

Response to Reviewers for “Exploring a lower resolution physics grid in CAM-SE-CSLAM” in consideration for the Journal of Advances in Modeling Earth Systems

General Remarks

I would like to thank the associate editor and reviewers for their constructive feedback. Below we address individual concerns, but here we would like to address a common concern regarding Figure 4. In general, the figure contained too much information and some of the panels were mislabelled. We have decided to split that figure up into two figures, one which addresses mapping errors of a smooth field (Figure 4), and one for errors associated with mapping steep gradients (slotted cylinder; Figure 5).

Associate Editor

Like both reviewers, I appreciate the clear explanation of the motivation for employing a physics grid at reduced resolution, the detailed description of the mapping algorithm, and the large amount of effort that went into the model development and evaluation presented in the paper. A few suggestions regarding the wording of the abstract, introduction, and conclusions are listed below.

"Lower resolution" in this paper has a very specific meaning, namely using 2x2 physics control volumes in each element of the dynamics grid instead of 3x3 in the default model that uses "equal" resolution for physics and dynamics. I'd suggest clarifying this in the abstract (and perhaps in the key points, too). This clarification would help putting the last sentence of the abstract into context ("... the coarser physics grid may allow for significant cost savings with little to no downside"). Presumably the validity of that statement will depend on how different the two resolutions are.

This is a very good point, and agree that the readers should be informed of the size of the lower resolution control volumes relative to the dynamical core grid, up front. The first sentence of the abstract has been amended to:

This paper describes the implementation of a coarser resolution physics grid into the Community Atmosphere Model (CAM), containing 5/9 fewer grid columns than the dynamics grid.

The first key point has been changed to:

A lower resolution finite-volume physics grid is implemented into CAM-SE-CLAM, containing 5/9 fewer grid columns than the dynamical core grid.

Apart from eliminating grid imprinting, one benefit of the pg2 grid that is mentioned several times in the paper is the reduction of computational cost of the physics parameterizations by about a half, especially in the context of the development of high-resolution, expensive Earth system models. It is worth noting, on the other hand, that in order to make good and efficient use of the next-generation supercomputers, the community has been making a substantial amount of efforts to explore super- or ultra-parameterization or the addition of sub-columns (e.g., CLUBB-SILHS) which effectively make the parameterization much more expensive. It would be useful if the authors could comment a bit more on these two seemingly opposite attempts (i.e., lower or higher costs for physics). I understand that super-parameterization is mentioned in the introduction following the mentioning of the work by Wedi (2014), but the authors viewpoint there does not seem to be very clear.

This is a very good point, and decided to include a sentence describing the usefulness of our approach for super-parameterization as the second to last sentence in the conclusions:

...This approach may also be useful in the computationally burdensome "super-parameterization" approach (Randall et al. 2003), in which a cloud resolving model could instead be embedded into the lower resolution physics grid, reducing computational overhead.

As a side note, the authors' have had some contact with the E3SM group regarding their development of an MMF model. Preliminary results using CAM-SE as the host model results in intolerable grid imprinting, more so than in a traditional CAM physics run. It is our understanding that the group is developing a similar approach to ours to address this grid imprinting issue.

Lastly, a very minor point: the statement "an isotropic representation of the numerics is provided to the physics" is used for the key points, in the abstract, and also in the conclusions. I think I understand what this refers to, but the phrase "isotropic representation of the numerics" seems very vague and hard to understand upon first read. It would be helpful if this could be rephrased to be more specific.

We understand and agree that that statement is ambiguous. It is hard to actually construct a concise statement making the point would like to make there, and decided that it's probably best not to stuff it into the key points. In the abstract and conclusions, the sentence was changed to be slightly more clear:

...The lower resolution physics grid provides a volume mean state to the physics that is computed from an equal sampling of the different types of nodal solutions arising from the spectral-element method...

Reviewer 1

Overview

This article describes the implementation and the model impact of a lower resolution grid for the CAM model. The motivation for this article is thoroughly and convincingly discussed in its introduction while the algorithms used to perform the mapping between the two grids are adequately presented. A lot of effort has been put to construct a grid-to-grid remapping which preserves accuracy, conservation, shape preservation, linear correlations. The expected benefit of reduced computational cost by using a lower resolution physics grid is achieved with negligible impact in the solution accuracy. Furthermore, the coarser grid is designed to eliminate grid imprinting which is manifested as noise in areas of steep orography such as the Andes and the Himalaya.

In my opinion this work is of high quality and of general interest to the climate but also weather modelling community. I would be happy to see the article published after some improvements have been made. I have listed a few comments and suggestions in the following section of this report.

Comments and suggestions

In page 24, in the last paragraph of 3.3 the question of which grid determines the characteristic forcing scale when physics and dynamics grid resolutions are different is posed. Although this question is addressed in 3.3.1, it takes very long to reach the conclusion raising the risk that readers may miss the point. I suggest to:

– bring upfront the answer to the question posed, for example after line 537 “We will show that the forcing scale is determined by the dynamics grid” and highlight if the answer depends on resolution. Ideally, this section can be made slightly shorter without loss of information.

Thank you for this feedback. We have brought the answer to this question upfront as suggested (see Line 553). We have done some cleaning up of section 3.3 to more clearly articulate our points.

The relation of the physics timestep with the horizontal resolution is discussed thoroughly in the article, however, vertical resolution aspects are completely left outside the discussion. In section 2.5 a set of test model configurations are specified but the vertical resolution and grid are not mentioned at all. Vertical resolution is very important for physics and although there is no widely used criterion such as CFL for explicit schemes, there are similar rules. A typical example is planetary boundary layer vertical diffusion. Conditionally stable schemes (explicit or partially implicit that keep constant the state dependent diffusion coefficient during a timestep) obey stability criteria that depend on the stiffness (and nonlinearity in the case of partially implicit schemes) of the problem. Increasing vertical resolution increases the stiffness of the problem and a shorter timestep would be required. Although vertical resolution seems to be beyond the

scope of this work, it still matters and at least some statement about it will be useful. For example consider two opposite situations: for very coarse, for physics processes, vertical resolution restricting timestep won't produce much impact. For fine resolution, timestep should have a bigger impact. Which category the presented simulations belong to? If timestep criteria are entirely based on horizontal resolution aspects, isn't there a risk that important vertical resolution issues which matter are ignored?

The reviewer rightly points out that we have omitted any discussion of vertical resolution in our manuscript. We have included a paragraph in the methods section (beginning at Line 429) describing the vertical grid, and the vertical dynamics. The vertical resolution is unchanged with changes in horizontal resolution. It might be wise to vary the horizontal and vertical resolutions together, to perhaps preserve an 'idea' aspect ratio for global models (note $dz \ll dx$ in all our simulations), but as the reviewer mentions, this is beyond the scope of this work.

The time-step used by the PBL scheme (CLUBB) is subcycled using a time-step of 300 s. If the physics time-step doesn't divide evenly, then it uses a time-step determined by dividing the physics time-step by the smallest integer such that the CLUBB time-step is < 300 s (at $ne120$, $dtphys = 450$ s, so CLUBB time-step is 225 s). If the physics time-step is less than 300 s, then CLUBB uses the physics time-step (this does not occur in any of our configurations). The model has not blown up in any of the presented simulations, so CLUBB is stable using the ~ 300 s time-step in our simulations.

Fig. 4 contains a very large amount of information, it is very complex and not easy to read while there aspects of it which I find confusing. Discussion related with this figure spreads from page 14 to page 16 and in two sections. It is difficult to relate the text with the picture while the caption doesn't give enough guidance to the reader to make sense of it as a separate entity. The labels in the sub-figures must be carefully checked, if necessary split the figure in two to make it more readable and to help its presentation in the text. Few points:

- In row 3, columns 1, 2 seem to describe the same mapping error but they are not the same. Is this plotted slotted cylinder the $pg2 \rightarrow pg3$ Error? It seems that column 1 is the analytical tendency while column 2 represents the mapped tendency, while the error of two different mapping methods is plotted at columns 3 and 4. The same applies for row 4 columns 1, 2. Furthermore, rows 3, 4 seem identical.
- Page 14, line 333-336: it is mentioned that the tendency is mapped from $p2$ to $pg3$ while the top label suggests $p3$ to $pg2$.
- Page 16, lines 365-368: in Figure 4, row 3 and 4 I cannot see the one order of magnitude reduction mentioned in the text because rows 3 and 4 are identical.

Thank you for pointing out our errors here. Figure 4 is addressed at the top of this document.

Fig. A1 has insufficient information in the caption and the legend is confusing. What is the difference between + and circles?

Apologies, the lack of information in that figure was an oversight. It has been edited to provide all the relevant information.

Reviewer 2

Overview

Herrington et al. have implemented a multi-grid approach in the CAM model to more consistently solve the physics at a uniform resolution instead of on the quadrature points of CAM's spectral element grid. After demonstrating the methodology conserves mass and gives a good solution relative to the default methodology, they show that their new methodology reduces issues of grid imprinting in regions of complex topography. They then examine the spectral characteristics of the solution to understand whether the model's effective resolution is determined from the dynamics grid or the physics grid. Ultimately, they find that the dynamics grid has the strongest influence.

Overall, this is a clearly written paper that, while heavy in the mathematics, is fairly readable. The authors have given much thought to the details of their methodology and walk the reader through how several "gotchas" are mitigated via careful algorithmic development. This reviewer recommends a minor revision for this paper.

Note that this reviewer is not fully versed in the mathematical notation used in this paper. Thus, not all formulas have been verified.

Specific Comments

(1) L352–3

The authors state that the dry pressure is not modified by the physics. What about for a convective scheme that includes a mass flux treatment? Would not this modify the dry pressure? This may or may not be relevant for the version of CAM being used. Is it relevant? Even if it is not, the authors should note this subtlety.

To our knowledge, the only convection scheme that actually modifies the grid-scale pressure field is Kuell et al. 2007, which is not used in CAM. Most AGCMs use a mass flux model that preserves the relationship $M_{\text{resolved}} = M_{\text{env}} + M_{\text{conv}}$ (M_{env} and M_{conv} is the parameterized 'compensating subsidence' and convective mass fluxes, respectively) and the convection scheme solves the right hand side so as to not change the resolved mass fluxes (M_{resolved}).

Kuell, V., A. Gassmann, and A. Bott: 2007, "Towards a new hybrid cumulus parametrization scheme for use in non-hydrostatic weather prediction models." QJRM, 133, 623, 479-490.

(2) L447

The text states a maximum relative error of 0.016%, but the caption of Fig. 6 states a relative error is contoured at 0.025%. Should not the maximum be greater than a contour level representing a value below the maximum?

We thank the reviewer for pointing out this inconsistency. As discussed in the text, the relative error is compiler dependent. The figure shows simulations using a different compiler (intel) that results in a larger maximum relative error (0.026%). The text has been edited (beginning at Line 467) to make this point clear.

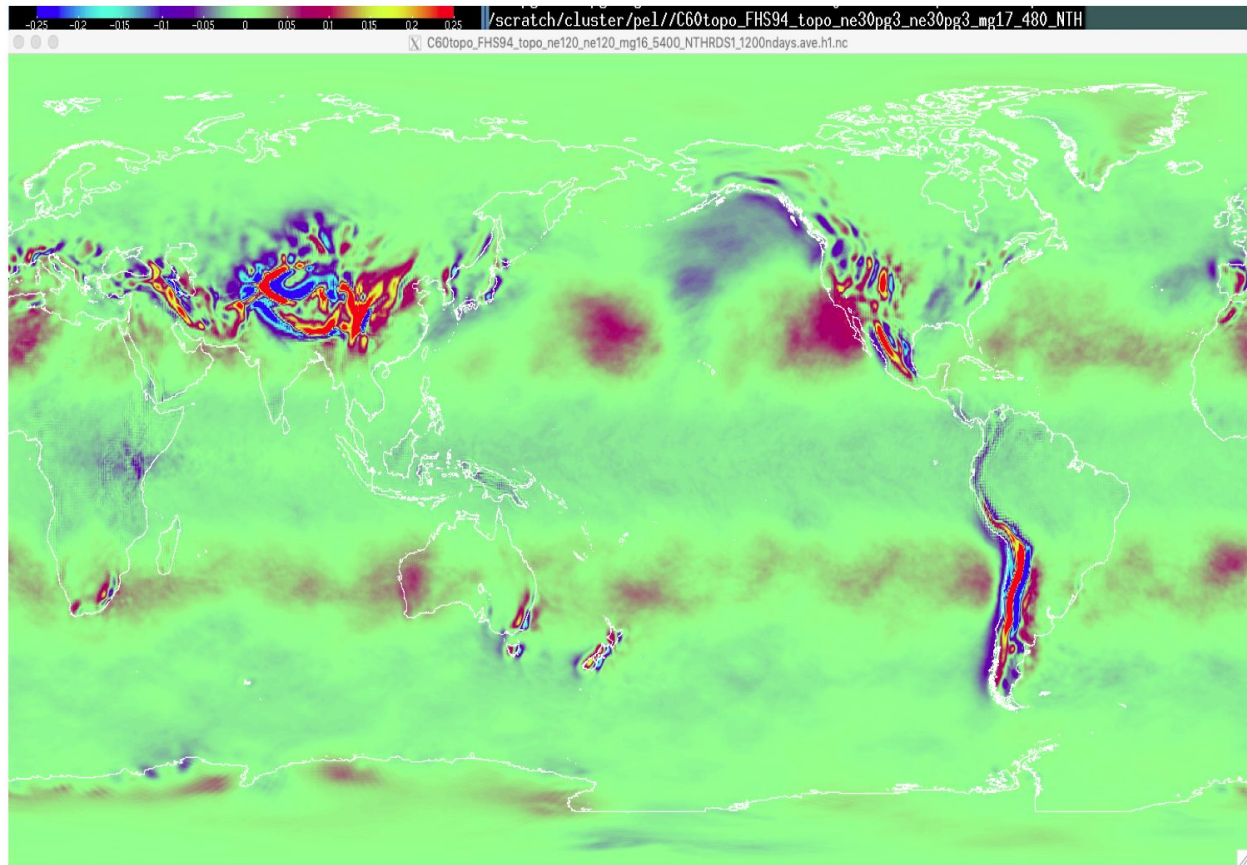
(3) L468ff

The authors discuss the waves emanating from the Himalayas as grid-imprinting- induced numerical noise. This particular case looks more like the Gibbs phenomenon, which would be induced from the flow around steep topography. Can the authors separate the two issues of Gibbs and grid irregularities and show that the Andes and Himalayas shown are really due to the location of the spectral element boundaries? For example, could one slightly shift the grid so that the element boundary is in a different location but still near the topographic features of interest to see if the waves form in the same location (likely Gibbs) or shift with the boundary (likely grid imprinting)?

It is plausible that waves emanating from the Himalayas are Gibbs ringing. We ultimately decided it would not be worth the effort to shift the grid one node north or south relative to the mountain, to test whether the location of the extrema would line up with an interior node. Our reasoning is that, in the spectral-element method, if you change a value at any node, be it an interior or boundary node, it changes the shape of the basis for the entire element. Had a local max coincided with an interior node instead, gradients near the element boundary would likely steepen in response and an extrema would be introduced there. This tendency for extrema to occur at element boundaries is discussed in detail in Herrington et al 2018 in MWR (in the references of this paper).

Additional experiments using CAM-SE (in which these oscillations are also pronounced), indicate that the oscillations disappear if the flow is sufficiently resolved. Below is a plot of the omega field at 500 hPa in a ne120 simulation, but using the ne30 topography (the smoothing radius of the raw topography is higher at lower resolutions):

10 month average OMEGA500; n120 with C60 topography



(4) L697ff

The authors make statements several times in the paper about potential computational cost savings due to the reduced number of physics columns to calculate for a given effective resolution. While idealized, the timing costs for the experiments presented in the paper could be used to quantify more specifically the actual impact. This has not been done. The reason this is brought up is because potential savings are often far from real savings due to parallelization issues. For example load imbalance would likely severely limit the potential speedup since all the physics columns related to a given spectral element are likely solved by the same node (I'm guessing on this part—correct me if I am wrong), and since clouds tend to cluster, there is likely to be a bimodal tendency for elements to contain all or none/few cloudy columns. This would lead to the cloud-free elements to have to wait for the cloudy elements.

Load balancing is not treated any differently in the pg2 configuration. We have not provided timings in this paper since at present, the pg3 code has been optimized by the software engineers whereas pg2 has not. We can not provide an apples-to-apples comparison at this time. Even though not optimized, pg2 is significantly cheaper than pg3 using CAM6 physics.

One issue that we included (see Line 719) that would limit the speed-up is that additional mappings required in the pg2 configuration (CSLAM \leftrightarrow pg2). Two additional communications are required to perform the high-order mapping, which is an added expense.

(5) Fig. 4

Are the data panels in row 4 incorrect? They appear to be a duplicate of row 3. And, the titles of row 3 appear to be what one would expect for row 4. Also, the labeling is too small to read.

Issues regarding Figure 4 are addressed at the top of this document.

(6) Fig. 7

Figure 7: The oscillations intended to be shown in the bottom-middle panel would be clearer if a divergent color palette were used. Likewise, the $2\Delta x$ noise along the Andes would more clearly show as numerical errors due to the oscillating sign by using a divergent palette.

At the suggestion of the reviewer, we did experiment with a banded color bar designed to bring out details. The banded color bar did not make the oscillations any more clear than in the original (we experimented with many).

(7) By having the physics and dynamics run at different resolutions, one reduces the issue of sub-effective-resolution noise from impacting the physics tendency calculations. However, does any adjustment need to be made to parameterizations that represent sub-grid motions (e.g., turbulence and convection) to account for possibly doubly treating sub-physics-grid motions through both the dynamics and the physics? This is not a fatal flaw of the methodology, but it is something that should be accounted for. Because the grid differences are not great, this likely is not a first-order issue. However, the authors should not it as an open problem.

This is an interesting point, that in theory, should matter. As the reviewer points out, mapping from the grid-scale, dx , to control volumes that are $1.5dx \times 1.5dx$ introduces additional sub-grid quantities to the grid-scale. Had the physics grid been larger than the effective resolution of the dynamical core (e.g., $10dx \times 10dx$), then the impact of double-counting sub-grid motions would presumably be important. Since explicit motions are not resolved at the $1-1.5dx$ scale, then the double containing is likely not a first-order issue.

Having said that, one result of our paper is that high-order mapping of the physics tendencies to the GLL grid reconstructs scales that the pg2 grid can't support. This has a noticeable impact on the solution on the GLL grid, and by extension the pg2 grid. (see Figure B.1) The high-order mapping may be viewed as a sort-of sub-grid parameterization for the pg2 grid.

The topography used in the model is native to the physics grid, and all the sub-grid topographic quantities used by the physics (e.g., gravity wave drag) are consistent with the physics grid.

(8) The interpolation between grids is handled level-by-level. In regions of complex topography, such as near the Andes, the use of terrain following coordinates implies (at least at low levels) the element over which the grid exchange occurs is slanted with one side higher than the other. How is this accounted for? Are there any implications of this in terms of the accuracy of the state presented to the physics grid for calculating the physics tendencies?

The reviewer brings up a very thoughtful and interesting point. In regions of steep topography it is certainly possible that the control volume mean is not consistent with the level it is at. This is mitigated to some extent through integrating the high-order basis / ppm reconstruction to compute the volume mean on the physics grid.

(9) The authors state that the model output used in the study is available from Github, which this reviewer has verified for some of the files. While not stated in the manuscript, the code to reproduce the figures is also part of the repository holding the data, for which the authors should be commended. Unless it was missed while reading the manuscript or looking through the repository, there seems to be no indication of where readers would be able to obtain the model code to reproduce the simulations in this paper. That is more fundamental than the actual model output.

The pg2 configuration was put onto the main CAM branch in March of this year, so it will be publicly available for the next release of CESM, CESM2.2, due to be released in June 2019.