Dear Prof. Xiao,

The authors would like to thank you for your thorough review of our manuscript. Below (in red font) is a point-by-point response to your comments.

This paper reports the inter-comparison results of some advection schemes used in different GCMs. The comparisons were conducted by using a set of benchmark tests to evaluate the accuracy of the participating schemes. It will of course provide useful information about the performance of each individual scheme, and enrich our knowledge about the current level of the advection schemes in GCMs. The paper can be accepted in its current form after corrections of typos and careless writings (there are quite a bit in the text).

The authors would be grateful if the reviewer could point out the careless writings (and typos). We have such a set of editorial suggestions from Reviewer #2 and will incorporate these.

Meanwhile, I have some reservations, which I think might be more appropriate to be raised in the interactive discussion stage, but I still hope an adequate response and possible reflections in the revised version to this peer review.

1) I wonder why this kind of experiments could not start with a Cartesian grid. I believe all developers should have begun with a Cartesian-grid version, and be ready to do so if a case suite in Cartesian grid is provided. This will largely remove the unnecessary factors, like the influence of different grids, and make the comparison more focused on the pure “numerical property” of the advection transport schemes.

The design philosophy of the Lauritzen et al. (2012) test case suite is to provide a “minimal” suite of tests to assess different aspects of accuracy in a setup similar to a “real-world” global model setup. This forces the scheme developer to make choices on the spherical discretization grid (and devise methods to deal with spherical geometry), address air and tracer mass coupling (which is necessary in “full” models), and shape-preservation (essential for physical parameterizations).

It is, of course, a logical step in the scheme development process to test algorithms in Cartesian geometry before proceeding to the sphere; so there would definitely be a use for a standard test case suite in Cartesian geometry and it would make scheme intercomparison more focused on the basic numerical method. However, the intent of the test case suite is to include all the non-trivial deliberations that the scheme developer must go through to develop a scheme for practical global models in spherical geometry.

The performance of an advection scheme can be assessed more rigorously by looking into the algorithms used in both spatial reconstruction and time marching. For example, the Taylor expansion and Fourier analysis provide well established tools to see the truncation order (convergence rate) and the numerical errors of linear schemes (without limiting). In principle, this kind of more rigorous analysis should be able to foretell the conclusion drawn from numerical inter-comparison. Inter-comparison can be more meaningful for complex models, but less interesting for purified solvers since we can see their performance from theoretical analysis even without “shed a light on ” by the numerical experiments.

The authors acknowledge that the separation of spatial and temporal errors is not carefully assessed in this study nor are rigorous analysis of errors at a theoretical level. The scope of the intercomparison is, however, to assess accuracy under the conditions they would be applied in real operations and a theoretical analysis will most likely not encompass all aspects of accuracy present in real applications.

For example, the authors agree that the asymptotic convergence rates for the unlimited schemes could have been estimated theoretically via a Taylor Series expansion or similar analysis. The authors are, however, not aware of a theoretical method to estimate at what resolution a scheme starts to converge at its theoretical convergence rate. For example, an 8th-order scheme that starts to convergence at 8th-order at very high spatial resolution may be considered less useful than a method that is formally lower order but starts to converge at coarser resolutions (and possibly with lower absolute errors at coarser resolutions) for some types of applications where the resolution can not easily be increased (such as climate model applications).

In all, the intent of the test case suite is to assess aspects of accuracy in a setting as similar as possible to how they would have been implemented in a full-physics model, and the authors believe that a theoretical analysis will not necessarily provide a complete picture of accuracy. Some of the published schemes here have already done this on an abstract 1D basis.

2) The schemes tested in this paper might not be the state-of-the-art. For example, the limiting projection cited is of old-fashion TVD style, which has at most second order. More advanced reconstructions, such as WENO (include the refined WENO-M and WENO-Z), are overlooked. Instead of “state-of-the-art”, “currently-in-use” might be more appropriate.

The schemes presented in this manuscript are basically the ones that participated in the NCAR tracer workshop in 2011. Although certain schemes make use of “old fashioned” methods, their application on the sphere may be new. All the schemes that participated in the intercomparison were schemes that were actively being developed by the participants. The authors believe that this qualifies the use of “state-of-the-art” in the manuscript title.

The authors would have liked to include WENO schemes and discontinuous Galerkin methods and recognize that these methods are unfortunately not part of the suite of schemes for which results are presented. The authors envision that new scheme developers will exercise this test case suite and publish the results so that readers can compare their schemes performance with the schemes presented here.

4) Without the computational cost as another account, it is hard for the readers to get a fair judge among the schemes. Usually, higher order schemes are more computation- ally expensive than low order schemes. In practice, there is always a tradeoff between accuracy and efficiency. The balance is a key point, which is worthy of more attention in the whole story. As the authors mentioned, the elapse time inevitably depends on the computing platform used, as well as the coding style. As the next follow-up, if it is possible for the organizer to provide a platform and measure the elapse time of all participating schemes in their Cartesian form.

The authors are aware that computational cost is not assessed rigorously in the manuscript and agree that the ratio between accuracy and cost is key. We provide a table with information on "Algorithmic considerations" that gives the reader an idea of maximum stable time-steps, computational halo width, etc. that all affect computational efficiency. We did consider porting all codes to one platform and report on computation time. However, we decided not to pursue this approach as it would bring in complications associated with, for example, different levels of coding optimization (we have operational codes and very basic research codes in the mix) and questions about efficiency on different architectures and methods of parallelization of the underlying full-physic code. The workload to rigorously assess efficiency versus accuracy seems daunting from a software engineering perspective. That said, it would indeed be a very interesting exercise!

3) As commented above, an "apple to apple" comparison among pure advection schemes can be made relatively easier by examining both spatial reconstruction and time updating algorithm. Rather than this, what we see here is how a scheme performs under the circumstance of a certain "culture" which heavily depends on the background of the developers. It is easy to say which advection scheme is more accurate, but it is not that easy to say which culture is better than others.

We believe the answers above also answers this comment.