[Reply on RC2]

We are very grateful to the referee #2 for the useful comments. They help to improve the paper quality. Below, we will reply to the specific comments.

[1]As I understand it the author approximates orthogonality and it is strictly satised for $m \leq 3$. 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 ?

Since we do not know how to perform simulations using a particular set of Hough modes, we have run Williamson's test case 6 to simulate Rossby-Haurwitz wave. Figure R3 shows the predicted height after a 14-day integration in Williamson test case 6. The error is similar among the old and new DFS models using Grid [0] and the SH model. The error in the new DFS model using Grid [1] is smallest. This may be because Grid [1] has grid points at the poles, where the minimum height exists, and on the equator, where the maximum height exists.

To set the Galewsky test simultaneously in the southern and the northern hemisphere, we have run the Galewsky-like test case using the north-south symmetric initial conditions created by adding the north-south opposite distribution of height and winds with perturbations in the southern and the northern hemisphere. Figure R4 shows the predicted vorticity after a 6-day integration in the Galewsky-like test case at 1.3 km resolution. The result in the new DFS model using Grid [0] is almost the same as in the SH model. Figure R5 shows Kinetic energy spectrum of horizontal winds after a 6-day integration in the Galewsky-like test case. The results are almost the same for the DFS model and the SH model, but small oscillations appear near the truncation wavenumber in the SH model. The differences between DFS and SH in the Galewsky-like test case are very similar to those in the Galewsky test case.

We will describe these results in the paper or in the supplement.

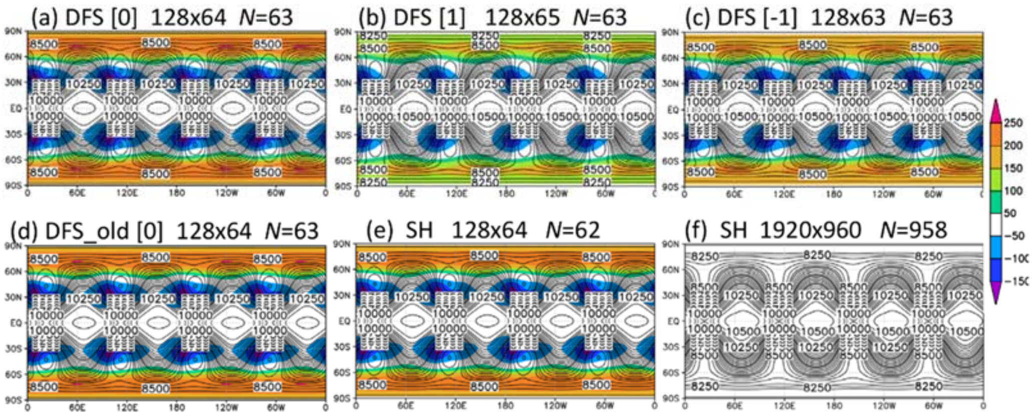

Fig. R3. Predicted height (m) after a 14-day integration in Williamson test case 6. (a) New DFS model with Grid [0]. (b) New DFS model with Grid [1]. (c) New DFS model with Grid [−1]. (d) Old DFS model with Grid [0]. (e) SH model. (f) SH model at high resolution, which is regarded as the reference solution. The number of longitudinal (I) and latitudinal (J) grid points is shown in the form $I \times J$. N is the truncation wavenumber. Color shading shows the error with respect to the reference solution.

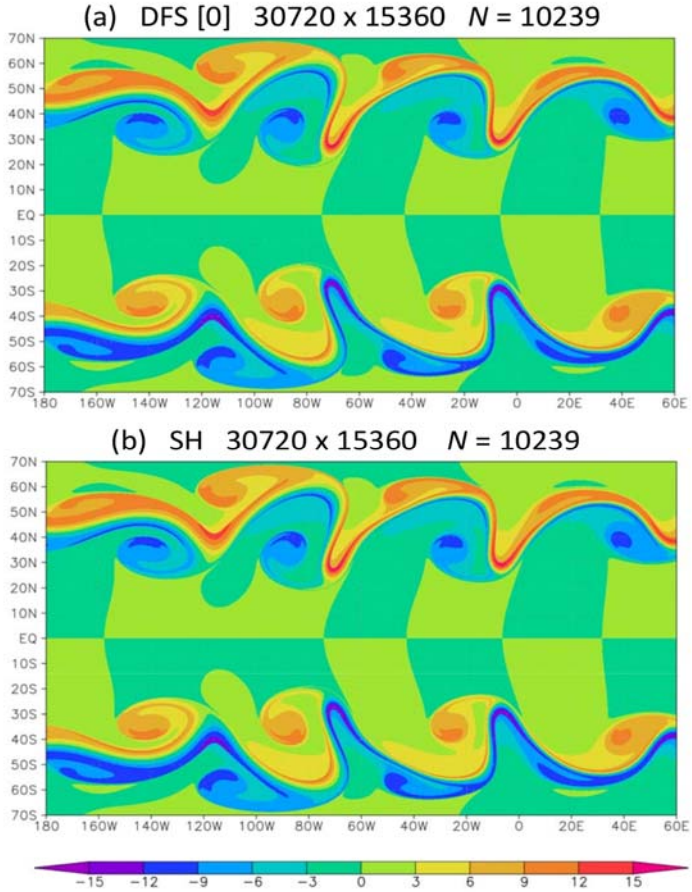

Fig. R4. Predicted vorticity (s^{-1}) after a 6-day integration in the Galewsky-like test case with north-south symmetry. (a) The new DFS model with Grid [0], and (b) the SH model at 1.3 km resolution with 30720 × 15360 grid points and $N = 10239$.

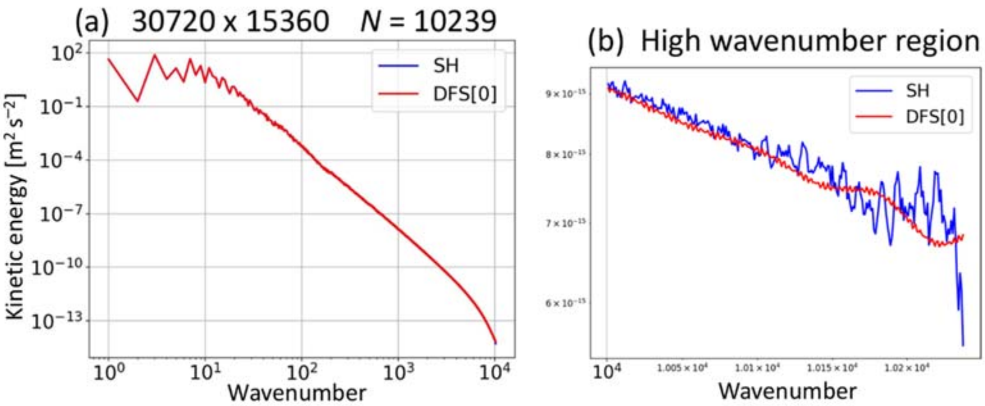

Fig. R5. Kinetic energy spectrum of horizontal winds $(m^2 s^{-2})$ after a 6-day integration in the Galewsky test case. (a) Results of the models with 30720 × 15360 grid points. The colors blue and red represent the models using SH and DFS with Grid [0], respectively. (b) As (a), but showing the high-wavenumber region.

[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.

Thank you for the information about FLT. I will describe in the paper that the FLT effectively reduces the memory usage. I have understood that the FLT does not compromises accuracy. I will describe instead that "the threshold parameter affecting the accuracy-cost balance in the FLT is chosen so that a loss of accuracy is sufficiently small." I will also cite Wedi (2014) in the paper.

[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 ?

Our DFS models do not need the spherical harmonics filter. In our atmospheric DFS model, we use the same fourth-order hyper-diffusion (see Sect. 2.14) as in our atmospheric SH model (Yoshimura and Matsumura, 2005; Yoshimura, 2012). In the paper, we choose a latitude-longitude grid with a zonal Fourier filter in the shallow water model for the simplicity of the source code. However, we can use a reduced grid instead of the latitude-longitude grid. In our atmospheric model using SH or DFS, we use the reduced grid. We will reflect this in the paper.

If we use an octahedral reduced grid, about 50 % of the grid point computations are saved. When the cubic truncation is used, the octahedral reduced grid probably does not reduce the accuracy in the DFS and SH models. The longitudinal grid distance near the pole at 1km resolution with the latitude-longitude grid is about 0.08 m. However, the resolution near the pole is not so high when the zonal Fourier filter is used.

[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 ?

Thank you for providing information about execution speed on the GPU. Since we have not performed our model on the GPU yet, we have described the computational order of complexity and the elapsed time on the CPU in the paper. We should enable the execution of our model on the GPU and test it in the future. The execution on the GPU is a big issue and is beyond the scope of this paper.

Our shallow water model in the paper is parallelized only with OpenMP. We think that when we perform the DFS or SH model with the truncation wavenumber N and K vertical levels using T threads with OpenMP and P processes with MPI, $T \times P$ can be up to $(N+1) \times K$. If the FFT, the Legendre transform, and the matrix calculations such as Eq. (26) are parallelized with OpenMP, $T \times P$ can be more than $(N+1) \times K$. These are the same between the DFS and SH models. In the future, we will write a paper about the DFS atmospheric model with MPI and OpenMP parallelization, where we will discuss the parallelization. We will describe this simply.

[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 ?

Thank you for the advice. We will move some of the repetition and the semi-implicit time-integration into the appendix. We will also move the discussion on the different grids 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.

Thank you for introducing us to a nice article. It is interesting to minimize some form of variational function instead of just the standard least-squares error in order to minimize ringing artifacts.

[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.

Thank you for the suggestion. In the paper, the SH model is the shallow water model using SH. The oscillations near the pole appear in the old DFS model, and do not appear in the new DFS model and SH model. In the SH shallow water model, small oscillations appear near the truncation wavenumber in the kinetic energy spectrum, but not in the new DFS model. This problem in the SH model can probably be solved by using the vector harmonic transform as described in the paper. We will modify "The new DFS model is faster than the SH model, especially at high resolutions, and gives almost the same results" in the abstract as follows:

> The shallow water model using the new DFS method is faster than the shallow water model using SH, especially at high resolutions, and gives almost the same results, except that small oscillations near the truncation wavenumber in the kinetic energy spectrum appear only in the SH model. This problem in the SH model can probably be solved by using the vector harmonic transform which is similar to the vector transform using the least-squares method in the DFS model.

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.

I will delete "much" and describe "The FFT ... and is faster than the fast Legendre transform" in the paper.

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

The new DFS meridional basis functions are not orthogonal but independent. Therefore, by using Gram-Schmidt orthogonalization, the basis functions can be converted to orthogonalized basis functions, where the latitudinal weight is constant. By using the least squares method with the orthogonalized basis functions, expansion coefficients are calculated uniquely by inner product similar to Eq. (34) except that the latitudinal weight is constant. Thus, the solution function like $T_m^{\mathrm{c,SH},N}(\theta)$ in Eq. (37) is obtained uniquely. This unique solution function is the same as that calculated by the least-squares method with the original non-orthogonal basis functions.

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

We use Eq. (84) obtained from the Galerkin method instead of the equation obtained from the least-squares method when calculating f from a given g. We always use the Galerkin method regardless of whether the equation obtained from the Galerkin method is the same as that obtained from the least-squares method. It does not mean “not strictly satisfying the differential relationships stated”. We will add the explanation in the paper.

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 ?

In test case 5, spectral ringing seen in the test caes 5 in Jakob-Chien et al. (1995) does not appear in our SH and DFS models. This is probably because the semi-Lagrangian scheme improves numerical stability.

Jakob-Chien, R., Hack, J. J., and Williamson, D. L.: Spectral transform solutions to the shallow water test set, J. Comput. Phys. 119, 164-187, doi:10.1006/jcph.1995.1125, 1995.

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.

The operation count of FLT is proportional to $N^2(\log N)^3$, and that of the usual Legendre transform is proportional to N^3. However, since we do not know their proportional coefficients, it is difficult to estimate the elapsed time when FLT is used. We will describe in the paper that the FLT will reduce the execution time of the Legendre transform.