

[Reply on Referee Comment 1]

## Comments to the manuscript

### General comments

This paper deals with the Double Fourier Series (DFS) on the sphere where new DFS functions are used to represent the variables on the global domain. Discretization procedures for the spatial differentiations, elliptic equations, and the shallow water models are shown in some detail using the trigonometric identities. Combining the DFS method with the semi-Lagrangian time differentiation, the paper provided simulation results for a couple of shallow water model test cases, including the standard test cases of Williamson et al. (1992). The author emphasizes that the DFS functions in the present study improves the simulation results over the DFS models in the previous studies. However, it is hard to be convinced of the improvement of the new DFS method due to less rigorous assessment of the solution method for differential operators such as elliptic equations, least square method, and also limited test case results. Specific comments are shown below.

We are very grateful to the referee #1 for the useful comments, which help to improve the paper quality.

We reply to the specific comments below.

### Specific major comments

[1] One of the most important aspect of the paper is that the DFS expansion coefficients ( $T_{n,m}^c$ ,  $T_{n,m}^s$ ) are calculated based on the least square method, as is shown section 2.3. I am afraid, however, that the derivation procedure does not seem to be the least square method which is required for determining the expansion coefficients. The residual function here is defined using the difference between the spectral representation of the function ( $T_m^c$ ,  $T_m^s$ ) two different sets of DFS. The fact that, for the spherical harmonics model (SHM), the spectral coefficients are determined in the least square sense on the spherical domain is explained as below:

$$E \equiv \int_0^\pi \left[ T_m^c(\theta) - \sum_{n=m}^N c_n P_n^m(\theta) \right]^2 \sin \theta d\theta \quad (\text{Q1})$$

$$\frac{\partial E}{\partial c_n} = 0 \dots \text{least squares error} \quad (\text{Q2})$$

$$2 \int_0^\pi P_n^m(\theta) \left[ T_m^c(\theta) - \sum_{n=m}^N c_n P_n^m(\theta) \right] \sin \theta d\theta = 0 \quad (\text{Q3})$$

$$\therefore c_n = \int_0^\pi P_n^m(\theta) T_m^c(\theta) \sin \theta d\theta \quad (\text{Q4})$$

$$\left[ \because \int_0^\pi P_n^m(\theta) P_n^m(\theta) \sin \theta d\theta = \delta_{nn'} \right] \quad (\text{Q5})$$

That is,  $c_n$  is obtained with least squared error on the sphere.

The least square method for the SHM in the present study is different from above equations.

We have modified the description about the derivation of Eq. (Q4) (Eq. (37) in the revised paper) using the least-squares method. The referee probably considers the use of  $\tilde{T}_m^{c,N}(\theta)$  instead of  $T_m^c(\theta)$  as a problem. In the SH method,  $\tilde{T}_m^{c,N}(\theta)$  can be used instead of  $T_m^c(\theta)$  (e.g., Sneeuw and Bun, 1996), although the values of  $T_{n,m}^{c,SH}$  calculated from  $\tilde{T}_m^{c,N}(\theta)$  are different from those calculated from  $T_m^c(\theta)$ . In the new DFS method, using  $\tilde{T}_m^{c,N}(\theta)$  instead of  $T_m^c(\theta)$  is not a problem, because the values of  $T_{n,m}^c$  calculated using  $\tilde{T}_m^{c,N}(\theta)$  are the same as those calculated using  $T_m^c(\theta)$ . This is important, and we have described this in Sect. 3.4 and Appendix D in the paper. Moreover, we have changed  $(\sum_{m=0}^M \tilde{T}_m^{c,N}(\theta) \cos m\lambda + \sum_{m=1}^M \tilde{T}_m^{s,N}(\theta) \sin m\lambda)$  to  $T(\lambda, \theta)$  in the residual in Eqs. (25) and (39).

[2] It looks like that the equations (24a)-(24d) are just algebraic equations resulted from simply multiplying  $\sin \theta$  or  $\sin^2 \theta$  or  $\sin^4 \theta$  to the same equation. For instance, in the case of odd  $m (\geq 3)$ , it follows:

$$\begin{aligned} T_m^c(\theta) &= \sum_n T_{n,m}^c \sin^2 \theta \sin n\theta \\ &= \sum_n \tilde{T}_{n,m}^c \sin n\theta \end{aligned} \quad (\text{Q6})$$

$$\begin{aligned} \sin^2 \theta T_m^c(\theta) &= \sum_n T_{n,m}^c \sin^4 \theta \sin n\theta \\ &= \sum_n \tilde{T}_{n,m}^c \sin^2 \theta \sin n\theta \end{aligned} \quad (\text{Q7})$$

$$\sum_n T_{n,m}^c \sin^4 \theta \sin n\theta = \sum_n h_{n,m}^c \sin n\theta \quad (\text{Q8})$$

$$\Rightarrow \underbrace{\begin{bmatrix} 5 - diagonal \\ matrix \end{bmatrix}}_{\text{lhs of (24d)}} \begin{bmatrix} T_{1,m}^c \\ T_{3,m}^c \\ T_{5,m}^c \\ \vdots \end{bmatrix} = \begin{bmatrix} h_{1,m}^c \\ h_{3,m}^c \\ h_{5,m}^c \\ \vdots \end{bmatrix} \quad (\text{Q9})$$

$$\sum_n \tilde{T}_{n,m}^c \sin^2 \theta \sin n\theta = \sum_n h_{n,m}^c \sin n\theta \quad (\text{Q10})$$

$$\Rightarrow \underbrace{\begin{bmatrix} \mathbf{3 - diagonal} \\ \mathbf{matrix} \end{bmatrix}}_{\text{rhs of (24d)}} \begin{bmatrix} \tilde{T}_{1,m}^c \\ \tilde{T}_{3,m}^c \\ \tilde{T}_{5,m}^c \\ \vdots \end{bmatrix} = \begin{bmatrix} h_{1,m}^c \\ h_{3,m}^c \\ h_{5,m}^c \\ \vdots \end{bmatrix} \quad (\text{Q11})$$

**It should be explained why above equations are the same as those the author derived.**

This is interesting. We have changed the derivation of Eq. (30), and we have explained in Appendix E in the paper why above equations are the same as those we derived.

**[3] The largest wavenumber (truncation wavenumber) in (8b) should be determined considering the grid structure, grid[0] or grid[1] or grid[-1] to make completeness of spectral expansion issue clear (refer to Cheong et al. 2004).**

Thank you for the advice. In the old DFS method using the Cheong's basis functions in Eq. (6) with Grid [0], the upper limit of  $N$  is  $J - 1$  for  $m = 0$ , and  $J$  for  $m \neq 0$  (Cheong et al. 2004). We have described in Sect. 1 in the paper as follows:

2. The meridional truncation with  $N = J - 1$  and  $N' = J$  is used, which enables to reconstruct accurately the given grid-data with the expansion coefficients (Cheong et al., 2004) and avoid the error due to the meridional truncation.

We have also described the upper limit of  $N$  in the new DFS method in Appendix C. We have also cited Cheong et al. (2004) in Sect. 2.

**[4] Section 2.12 presents the Laplacian operator and the Poisson's equation. The accuracy of the new DFS method for these basic operators and others such as biharmonic diffusion operator should be addressed with detailed error magnitude. Also important is the global mean associated with the Poisson's equation.**

This kind of accuracy test is important, and we have described the results of the test in Sect. 5. The global mean associated with the Poisson's equation is also important, and we have described in Sect. 3.7 as follows:

In Eq. (65) ( $g = \nabla^2 f$ ), the global mean of  $g$  must be zero because the global mean of the right-hand side of Eq. (65) is zero. Before calculating  $f$  from a given  $g$  in the Poisson equation, we should subtract the global mean from  $g$  (Cheong 2000b). See Eq. (G1) about the calculation of the global mean.

**[5] One of the most basic test case is the cosine-bell advection, which is not included in this manuscript. The test case is simple but useful to demonstrate the advantage and disadvantages of a numerical method.**

Thank you for the advice. Since the advection equation is highly non-linear, it is challenging to solve the equation in the longitude-latitude grid using the Eulerian scheme with the DFS spectral method instead of the semi-Lagrangian scheme. We have run the Williamson test case 1 simulating the cosine-bell advection in the Eulerian and semi-Lagrangian models to examine the advantage and the disadvantage of the DFS methods. The Eulerian advection model using the new DFS method is stable without horizontal diffusion, but that using the old DFS method is not stable. The results of this test are important, and we have described the results in Sect. 6.3. We also change the title of this paper to “Improved double Fourier series on a sphere and its application to a semi-implicit semi-Lagrangian shallow water model and an Eulerian advection model”.

**[6] It is very nice to see that the simulations are carried out without numerical instability even without horizontal diffusion. The author may address why it is possible. Is it due to the diffusive property of the semi-Lagrangian?**

One reason is due to the stability of the semi-Lagrangian scheme, which avoids numerical instability due to the nonlinear advection term. Especially, the old DFS method needs to use the semi-Lagrangian scheme for stability. We have described this in Sect.1.

We have shown in Sect. 4 that the error due to the wavenumber truncation is small in the new DFS method because the expansion coefficients are calculated using the least-squares method (or the Galerkin method) to minimize the error. This improves the stability of the model using the new DFS method. In the new DFS method, probably it is not always necessary to use the semi-Lagrangian scheme from the results of the Williamson test case 1. We have added the following description in Sect. 7:

In the old DFS model, the use of the semi-Lagrangian scheme is important for numerical stability. On the other hand, the advection model using the new DFS method is stable, even when the Eulerian scheme is used instead of the semi-Lagrangian scheme. The Eulerian shallow water model using the new DFS method is also likely to be stable, although we have not tested it yet.

**[7] Figure 2. The problem setting is quite strange. In principle, any scalar function with  $m > 0$  should vanish at poles. Nevertheless, the ‘original’ function is given to have value of unity at north pole. Therefore, the computation and comparison are not meaningful.**

In Figure 2, Grid [0] is used and there are no grid points at the poles. The original (initial) values are set to one at the grid points north of 30°N (except for the North pole). Since the values at the North and South poles are zero due to the pole conditions in Eq. (3), the initial values abruptly change around the North pole. This explanation is necessary, and we have modified the explanation in Sect. 4.

**[8] Figure 5. Result of DFS0 appears to be too much smooth compared to DFS\_old. Why is it?**

This is because the least squares method (or the Galerkin method) is used to calculate the expansion coefficients in new DFS [0]. In DFS\_old, the latitudinal derivative at the grid points can be unrealistically large even when  $N' = J$  for even  $m (\geq 2)$  as described in Sect. 4 (Fig. 3b). The new DFS expansion method with the least-squares method does not have this problem. We have described this in Sect. 6.5.

**Specific minor comments**

**[1] The right hand side of (25) should be represented with matrix-vector multiplication as in the left hand side.**

Thank you for the advice. We have modified it.

**[2] Terms associated with  $\tilde{T}_{1,m}^{cJ}$  and  $\tilde{T}_{1,m}^{sJ}$  in (36) do not appear in (37). The reason should be explained.**

I am sorry this is a typo.  $\tilde{T}_{m=0}^{c,SH,J}(\theta)$  should be  $\tilde{T}_{m=0}^{c,J}(\theta)$ , and  $\tilde{T}_{m=0}^{s,SH,J}(\theta)$  should be  $\tilde{T}_{m=0}^{s,J}(\theta)$ . We no more use this equation in the revised paper.

**[3] Equation (B1) can be found in Cheong 2000a.**

Thank you for the information. We forgot to cite Cheong 2000a. We have cited Cheong 2000a in Appendix G.

[Reply on Referee Comment 2]

**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.**

We are very grateful to the referee #2 for the useful comments, which help to improve the paper quality. We reply to the specific comments below.

**[1]As I understand it the author approximates orthogonality and it is strictly satisfied 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 Vasylyevych & 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 to 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 test case 6 to simulate Rossby-Haurwitz wave. We have described the results of the test case 6 in Sect. 6.6.

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. We have described the results of the Galewsky-like test case in the supplement.

The results of the Galewsky-like test case are similar to those of the Galewsky test case.

**[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 have described in the paper that the FLT effectively reduces the memory usage. I have understood that the FLT does not compromise accuracy. I have described instead that “In the fast Legendre transform, the threshold parameter affecting the accuracy-cost balance is chosen so that a loss of accuracy is sufficiently small.” I have also cited 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 horizontal diffusion as in our atmospheric SH model (Yoshimura and Matsumura, 2005; Yoshimura, 2012). We have described this in Sect. 1.

We can use a reduced grid instead of the latitude-longitude grid. In our atmospheric DFS model, we use the reduced grid. In this study, we use the longitude–latitude grid with the zonal Fourier filter, not the reduced grid, for the simplicity of the source code. We have described this in APPENDIX F.

If we use an octahedral reduced grid (Malardel, 2016), 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. We have cited Malardel (2016) in APPENDIX F.

The longitudinal grid distance near the pole at 1km resolution with the latitude-longitude grid is about 0.08 m. The resolution near the pole is not so high when the zonal Fourier filter is used. Numerical instability occurs in some test cases in the old DFS model without the zonal Fourier filter, but stable integration is possible in all test cases shown here in the new DFS model even without the zonal Fourier filter as described in Sect. 6.2.

**[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  $c \cdot n^2$  where  $c \rightarrow 0$ . 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. We have described in Sect. 7 as follows:

We have executed our shallow water models on Intel CPUs. The execution on GPUs is one important topic, but we have not tested our models on GPUs because the execution on GPUs is not an easy task.

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. (31) 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 have described in Sect. 7 as follows:

MPI parallelization is another important topic. However, in our shallow water models, we use only OpenMP parallelization, not MPI parallelization for the simplicity of the source code.

We will describe the nonhydrostatic DFS model and the MPI parallelization in another paper after improving the nonhydrostatic dynamical core as needed.

**[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 too 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 have moved the description about the semi-implicit time-integration scheme to the appendix I. We have moved the description about the different grids to Sect. 2. Some topics, such as the zonal Fourier filter and the Fourier discrete cosine and sine transforms, are also moved into the appendices. To avoid repetition, we use  $S_{n,m}(\theta)$  to show the basis functions of the new DFS method, and use  $T_m^{c(s)}(\theta)$  instead of  $T_m^c(\theta)$  and  $T_m^s(\theta)$ . We also use the following expression to reduce the number of lines.

$$\begin{bmatrix} \chi(\lambda, \theta) \\ \psi(\lambda, \theta) \end{bmatrix} \cong \sum_{m=0}^M \begin{bmatrix} \chi_m^c(\theta) \\ \psi_m^c(\theta) \end{bmatrix} \cos m\lambda + \sum_{m=1}^M \begin{bmatrix} \chi_m^s(\theta) \\ \psi_m^s(\theta) \end{bmatrix} \sin m\lambda, \quad (45)$$

**[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](http://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 (or the advection 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. The small oscillations in the SH model can probably be eliminated 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 that 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 shallow water model using SH. This small oscillations in the SH model can probably be eliminated by using the vector harmonic transform which is similar to the vector transform using the least-squares method (or the Galerkin method) in the model using the new DFS method.

**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 ?**

We have described the reason that the solution is unique in the following part in Sect. 3.4:

The new DFS meridional basis functions  $S_{n,m}(\theta)$  for each  $m$  are not orthogonal but independent. Therefore, by using Gram-Schmidt orthogonalization, the basis functions can be converted to orthogonalized basis functions  $S_{n,m}^0(\theta)$ , ...  
 ... , we derive

$$T_{n,m}^{c(s),0} = \frac{1}{\pi} \int_0^\pi S_{n,m}^0(\theta) \tilde{T}_m^{c(s),N}(\theta) d\theta. \quad (43)$$

Thus,  $T_{n,m}^{c(s),0}$  and  $T_m^{c(s),0,N}(\theta)$  in Eqs. (43) and (42) are calculated uniquely. This unique solution  $T_m^{c(s),0,N}(\theta)$  is the same as  $T_m^{c(s),N}(\theta)$  in Eq. (29) obtained by the least-squares method with the non-orthogonal basis functions  $S_{n,m}(\theta)$ , because  $S_{n,m}^0(\theta)$  ( $n = N_{\min,m}, \dots, N_{\max,m}$ ) are the linear combination of  $S_{n,m}(\theta)$  ( $n = N_{\min,m}, \dots, N_{\max,m}$ ) for each  $m$ , and vice versa.

**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) (Eq. (70) in the revised paper) obtained from the Galerkin method instead of the equation obtained from the least-squares method when calculating  $f$  from a given  $g$ . Both the equations obtained from the Galerkin method and from the least-square method strictly satisfy  $g^{N,M}(\lambda, \theta) = \nabla^2 f^{N,M}(\lambda, \theta)$  for  $0 \leq m \leq 3$ , but approximately satisfy the equation for  $m \geq 4$ .

Generally, it cannot be said that the least-squares method is superior to the Galerkin method or vice versa. We have described in Sect. 3.7. as follows.

Generally it cannot be said that the least-squares method is superior to the Galerkin method or vice versa, and here we choose to use the Galerkin method because of the reason described above.

**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 have added the description below in Sect. 6.8.

If the fast Legendre transform, which requires only  $N^2(\log N)^3$  operation, is used instead of the usual Legendre transform in the SH model, the difference of the elapsed time between the models will be reduced at high resolutions. We have not tested the fast Legendre transform yet because we do not have subroutines for the fast Legendre transform.

[Reply on Referee Comment 3]

**Comments on “Improved double Fourier series on a sphere and its application to a semi-implicit semi-Lagrangian shallow water model” by Hiromasa Yoshimura.**

**General comments**

**The author proposes an alternative formulation to alleviate the pole problem used in a shallow water model using double Fourier series. The paper presents specific forms for the gradient of a scalar, vectors, and Laplacian and Helmholtz operators. The comparison against spherical harmonics and the new double Fourier series provides insights into the properties of the former. I recommend the publication of the paper provided that the following concern is appropriately addressed.**

We are very grateful to the referee #3 for the useful comments, which help to improve the paper quality. We reply to the specific comments below.

**Major comments**

**As the other reviewers noted, the tests seem to be randomly chosen. I recommend the author to investigate at least the convergence (error vs horizontal resolution) and conservation (energy, vorticity, etc.). Avoid visual comparison where possible and conduct quantitative evaluation. The errors should be given in standard error norms by numerical values.**

Thank you for the advice. We have given the errors by numerical values at several horizontal resolutions in the tables of the paper. We have also given the conservation (height, energy, vorticity) in the tables of the supplement.

**The presentation of the paper can be improved. For example, Subsection 5.2 and 5.4 are short and 5.3 is of two long paragraphs. Author’s intention for the tests should be provided. The author discusses the pole problem, but Figures 4, 7 and 9 are shown in longitude–latitude; Figure 9 omits the polar region. It would be nice to add a diagram to show the differences of expansion visually.**

Thank you for the advice. We have described our intention for the tests in the first part of Sect. 6 as follows:

We ran the Williamson test cases 1, 2, 5 and 6 (Williamson et al., 1992), and the Galewsky test case (Galewsky et al., 2004) in the model using the new DFS method described in Sect. 3, the model using the old DFS method of Yoshimura and Matsumura (2005), and the model using the SH method. By comparing the results of these model, we evaluated the old and new DFS methods. We have shown the northern hemisphere region including the polar region in Fig. 12, and the global

region including the polar region in Fig. S1 in the supplement. Figure 2 explains the differences between the DFS expansion methods. We have also given the same figures for the SH expansion method for comparison. Sections 6.3 and 6.5 have become long because we have a lot of things to describe.

#### **Minor comments**

##### **Page 1, Line 21: more accurate rather than good**

We have changed “the accuracy in horizontal derivatives is good” to “the horizontal derivatives are accurate”.

##### **Page 1, Line 25: $O(N^3)$ memory usage, unless calculated on-the-fly**

We have added “(unless the fast Legendre transform or on-the-fly computation of the associated Legendre functions shown below is not used)”.

##### **Page 2, Line 4: Alternatively (avoid repeating the same word, another).**

We have changed the sentence to “Alternatively, we can use double Fourier series (DFS) as basis functions to reduce the operation count and the memory usage in the global spectral model.”