Review of *The Fully Coupled Regionally Refined Model of E3SM Version 2: Overview of the Atmosphere, Land, and River*

Qi Tang and Coauthors

General Impressions

This manuscript provides the first documented regional refinement model (RRM) configuration fully coupled to an ocean model. Further, the ocean is regionally-refined as well, although the details are left to a paper currently in preparation. In accordance with the CMIP6 protocol, a suite of DECK experiments were carried out on a grid with 25 km refinement over North America (NARRM). Extensive analysis illustrates clears improvements in the climatology over the standard low resolution (LR) control. For this reason, I believe this work should be documented and communicated to the broader community.

However, I believe major revisions are required before this manuscript is in an acceptable form for publication. The paper is generally too long. It is clear that specialized subgroups were assigned a particular results section and the lead authors were responsible for stringing those sections together into a logical progression and coherent manuscript. I believe more needs to be done in this area. Here are some suggestions:

- Reduce non-results text in Introduction, Methods and Discussions and Conclusions section (generally referring to my a, b complaints below).
- Combine Section 5.1.5 on Cloud Feedback with Section 5.1.3 on North Polar Clouds (also Figure 12 seems like it should be in this combined section as well).
- Combine Section 5.2.1 on Hydrology and 5.4 River, as both sections contains runoff plots suggesting the analysis can be seamlessly combined.

Length and readability issues aside, I have two major concerns (a, b).

(a). Early on the authors develop "two criteria should be satisfied for the RRM to be widely adopted for global ESM releases: (1) reasonable global climate, and (2) minimal effort of retuning based upon the low-resolution counterpart." It's not clear what the authors mean in (2). They describe it as "Regarding (2), some physics parameterizations (e.g., deep convection) suffer from poor scale-awareness and hence require retuning as the model resolution increases (e.g., Xie et al., 2018). This implies significant model calibration efforts that may be unaffordable in addition to tuning the low-resolution model." This last sentence stating that tuning may be unaffordable does not constitute a "criteria" as the authors have chosen to characterize (2). This all depends on the resources available to the modeling institution to engage in an additional tuning effort for RRM — there is no requirement to use "minimal" resources as stated in (2).

Later on the authors state that their proposed solution to (2) "simplifies the RRM development as it naturally avoids further tuning the RRM beyond what was done for LR" (Line 147). This I think is the correct point that authors should flesh out. They simplified the process by using a hybrid time-step and therefore not requiring additional resources for tuning. They could have gone the hard way, full tuning, but they choose to use minimal resources.

The authors state that the usual practice of using a shorter physics time-step globally in RRM "cannot satisfy the two criteria above for the purpose of global climate production simulations mainly because the ZM deep convection scheme and other cloud parameterizations used by EAM are by design not scale-aware (Xie et al., 2018)." Here begins a pattern throughout the paper of conflating scale-awareness with time-step awareness in the cloud schemes. (2) is not a true criteria, and is more of a statement that it will take additional resources if you want to tune the physics at the shorter-physics time-step to produce a reasonable climate, i.e., criteria (1). The justification authors provide that they didn't tune because the physics aren't scale-aware does not make sense, as they are still running non scale-aware physics, just with a longer time-step. In many places throughout the paper they refer to how this hybrid time-step strategy is responsible for the success of satisfying both criteria. And then they also make statements like at line 678 where they say the hybrid time-step "bypasses the persistent poor scale-aware problem of atmospheric physics in a multi-scale framework," which is just not true.

(b). Hybrid time-step strategy. I have more details in the line by line comments, but my main issue here is the lack of discussion on the downsides of the hybrid time-step strategy. It is the resolved updrafts at higher-resolution that are responsible for the improvements in precipitation rates and extremes O'Brien et al. (2016). In GCMs, updrafts manifest through an instability between the physics (condensation and buoyancy) and the dynamics (vertical motion, adiabatic cooling and super-saturation). As the 25 km dynamics have larger vertical velocities and hence fast time-scales, it requires the physics time-step to be reduced in order to keep up with the the now faster evolving instability. These time-truncation errors due to the hybrid time-step approach can reduce vertical velocities by over half according to Appendix A in Herrington et al. (2019) or Figure 6 in Herrington and Reed (2018). Therefore by choosing the hybrid time-step strategy, one is choosing not to fully embrace the resolved updrafts that come with 25 km, and arguably not taking full advantage of all that 25 km can offer.

Comments

Line 69. "The RRM high-resolution results are robust regardless of where the fine-grid patch is located ..." I disagree with this statement based on the studies of Rauscher et al. (2013) (also the CAM4 simulations in the Zarzycki et al. (2014) paper you've cited). It's clear that putting the refinement region over the ITCZ can facilitate grid induced circulations that degrade the simulation quality over the low-res uniform resolution solution.

Line 86. Exactly what is meant by the (2) requirement "minimal effort of retuning based upon the low-resolution counterpart." Are the authors trying to convey that the tunings are based on the low resolution counterpart, and therefore we need to spend a "minimal" amount of time retuning the RRM? Why shouldn't a modeling center spend just as much time tuning a RRM as their low resolution counterpart? I mean, the underlying issue is that it's expensive to tune the RRM, which I agree with, but that doesn't mean the modeling centers shouldn't try to do their due diligence by using a similar level of resources as used in the low resolution effort. Of course resources are limited, and it may not be practical to provide the same level of support to two versions (low-res and RRM). But if that's the case it's institution specific and needs to be stated as the justification for the authors design choices, not a blanket requirement for all modeling centers. I don't think it's out of the realm of possibilities that a modeling institution can rearrange their priorities to provide the same level of human support it provides to the low-resolution effort, plus the additional computational resources required to tune RRM.

Line 118. Change "with additional tuning" to "with additional modifications" as you cite enhancements, not just tunings.

Line 138. What does the lack of scale-awareness in the physics have to do with either of the two requirements? I suspect you mean that the lack of scale-awareness would require extensive tuning, and that's inconsistent with (2). As I've mentioned, I don't think (2) as it's written is an actual constraint.

Line 142. The claim that you can't get "optimal climate" even if you retuned the model because the physics aren't scale-aware. Have the authors tried? This is possible but it is not a foregone conclusion. The extent that a RRM solution is tainted by the lack of scale-aware physics depends on the region chosen for refinement, e.g., refining the ITCZ is more problematic (Rauscher et al. (2013)) than refining high-latitude regions where large amounts of buoyancy and convection are not being generated.

Line 144. This hybrid time-stepping strategy is not new (see Zarzycki et al. (2014) you've cited). While I agree with the benefits of this approach as the authors describe, what are you losing? You're increasing your truncation error substantially. See for example the truncation errors derived from moist bubble experiments in Appendix A of Herrington et al. (2019) (full citation below). A more more balanced discussion is needed here.

Line 187. Can you be more clear and explicitly state you are substituting out the smoothing algorithm in Lauritzen for an internal algorithm?

Figure 4. Could the authors add an outline of where the transition region is in this plot? It's a nice result, and I think that addition would better convey the result.

Line 250. What is the production i/o? Is there high-frequency output?

Figure 6. Are you using a fixed low-res reference for the dashed lines? Or are you changing the low-res reference for each processor count to be the equivalent low-res processor count?

Line 289. Maybe put the number of column in NARRM in parenthesis after NARRM, so the 14k number doesn't come out of nowhere?

Line 298. RMSE, not RSME.

Line 299. “for the first historical member ...” why not use the ensemble means in Figure 7?

Line 358. “highlights the advantage of coupled RRM over a single-component RRM ...” How can you conclude this? You are not comparing to RRM AMIP, you are comparing to LR coupled. In RRM AMIP, you can easily select to have a high-resolution coastline, and the ocean grid is nominally on the RRM grid.

Line 367. Is it correct to compare min/max values of field on different resolution grids?

Line 375. Can the authors provide more details on resolution dependent emission factors? Why is this a thing?

Line 381. “the same input datasets.” Are the datasets coarser or finer than the RRM? If so, are the mapped emissions dataset aliased to their coarser grid?

Figure 16. I suspect this increase in AOD would survive remapping to a common grid before computing the histogram, but I still think this should be done to make it a more apples-to-apples comparison.

Line 537. “which is dominated by the reduced SW cloud feedback over the Northeastern US ... ” Just a suggestion, but it would be useful if the cloud analysis in 5.1.3 could be extended beyond the North Polar Region so that we could compare the cloud feedback’s in this section with existing cloud bias in the base state in LR and NARRM (and also extend the cloud feedbacks to the North Polar Region?).

Section 5.3. Why are the latent heat fluxes larger in LR(H1-5), and has systemically lower LCL on clear-sky days? Does NARRM introduce a dry bias in the atmosphere?

Figure 31. no corresponding a, b, c labels in the plot. What should be “a” is labeled as NARRM when it should be LR?

Figure A1. As topography is responsible for so many of the differences between LR and NARRM, I would suggest bringing this figure into the main paper.

References

A. Herrington and K. Reed. An idealized test of the response of the community atmosphere model to near-grid-scale forcing across hydrostatic resolutions. *J. Adv. Model. Earth Syst.*, 10(2):560–575, 2018.

A. R. Herrington, P. H. Lauritzen, K. A. Reed, S. Goldhaber, and B. E. Eaton. Exploring a lower resolution physics grid in cam-se-cslam. *Journal of Advances in Modeling Earth Systems*, 11, 2019.

T. A. O’Brien, W. D. Collins, K. Kashinath, O. Rübel, S. Byna, J. Gu, H. Krishnan, and P. A. Ullrich. Resolution dependence of precipitation statistical fidelity in hindcast simulations. *J. Adv. Model. Earth Syst.*, 8(2):976–990, 2016. doi: 10.1002/2016ms000671. URL `http://dx.doi.org/10.1002/2016ms000671`.

S. A. Rauscher, T. D. Ringler, W. C. Skamarock, and A. A. Mirin. Exploring a global multiresolution modeling approach using aquaplanet simulations. *Journal of Climate*, 26(8):2432–2452, 2013. doi: 10.1175/jcli-d-12-00154.1.

C. M. Zarzycki, M. N. Levy, C. Jablonowski, J. R. Overfelt, M. A. Taylor, and P. A. Ullrich. Aquaplanet experiments using cam's variable-resolution dynamical core. *J. Climate*, 27(14):5481–5503, 2014. doi: 10.1175/JCLI-D-14-00004.1.