Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-252-RC1, 2019

Geoscientific Model Development Discussions

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.

Gill (Referee)

gill@ucar.edu

Received and published: 19 January 2019

Comments

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.

General Comments

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.

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.

Specific Comments

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.

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.

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.

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.

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?

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.

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.

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.

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.

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.

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.

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.

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.

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.

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).

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)”

Given that 30 days * 24 hr/day * 60 min/hr * 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.

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.

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.

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."

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."

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

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 âŮęC (CPL), -1.66 âŮęC (ATM.STA), and -1.70 âŮęC (ATM.DYN) after 36 hours, and -0.99 âŮęC (CPL), -1.10 âŮęC (ATM.STA), and -1.12 âŮęC (ATM.DYN) after 48 hours."

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?

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

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?

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

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?

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

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.

Technical Corrections

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

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

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

Figure 5 would benefit from having some highlight that indicated the four heatwave periods.

A number of figures would benefit from smaller color bar ranges. For example 10a has a range from about -500 W/mˆ2 to 500 w/mˆ2. 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".

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

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.

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

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-252, 2018.