Review: Improved double Fourier series on a sphere ...

General Comment:

The Double Fourier Series on the sphere approach historically is compared to the spherical harmonic representation, e.g. Boer & Steinberg (1975) state: "the ease of calculation using the FFT must be weighed against the "pole problem" and the fact that the expansion functions are not orthogonal with respect to area weighting on the sphere." In this sense, the author makes a very good case resolving the "pole problem" with the improved DFS, and the result of the very high resolution Galewsky test is convincing. I like the approach and comparison to spherical harmonics and I certainly recommend publication. But I have a few comments below that should be addressed.

[1]As I understand it the author approximates orthogonality and it is strictly satisfied for $m <= 3$. We know that the spherical harmonics are eigensolutions of the barotropic vorticity equation on the sphere. Hough functions as eigensolutions of Laplace's tidal equation go further in providing eigensolutions for atmospheric Rossby and gravity wave dynamics of the linearised primitive equations (see also a recent article Vasylkevych & Zagar, Q J R Meteorol Soc. 2021;147:1989?2007). So the DFS approach is still a deviation from the "normal mode approach", although this does not mean of course that the fundamental modes of predictability are not well captured. It would nevertheless be interesting so see Rossby and gravity waves somehow in separation and the effect of the numerical method on these (and in combination with the time-stepping on propagation speed). For example, one could force a particular set of Hough modes for the shallow water equations and test this ? Also, if the Galewsky test was set simultaneously in the southern and the northern hemisphere (possibly with an onset delay between the two), would one expect more differences between DFS and SH ?

[2] The author focusses on computational performance and memory requirement as a primary reason for the advantage of the DFS to the SH. However, while memory may be an issue in the short term, there is a large trend towards very high memory nodes in the future. The fast Legendre transform (FLT) also effectively reduces the memory requirement. It is wrong to state that the FLT compromises accuracy (see also Wedi, 2014 Phil. Trans. R. Soc. A 372: 20130289). The FLT is not very sensitive to the threshold epsilon which is essentially a 'selection of zeros (that do not need to be computed)' threshold parameter.

[3] In above paper there is also the case made for cubic truncation with increasing resolution, which aligns much better the cost of grid point and spectral calculations in global numerical weather prediction (NWP) and climate models. In terms of grid point calculations, the DFS operates on a latitude-longitude grid (and associated area weighting of the basis functions is similar) and the author makes a case for applying the spherical harmonics filter in practical applications (even if in the idealised cases shown this may not be necessary). But in today's models 50 percent of the computations are done in grid point space (e.g. physics computations, SL advection). Can the DFS be applied on a reduced grid saving 50 percent of these grid point computations ? What would it do to the accuracy ? What is the average grid distance near the pole at 1km resolution with the

latitude longitude grid ?

[4] The differentiation and advantage of methods will not be decided, as done in this paper, on the computational order of complexity (n^3 or $n^2 \log n$) but rather on how well a method computes on accelerators such as GPUs and how well the method parallelises across MPI nodes. In my experience on GPUs the matrix-matrix multiplies, regardless of complexity, are so fast that these parts of the computation in practice reduce to $c{*}n^2$ where $c- > 0$. Can the author say more about the inherent parallelism that may be exploited in the DFS method in comparison to SH models ?

[5] Necessarily this paper needs the mathematical detail to be able to reproduce the results. This is good. However, it would be much more readable by potentially moving some of the repetition (scalar/vector) into the appendix and pointing just out where there are differences. The semi-implicit time-integration is fairly standard now and could also be in the appendix or supplementary material ? I think it otherwise distracts to much from what is new and what is already elsewhere in the literature. I think this will improve readership of this article. Also the discussion on the different grids is a little confusing and may be moved, eg to the beginning of the article ?

[6] On the least squares approach for spherical harmonics, the author may want to refer to appendix A6 in www.ppsloan.org/publications/StupidSH36.pdf, a nice article on SH.

[7] The author compares with a specific implementation of the SH, and refers to the oscillations near the pole and the cost benefit as advantages of the DFS method. This may be read as rather general statements, e.g. in the abstract: "The new DFS model is faster than the SH model, especially at high resolutions, and gives almost the same results." Is the SH model the one used operationally at JMA ? The author states that the oscillations near the pole can be overcome in the SH method, so how relevant is Figure 10 in comparing the two methods ? I would also suggest to slightly rephrase the abstract in light of this comment.

Specific comments:

page 2, line 7 "The FFT ... and is much faster than the fast Legendre transform", this is not necessarily true (e.g. with GPUs) and is purely judged on computational complexity, I would delete this phrase or qualify.

page 10, on the least-squares approach, how do you know the solution found is unique ?

page 25, line 4-5, what does the choice eq 84 imply more generally and thus not strictly satisfying the differential relationships stated ?

page 33, Test case 5 topography can give rise to spectral ringing in the SH model, what happens in the DFS model, did the author test this ?

Figure 11, Does the SH model employ the FLT, what would this look like if it had (e.g. based on complexity arguments?) ? This could be stated explicitly in the caption.