A mimetic, semi-implicit, forward-in-time, finite volume shallow water model: comparison of hexagonal-icosahedral and cubed sphere grids

Reply to reviewers

J. Thuburn[1] C. J. Cotter[2] T. Dubos[3]

We repeat here the responses to reviewers given in the interactive discussion.

Reviewer 1

We thank the reviewer for their positive and constructive comments on the manuscript.

1. *Special handling on the cube edges...* We use no special handling near the cube edges, or near the pentagons on the hexagonal-icosahedral grid. (We have emphasized this in the revised manuscipt at the end of section 2.) The scheme presented is essentially an unstructured grid scheme; part of the purpose of the work is to investigate the intrinsic performance of the scheme on different grids with no special handling of particular grid regions.

2. *The necessary globally-implicit operation would then be very inefficient in parallel...* It has been widely feared that the solution of global elliptic problems needed for implicit time integration would be inefficient and scale poorly on massively parallel computers, and this has led most recent model development efforts to opt for a HEVI (horizontally explicit vertically implicit) approach. However, there is growing evidence (e.g., Heikes et al. 2013, Müller and Scheichl 2014) that elliptic problems can be solved efficiently and scalably in parallel. We have mentioned this in the revised manuscript (section 4, penultimate paragraph).

3. *Are there any plans for some sort of monotonicity or positivity enforcement?* There exist limiters, guaranteeing local boundedness of advected tracer mixing ratio, that could be used with the advection scheme presented in section 5 and that would be straightforward to implement (e.g. Thuburn 1996 and improvements by Miura 2013). We have not used them here simply because there was no evidence that they were necessary in these tests. Nevertheless, we have added these two references to section 5.7.

...is there any explicit artificial dissipation,...? The only dissipation mechanism in the model is the inherent dissipation in the upwind advection scheme; in section 2 we have referred the reader to work by Kent and colleagues for a discussion of the effects of this dissipation on energy and potential enstrophy cascades. Apart from this, no other dissipation mechanism (either explicit or inherent in the numerics) is needed to maintain stability or to control dispersion errors or other numerical sources of noise. This is now emphasized in section 2 of the revised manuscript, after the list of desirable properties

Minor comments

4. *Many figures are hard to read...* Apologies for this. This was partly due to the change in format between the preprint and the online version. We have improved figure 6 as suggested.

5. *(Specification of grid resolution)* This is largely a matter of what you're accustomed to.

[1]University of Exeter, College of Engineering, Mathematics and Physical Sciences, Exeter, UK
[2]Imperial College, Department of Mathematics, London, UK
[3]IPSL/Laboratoire de Météorologie Dynamique, École Polytechnique, Palaiseau, France

Some of us find the “C90” notation for the cubed sphere non-intuitive! The important point is that Table 1 allows the reader to quickly covert between number of cells, degrees of freedom, and grid spacing.

6. *Why is the KE term treated as a backwards term?* Apart from the advective fluxes, the scheme is essentially a Crank-Nicolson scheme. Thus, the KE term is treated in a centred time-averaged way (not really backwards). In particular, the scheme is actually quite different in detail from the Lin-Rood (1997) scheme, though it shares the important property of computing the nonlinear Coriolis term as a vorticity or potential vorticity flux.

7. *Why build the stencils iteratively?* One could write down the stencil explicitly by hand for particular cases. However, for higher degree polynomial fits (requiring large stencils) near pentagons or cube corners, there are many cases that would need to be considered, and the process would become tedious (and error prone!) The iterative scheme is very general and easily automated, and appears to produce reasonable stencils in all cases tested. Note that the stencils are generated once at the start of a run, so cost is not an issue.

8. *Have you done the test of a spatially-uniform tracer field in a non-divergent flow?* Yes, we have done the test. An initially uniform tracer density remains uniform in a non-divergent flow, and an initially uniform tracer mixing ratio remains uniform in any flow, on both the primal and dual grids. In the revised manuscript (section 5.8) we mention that these two design requirements are verified in tests. See also the reply to Reviewer 3 in the interactive discussion.

9. *Tables 6 and 7, ...convergence rate...* Qualitative rough estimates of the convergence rate are already given in the text for both cases. The resolutions tested are not high enough to obtain clean estimates of the asymptotic convergence rate.

10. *Could a couple of different resolutions be shown?* Table 6 already shows how various error norms depend on resolution. Note that grid imprinting errors are nothing but numerical truncation errors whose pattern happens to reflect the underlying grid structure.

11. *Angle of the flow in test case 2.* The default grid orientation (as noted in section 1.2) is used; i.e., pentagons at the poles on the hexagonal grid, and cube corners at latitude $\pm\pi/4$ on the cubed sphere, and the flow parallel to the equator.

12. *... measures of the error ... may have little meaning, ...* All the evidence (from our models and others in the literature) implies that the solution *does* converge with finer space and time resolution, (and to the same solution for different models), implying that the error measures *are* meaningful.

13. *... is the maintenance of sharp filaments and gradients truly a consequence of mimetic properties?* First and foremost, these properties come from the advection scheme; as you note, such behaviour is typical of finite volume schemes. (The mimetic properties do help to ensure that when we insert the PV fluxes in the momentum equation, the PV we diagnose does indeed evolve in a way consistent with those fluxes.)

14. *Does the mimetic scheme conserve energy better than (say) ENDGame?* The results for available energy and potential enstrophy are very similar in ENDGame (for interest, see figure 1 of the reply to Reviewer 1 in the interactive discussion). The inherent dissipation due to the semi-Lagrangian advection scheme used in ENDGame is expected to be similar in its magnitude and scale dependence to that due to the finite volume advection scheme: both are (quasi-) third-order upwind schemes. The similarity to ENDGame is mentioned in the revised manuscript at the end of section 6.6.

15. *Use same contour interval in all panels.* Figure 6 has been replotted to use the same contour interval in all panels and to indicate the position of the mountain.

16. *I would be interested in seeing the Rossby-Haurwitz wave,...* In fact the mimetic properties do not really help the model to hang on to a steady (or steadily propagating) but dynamically unstable solution such as the Rossby-Haurwitz wave. The perturbation introduced by grid imprinting is enough to trigger the dynamical instability. A similar thing happens in the Galewsky test case, so we decided against showing results from the Rossby-Haurwitz wave. In the revised manuscript we have expanded the discussion of the Galewsky test (end of section 6.7) and mentioned that similar grid-triggering of instability happens for the Rossby-Haurwitz wave.

17. *... initializing with a fully-backward method ... yields a better result...* We must be careful with our interpretation here. The test case as those authors defined it does include large amplitude gravity waves (which are deterministic and reproducible at high enough resolution) generated by the initial condition. The correct result therefore includes those gravity waves. However, here we want to know whether there is any spurious generation of gravity waves by the numerics. To see this, we must modify the test case to damp the initial condition gravity waves, which would otherwise mask any numerically generated ones. We do agree that it is important to understand the degree of initial balance or imbalance, even in such idealized test cases; we have seen the consequences of imbalance both in this test and in the isolated mountain test case!

Reviewer 2 (Dr Gaßmann)

We are grateful to Dr Gaßmann for her comments on the manuscript.

We agree that the extension of the proposed approach to three dimensions is far from easy (though this is probably true of all suitable numerical methods). Dr Gaßmann highlights two particularly interesting and challenging aspects.

1. We are aware of Dr Gaßmann's very interesting work on the Hollingsworth instability (reference given in the interactive discussion for others following the discussion). The instability can affect schemes, like ours, that are based on the vector invariant form of the equations. In particular, our scheme has much in common with the TRSK scheme that manifests the instability presented in her paper. However, as she says, unfortunately the instability does not arise in the shallow water case; at the same time, a complete and convincing analysis on paper remains elusive. Thus, the issue is hard to explore except through numerical experimentation with a three-dimensional dynamical core.

We are currently working on extending our approach to 3D, and we will certainly be looking out for signs of this instability. It is possible that ensuring accurate PV advection will eliminate or minimize the problem. Alternatively, it is possible that the mimetic finite element approach (which has much in common with the scheme we describe here, but improved accuracy) will eliminate or minimize the problem. However, if the instability does arise we will certainly explore the modifications suggested by Gaßmann (2013).

2. The second topic raised is extremely difficult and complex. At least conceptually, we can distinguish between (a) dissipation mechanisms needed to keep a model stable, and (b) dissipation mechanisms intended to represent real physical processes on subgrid scales. (In practice it may not be so easy to separate these two.) Regarding (a), our use of a linearly-energy-conserving

spatial discretization combined with a Crank-Nicolson-based time scheme gives us stability without the need for additional ad hoc dissipation.

Issue (b) can be further dissected into the related questions of (i) what is an appropriate form of the subgrid model, and (ii) whether and how to conserve total energy.

On (i), the results of Kent et al. indicate that upwind advection of (potential) vorticity gives us least a partial implicit sugbrid model in the vortex-dominated enstrophy cascade regime, and it is plausible that this will remain true in 3D; this is the approach we have used. Smagorinsky-type schemes are widely used and accepted for LES of 3D turbulence such as that in the planetary boundary layer. Although they are also used on larger scales in atmospheric modelling, their justification is less clear; they might not be sufficiently scale-selective, and they might not capture processes such as gravity wave energy cascade or frontal collapse, which might be relevant in the $k^{-5/3}$ energy spectrum range. We think the choice of suitable subgrid models in these regimes is an open research question. Nevertheless, some form of eddy momentum flux tensor τ is a plausible approach. If τ is to depend on the rate of strain tensor then the referee makes the valid point that, on the hexagonal C-grid (and other C-grids in general), although the divergence and vorticity have natural, simple and compact, approximations in terms of the velocity, the other components of the rate of strain tensor do not; then the most suitable form for use in estimating τ is not obvious.

On (ii), as the referee says, some climate modellers consider it highly desirable to be able to close the energy budget. To do this, we must fully include all energy source and sink terms in our governing equations (the $-\mathbf{v}\cdot\nabla\tau$ and $-\tau\cdot\cdot\nabla\mathbf{v}$ terms mentioned by the referee). We must also either use fully energy conserving numerics (which may come at a price, e.g. in terms of advective dispersion errors), or keep track of all energy dissipated by the numerics (which is difficult in practice) and restore it somehow. This is complicated by the fact that numerical dissipation (and also dissipation by subgrid models) is often excessive and at too large scale, raising the question of in what form to restore the lost energy (internal or kinetic, rotational or divergent ...?) and at what scales.

Moreover, energy is a nonlinear function of the prognostic variables usually used, and so has unresolved as well as resolved contributions. The equations mentioned in the referee's comment ($\rho\partial_t K = \ldots - \mathbf{v}\cdot\nabla\tau$, $\rho\partial_t E_{int} = \ldots - \tau\cdot\cdot\nabla\mathbf{v}$) assume that sources and sinks of the unresolved contribution are instantaneously in balance; a more complete treatment would carry a prognostic equation for the unresolved contribution with sources and sinks that need not balance instantaneously.

The referee suggests that we might mention possible 3D development of the scheme in the manuscript, particularly the Smagorinsky diffusion term. However, as is clear from the above (and we have barely scratched the surface!), we could not possibly do justice to such a vast and complex topic. We therefore prefer to leave such discussion to another place where it can be treated more completely.

Reviewer 3 (Dr Ullrich)

We thank Dr Ullrich for his positive comments on the manuscript.

1. *Scalability.* Yes indeed, all serious model development efforts are concerned about scalability! Something of the order of 1000 columns per processor are the sort of numbers we hear

talked about before models start to run out of scalability (whether or not they require elliptic solves). We are encouraged by the results of Müller and Scheichl (2014) for geometric multigrid, particularly as the elliptic problem that arises from implicit time stepping has an inherent length scale $c_s \Delta t$ (c_s is the sound speed), so that, unlike the Poisson problem solved by Heikes et al. (2013), we only need ~ 3 levels of coarsening, thus avoiding the need to gather the coarsened domains onto a smaller number of processors.

2. *Stability.* In short, we don't have a simple answer to this question. Conversely, one could ask why is the scheme stable at all. Although a Crank-Nicolson treatment of linear fast waves is unconditionally stable, and the advection scheme in 1D is stable for Courant numbers up to 1, it is not immediately obvious that the two, combined in the way we have done, should be stable – that is why we did the normal mode stability analysis. We should emphasize that for advective Courant numbers between 0.75 and 1 the instability growth rate is extremely small. Moreover, in 2D on irregular grids there are no analytical guarantees about the stability limit of the advection scheme on its own, but practical experience suggests it is very close to 1. (If a flux limiter were used then stability of the advection scheme could be guaranteed for advective Courant numbers up to 1).

3. *Advection and grid imprinting.* We agree that the cosine bell test is not the most challenging for advection schemes. However, the test suggested by the referee is actually just a consistency check for the way we have formulated our advection scheme (sections 5.7 and 5.8). A tracer initialized on the primal grid to be the same as ϕ remains the same as ϕ thereafter. Initializing a tracer on the dual grid by averaging the primal grid ϕ using equation (28) and then advecting it gives the same result as advecting the primal grid ϕ (or the ϕ-like tracer) and then averaging the result to the dual grid using (28). Figures 1-3 in the interactive discussion confirm this for the barotropic instability test. Since Reviewer 1 also raised a point about tracer advection and preservation of constant mixing ratio, we have added a reference to the importance of mass-tracer consistency at the end of section 5.8.

Around the time we submitted the paper we were concerned that accuracy of the advection scheme, linked to grid imprinting errors, was the primary cause of the errors seen in in the barotropic instability test case. In particular, the parallelogram approximation of the swept areas might be inaccurate in strongly sheared flow, as suggested by Ullrich et al. (2013). We therefore extended the advection scheme to use more general quadrilateral swept areas (but still with straight edges). It made no difference to the results! Further investigation (switching off the initial height perturbation and looking at the step 1 errors) revealed that the primary source of errors was the perp operator, which gives the mass fluxes used to advect potential vorticity. Thus the errors do take on a grid imprinting pattern, but are coming not from the advection scheme itself but from the mass fluxes input to the advection scheme. These tests were actually done with a finite element model that uses the same advection scheme but has a consistent perp operator; even so, grid imprinting in the perp operator appears to be the factor limiting accuracy in the bartropic instability test. For the finite volume model discussed in the paper, with its inconsistent perp operator, it would be surprising if this did not remain true.