Topic editor decision: Publish subject to minor revisions (review by editor)
by Peter Caldwell
Public justification (visible to the public if the article is accepted and published):

Like the reviewers, I thought this was a great paper. I also thought the authors did a decent job of addressing the reviewer comments. In my own read-through of the most recent draft, I found a few things that could make the paper even better. Thus I expect the paper will be accepted after some minor corrections.

Reply: We would like to express our gratitude to Dr. Peter Caldwell for his handling, detailed review, and constructive feedback on our manuscript. We appreciate your insightful comments and suggestions. The issues you highlighted have been addressed in the revised manuscript. They help to further improve the clarity and quality of the paper.

1. **I was a bit disappointed that many of the authors' reviewer responses were just that - they didn't actually change the text to address the question. This is unfortunate because many readers will have the same questions. I was also a bit disappointed that the authors just conjectured in a lot of these cases rather than actually doing any analysis. I think the paper is in great shape as-is so I don't insist that more analysis is done, but I thought some of the reviewers' questions were quite good. I'm thinking particularly of reviewer 1 questions 1-2, 4, and reviewer 2 question 1.**

Reply: We acknowledge that these questions warrant more detailed investigation. The reviewers' questions have been now discussed in the main text. Specifically, we **have revised lines** 119-122, 407-413, and 200-224 to **address** Reviewer 1's questions 1-2 and 4, and lines 122-125 to **address** Reviewer 2's question 1.

2. **In lines 16-17 of the abstract, the authors refer to the *solver* for hydrostatic, non-hydrostatic, and advection but I think they mean that these are timing reductions for the whole dycore (or the advection part). In particular, they say later that the dycore solver is actually slower in mixed precision. Another issue is that around line 186 the timings are described for the *dry* dycore but the moist dycore is used for most of the physics tests. Is there a difference in timing?**

Reply: Yes, in lines 16-17 of the abstract, the hydrostatic, non-hydrostatic, and advection solvers refer to the timing reductions for the whole dycore (in our term, dycore means dry dynamical core, not with tracer transport) or the advection part (tracer transport). The mixed precision dycore solver is slower only when compared to the single-precision dycore solver, because in the mixed precision dycore solver, the precision-sensitive calculations remain in double precision. This is described around lines 195-196 of the manuscript.

The computational time is diagnosed independently for the dycore (hydrostatic or nonhydrostatic) and tracer transport module, i.e., 24%, 27%, 44% in the abstract. The physical performance (Section 4.1-4.5) focuses on the moist dynamical core (dycore + tracer). This is to give a clean comparison.

I'd also like to know how much faster the full model is when the dycore uses mixed precision for the dycore, both in AMIP and global storm-resolving

model (GSRM) configuration. Since these simulations were used for physics analysis, it should be trivial to report their timings. In my own GSRM, we spend ~90% of our run time in the dycore so I suspect GRIST total performance improvement will be O(20%) in this case.

Reply:

We have now mentioned "how much faster the full model is" in the AMIP and GSRM simulations. The GSRM tests have an impressive time reduction. Your estimation (~20%) is rather accurate. The actual number derived from the DBL and MIX tests in Section 4.5 was 25.5%, diagnosed for the dynamics and physics procedures, including physics-dynamics coupling.

The AMIP test was performed at 120-km resolution using 640 MPI processes distributed across 32 nodes. The MIX code did not witness very impressive time reduction (~12% for the hydrostatic model), because the memory-bound issue may not be significant for such low-resolution simulations.

This is further detailed at the 3rd and 4th paragraphs of Section 3.

3. It would be nice if the authors could say whether this project has been deemed a success internally: will all future users be told to use the mixed-precision mode, or do you think the single- versus double precision differences are too big to be used for operations and/or research?

   As an aside, I was surprised this paper just looks at differences between simulations rather than checking whether model skill is significantly affected (as in https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/qj.4181, for example). If mixed-precision isn't good enough, would being more stringent in deciding which variables and equations needed double precision have resulted in a more acceptable solution (and is this easy enough that the authors could try this if need be)?

Reply: We consider this project to be a success, at least in its current phase. From idealized experiments and AMIP simulations to GSRM simulations, the mixed-precision simulations have well reproduced solutions as the double-precision simulations and saved the computational time. Based on these test cases, the mixed-precision model code can be said largely successful and has been used for many ongoing efforts.

Safely speaking, for quality operational runs, more testing efforts are required. The examination may further include checking operational metric, skill score card, etc, as the paper you mentioned. This requires longer term efforts.

Mixed-precision computing will maximize its value in the kilometer-scale applications, and future applications and studies will more focus on this scale, including metric and skill score related to the k-scale forecast. One plan is to merge the mixed-precision code and a limited-area model recently developed for GRIST, to enable efficient and skillful customized model applications at the convection-permitting scale.

We have added a paragraph at the end of Section 5 to address this issue.

4. Around line 110, it might be useful to clarify that error is calculated relative to the double-precision results (rather than observations).

Reply: We have **clarified** in the manuscript that the error calculations are relative to the double-precision results.

5. typos (please do one more careful read through before you resubmit!):
a. L253: "appears, expectED for..."
b. L320: "may DIVERGE MORE from..."
c. L354: "a few grid spacES"
d. L397: "It also needs to BE recognizeD that"
e. Fig 7 caption: fix "from the and deterministic"
f. Fig 9 caption: I think you mean the continuity equation rather than the continuous equation.

Reply: The typos mentioned have been individually **corrected**. We have **carefully read** the manuscript again to avoid potential mistakes. We have also slightly improved some sentences to improve the clarity of the paper. Thank you!