General response

We were happy to see that the reviewers found the proposed test case useful for model developers. We appreciate the reviewers' comments and found their suggestions helpful in making our proposed test case even more useful. We address all their comments below and have already implemented most of their suggestions in a revised manuscript. Having said that, we are unable to meet the journal's standards with regard to the supplied codes (e.g. "reference implementation") at this time. Our original intent in submitting the codes for computing the initial conditions was to provide a "first aid" rather than a "commercial grade" software package. The code essentially evaluates algebraic relations that describe the initial (u, v, Φ) fields and requires about 150 lines of code to do so. Evidently, the developers can write the required code better than us and do it in a way that better suits their particular needs, e.g. unstructured meshes. Therefore, we will not include the Fortran and Python codes in the assets of the paper and leave the Matlab code only for the sake of reproducibility. We will accept the editor's decision on this matter. In light of the referees' overall favorable reviews, and our efforts to accommodate their suggestions, we hope that manuscript is accepted for publication in GMD.

Following the reviewers' suggestions (see below), we have implemented the changes summarized below in the revised manuscript.

Added subsections in the Results section:

1. A demonstration of the applicability of the test case to a global-scale model (GFDL, RSW model).
2. An examination of the stability of the waves along the lines used in Thuburn and Li (2000).
3. A demonstration of the use of the proposed test case a linear convergence test (Ref. 1).

Removed items:

1. The H=0.5 m waves were removed. The results of the simulations with this value were less robust. In addition, adhering to a single value of H simplifies the test case.

Replaced items:

1. The spectral analyses were found to be too sensitive to be used for assessment purposes and were therefore replaced with the difference between the global means, which unlike the L2 norm employed by Williamson et al (1992) is insensitive to phase speed errors.

Response (in blue) to the referees' comments (in black)

<u>Anonymous referee #1</u>

General comments:

The Rossby-Haurwitz wave described in test case 6 of Williamson et. al. (1992) is known to be problematic. Thuburn and Li (2000) describes these issues and I think that paper should be referenced here, as it was in the previous paper (Shamir and Paldor, 2016).

The findings of Thuburn and Li (2000) on the Rossby-Haurwitz wave are discussed in the revised Introduction. In addition, we added a new subsection to the Results section, where we examine the generation of small-scale features and the stability of the proposed test case, similar to the way it is done in Thuburn and Li (2000).

One issue is that the original initial conditions as specified in Williamson et. al. (1992) lead to wrapping up of potential vorticity contours and the associated generation of small scale features and potential enstrophy cascade. A figure showing the potential vorticity at several times throughout the simulations would be appreciated here to show that this does not happen for this test case.

For the small wave amplitude, A=1e-5 ms^{-1}, used in our test case the potential vorticity is dominated by the planetary vorticity. Therefore, we added Hovmöller diagrams of the relative vorticity, instead of the potential vorticity, which show that there is no generation of small-scale features during the last wave-period of the simulation. See new section 4.3 and 4^{th} columns in new Figs 1,2,4,5,7 and 8.

The other issue with the original test case is that it is unstable. This is demonstrated numerically in Thuburn and Li (2000) by adding some small noise to the initial conditions after they noticed that the solution they computed using a finite volume model on a grid of hexagons and pentagons (i.e. their only non latitude-longitude model) broke down. The errors related to the structure of the underlying grid triggered the dynamical instability. The solutions in this paper have been computed using a regular latitude-longitude grid so I wonder if a similar issue could occur with this test case. I suggest that the authors could check this by either adding some noise to the initial conditions, as in Thuburn and Li (2000), or by running their code on a rotated grid (i.e. with the poles in the midlatitudes).

Similar to Thuburn and Li (2000) we added a small (5% in our MS) uniformly distributed random noise (perturbation) to the initial conditions (IC). The simulations with the perturbed IC demonstrate that after 100 wave-periods the simulated solutions preserve the initial wave structure. In particular, the small-scale features in the initial u, v, Φ and ξ fields smooth out and do not generate smaller-scale features. See new section 4.3.

some papers use the Rossby wave test as a convergence test, using a reference solution from either a higher resolution run or from a different model. Could the analytical solution here be used to test the convergence of a linear shallow water model? This would provide a useful test in between the steady state test case 2 and the other tests that require a reference solution from a higher resolution run.

We added a convergence test of the linear shallow water model, which demonstrates that the "error" decreases "exponentially" as the resolution increases. Following one of the other comments by the reviewer on the sensitivity of the spectral analyses we adopted a different assessment criterion which is also used to estimate the error here. See the revised section 3.2 and new section 4.4.

Also, if the wave is indeed stable it would be a fantastic replacement for test case 6, especially for unstructured grid models, or models that use adaptive mesh refinement, since truncation errors related to mesh topology will have no dynamic instability to trigger.

We hope that the revised version of the manuscript, and, in particular, the addition of noise in section 4.3, is more convincing than the previous version. We, too, view the proposed Matsuno test case as a substitute for test case 6 and hope it is adopted by the community.

Specific comments:

1. pg 4, lines 14-15: I am concerned that different pre-factors lead to less stable solutions - it makes me wonder if the version chosen in this paper is indeed stable to differences in grid alignment.

This is a subtle question. There is no reason why Matsuno's expressions should be more stable. We imagine that the most optimal choice of pre-factors can depend on considerations e.g. the prognostic variables used. For example, it is quite possible that different choices are more optimal for models that use vorticity-divergence. Note also that the different pre-factors originate from the use of the normalized Hermite functions whose amplitudes are bounded (Cramér's inequality), as oppose to the amplitudes of the non-normalized Hermite functions that grow indefinitely as *n* increases. Thus, for large *n* we expect Matsuno's expressions to be less stable numerically. On the other hand, for the chosen n=1 it is unclear whether the difference between the two forms has any effect on there dtability. Finally, while the present choice might not be the most optimal, the simulated solutions seem to be stable for 100 wave-periods.

2. Figure 2:

Is there any reason why the Rossby wave with H=0.5 is less regular than the other solutions?

We are unsure but the wave modes of H=0.5 m were deleted altogether from the revised manuscript.

Why do some of the contour plots have white regions when the values have been normalised so should lie in the range [-1, 1]?

As is stated in the figure caption, the fields are normalized on the global maximum at t=0. Therefore, the white regions correspond to times when the field's global extrema temporarily exceed the [-1,1] contour range. In our opinion normalizing on the global maximum at t=0 and keeping the contour range fixed is the better option. We added a clarification in the text in the paragraph discussing Figure 1.

3. Power spectra: Are these at all sensitive to the sampling frequency? My experience is that the spectra can be very sensitive to this but maybe that is for more turbulent simulations.

The reviewer is right. By sub-sampling our results by factors of 2 or 4 (so as to insure there are at least 2.5 samples per wave-period) it was evident that while the power spectra were generally similar, the results can be too sensitive to be used for assessment purposes. Therefore, we adopted a different assessment criterion, which is also simpler than the spectral analyses. See the revised section 3.2.

4. Supplement: The code provided to compute the initial conditions, while appreciated, could be improved. The authors state that the code will compute the analytic fields on arbitrary latitude-longitude grids but they have assumed that these grids are structured. These codes will not work as written for unstructured meshes, which are becoming more common in the community. The test case is much more likely to be used if these codes could be amended (i.e. they return values given a list of latitude-longitude values). In addition to this, there are some unnecessarily confusing aspects of the code. For example, there is no need to capitalise variable names so the radius of the Earth, which is called
a in the paper could be a rather than A in the code. The is especially confusing since there is also an A in the equations described in the paper. It would also make sense to have H as an input parameter, since this can be varied.

Thank you, but we have decided to leave the computation of the initial conditions to the developers that can do it better than us and do it in a way that suits their particular needs, e.g. unstructured meshes.

Technical corrections:

1. Equation 3b: This is different to that in the code matsuno.py (and I think the code is correct).

Equation 3b and the code are consistent and both are correct! Note that, in the Fortran code for example, in addition to the different pre-factor in line 200, the expressions in lines 193-194 are also different from the text. The expressions in the code are obtained from the ones in the text by taking another $(gH)^{0.5}$ factor outside of the square brackets, so that the pre-factors of $\hat{u}$ and $\hat{\Phi}$ both have (gH) in the numerator, but ω in 3b is divided by $(gH)^{0.5}$. This was also flagged by Referee #2. Clearly, the difference between the code and the text is confusing. In the revised manuscript we change 3b to match the expressions in the codes.

2. Equation 3c: I think this is missing a sqrt around the gH.

Again, Equation 3c is correct! Dimensional consideration suggests that the referee’s suggestion cannot be correct.

With regard to the last two comments, we have repeated the derivation of the expressions in Equation 3 from scratch and derived the same expressions as in the previous version. Also, we encourage the community to implement the test, including different pre-factors and/or different powers of (gH).

Anonymous referee #2

General comments:

1. The manuscript claims (e.g. at the bottom of page 4) that this test case can be used for tropical-channel shallow water models (as presented in this manuscript) and global-scale models. From the manuscript it is not entirely clear that the test will work for global models due to the use of the equatorial beta-plane approximation in the derivation for e.g. the transformations of (x,y) and the wavenumber k. The modeling community (as a 'customer' of this test case) generally works with global shallow water models and tropical-channel model in spherical geometry are extremely rare. It therefore would have been more valuable (or convincing) to present example solutions for a global shallow water model instead of a tropical-channel model. Can the tropical-channel shallow water model also be configured as a global model to demonstrate that the test case works for the whole sphere? Please provide extended explanations or ideally results from a global model.

We added a new subsection to the Results section where we repeat the simulations using a global-scale model (the GFDL, RSW Model). The equatorial channel model cannot be easily adopted to the entire sphere due to the convergence of longitudinal lines at the poles. Therefore, we used GFDL's global-scale model which is spectral. Please see the new section 4.2.

2. Model developments with regular latitude-longitude grids have become very rare over the last decade. More typical grids are now cubed-sphere, hexagonal or icosahedral grid with built-in grid irregularities. The manuscript states that the solutions of this test case are very stable for at least 10 wave periods, which is demonstrated on a regular lat-lon grid. This triggers the question whether this statement will hold for today's models with non-latitude-longitude grids. Another question is whether small perturbations of the initial conditions will disrupt or shorten the stability of the test case. Please provide information on these aspects.

Unfortunately, we are unable to provide results with a non-latitude-longitude grid model. We hope the community will employ the Matsuno test case with such models and comment on the subject.

With regard to the perturbations, we added a new subsection to the Results section where we examine the stability of the chosen waves. As in Thuburn and Li (2000) we added a small (5% in our MS) uniformly distributed random noise (perturbation) to the initial conditions (IC). The simulations with the perturbed IC demonstrate that after 100 wave-periods the simulated solutions preserve the initial wave structure. In particular, the small-scale features in the initial u, v, Φ and ξ fields smooth out and do not generate smaller-scale features. See new section 4.3.

3. As detailed below (points 5-7), the description of the initial conditions is incomplete. In addition, the analytic equations (Eq.(3)) differ slightly from the implementation in the Fortran, Matlab and Python codes. The test is therefore not usable by others in its current form, and the manuscript/codes need to be corrected.

All the required information can be found in the original manuscript, and Equation (3) and the code are consistent and are both correct! Evidently, the original version was not clear/organized enough. We hope that the revised version does a better job at conveying the information. Please see detailed response to points 5-7 below.

Technical comments:

1. Page 1, line 9, also page 2, line 32: Please describe the model as an 'equatorial channel' model.

We now refer to the model as an 'equatorial channel' model as requested.

2. Page 1 line 12, page 2 line 2, page 5 lines 3&7: Generalize the description of the grids. A test case for only 'latitude-longitude' grids will have rather limited use. I think you meant to say that given the location of a latitude and longitude, the initial conditions and analytic solutions can be computed on any grid.

Fixed

3. Page 2, line 24: It is incorrect to say that the term 'baroclinic' is associated with density variations in the vertical directions. A flow with identical density and pressure variations (e.g. for isothermal conditions) is still barotropic. Density and pressure variations need to vary independently of each other.

Rephrased in the revised version

4. Page 3, line 15: What is meant by 'reduced gravity'. The initialization of the test case uses the regular Earth's gravity. Modify.

The reviewer is right. As stated in page 5, line 24 of the original manuscript, we control the speed of gravity waves $(gH)^{0.5}$ by holding g fixed and equal to the Earth gravitational acceleration and varying H. The use of the term 'reduced gravity' originates from the fact that the linearized shallow water equations can also be derived as the horizontal structure equations in a stratified layer (in the linear case with a motionless mean flow), in which case Earth gravity is replaced by the reduced gravity and the layer depth by the equivalent height. In order to avoid confusion, we removed these two terms and in the revised manuscript we now adhere to a "single layer" fluid.

5. Page 3, line 17, also page 5&6 section 3.1: The wave mode n=1 is selected which leads to three distinct real roots in Eq. (2). Two of these roots are selected for the example results, but no equations are given for the Rossby wave root and EIG root. Without this information, the description of the initial conditions is incomplete. Add this information to Section 3.1.

This information was provided in Appendix A of the original manuscript. In the revised version this information is moved to the main text after Equation (2) in Section 2, which is more suitable than sec. 3.1.

6. Page 4, Eq. (3) and text: The manuscript fails to explain the meaning and definition of ψ n . What is the relationship between ψ n and the normalized Hermite polynomials H n ? Without the definition of ψ n the description of the initial conditions is incomplete.

ψ_n equals $\hat{v}_n$. Thus, in the revised manuscript we have decided to remove ψ_n altogether and adhere to $\hat{v}_n$, which is just the latitude-dependent amplitude of the meridional velocity.

7. Page 4, Eq. (3): Eq. (3) seems to be correct, but the Fortran/Matlab.Python scripts use a wrong u_hat calculation. E.g. the Fortran code in line 200 needs to read sqrt(g*H0) instead of just 'g*H0'.

Equation 3 and the code are consistent and both are correct! Note that, in the Fortran code for example, in addition to the different pre-factor in line 200, the expressions in lines 193-194 are also different from the text. The expressions in the code are obtained from the ones in the text by taking another $(gH)^{0.5}$ factor outside of the square brackets, so that the pre-factors of $\hat{u}$ and $\hat{\Phi}$ both have (gH) in the numerator, but ω in 3b is divided by $(gH)^{0.5}$. This was also flagged by Referee #1. Clearly, the difference between the code and the text is confusing. In the revised manuscript we change 3b to match the expressions in the codes.
We have repeated the derivation of the expressions in Equation 3 from scratch and derived the same expressions as in the previous version. Also, we encourage the community to implement the test, including different pre-factors and/or different powers of (gH).

8. Page 4, line 10: State that the amplitude A needs to have units of m/s.

Added – Thank you

9. Page 4, line 25: you imply that the planar wavenumber k is unitless, so that that spherical wavenumber k/(a cosφ 0) has units of 1/m. Please comment and clarify.

The planar wave-number has units of 1/length, while the spherical wave-number is dimensionless. To avoid any confusion we added a subscript ‘s’ for spherical variable and a comment in the text.

Correct typo, should be ‘replaced’. Corrected. Thank you

10. Page 8, line 1: What is meant by the ‘transport form’ of the SWEs? This seems to imply the ‘advective form’. However, the provided equations are in ‘conservation form’.

The reviewer is right, the equations are in ‘conservation form’ - corrected.

11. Page 8, line 10: Explicitly state whether the example model uses diffusion or smoothing/filtering operations for the computations, and if yes, which ones. Should users of the test case try to omit all diffusion/filtering operations in their models when using this test case? E.g. the provided shallow water code contains provisions for a temporal Asselin filter.

The equatorial channel model has no diffusion/viscosity terms. It does contain provisions for a Robert-Asselin filter, but in our implementation the coefficient is set to zero. The global model also contains hyperdiffusion terms, but the coefficient was also set to zero. Please see the revised model descriptions. As is stated in the first paragraph of section 3.1, we consider the choice of diffusion/viscosity terms a modeling choice, but we acknowledge the other approach of specifying them as part of the test case.

12. Page 9, Fig. 2: The value for the symbol φ f is not provided. Add this information.

φ_f is removed from the text of the revised version.

13. Page 9, Fig. 2: It is highly unusual and confusing to see and interpret the flipped Hovmoeller diagrams. Typically, Hovmoeller diagrams list the position along the x-axis and time along the y-axis. I recommend flipping the axes in Fig. 2 to make the interpretation of the Hovmoeller diagrams easier.

To conform to common practice we changed the longitude-time diagrams into time-longitude diagrams.

14. Supplemental material: Please add Fortran/Matlab wrapper codes that will enable the user to create/test the initial conditions. In addition, the codes should not expect to receive regular longitude and latitude arrays, but should be callable for any longitude and latitude position.

Thank you, but we have decided to leave the computation of the initial conditions to the developers that can do it better than us and in a way that suits their particular needs, e.g. unstructured meshes.