiations, with several options available for each process.

Similar with the implementations in MITgcm, WRF is also separated into initialization, run, and finalization subroutines to enable the WRF–ESMF interface to control the atmosphere model during the coupled simulation, shown in Fig. ~~1(b)~~2. The implementation of the present WRF–ESMF interface is based on the prototype interface (Henderson and Michalakes, 2005). In the present work, the prototype WRF–ESMF interface is updated to a modern version of WRF-ARW and a modern version of ESMF, based on the NUOPC layer. This prototype interface is also expanded to interact with the ESMF/NUOPC coupler to receive the ocean surface variables and send the atmosphere surface fluxes. The surface boundary condition fields are registered in the coupler following the NUOPC consortium with timestamps. The WRF grid information is also provided for online re-gridding by ESMF. To carry out the high-resolution simulation, the WRF–ESMF interface also runs in parallel via MPI communications.

2.4 ESMF/NUOPC Coupler

The coupler is implemented using ESMF version 7.0.0. The ESMF is selected because of its high-performance and flexibility for building and coupling weather, climate, and related Earth science applications (Collins et al., 2005; Turuncoglu et al., 2013; Chen and Curcic, 2016; Turuncoglu and Sannino, 2017). It has a superstructure for representing the model and coupler components and an infrastructure of commonly used utilities, including conservative grid remapping, time management, error handling, and data communications.

The general code structure of the coupler is shown in Fig. ~~1(b). In the main program, an~~ 2. To build the ESMF/NUOPC driver~~is created and controls the ESMF component ‘coupledModel’, which is also the parent component. When the main program calls the~~ , a main program is implemented to control an ESMF parent component, ~~the parent component cascades the calls to the child componentsand the coupler components~~which controls the child components. In the present work, ~~the ESMF child gridded ocean and atmosphere components are connected via the ESMF gridded couplercomponent. The~~ three child components are implemented: (1) the oceanic component; (2) the atmospheric component; and (3) the ESMF coupler. The coupler is used here because it performs the two-way interpolation and data transfer~~are performed using the coupler component, and ESMF supports different re-gridding approaches or unit conventions~~ (Hill et al., 2004). ~~The ESMFgridded~~

~~and coupler~~ In ESMF, the model components can be run in parallel as a group of Persistent Execution Threads (PETs), which are single processing units (i.e. CPU, GPU) defined by ESMF~~, and the PETs can be created by the user in a flexible way for parallelization~~. In the present work, the PETs are created according to the grid decomposition, and each PET is associated with an MPI process running on a separate processor.

The ESMF also allows the PETs running in sequential mode, concurrent mode, or a mixed mode. We selected the sequential mode in the implementations, shown in Fig. 2. In sequential mode, a set of ESMF gridded/coupler components runs in sequence on the same set of PETs~~;~~. At each coupling time step, the oceanic component is executed when the atmosphere component is completed or vice versa. However, in concurrent mode, the gridded components are created and run on mutually exclusive sets of PETs~~, and are coupled by a coupler component. We selected the sequential mode in the implementations.~~ . There are some advantages of concurrent mode, however the simplicity of sequential mode makes it a natural starting point (Collins et al., 2005) , and it is chosen for this work.

In ESMF, the gridded components are used to represent models and coupler components are used to connect these models. The interfaces and data structures in ESMF have few constraints, providing the flexibility to be adapted to many modeling systems. However, the flexibility of the gridded components can limit the interoperability across different modeling systems. To address this issue, the NUOPC layer is developed to provide the coupling conventions and the generic representation of the model components (e.g. drivers, models, connectors, mediators). The NUOPC layer in the present coupled model is implemented according to the documentations (Hill et al., 2004; Theurich et al., 2016), and the ~~ocean~~oceanic/~~atmosphere~~ atmospheric component each has:

1. Prescribed variables for NUOPC to link the components;
2. The entry point for registration of the components;
3. An *InitializePhaseMap* which describes a sequence of standard initialization phases, including advertising the fields that a component can provide, checking and mapping the fields to each other, and initializing the fields that will be used;
4. A *RunPhaseMap* that checks the incoming clock of the driver, examines the timestamps of incoming fields, and runs the component;
5. Timestamps on exported fields consistent with the internal clock of the component;
6. The *finalization method* to clean up all allocations.

The subroutines that handle initialization, running, and finalization in MITgcm and WRF will be included in the *InitializePhaseMap*, *RunPhaseMap*, and *finalization method* in the NUOPC layer, respectively.

3 Experiment Design and Observational Datasets

To test the coupled model, we applied it to study a series of heat wave events in the Red Sea region. We selected the extreme heat wave events because of their societally relevant impacts. The simulation of the Red Sea extends from 0000 UTC 01 June

2012 to 0000 UTC 01 July 2012. We select this month because of the record-high surface air temperature observed in the Makkah region, located 70 km inland from the eastern shore of the Red Sea (Abdou, 2014).

The computational domain and bathymetry are shown in Fig. 3. The model domain is centered at 20° N and 40° E, and the bathymetry is from the 2-minute Gridded Global Relief Data (ETOPO2) (National Geophysical Data Center, 2006). WRF is implemented using a horizontal grid of 256×256 points and grid spacing of ~~8 km, with~~ 0.08°, using cylindrical equidistant map (latitude-longitude) projection. There are 40 terrain-following vertical levels, more closely spaced in the atmospheric boundary layer. The time step for atmosphere simulation is 30 seconds. The Morrison 2-moment scheme (Morrison et al., 2009) is used to resolve the microphysics. The updated version of the Kain–Fritsch convection scheme (Kain, 2004) is used with the modifications to include the updraft formulation, downdraft formulation, and closure assumption. The Yonsei University (YSU) scheme (Hong et al., 2006) is used for the planetary boundary layer (PBL), and the Rapid Radiation Transfer Model for GCMs (RRTMG; Iacono et al. (2008)) is used for longwave and shortwave radiation transfer through the atmosphere. The Rapid Update Cycle (RUC) land surface model is used for the land surface processes (Benjamin et al., 2004). The MITgcm uses the same horizontal grid spacing as WRF, with 40 vertical z-levels that are more closely spaced near the surface. The time step of the ocean model is 120 seconds. The horizontal sub-grid mixing is parameterized using nonlinear Smagorinsky viscosities, and the K-profile parameterization (KPP) (Large et al., 1994) is used for vertical mixing processes.

In the coupling process, the ocean model sends SST and ocean surface velocity to the coupler, and they are used directly as the boundary conditions in the atmosphere model. The atmosphere model sends the surface fields to the coupler, including (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, including (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 (?) . 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.

To study the air–sea interactions, the following sets of simulations using different surface forcings are performed:

1. Run CPL: a two-way coupled MITgcm–WRF simulation. The coupling interval is 20 minutes to capture the diurnal cycle (Seo et al., 2014) . This run tests the performance of the ~~high-resolution~~ two-way coupled ocean–atmosphere model. ~~The atmosphere is initialized using the ECMWF ERA5 reanalysis dataset, with a grid resolution of 30 km (Hersbach, 2016) . The same data also provide the boundary conditions for air temperature, wind speed, and air humidity every six hours. The ocean model uses the assimilated HYCOM/NCODA 1/12° global analysis data () as initial and boundary conditions for ocean temperature, salinity, and horizontal velocities. The boundary conditions for the ocean is updated on a daily basis. The initial condition, boundary condition, and forcing terms of this run are summarized in Table 1.~~

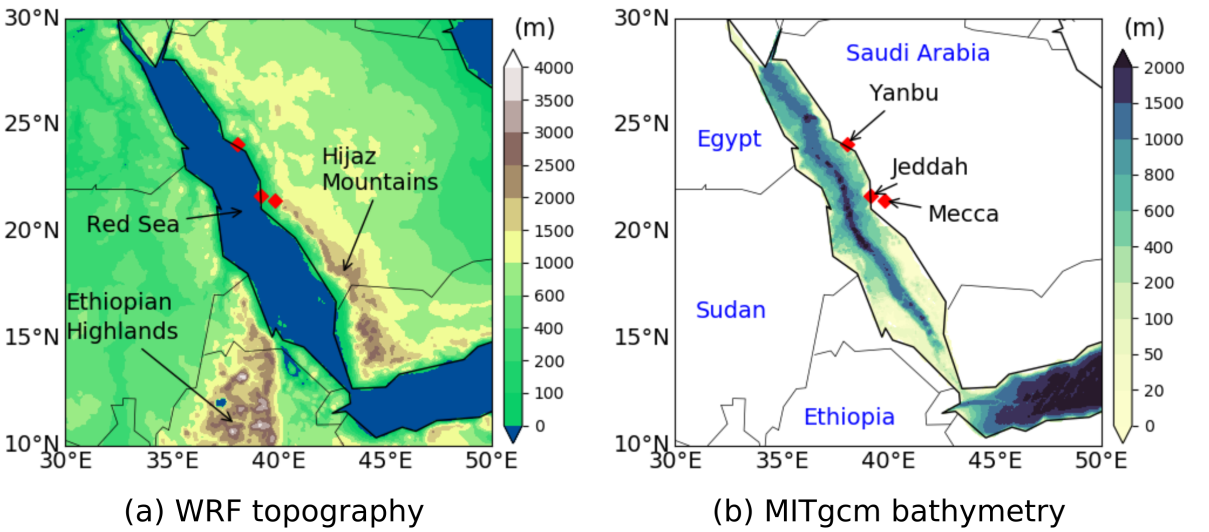

Figure 3. The WRF topography and MITgcm bathymetry in the simulations. Three major cities near the eastern shore of Red Sea are highlighted.

2. Run ATM.STA: a stand-alone WRF simulation with its initial ~~HYCOM/NCODA~~ SST kept constant throughout the simulation. This run allows ~~to access~~ assesment of the WRF model behavior with realistic, but ~~static SST, and~~ persistent SST. This case serves as a benchmark to highlight the difference between coupled and uncoupled runs.

3. Run ATM.DYN: a stand-alone WRF simulation with ~~the SST forcing prescribed using daily HYCOM/NCODA SST. The HYCOM/NCODA SST dataset is selected because it also provides the oceanic boundary condition in the CPL run. This allows accessing~~ a varying, prescribed SST. This allows assessing the WRF model behavior with updated sea surface temperature. The ocean's effect on the atmosphere is considered in the ATM.DYN run. In practice an accurately evolving SST would not be available for forecasting, however the comparison between ATM.DYN and CPL runs is used to demonstrate skill in the coupled model.

4. Run OCN.DYN: a stand-alone MITgcm simulation forced by the ERA5 dataset. The bulk formula in MITgcm is used to derive the turbulent heat fluxes. ~~The~~ This run assesses the MITgcm model behavior with prescribed lower-resolution atmospheric surface forcing, and like the ATM.DYN run is used to show the skill of the coupled model.

The ocean model uses the assimilated HYCOM/NCODA 1/12° global analysis data as initial and boundary conditions for ocean temperature, salinity, and horizontal velocities (http://hycom.org/data-server/glb-analysis). The boundary conditions for the ocean are updated on a daily basis and linearly interpolated between two simulation time steps. 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 CPL, ATM.STA, and ATM.DYN runs, we used the same initial condition and lateral boundary condition for the atmosphere. The atmosphere is initialized using the ECMWF ERA5 ~~dataset is used because it~~ reanalysis dataset, which has a grid resolution of approximately 30 km (Hersbach, 2016) . The same data also provide the boundary conditions for air temperature, wind speed, and air humidity every 6 hours. The atmosphere boundary conditions are also linearly interpolated between two simulation time steps. 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. The pressure at the top of the atmosphere is 50 hPa. In ATM.STA run, the SST from the HYCOM/NCODA data is used as initial and persistent SST. The time-varying SST in ATM.DYN run is also generated using HYCOM/NCODA data. We selected HYCOM/NCODA data because the ocean model initial condition and boundary conditions are generated using it. For the OCN.DYN run we select the ERA5 dataset for prescribed atmospheric state because it also provides the atmospheric boundary ~~condition~~ conditions in the CPL run. ~~This run accesses the MITgcm model behavior with prescribed lower-resolution atmospheric surface forcing .~~ The initial condition, boundary condition, and forcing terms of this run are summarized in Table 1.

Table 1. The initial condition, boundary condition and forcing terms used in present simulations.

run	initial and boundary conditions	ocean surface conditions	atmospheric forcings
CPL	ERA5 (atmosphere) HYCOM/NCODA (ocean)	from MITgcm	from WRF
ATM.STA	ERA5	HYCOM/NCODA initial condition kept constant	N.A.
ATM.STA	ERA5	HYCOM/NCODA updated every 24 hours	N.A.
OCN.DYN	HYCOM/NCODA	N.A.	ERA5 + MITgcm bulk formula

The analysis of the results focuses on temperature, heat flux, surface wind, and evaporation. The simulated SST data are validated against the OSTIA (Operational Sea Surface Temperature and Sea Ice Analysis) system in GHRSST (Group for High Resolution Sea Surface Temperature) (Donlon et al., 2012; Martin et al., 2012), and the simulated 2-meter air temperature (T2) is validated against the ECMWF ERA5 dataset. To evaluate the modeling of the heat wave event in three major cities near the eastern shore of Red Sea, the diurnal temperature variation is compared with observed daily maximum and minimum temperatures from NOAA National Climate Data Center (NCDC climate data online at http://cdo.ncdc.noaa.gov/CDO/georegion). Surface heat fluxes (e.g., latent heat, sensible heat, longwave and shortwave radiations), which are important for ocean–atmosphere interactions, are compared with MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, version 2) datasets (Gelaro et al., 2017). 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 ~~data used for validation are outlined~~ MERRA-2 dataset also provides a $0.625^o \times 0.5^o$ (lon $\times$ lat) resolution reanalysis fields of turbulent heat fluxes. To compare with validation data, we interpolated the validation data on the lower resolution grid to the higher resolution grid of the regional model. The validation data are summarized in Table 2.

Table 2. The dataset used to validate the simulation results.

variable	validation data
sea surface temperature (SST)	GHRSST and HYCOM
2-meter air temperature (T2)	ERA5 and NCDC climate data
turbulent heat fluxes	MERRA-2
solar radiations	MERRA-2
surface wind	MERRA-2
surface evaporation	MERRA-2

4 Results and Discussions

The Red Sea is an elongated basin covering the area between 12-30°N and 32-43°E. The basin is 2250 km long, extending from the Suez and Aqaba gulfs in the north to the strait of Bal el-Mandeb in the south, which connects the Red Sea and the Indian Ocean. Since the global models with coarse resolution cannot properly resolve local features in the narrow basin of the Red Sea (Yao et al., 2014b, a; Zhan et al., 2014), regional models with relatively higher resolutions can be used as dynamical downscaling tools for extreme temperature studies (Li et al., 2018). In this section, results of ~~high-resolution~~ the simulations using different model configurations will be presented and examined to assess the performance of the coupled model in simulating the heat wave events in the Red Sea region.

4.1 2-meter Air Temperature (T2)

We begin our analysis by examining the simulated T2 from various experiments. The simulation results obtained from coupled (CPL) run, the ERA5 data, and their associated difference are shown in Fig. 4 after 36 hours and 48 hours. It can be seen in Fig. 4(I) that the CPL run captures the heat wave event in the Red Sea region on June 2nd, compared with the ERA5 dataset in Fig. 4(II). Since ERA5 air temperature data are in good agreement with the NCDC ground observation data in the Red Sea region (comparison not shown), we use ERA5 data to validate the simulation results. The difference between the CPL run and ERA5 dataset is shown in Fig. 4(III). The ATM.STA and ATM.DYN simulation results have consistent patterns with the CPL run results and thus are not shown, but their differences with respect to the ERA5 data are shown in Fig. 4(IV) and 4(V), respectively. Fig. 4(VI) to 4(X) show the same results after 48 hours. It can be seen in Fig. 4 that all simulations reproduce the T2 patterns over the Red Sea region reasonably well compared with the ERA5 data. The mean T2 differences over the sea are -1.55 °C (CPL), -1.66 °C (ATM.STA), and -1.70 °C (ATM.DYN) after 36 hours, and -0.99 °C (CPL), -1.10 °C (ATM.STA), and -1.12 °C (ATM.DYN) after 48 hours. ~~The~~ The mean T2 ~~over the Hijaz Mountains (see Fig. 3(a)) is under-estimated by~~

~~more than 4~~differences in all simulations are mostly the same compared with the mean and standard deviation of T2 (31.01 °C and 1.93 °C ~~in all simulations after 48 hours. This is likely because the simulations on T2 suffer from the mismatches between the model terrain and the actual terrain, especially over complex mountains (Zhang et al., 2013a) . The diurnal T2 variation in the simulations is also shown in the snapshots.All~~ after 36 hours; 30.25 °C and 1.36 °C after 48 hours). Fig. 4 also shows that all simulations can capture the ~~diurnal variation of the T2~~ T2 diurnal variation in the Red Sea region, and this will be further discussed later in this section.

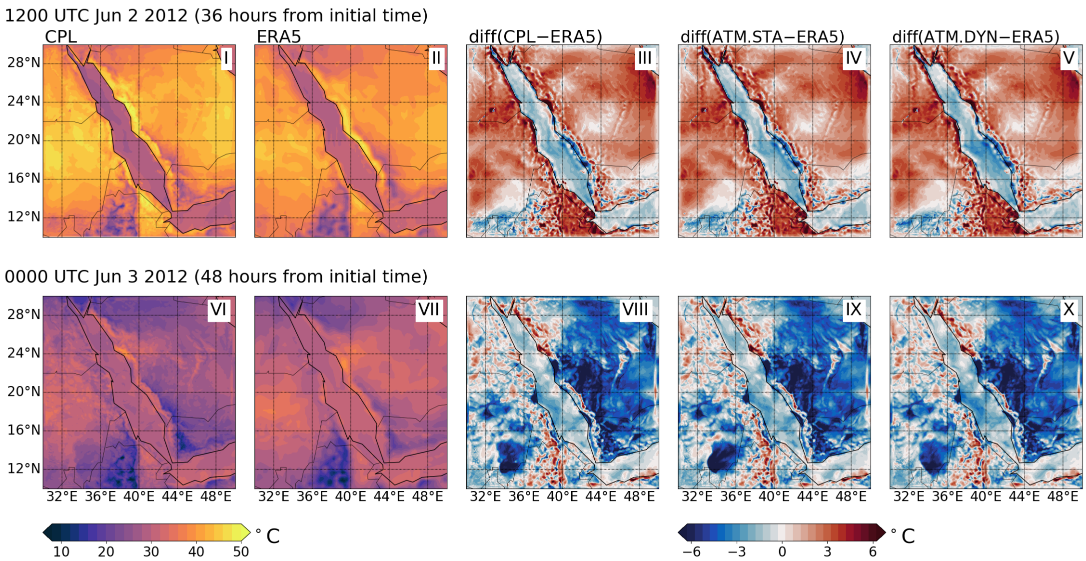

Figure 4. The surface air temperature as obtained from the CPL run, the ERA5 data, and their difference (CPL−ERA5). The difference between ATM.STA and ATM.DYN with the ERA5 data (i.e., ATM.STA−ERA5, ATM.DYN−ERA5) are also presented. The simulation initial time is 0000 UTC Jun 01 2012 for both snapshots. Two snapshots are selected: (1) 1200 UTC Jun 02 2012 (36 hours from initial time); (2) 0000 UTC Jun 03 2012 (48 hours from initial time).

The simulation results for the heat wave events on June 10th and 24th are shown in Fig. 5 to demonstrate the performance of the coupled model over longer periods of time. It can be seen in Fig. 5(III) and 5(VIII) that the T2 patterns simulated by the coupled run are consistent with the ERA5 dataset. The differences between ATM.STA and ATM.DYN simulation results with respect to the ERA5 data are shown in Fig. 5(IV), 5(V), 5(IX), and 5(X), respectively. It can be seen that the T2 over the sea in CPL simulation has a much smaller difference with the validation ERA5 data (10th: -1.02 °C; 24th: -0.84 °C) compared with the ATM.STA run (10th: -1.56 °C; 24th: -2.13 °C). Although the difference is still very small compared with the mean T2 (31.12 °C on 10th; 32.09 °C on 24th), the improvement of the coupled run is comparible to the standard deviation of T2 (2.14 °C on 10th; 2.02 °C on 24th). The CPL run results are closer to the ERA5 dataset because the ~~ocean~~ oceanic component (MITgcm) is providing updated SST, which warms the T2; the ATM.STA run uses a constant cooler SST from June 1st, and the T2 is determined by the constant cooler SST. On the other hand, when comparing the CPL run with the ATM.DYN run on

June 24th, the difference is very small (-0.10 °C on June 24th). This is because the SST fields from CPL and ATM.DYN runs are similar, which means that the SST in CPL run is tending to be similar to the realistic.

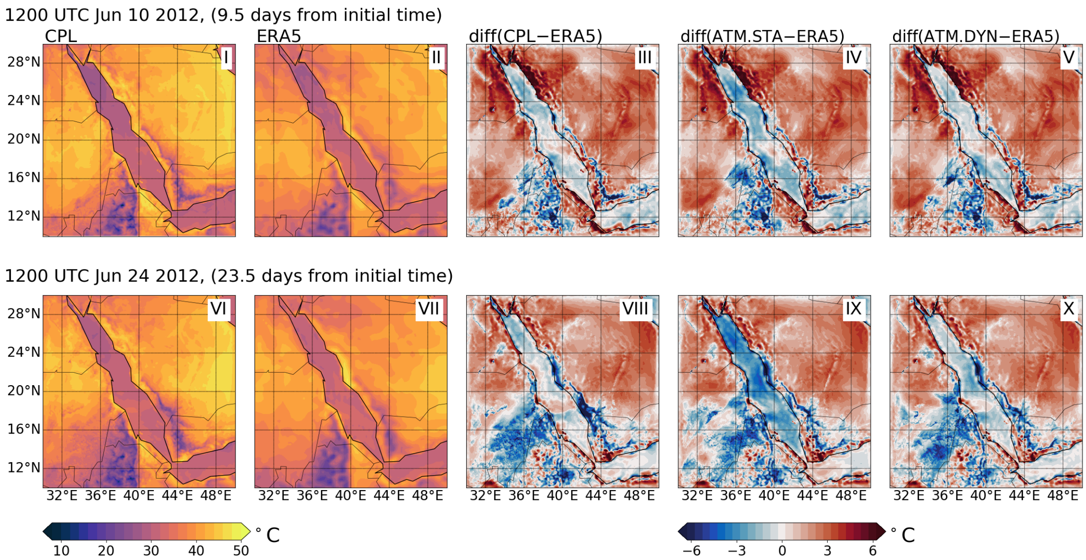

Figure 5. The surface air temperature as obtained from the CPL run, the ERA5 data, and their difference (CPL−ERA5). The difference between ATM.STA and ATM.DYN with the ERA5 data (i.e., ATM.STA−ERA5, ATM.DYN−ERA5) are also presented. The simulation initial time is 0000 UTC Jun 01 2012 for both snapshots. Two snapshots are selected: (1) 1200 UTC Jun 10 2012 (9.5 days from initial time); (2) 1200 UTC Jun 24 2012 (23.5 days from initial time).

To investigate the diurnal T2 variation in Fig. 4, the time series of T2 in three major cities as simulated in CPL and ATM.STA runs are plotted in Fig. 6, starting from June 1st~~.~~; the mean and standard deviation are shown in Fig. 7. The ATM.DYN run results are similar with the CPL run results and thus are not shown. To validate the simulation results, the time series in ERA5 data and the daily observed high/low temperature data from NOAA National Climate Data Center are also plotted. It can be seen that four major heat waves (i.e., June 2nd, 10th, 17th, and 24th) and the T2 variations during the 30-day simulation are all captured by the simulations. Before June 17th (lead time $<$ 16 days), the CPL and ~~WRF~~ATM.STA runs results are in good agreement with the ground observation and ERA5 dataset. The root mean square error (RMSE) between the simulations and ground observation are 2.79 °C and 2.83 °C for CPL and ~~WRF~~ATM.STA runs, respectively. However, the error after June 18th (simulation lead time $>$ 17 days) is larger for both CPL (3.42 °C) and ~~WRF~~ATM.STA (3.94 °C) runs. It can be also seen that the CPL run better captures the daily high temperatures in Yanbu (RMSE difference: 2.77 °C) than ERA5 dataset (RMSE: 5.59 °C), which is probably because ERA5 uses a lower resolution grid and is unable to capture the T2 in the coastal city. This is one of the advantages when employing ~~high-resolution regional simulations~~ regional simulations using higher resolution. It should be mentioned that both the present simulations and ERA5 dataset reported a T2 that is 4.5 °C lower than observed T2 in Mecca on June 2nd, though the heat wave events in the other cities are still captured. This may be due to the errors in initial conditions, or WRF ~~micro-physics models~~ physics schemes (e.g., land surface model, the PBL model) are unable to parameterize this extreme event. It can be also seen in the results that taking into account ocean–atmosphere coupling can improve the simulation of T2 in the CPL run. In Fig. 6, the CPL run can better reproduce the evolution of the T2 compare to ATM.STA run during the 30-day simulation: the CPL run better captures the daily high/low temperature in Yanbu and Jeddah (RMSE: 2.69 and 2.81 °C) than ATM.STA run (RMSE: 3.04 and 3.28 °C). However, the difference of T2 in Mecca is negligible (0.05 °C) between CPL and ATM.STA runs~~.~~, shown in Fig. 7. We hypothesize that Mecca is much further away from the Red Sea than Yanbu and Jeddah, which indicates that the influence of air–sea coupling is strong near the coast.

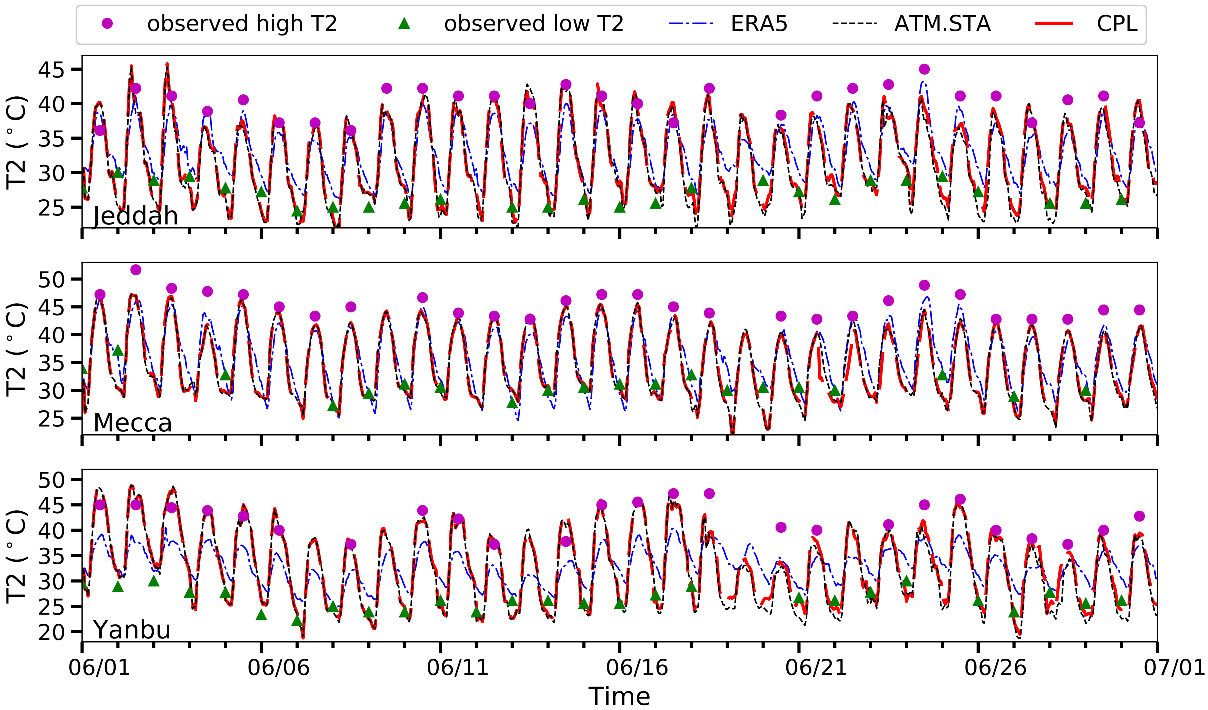

Figure 6. ~~Comparison of~~ Temporal variation the surface air temperature at three major cities near the eastern shore of Red Sea (Jeddah, Mecca, Yanbu) as resulting from CPL and ~~WRF~~ATM.STA runs. The ATM.DYN run results are similar with the CPL run results and thus are not shown. The temperature data are compared with the time series in ERA5 dataset and daily high/low temperature in the NOAA national data center dataset. Note that some surface air temperature data gaps exist in the NCDC ground observation dataset.

The simulation error of T2 also oscillates diurnally in the present simulations. To demonstrate the diurnal variation of the simulation error quantitatively, the mean ~~deviation~~ bias and RMSE of T2 between the simulations (i.e., ATM.STA, ATM.DYN, and CPL) and ERA5 data are shown in Fig. 8. To highlight the air–sea interactions in the simulations, only the temperature over the Red Sea is compared. It can be seen in Fig. 8 that the ATM.STA run using the static SST can still capture the T2 patterns in the first week, but it under-predicts T2 by 2.5 °C because of ignoring the SST evolution. On the other hand, CPL run has

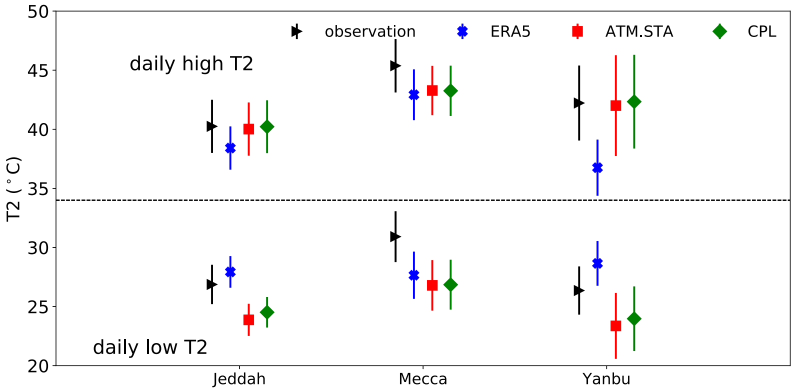

Figure 7. The mean and standard deviation of the surface air temperature (T2) at three major cities near the eastern shore of Red Sea (Jeddah, Mecca, Yanbu) as resulting from CPL and ATM.STA runs. Both daily high and low T2 are presented. The ATM.DYN run results are similar with the CPL run results and thus are not shown. The T2 data in all simulations are not used if they are missing in NCDC ground observation.

much smaller ~~mean deviation~~ bias (-0.49 °C) and root mean square error (1.46 °C) compared with those in ATM.STA run (~~deviation~~bias: -1.34 °C; RMSE: 2.04 °C) during the 30-day simulation as the SST evolution is considered. The ATM.DYN run uses the prescribed SST and its results are consistent with those in CPL run (~~deviation~~bias: -0.58 °C; RMSE: 1.40 °C), indicating that the coupled model captures the SST revolution. The ~~mean deviation~~ bias and RMSE of T2 in the present work are similar to those in the benchmark WRF-ARW simulations (Xu et al., 2009; Zhang et al., 2013a). The differences ~~between the present~~ of the mean bias and RMSE between the simulations and ERA5 data are ~~plotted below the mean deviation and RMSE~~ also plotted to demonstrate the improvement of the CPL run over ATM.STA and ATM.DYN runs. It can be seen that the CPL run captures improved T2 patterns in both ~~mean deviation~~ bias and RMSE than the ATM.STA run throughout the entire simulation. The ~~deviation~~ bias and RMSE between CPL run and ATM.DYN runs are consistent within 0.5 ~~o~~°C. This demonstrates the capability of the coupled model in performing realistic regional ocean–atmosphere simulations.

4.2 Sea Surface Temperature

The simulated SST patterns are compared to the validation data to demonstrate the performance of the coupled model in capturing the ocean surface state. The daily SST fields from CPL run on June 2nd and 24th are shown in Fig. 9(I) and Fig. 9(VI). To validate the CPL run results, the SST fields obtained in OCN.DYN runs are shown in Fig. 9(II) and 9(VII) and the GHRSST fields are shown in Fig. 9(III) and 9(VIII). It can be seen that both OCN.DYN and CPL runs are able to reproduce the SST patterns reasonably well in comparison with the satellite observations. Though the CPL run uses the surface forcing fields with a higher resolution, the SST patterns obtained in both simulations are very similar after two days. On June 24th, the SST patterns in both runs are less similar, but both simulation results are still consistent with GHRSST (RMSE $<$ 1°C). Both simulations under-estimate the SST in the northern Red Sea. The CPL run over-estimates the SST in the central and southern Red Sea on

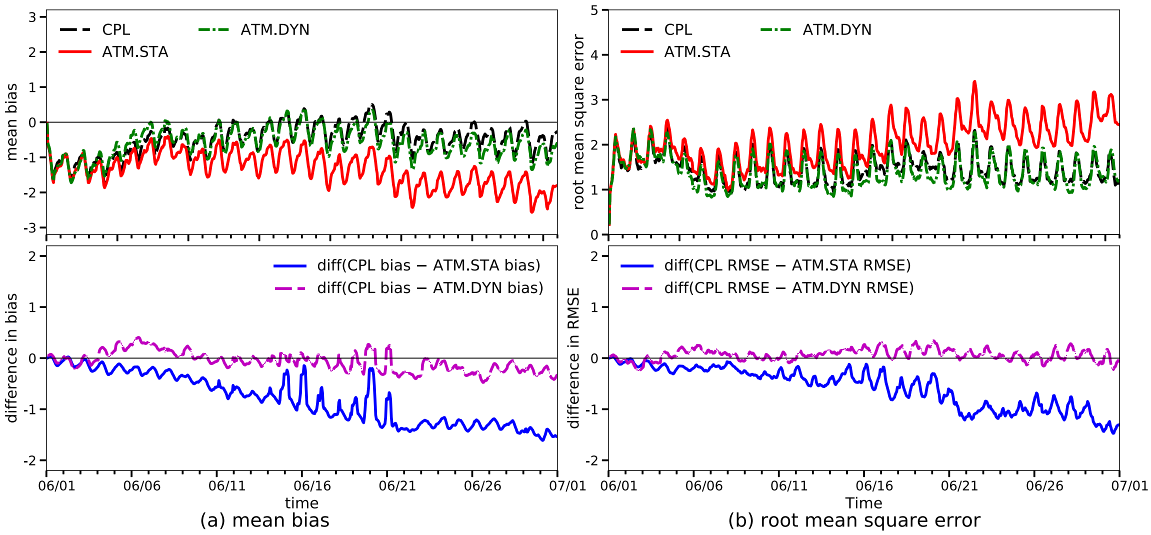

Figure 8. The ~~mean deviation~~ bias and root mean square error (RMSE) between the surface air temperature obtained by the simulations (i.e., ATM.STA, ATM.CPL, and CPL) in comparison with ERA5 data. Only the errors over the Red Sea are considered. The differences between the simulation errors from CPL run and stand-alone WRF simulations are presented below the mean ~~deviation~~ bias and the root mean square error. The initial time is 0000 UTC Jun 01 2012 for all simulations.

June 24th, while the OCN.DYN run under-estimates the SST in the central Red Sea. ~~The difference may be because the CPL run uses the cloud information from the atmosphere component when calculating the surface radiation fluxes, although the uncertainty in the cloud modeling can be significant.~~

To quantitatively compare the errors in SST results, the time history of the SST in the simulations (i.e., OCN.DYN and CPL) and validation datasets (i.e., GHRSST and HYCOM data) are shown in Fig. 10. The mean ~~deviation~~ bias and RMSE between simulation results and validation datasets are also plotted. Again, only the errors between daily SST fields are presented because both observational datasets only provide daily data. It can be seen in Fig. 10 that the ~~mean deviation~~ bias and RMSE of SST in CPL run (~~deviation~~bias: -0.26 °C; RMSE: 0.74 °C) is smaller than that of T2 (~~deviation~~bias: -0.47 °C; RMSE: 1.42 °C) shown in Fig. 8. Generally, the OCN.DYN and CPL runs have a similar range of error compared to both validation datasets, ~~as both simulations are driven by realistic atmospheric forcing~~which shows the skill of the coupled model in simulating the ocean SST. Compared with the HYCOM dataset, the ~~mean deviations~~ bias of CPL and OCN.DYN runs are small (CPL: -0.12 °C; OCN.DYN: -0.04 °C) before June 10th. After June 11th, the CPL run slightly over-estimated the SST (0.37 °C), but the OCN.DYN run slightly under-estimated it (-0.05 °C). In addition, the RMSEs of both simulations increase in the first 10 days, but the increase is not significant after that. On the other hand, when comparing with the GHRSST, the initial SST patterns in both runs are cooler by 0.8 °C. This is because the HYCOM data is cooler than GHRSST at the start of the simulation. After the first 10 days, the difference between GHRSST data and HYCOM decreases, and likewise the difference between the simulation results and GHRSST also decreases. Before June 10th, both CPL and ATM.STA runs under-estimated the SST

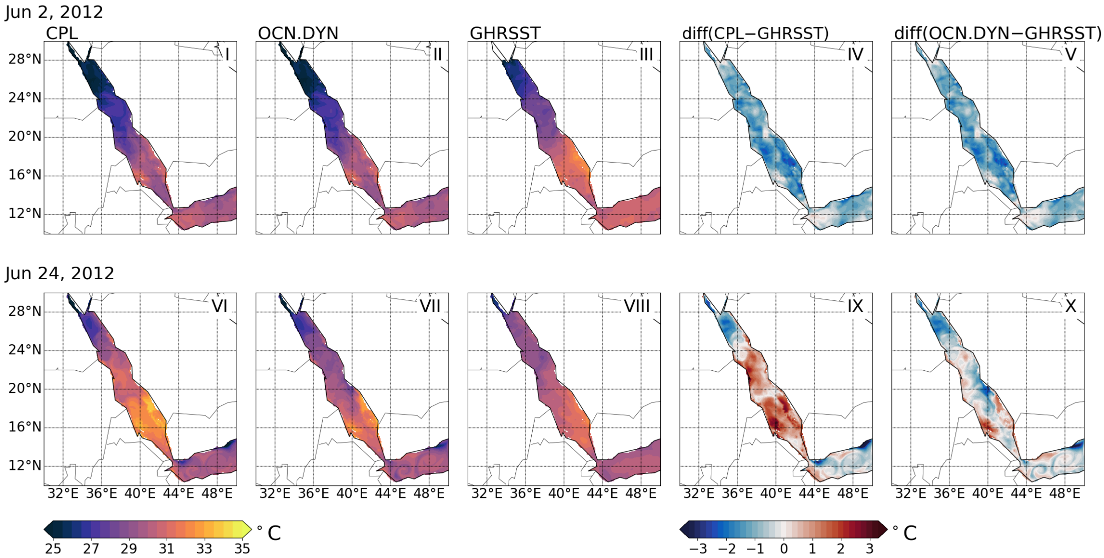

Figure 9. The daily SST patterns obtained by OCN.DYN and CPL runs, and GHRSST dataset. The corresponding differences between the simulations and the GHRSST dataset are also plotted. Two snapshots are selected: (1) Jun 02 2012; (2) Jun 24 2012. The simulation initial time is 0000 UTC Jun 01 2012 for both snapshots.

(CPL: -0.73 °C; OCN.DYN: -0.66 °C). It should be noted that the mean SST in CPL run (-0.01 °C) is closer to GHRSST than OCN.DYN (-0.34 °C) after June 11^{th}.

4.3 Surface Heat Fluxes

The surface heat budget strongly influences the forecast of the surface temperature fields in the simulations. Here we evaluate the performance of the coupled model in capturing the heat fluxes, as compared to the stand-alone simulations. The results are also compared to the MERRA-2 dataset and their differences are plotted.

The turbulent heat fluxes (THF), including the latent heat and sensible heat, and their differences with the validation dataset are shown in Fig. 11. The snapshots of the turbulent fluxes in the heat wave events on June 2^{nd} and 24^{th} are presented. It can be seen that all simulations reproduce the turbulent heat fluxes reasonably well in comparison with the MERRA-2 dataset. On June 2^{nd}, all simulations exhibit similar THF patterns since they have the same initial conditions and air–sea interactions do not significantly impact the THF within two days. On the other hand, for the heat wave event on June 24^{th}, CPL and ATM.DYN runs exhibit more latent heat fluxes coming out of the ocean (157 and 131 W/m^2) than that in ATM.STA run (115 W/m^2). ~~This is because the~~ The mean biases in ATM.STA~~run is forced by a cooler SST pattern~~, ATM.DYN, and CPL runs are -9.8 w/m2, 5.9 w/m2, and 31.8 w/m2, respectively. This is because the SST fields in stand-alone WRF runs are cooler compared with CPL run. When forced by cooler SST, the evaporation decreases and thus the latent heat is smaller. Compared with the

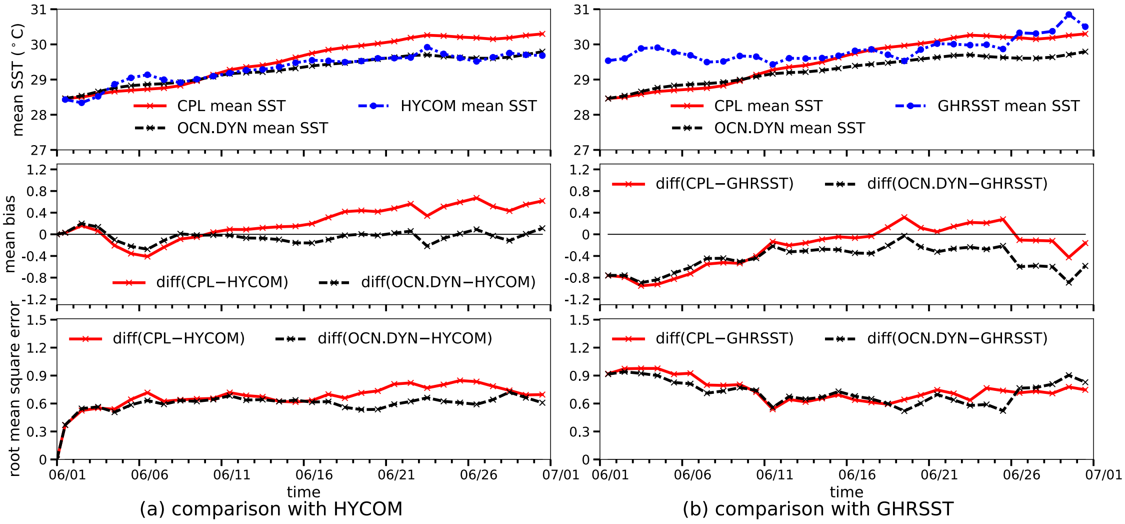

Figure 10. The ~~mean deviation~~ bias and mean-root-square-error between the daily SST as resulting from the simulations (i.e., OCN.DYN and CPL) in comparison with the observational dataset. Panel (a) shows the comparison with HYCOM dataset and Panel (b) shows the comparison with GHRSST dataset. The initial time is 0000 UTC Jun 01 2012 for all simulations.

latent heat, the sensible heat in the Red Sea region is much smaller in all simulations (10 W/m^2). It should be noted that the MERRA-2 dataset has unrealistically large sensible heat in the coastal regions because its resolution is not adequate to resolve the coastline in the Red Sea region (Gelaro et al., 2017).

The net downward shortwave and longwave heat fluxes are shown in Fig. 12. Again, all simulations reproduce the shortwave and longwave radiation fluxes reasonably well. For the shortwave heat flux, all simulations show similar patterns on both June 2nd and 24th as the air–sea interactions do not significantly impact the solar radiation. 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 ~~can be observed in the longwave radiation~~ on June 24th. This is because these two simulations are ~~forced~~ driven by realistic SST and thus can capture longwave radiation according to the bulk formula. The total downward heat fluxes, which is the sum of the results in Figs. 11 and 12, are shown in Fig. 13. It can be seen that the present simulations over-estimated the total downward heat fluxes (CPL: 646 W/m^2; ATM.STA: 674 W/m^2; ATM.DYN: 663 W/m^2) for both heat wave events compared with MERRA-2 dataset (495 W/m^2), especially in the central Red Sea, the southern Red Sea and the coastal regions. In the central and southern Red Sea, the over-estimation is due to the discrepancies in shortwave solar radiation. To improve the forecast of shortwave radiation, a better understanding of the cloud and aerosol in the Red Sea region is required. In the coastal region, the discrepancy is because MERRA-2 data are only available on a lower resolution grid and do not resolve heat fluxes in the coastal regions. It should be noted that ATM.STA run has the largest discrepancy on June 24th when using a constant SST field. Overall, the present CPL simulations are capable of well capturing all the components of the surface heat fluxes during the heat wave events.

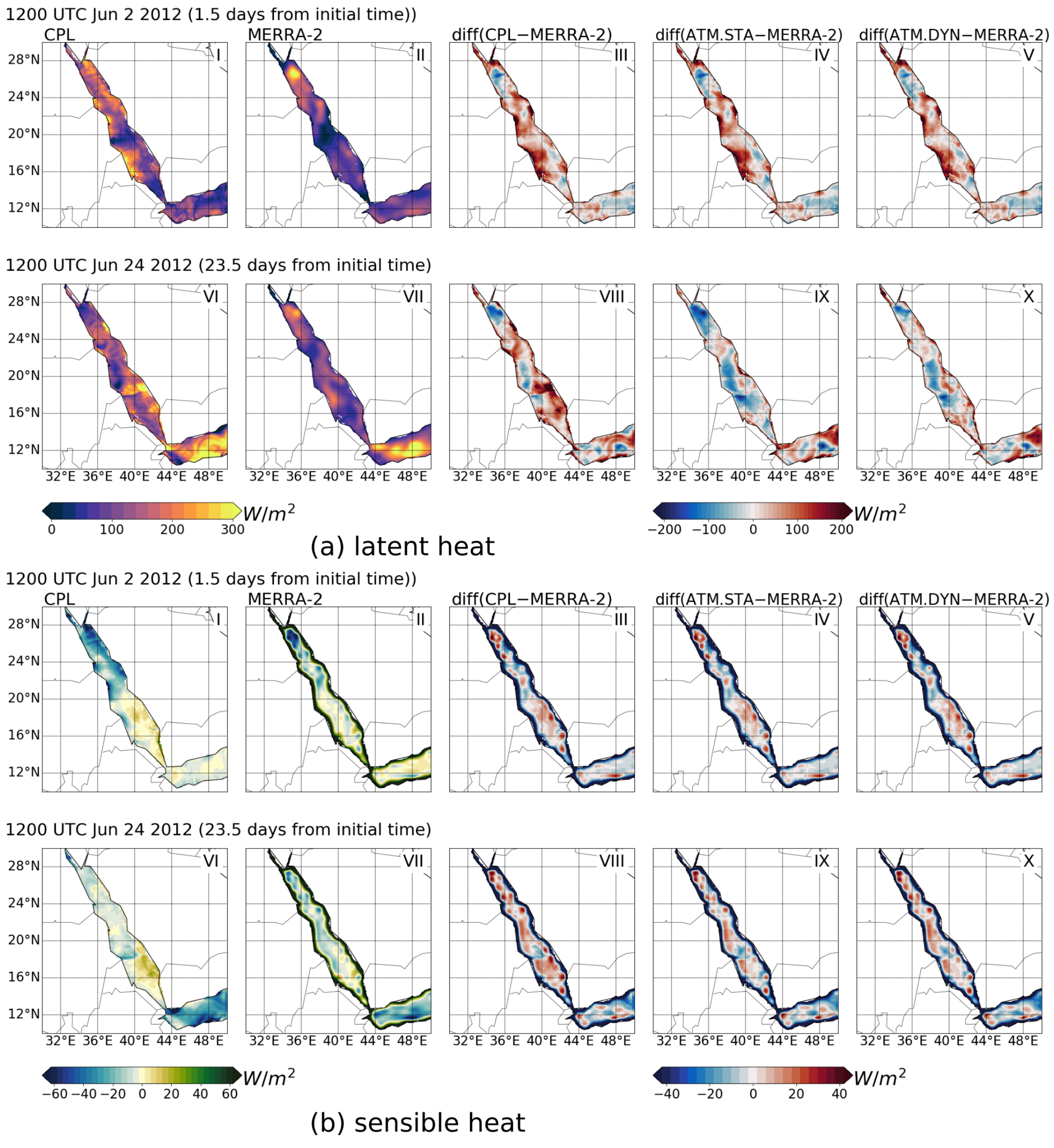

Figure 11. The turbulent heat fluxes out of the sea obtained in CPL run, MERRA-2 data, and their difference (CPL−MERRA-2). The difference between ATM.STA and ATM.DYN with the MERRA-2 data (i.e., ATM.STA−MERRA-2, ATM.DYN−MERRA-2) are also presented. Two snapshots are selected: (1) 1200 UTC Jun 02 2012; (2) 1200 UTC Jun 24 2012. The simulation initial time is 0000 UTC Jun 01 2012 for both snapshots. Only the heat fluxes over the sea is shown to highlight the air–sea interactions.

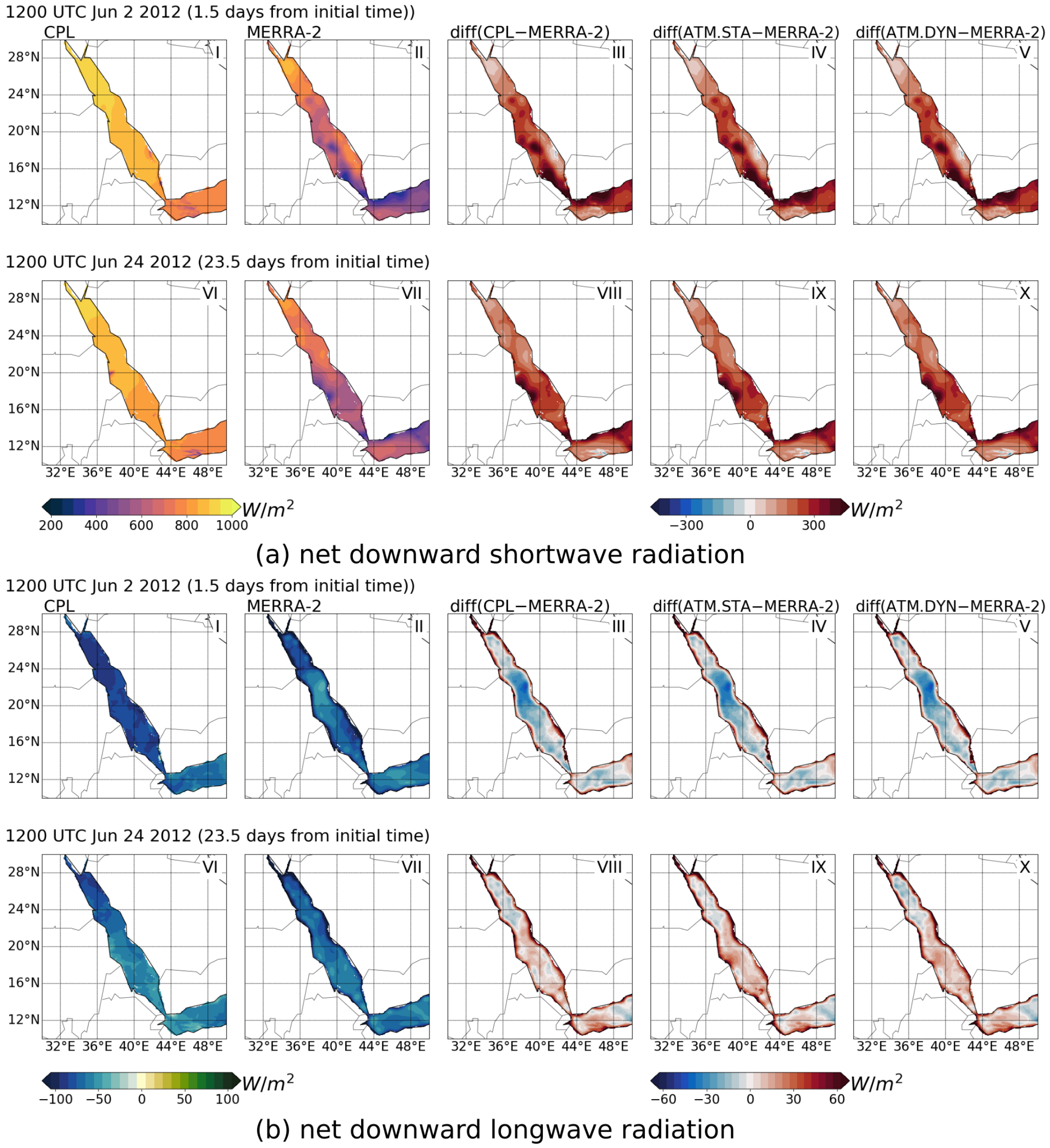

Figure 12. The net downward shortwave and longwave heat fluxes obtained in CPL run, MERRA-2 data, and their difference (CPL−MERRA-2). The difference between ATM.STA and ATM.DYN with the MERRA-2 data (i.e., ATM.STA−MERRA-2, ATM.DYN−MERRA-2) are also presented. Two snapshots are selected: (1) 1200 UTC Jun 02 2012; (2) 1200 UTC Jun 24 2012. The simulation initial time is 0000 UTC Jun 01 2012 for both snapshots. Only the heat fluxes over the sea is shown to highlight the air–sea interactions.

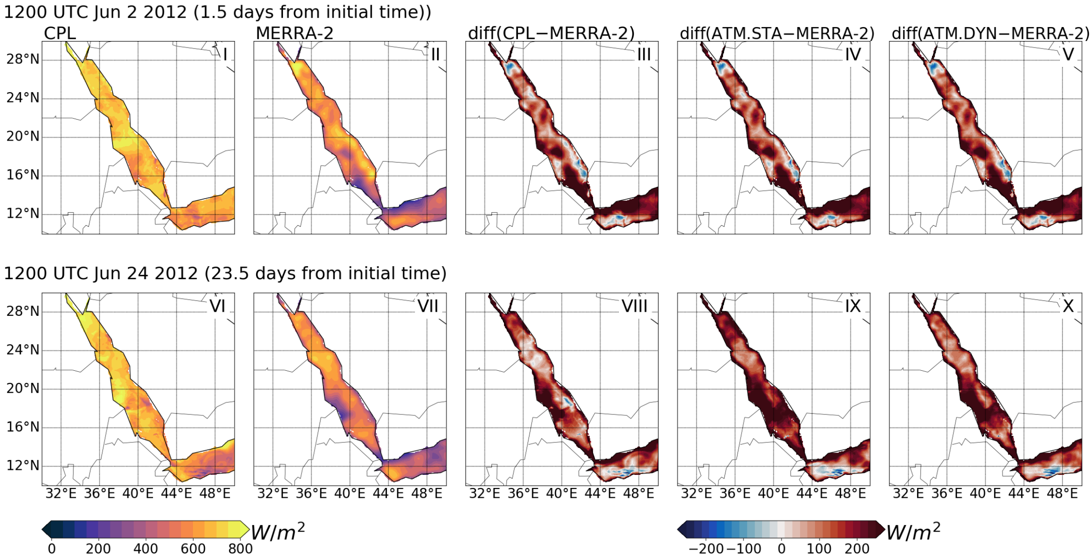

Figure 13. Comparison of the total downward heat fluxes obtained in CPL run, MERRA-2 data, and their difference (CPL−MERRA-2). The difference between ATM.STA and ATM.DYN with the ERA5 data (i.e., ATM.STA−MERRA-2, ATM.DYN−MERRA-2) are also presented. Two snapshots are selected: (1) 1200 UTC Jun 02 2012; (2) 1200 UTC Jun 24 2012. The simulation initial time is 0000 UTC Jun 01 2012 for both snapshots. Only the heat fluxes over the sea is shown to highlight the air–sea interactions.

4.4 Surface Wind and Evaporation

To evaluate the simulation of the surface momentum and freshwater fluxes by the coupled model, the surface wind and evaporation patterns obtained from ATM.STA, ATM.DYN, and CPL runs are presented. The MERRA-2 data is used to validate the simulation results.

The simulated surface wind velocity fields are shown in Fig. 14. ~~They show that~~ The RMSE of the wind velocity magnitude ~~and direction in~~ between the CPL run ~~agree well (RMSE:~~ and MERRA-2 data is 2.17 m/s ~~) with the MERRA-2 data~~ when using the selected ~~micro-physics schemes .~~ WRF physics schemes presented in Section 3. On June 2^{nd}, high-speed wind is observed in the northern and central Red Sea, and the ~~high-resolution~~ CPL run successfully captures the small-scale features of wind speed patterns. On June 24^{th}, the differences between the simulations are larger than those on June 2^{nd}, especially in the central Red Sea and the southern Arabian Peninsula. It should be mentioned that although the SST in the ATM.STA run is lower than the ~~other simulations, the difference~~ CPL run, the RMSE in the wind velocity magnitude is small ~~(RMSE,~~ than 1 m/s (June 2^{nd}: 0.15 m/s; June 24^{th}: 0.74 m/s). ~~This suggests that the ocean–atmosphere coupling does not significantly influence the wind field in the Red Sea region during the heat wave events.~~

The surface evaporation results are shown in Fig. 15. All simulations reproduce the overall evaporation patterns in the Red Sea. The CPL run is able to capture the relatively high evaporation in the northern Red Sea and the relatively low evaporation

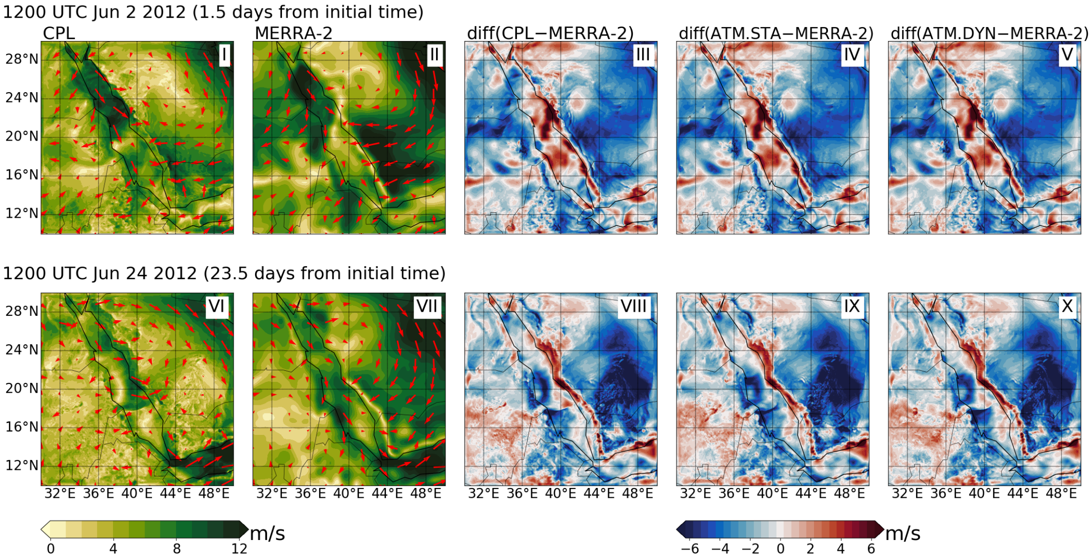

Figure 14. The magnitude and direction of the surface wind obtained in the CPL run, the MERRA-2 data, and their difference (CPL−MERRA-2). The difference between ATM.STA and ATM.DYN with the MERRA-2 data (i.e., ATM.STA−MERRA-2, ATM.DYN−MERRA-2) are also presented. Two snapshots are selected: (1) 1200 UTC Jun 02 2012; (2) 1200 UTC Jun 24 2012.~~Only the heat fluxes over the sea is shown to highlight the air–sea interactions.~~

in the southern Red Sea in both snapshots, shown in Fig. 15(I) and 15(VI). Again, all simulation results are consistent on June 2nd because they are driven by the same initial condition~~and the air–sea interactions do not significantly influence the evaporation fields within two days~~. However, after 24 days, the CPL run agrees better with MERRA-2 dataset (~~deviation~~bias: 4 cm/year; RMSE: 64 ~~/~~cm/year) than the ATM.STA run (~~deviation~~bias: -34 cm/year; RMSE: 69 cm/year) by better reproducing the realistic ocean–atmosphere coupling. Although the CPL run results are ~~still~~ consistent with that of the ATM.DYN run, the ~~CPL run~~ coupled model over-estimates the evaporation in the southern Red Sea. This is because the CPL run ~~slightly~~ over-estimated the SST than the ATM.DYN run, shown in Fig. 9(IX). Since there is no precipitation in three major cities (Mecca, Jeddah, Yanbu) near the eastern shore of the Red Sea during the month according to NCDC climate data, the precipitation results are not shown.

5 Scalability Test

~~The parallel~~ Parallel efficiency is crucial for coupled ocean–atmosphere models for simulating large and complex problems. In this section, the parallel efficiency in the coupled simulations is investigated~~and presented. We investigate the scalability~~. This aims to demonstrate the implemented ESMF/NUOPC driver and model interfaces are able to simulate parallel cases effectively. The parallel speed-up of the model is investigated to evaluate its performance for a constant ~~sized~~ size problem simulated using

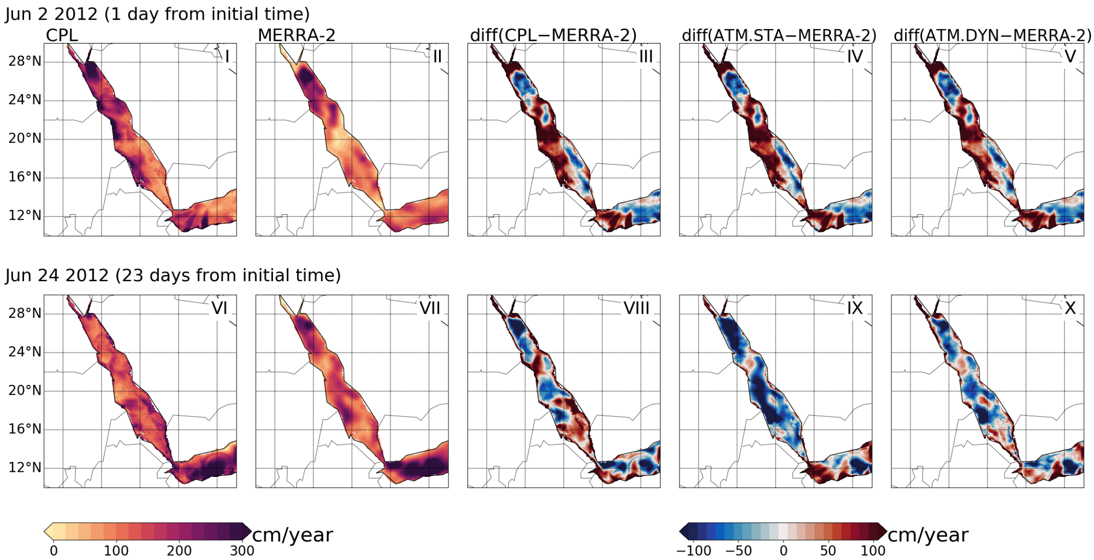

Figure 15. The surface evaporation patterns obtained in the CPL run, the MERRA-2 data, and their difference (CPL−MERRA-2). The difference between ATM.STA and ATM.DYN with the MERRA-2 data (i.e., ATM.STA−MERRA-2, ATM.DYN−MERRA-2) are also presented. Two snapshots are selected: (1) 1200 UTC Jun 02 2012; (2) 1200 UTC Jun 24 2012. Only the evaporations over the sea is shown to highlight the air–sea interactions.

different numbers of processors ~~.~~ (i.e. strong scaling). Additionally, the CPU time spent on different parts of the coupled model is detailed. The parallel efficiency tests are performed on the COMPAS (Center for Observations, Modeling and Prediction at Scripps) cluster in Scripps Institution of Oceanography (http://www.compas.ucsd.edu/). The COMPAS cluster is composed of 1192 Intel 5400 and 5500 series CPUs and has a theoretical peak speed of 12.6 TeraFlops. The cluster uses Myrinet for its high-performance network.

The parallel efficiency of the scalability test is $N_{p0}t_{p0}/N_{pn}t_{pn}$, where N_{p0} and N_{pn} are the number of processors employed in the simulation of the baseline case and the test case, respectively; t_{p0} and t_{pn} are the CPU time. The speed-up is defined as t_{p0}/t_{pn}, which is the relative improvement of the CPU time when solving the problem. The scalability tests are performed by runing 6-hour simulations for ATM.STA, OCN.DYN, and CPL cases. The results obtained in the scalability test of the coupled model are shown in Fig. 16. It can be seen that the parallel efficiency is close to 100% when employing less than ~~64~~ 128 processors and is still as high as ~~65~~70% when using 256 processors. ~~The decrease in parallel efficiency is because when~~ When using 256 processors, there are ~~only~~ 20480 cells (16 lat×16 ~~grid points in the horizontal plane. The boundary tiles in each processorare~~ lon×80 vertical levels) in each processor, but there are 5120 overlap cells (4 sides×16 tiles per side×80 vertical levels), which is 25% of the total ~~, and the parallel communication cost increases significantly. From the~~ cells. From results reported in ~~the~~ previous literature, the parallel efficiency of the coupled model is comparable to other ~~models~~ocean-alone or atmosphere-alone models when having similar number of grid points per processor (Marshall et al., 1997; Zhang et al., 2013b). 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 ~~attributed to~~ because of the fluctuation of the ~~CPU timewhen 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. This fluctuation may also be due to the variation of CPU cache or memory.~~ 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.

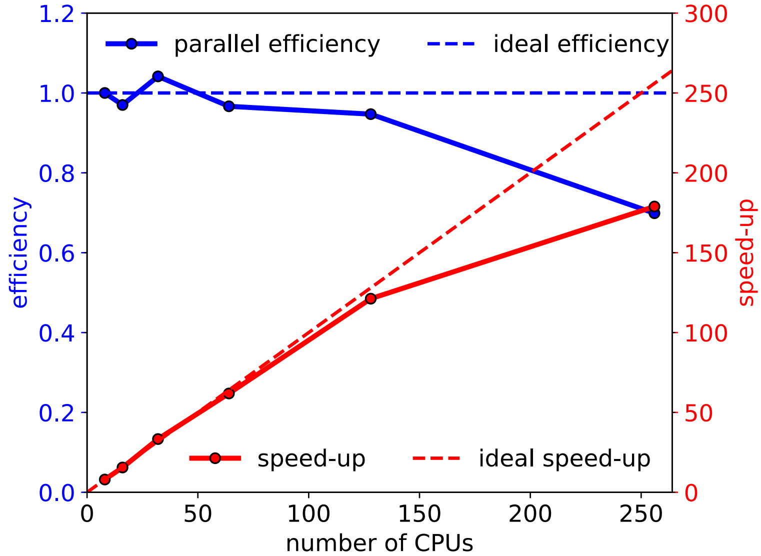

Figure 16. The parallel efficiency test of the coupled model in the Red Sea region. The test cases employ up to 256 CPU cores. The simulation with the smallest case is regarded as base case when computing the speed-up. Tests are performed on the COMPAS cluster in ~~UCSD~~Scripps Institution of Oceanography.

The CPU time spent on coupled ~~run~~ and stand-alone runs is shown in Table. 3. The time spent on the coupler is estimated by subtracting the time spent on stand-alone simulations from the coupled run. The most time-consuming process is the atmospheric model integration, which accounts for 76% to 93% of the total costs. The ocean model integration is the second most time-consuming process, which is 7% to 14% of the total computational costs. The atmospheric model is much more time-consuming ~~than the ocean model because it~~ 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. ~~Moreover, the atmospheric model solves the atmosphere in the entire computational domain, while the ocean model only solves the Red Sea~~If a purely marine region is selected in an ideal case, the cost of ocean and atmosphere models would be more equal. The coupling process takes less than

5% of the total costs when using fewer than ~~64 processors~~ 128 processors (40960 grid points per processor). However, when using 256 processors (20480 grid points per processor), the proportion of this cost increases to 10%~~because of the increase of inter-processor communication with more processors~~, though the amount of time spent on the ESMF/NUOPC coupler is similar with using 128 processors. We hypothesis that the cost of the ESMF/NUOPC coupler is communication cost and it becomes important as the amount of computation work is reduced with the number of grid cells in these strong scaling tests. In summary, the scalability test results ~~demonstrate the coupled model can be applied for high-resolution~~ suggest that the ESMF/NUOPC coupler will not be a bottleneck for using SKRIPS in coupled regional modeling studies.

Table 3. Comparison of CPU time spent on the coupled run and stand-alone simulations. The CPU times presented here are normalized by the time spent on the coupled run using 256 processors. The CPU time spent on the ESMF/NUOPC coupler is obtained by subtracting two stand-alone simulation time from the CPL run time.

	$N_p = 8$	16	32	64	128	256
CPL ~~run~~	22.36	11.52	5.37	2.89	1.48	1.00
~~WRF stand-alone~~ ATM.STA	20.42(91%)	10.41(90%)	4.97(93%)	2.57(89%)	1.27(86%)	0.76(76%)
~~MITgcm stand-alone~~ OCN.DYN	1.76(8%)	0.93(8%)	0.36(7%)	0.20(7%)	0.14(9%)	0.14(14%)
ESMF/NUOPC coupler	0.17(1%)	0.18(2%)	0.03(1%)	0.11(4%)	0.07(5%)	0.10(10%)

6 Conclusion and Outlook

This study describes the development of ~~a regional coupled ocean–atmosphere numerical framework (MITgcm–WRF)based on the ESMF coupler, with an example of a specific~~ the Scripps–KAUST Regional Integrated Prediction System (SKRIPS). To build the coupled model, the ESMF coupler is implemented according to NUOPC consortium. The ocean model MITgcm and the atmosphere model WRF are split into initialize, run, and finalize sections, with each of them being called as subroutines of the main function.

The development activities has been focused on providing a useful coupled model for realistic application to simulate the heat wave events in the Red Sea region. Results from the coupled and stand-alone simulations are compared to a wide variety of available observational and reanalysis datasets, aiming to demonstrate the overall performance of the coupled model with respect to stand-alone models. The results obtained from various configurations of coupled and stand-alone model simulations all realistically capture the basic characteristics of the ocean–atmosphere state in the Red Sea region over a 30-day simulation period. The surface air temperature variations in three major cities are consistent with the ground observations and the heat wave events are also well captured in the CPL run. The surface flux fields (e.g., surface air temperature, surface heat fluxes, surface evaporations, surface wind) in the CPL run are consistent with the reanalysis data over the simulation period. The SST fields in CPL run are also consistent with the satellite observation data. 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. ~~On the other hand, the difference between coupled simulation and stand-alone simulations with updated forcings is also discussed.~~

The parallel efficiency of the coupled model is examined by simulating the Red Sea region using increasing number of processors. The coupled model scales linearly for up to 128 CPUs and the parallel efficiency remains about 70% for 256 processors. The CPU time associated with different parts of the coupled simulations is also presented, suggesting good parallel efficiency in both model components and ESMF coupler. Hence the coupled model can be applied for high-resolution coupled regional modeling studies on massively parallel processing supercomputers.

These preliminary results motivate further studies in evaluating and improving this new regional ~~high-resolution~~ coupled ocean–atmosphere model for investigating dynamical processes and forecasting applications in regions around the globe where ocean–atmosphere coupling is important. This regional coupled model can be further improved by developing coupled data assimilation capabilities on initializing coupled forecasts from an assimilated ~~high-resolution~~ analysis state. In addition, the model physics and model uncertainty representation in the coupled system can be enhanced using advanced techniques, such as stochastic physics parameterizations. Future work will involve exploring these and other aspects of developing a ~~high-resolution regional~~ regional coupled modeling system that is best suited for forecasting and process understanding purposes.

Code and data availability. The coupled code, documentation, and tutorial cases used in this work are available at https://github.com/iurnus/scripps_kaust_model. ECMWF ERA5 dataset is used as the atmospheric initial and boundary conditions. The ocean model uses the assimilated HYCOM/NCODA $1/12°$ global analysis data as initial and boundary conditions. To validate the simulated SST data, we use the OSTIA (Operational Sea Surface Temperature and Sea Ice Analysis) system in GHRSST (Group for High Resolution Sea Surface Temperature). The simulated 2-meter air temperature (T2) is validated against the ECMWF ERA5 dataset. The observed daily maximum and minimum temperatures from NOAA National Climate Data Center is used to validate the T2 in three major cities. Surface heat fluxes (e.g., latent heat, sensible heat, longwave and shortwave radiations), which are important for ocean–atmosphere interactions, are compared with MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, version 2) datasets.

Author contributions. RS worked on the coding tasks for coupling WRF with MITgcm using ESMF, wrote the code documentation, and performed the simulations for the numerical experiments. RS and ACS worked on the technical details for debugging the model and drafted the initial manuscript. All authors designed the computational framework and the numerical experiments. All authors discussed the results and contributed to the writing of the final manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We appreciate the computational resources provided by COMPAS (Center for Observations, Modeling and Prediction at Scripps) used for this project. We gratefully acknowledge the research funding from KAUST (grant number: OSR-2-16-RPP-3268.02). We are immensely grateful to Caroline Papadopoulos for helping with installing software, testing the coupled code, and using the COMPAS cluster. We appreciate Professor U.U. Turuncoglu sharing part of their ESMF/NUOPC code on GitHub which helped our code development. We wish to thank Dr. Ganesh Gopalakrishnan for setting up the stand-alone MITgcm simulation (OCN.DYN) and providing the external forcings. We ~~also~~ thank Drs. Stephanie Dutkiewicz, Jean-Michel Campin, Chris Hill, Dimitris Menemenlis ~~in~~ for providing their ESMF–MITgcm ~~coupling~~interface. We also wish to thank Dr. Peng Zhan for discussing the simulations of the Red Sea. We also thank the reviewers for their insightful review suggestions.

References

Abdou, A. E. A.: Temperature trend on Makkah, Saudi Arabia, Atmospheric and Climate Sciences, 4, 457–481, 2014.

Aldrian, E., Sein, D., Jacob, D., Gates, L. D., and Podzun, R.: Modelling Indonesian rainfall with a coupled regional model, Climate Dynamics, 25, 1–17, 2005.

Barbariol, F., Benetazzo, A., Carniel, S., and Sclavo, M.: Improving the assessment of wave energy resources by means of coupled wave-ocean numerical modeling, Renewable Energy, 60, 462–471, 2013.

Bender, M. A. and Ginis, I.: Real-case simulations of hurricane–ocean interaction using a high-resolution coupled model: effects on hurricane intensity, Monthly Weather Review, 128, 917–946, 2000.

Benjamin, S. G., Grell, G. A., Brown, J. M., Smirnova, T. G., and Bleck, R.: Mesoscale weather prediction with the RUC hybrid isentropic–terrain-following coordinate model, Monthly Weather Review, 132, 473–494, 2004.

Boé, J., Hall, A., Colas, F., McWilliams, J. C., Qu, X., Kurian, J., and Kapnick, S. B.: What shapes mesoscale wind anomalies in coastal upwelling zones?, Climate dynamics, 36, 2037–2049, 2011.

Chen, S. S. and Curcic, M.: Ocean surface waves in Hurricane Ike (2008) and Superstorm Sandy (2012): Coupled model predictions and observations, Ocean Modelling, 103, 161–176, 2016.

Chen, S. S., Price, J. F., Zhao, W., Donelan, M. A., and Walsh, E. J.: The CBLAST-Hurricane program and the next-generation fully coupled atmosphere–wave–ocean models for hurricane research and prediction, Bulletin of the American Meteorological Society, 88, 311–318, 2007.

Christidis, Z.: Performance and Scaling of WRF on Three Different Parallel Supercomputers, in: International Conference on High Performance Computing, pp. 514–528, Springer, 2015.

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

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The operational sea surface temperature and sea ice analysis (OSTIA) system, Remote Sensing of Environment, 116, 140–158, 2012.

Doscher, R., Willén, U., Jones, C., Rutgersson, A., Meier, H. M., Hansson, U., and Graham, L. P.: The development of the regional coupled ocean-atmosphere model RCAO, Boreal Environment Research, 7, 183–192, 2002.

Evangelinos, C. and Hill, C. N.: A schema based paradigm for facile description and control of a multi-component parallel, coupled atmosphere-ocean model, in: Proceedings of the 2007 symposium on component and framework technology in high-performance and scientific computing, pp. 83–92, ACM, 2007.

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

Fang, Y., Zhang, Y., Tang, J., and Ren, X.: A regional air-sea coupled model and its application over East Asia in the summer of 2000, Advances in Atmospheric Sciences, 27, 583–593, 2010.

Fowler, H. and Ekström, M.: Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes, International Journal of Climatology, 29, 385–416, 2009.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, 2017.

Gualdi, S., Somot, S., Li, L., Artale, V., Adani, M., Bellucci, A., Braun, A., Calmanti, S., Carillo, A., Dell'Aquila, A., et al.: The CIRCE simulations: regional climate change projections with realistic representation of the Mediterranean Sea, Bulletin of the American Meteorological Society, 94, 65–81, 2013.

Gustafsson, N., Nyberg, L., and Omstedt, A.: Coupling of a high-resolution atmospheric model and an ocean model for the Baltic Sea, Monthly Weather Review, 126, 2822–2846, 1998.

Hagedorn, R., Lehmann, A., and Jacob, D.: A coupled high resolution atmosphere-ocean model for the BALTEX region, Meteorologische Zeitschrift, 9, 7–20, 2000.

Harley, C. D., Randall Hughes, A., Hultgren, K. M., Miner, B. G., Sorte, C. J., Thornber, C. S., Rodriguez, L. F., Tomanek, L., and Williams, S. L.: The impacts of climate change in coastal marine systems, Ecology letters, 9, 228–241, 2006.

He, J., He, R., and Zhang, Y.: Impacts of air-sea interactions on regional air quality predictions using WRF/Chem v3.6.1 coupled with ROMS v3.7: southeastern US example, Geoscientific Model Development Discussions, 8, 9965–10 009, 2015.

Henderson, T. and Michalakes, J.: WRF ESMF Development, in: 4th ESMF Community Meeting, Cambridge, USA, Jul 21, 2005.

Hersbach, H.: The ERA5 Atmospheric Reanalysis., in: AGU Fall Meeting Abstracts, San Francisco, USA, Dec 12-16, 2016.

Hill, C., DeLuca, C., Balaji, Suarez, M., and Silva, A.: The architecture of the Earth system modeling framework, Computing in Science & Engineering, 6, 18–28, 2004.

Hill, C. N.: Adoption and field tests of M.I.T General Circulation Model (MITgcm) with ESMF, in: 4th Annual ESMF Community Meeting, Cambridge, USA, Jul 20-21, 2005.

Hodur, R. M.: The Naval Research Laboratory's coupled ocean/atmosphere mesoscale prediction system (COAMPS), Monthly weather review, 125, 1414–1430, 1997.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Monthly weather review, 134, 2318–2341, 2006.

Huang, B., Schopf, P. S., and Shukla, J.: Intrinsic ocean–atmosphere variability of the tropical Atlantic Ocean, Journal of Climate, 17, 2058–2077, 2004.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models, Journal of Geophysical Research: Atmospheres, 113, 2008.

Kain, J. S.: The Kain–Fritsch convective parameterization: an update, Journal of applied meteorology, 43, 170–181, 2004.

Kharin, V. V. and Zwiers, F. W.: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM, Journal of Climate, 13, 3760–3788, 2000.

Large, W. G., McWilliams, J. C., and Doney, S. C.: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization, Reviews of Geophysics, 32, 363–403, 1994.

Li, D., Zou, L., and Zhou, T.: Regional air–sea coupled model simulation for two types of extreme heat in North China, Climate Dynamics, 50, 2107–2120, 2018.

Loglisci, N., Qian, M., Rachev, N., Cassardo, C., Longhetto, A., Purini, R., Trivero, P., Ferrarese, S., and Giraud, C.: Development of an atmosphere-ocean coupled model and its application over the Adriatic Sea during a severe weather event of Bora wind, Journal of Geophysical Research: Atmospheres, 109, 2004.

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

Martin, M., Dash, P., Ignatov, A., Banzon, V., Beggs, H., Brasnett, B., Cayula, J.-F., Cummings, J., Donlon, C., Gentemann, C., et al.: Group for High Resolution Sea Surface Temperature (GHRSST) analysis fields inter-comparisons. Part 1: A GHRSST multi-product ensemble (GMPE), Deep Sea Research Part II: Topical Studies in Oceanography, 77, 21–30, 2012.

Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes, Monthly Weather Review, 137, 991–1007, 2009.

National Geophysical Data Center: 2-minute Gridded Global Relief Data (ETOPO2) v2, 2006.

Powers, J. G. and Stoelinga, M. T.: A coupled air–sea mesoscale model: Experiments in atmospheric sensitivity to marine roughness, Monthly weather review, 128, 208–228, 2000.

Pullen, J., Doyle, J. D., and Signell, R. P.: Two-way air–sea coupling: A study of the Adriatic, Monthly weather review, 134, 1465–1483, 2006.

Ricchi, A., Miglietta, M. M., Falco, P. P., Benetazzo, A., Bonaldo, D., Bergamasco, A., Sclavo, M., and Carniel, S.: On the use of a coupled ocean–atmosphere–wave model during an extreme cold air outbreak over the Adriatic Sea, Atmospheric Research, 172, 48–65, 2016.

Roessig, J. M., Woodley, C. M., Cech, J. J., and Hansen, L. J.: Effects of global climate change on marine and estuarine fishes and fisheries, Reviews in fish biology and fisheries, 14, 251–275, 2004.

Seo, H.: Distinct influence of air–sea interactions mediated by mesoscale sea surface temperature and surface current in the Arabian Sea, Journal of Climate, 30, 8061–8080, 2017.

Seo, H., Miller, A. J., and Roads, J. O.: The Scripps Coupled Ocean–Atmosphere Regional (SCOAR) model, with applications in the eastern Pacific sector, Journal of Climate, 20, 381–402, 2007.

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

Sitz, L., Di Sante, F., Farneti, R., Fuentes-Franco, R., Coppola, E., Mariotti, L., Reale, M., Sannino, G., Barreiro, M., Nogherotto, R., et al.: Description and evaluation of the Earth System Regional Climate Model (RegCM-ES), Journal of Advances in Modeling Earth Systems, 2017.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 2, Tech. rep., National Center For Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology Div, 2005.

Somot, S., Sevault, F., Déqué, M., and Crépon, M.: 21st century climate change scenario for the Mediterranean using a coupled atmosphere–ocean regional climate model, Global and Planetary Change, 63, 112–126, 2008.

Theurich, G., DeLuca, C., Campbell, T., Liu, F., Saint, K., Vertenstein, M., Chen, J., Oehmke, R., Doyle, J., Whitcomb, T., et al.: The earth system prediction suite: toward a coordinated US modeling capability, Bulletin of the American Meteorological Society, 97, 1229–1247, 2016.

Turuncoglu, U., Giuliani, G., Elguindi, N., and Giorgi, F.: Modelling the Caspian Sea and its catchment area using a coupled regional atmosphere-ocean model (RegCM4-ROMS): model design and preliminary results, Geoscientific Model Development, 6, 283, 2013.

Turuncoglu, U. U.: Toward modular in situ visualization in Earth system models: the regional modeling system RegESM 1.1, Geoscientific Model Development, 12, 233–259, 2019.

Turuncoglu, U. U. and Sannino, G.: Validation of newly designed regional earth system model (RegESM) for Mediterranean Basin, Climate dynamics, 48, 2919–2947, 2017.

Van Pham, T., Brauch, J., Dieterich, C., Frueh, B., and Ahrens, B.: New coupled atmosphere-ocean-ice system COSMO-CLM/NEMO: assessing air temperature sensitivity over the North and Baltic Seas, Oceanologia, 56, 167–189, 2014.

Warner, J. C., Armstrong, B., He, R., and Zambon, J. B.: Development of a coupled ocean–atmosphere–wave–sediment transport (COAWST) modeling system, Ocean modelling, 35, 230–244, 2010.

Xie, S.-P., Miyama, T., Wang, Y., Xu, H., De Szoeke, S. P., Small, R. J. O., Richards, K. J., Mochizuki, T., and Awaji, T.: A regional ocean–atmosphere model for eastern Pacific climate: toward reducing tropical biases, Journal of Climate, 20, 1504–1522, 2007.

Xu, J., Rugg, S., Byerle, L., and Liu, Z.: Weather forecasts by the WRF-ARW model with the GSI data assimilation system in the complex terrain areas of southwest Asia, Weather and Forecasting, 24, 987–1008, 2009.

Yao, F., Hoteit, I., Pratt, L. J., Bower, A. S., Köhl, A., Gopalakrishnan, G., and Rivas, D.: Seasonal overturning circulation in the Red Sea: 2. Winter circulation, Journal of Geophysical Research: Oceans, 119, 2263–2289, 2014a.

Yao, F., Hoteit, I., Pratt, L. J., Bower, A. S., Zhai, P., Köhl, A., and Gopalakrishnan, G.: Seasonal overturning circulation in the Red Sea: 1. Model validation and summer circulation, Journal of Geophysical Research: Oceans, 119, 2238–2262, 2014b.

Zhan, P., Subramanian, A. C., Yao, F., and Hoteit, I.: Eddies in the Red Sea: A statistical and dynamical study, Journal of Geophysical Research: Oceans, 119, 3909–3925, 2014.

Zhang, H., Pu, Z., and Zhang, X.: Examination of errors in near-surface temperature and wind from WRF numerical simulations in regions of complex terrain, Weather and Forecasting, 28, 893–914, 2013a.

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

Zou, L. and Zhou, T.: Development and evaluation of a regional ocean-atmosphere coupled model with focus on the western North Pacific summer monsoon simulation: Impacts of different atmospheric components, Science China Earth Sciences, 55, 802–815, 2012.