

Answers to Reviewer 1

We would like to thank the reviewer for his detailed and thorough review and very informative and helpful comments and hints. These allowed us to significantly improve our paper at several places. Below please find our point by point responses. **In the revised .pdf version our changes are marked in red color (green color for changes with respect to both reviewers).**

(i) The takeaway of this paper is a bit vague: they show that for the normal mode analysis, the energy form yields the expected results whereas the theta form does not. This then allows them to justify having to use the STF stabilization for the theta form. However, when they run the baroclinic instability, they find that both equation sets are unstable for this test case but when filtering is added, the theta form is stable for at least 40 days whereas the energy form is not. They mention nonlinear instabilities but no more explanation is given.

Answer: We agree that our results are a bit ambivalent. So, we only could list the pro's and con's of both equation sets we have found by the shown test cases. In the revised version of the paper, we now tried to give a clearer perspective at least about our plans and the reasonings for them. Essentially, we think that it is worth to follow equation set 'E' further on, due to its otherwise advantageous properties and potential means to overcome the instabilities. Also see our answer about similar concerns of reviewer 2 and our related changes in the conclusion section.

(ii) Could you perform an analysis to get to the bottom of this like you did in the normal mode analysis?

Answer: We have performed several linear stability analysis by a von-Neumann analysis to check basic stability properties of our solver (apart from the normal mode analysis shown in appendix B). These mainly led us to the Courant criteria that we have used (horizontally, vertically, explicit, implicit). These studies would probably need another article for its own and lies clearly beyond this paper. Since we essentially fulfil those (linear) Courant criteria, the instabilities, which occur relatively late during the simulations, lead us to the assessment of 'nonlinear instabilities'. They are probably induced by an accumulation of energy at the smallest scales. A true stability analysis for nonlinear instabilities probably would need an elaborate energy method and is out of scope of our paper, too.

(iii) I also find this baroclinic instability test (the Jablonowski-Williamson) as non-standard for a nonhydrostatic model. The Ullrich test case is designed for nonhydrostatic deep and shallow planet equations (Ullrich et al QJRM 2014). Moreover, this test case should remain stable for a well-constructed atmospheric model.

Answer: we now have implemented and performed this test case, see our answer below to l. 685.

(iv) I also found the use of filters to stabilize the model as non-standard. Why not use either hyper-diffusion (as most models do) or the same stabilization techniques used in ICON to perform a comparison? The issue with filters is that if they are non-idempotent (this is not discussed in the manuscript) then it is problematic comparing models that require different time-steps as the dissipation is not equal (this is not the case for artificial diffusion strategies).

Answer: we agree that filtering is not the most optimal stabilization strategy for DG schemes. Therefore, we had tried to improve it by standard oscillatory sensors. We give an outlook to future targeted improvements in our conclusions. Apart from that see our answers below to lines 571 and 710.

(v) I also found the discussion difficult to follow with too many details that seem less important while omitting references that could have helped the authors either reduce the amount of text and/or

assist them with strategies for analyzing/improving their models.

Answer: We have included most of the citations proposed by the reviewer during the text. We also have partly reformulated our conclusion section (in particular see our answers to lines 743-764)

(vi) It would also be helpful to the reader if they wrote early on what exactly is the novelty of the paper.

Answer: We have added some items that we think are new in the context of atmospheric models for targeted numerical weather prediction and climate research using DG methods at the end of the introduction: mainly the use of prismatic grid cells, consistent use of the covariant formalism on manifolds, and details of the linearization.

(vii) The last point is that the number of degrees of freedom is never mentioned. This is important because it helps to understand exactly how this model compares to other models (e.g., ICON) for the same number of degrees of freedom. Of course it is better to do so for the same error but this requires constructing work-precision diagrams which requires: (1) knowing the analytic solution and (2) running BRIDGES and comparing them to (3) another model (e.g., ICON). It is unclear from just reporting timing whether this is fast or slow.

Answer: We now have added a work-precision diagram for the linear wave expansion test together with ICON runs and a short discussion. See our answer below for l. 720. Related to 'degrees of freedom', see our answer below to l. 715.

Specific Comments

L. 15: "partially solved" should be "be partially solved". Otherwise, the paper is well-written with few (if any) grammatical errors.

Answer: done

L. 130: why does it matter that it has only linear dependence in the 1st and 3rd terms? It is still nonlinear in the 2nd.

Answer: Right, this is not a good place for this remark. And in fact it only doubles the information given in section 3.1.2 (after the equation for $\tilde{p}$) about the linearisation of the pressure. Therefore, we simply skip this sentence.

L. 143: I don't understand why you need a (constant) reference state. Aren't you using a HEVI method? To me HEVI methods imply that the implicit problem is nonlinear, unless you linearize. Please elaborate on this point as this is confusing and permeates throughout the paper. In Giraldo et al JCP 2024 (which you reference), they show the pros and cons of using fully nonlinear HEVI versus various options for linearization - linearization by a constant reference state is not an option because it is not guaranteed to remain stable given large perturbations, real atmospheric states, etc.

Answer: Yes, this needs a deeper explanation as the linearisation approach is also mentioned by reviewer 2. The reference state is only used to improve (to a certain extent) the well-balancing properties of the solver, as mentioned in section 2.2. Now, we add a remark there that it is *not* used for the linearisation in the HEVI scheme. That the reference state is not used for linearisation can also be recognized by the linearisation coefficients given in sec. 3.1.2: they do depend from the actual or at least recent state of the atmosphere but do not depend (in general) from the reference state alone. This is now also mentioned in sec. 3.1.2 by a short sentence. Our methodology is exactly what is described in B21 for the HEVI scheme and is e.g. denoted in Giraldo et al. (2024) as 'LHEVI-PS'. We now have added this information and their assessment of several HEVI schemes in sec. 3.1. Additionally, we cite here (as was done in B21) the article by Giraldo et al. (2013), where the division into implicit linear and explicit nonlinear terms in a consistent manner is mentioned. Around eq. (21) we also emphasize that the linearisation coefficients can depend

from the prognostic variables, generally taken from a previous time step (but we omit the term 'depend only weakly', since this is not entirely correct) and that the remaining, explicitly treated, terms are mainly nonlinear.

L. 180: to my knowledge using a local reference coordinate on triangular DG elements is first discussed in the context of a spherical shallow water model in Laeuter et al. JCP 2008 (A discontinuous Galerkin method for shallow...).

Answer: Yes, this is probably the first occurrence of this idea, at least in a DG context. This work, of course, is cited in B20, but it should definitely be mentioned here again (is done now).

L. 197: why do you choose shallow planet? Deep planet is straightforward for you to use, is it not? And more general.

Answer: Right, a deep atmosphere would be more straightforward. We decided to concentrate on the shallow atmosphere for several reasons. First, due to efficiency reasons, it is still the standard choice in our forecast model ICON and (as far as we know) still used by the most global weather forecast models worldwide (we now mention this aspect at the end of sec. 2.3) Therefore, as a demonstration step towards a later ICON implementation we want to 'offer' this approach, at least to our decision makers, too. Secondly, some of the test cases, and most important, the analytic solution by Baldauf et al. (2014), which we use for convergence tests, explicitly uses the shallow atmosphere. Thirdly, we wanted to show how simple the shallow atmosphere approach is by just changing metric terms (instead of *omitting* metric terms as it is usually done in the literature). Simulations using the deep atmosphere are planned for later publications.

L. 210: Section 3.1 is confusing to me. E.g., how is the nonlinear part treated linearly? E.g. in line 248 it is mentioned that "practical simulations reveal" that you only need half of the KE contributions. What do you mean by "practical" and how did you come up with this factor? What happens if you use the complete factor and what savings are you getting by omitting them?

Answer: As said above to our answer to l. 143, our linearization strategy is exactly what is done in B21 and is e.g. denoted in Giraldo et al. (2024) as 'LHEVI-PS'. We admit that our description should be improved. We now have skipped the whole sentence with '... practical simulations reveal ...' around l. 248, because it is essentially repeated later on. We have partly reformulated the paragraph after the linearisation formula (old lines 259-265), where we clearer state which neglects are just due to efficiency reasons and which neglects are due to numerical necessity. The omission of these terms and this factor 1/2 are necessary to avoid artifacts (overshooting of vertical velocity) near the top and bottom boundaries. We have not really a proof that '1/2' is exactly correct, but it just solves any boundary problems and at least can be motivated by stemming from an essentially quadratic term, namely the kinetic energy.

L. 260-265: I'm confused by this discussion. If you are indeed using a nonlinear HEVI method then you need all the terms in order to satisfy the PDE. However, if you are using linear HEVI and only add some of the stiff terms into the implicit linear operator (and have the full nonlinear operator in the explicit part) then your omission of terms makes sense. Please elaborate on whether the HEVI method is linear or nonlinear or better said, whether you first build the full nonlinear operator explicitly and then add/subtract linear terms. See, for example, the HEVI method described in Giraldo et al JCP 2024 (for the nonlinear form) versus the linear form in Giraldo et al SISC 2013.

Answer: as mentioned above in our answer to 'l. 143' we use 'linear HEVI' (denoted as 'LHEVI-PS' in Giraldo et al (2024)) and do not treat nonlinear terms in an implicit manner (unless they are linearised). And, in fact, this is just what we do in the BBRIDGE code: we first build the full nonlinear flux and source terms and subtract from these the linear implicitly treated terms as in

eq. (21). In the manuscript we add the information in sec. 3.1 that the symbols f and S without lower index are the complete, fully nonlinear flux and source terms, respectively.

L. 280: is your choice of how often to construct the linear coefficients in the same vein as what is discussed in Giraldo et al JCP 2024? Also does including the terms that you omit keep your solution stable for longer? Would be an interesting experiment although performing the analysis to see which terms are more important would be better. E.g., you can see which terms have large growth. Can you be sure that the differences in the two equation sets is not due to your linearization choices?

Answer: Yes, the construction frequency of the linear coefficient should be the same as is done in Giraldo et al. (2024). It is also our experience that updates every 50 time steps is sufficient enough (as was found in B21). However, for longer time steps some reduction seems to be needed (also shown in Giraldo et al., see their table 2 for one of the LHEVI simulations). We have performed such a sensitivity analysis as you suggest and have updated the paragraph after the linearization coefficients (old lines 259-265) accordingly. We also have performed an explicit simulation for the baroclinic instability test (with a rather coarse resolution and a quite small timestep) and compared it with an analogous HEVI simulation: both runs crashed roughly after the same time (although very different time steps are used). This also shows us that the simplifications in our linearisation should be acceptable. Also see our answer to a related question by reviewer 2.

L. 305: Perhaps the crux of my confusion on your time-integration stems from your discussion on linearization of the fluxes, etc. In Reddy et al JCP 2023 on IMEX forms for DG methods, the same equation sets that you study here are analyzed there for atmospheric flows. The focus of that paper is exactly on constructing the numerical fluxes in the right way to handle IMEX DG methods. Might be worth reading this paper. Maybe it's not relevant to your discussion but I am not sure I understand this section of your paper.

Answer: Thank you very much for suggesting this article to us. The discussion there of the two equation sets is a good complement to our introduction and we now mention it there. The linearisation strategy is a bit different to ours, e.g. we treat parts of the source term in an implicit manner, too. This probably leads us in general to a slightly different approach and denotations. Treating buoyancy implicitly helps to stabilize vertical contributions of gravity waves, too. This is a property that we found helpful in our old COSMO model. They also seem to linearize around a constant reference state, which we do not. Very interesting is their observation that the non-dissipative nature of an energy-conserving scheme leads to potential sources of instability. In this respect we have cited this work in two places, too (in sec. 6.6.1 and in the conclusions).

L. 318: You finally write out BR1 here but use BR1 previously. Please define it the first time you use BR1 in the paper.

Answer: Thank you for pointing this out. In fact, we had already introduced the denotation in line 164. So we can skip this additional reference in l. 318.

L. 325-330: Why did you choose SSP time-integrators? Without limiters in the spatial part with DG I'm not sure you will get much benefit. In addition, you may be able to find non-SSP time-integrators with larger stability regions that would then allow your models to likely be faster. Not a major point but might be worth considering.

Answer: It is right that we don't really benefit from the SSP property. This has been mentioned in the previous paper B20 (sec. 3.2, there). However, we agree that one should mention this here again. Our choice of SSP3(3,3,2) (or sometimes even SSP3(4,3,3)) lies in the fact that its stability properties are very good. This can be seen if one performs an analogous analysis as has been done in Giraldo et al. (2024), Fig. 3, 5 (not shown here).

L. 331: best to explain better. By block tridiagonal you mean that a point on a column is dependent on the element above and below (3 elements) and the full polynomial space in the vertical. Please add the bandwidth to be clear. It should be something like $2N+1$ (for a nodal DG method, where N is the degree of the polynomial).

Answer: Yes, this is what we mean. As you suggest, we now add the size of each block matrix. However, to introduce the denotations properly, we decided to first move the paragraph in (old) lines 332-338 into a new subsection 3.3.3 and made text changes there. In B21, in fact a band-diagonal solver was used (together with a complexity analysis using the bandwidth which in fact is something like $2N+1$). Here, we use a true block-tridiagonal solver, therefore, we think we shouldn't give a bandwidth but rather the block size.

L. 332-335: The discussion in this part of the paper is confusing. I looked at the B21 paper which states that you can use either quadrilateral or triangular nodal bases. In the current paper, you are only using triangles and if they are nodal then you are likely using Proriot-Koornwinder-Dubiner orthogonal polynomials to construct the Lagrange polynomials. To do so, you need interpolation points (usually Fekete points) and then need a separate set of points for integration (Gauss). My confusion is your mention of collocation: there are no set of points that are both good for interpolation and integration on the triangle therefore collocation is not possible on simplices (see, e.g., Helenbrook SINUM 2009), only on tensor-product bases derived from the one-dimensional Lagrange polynomials which permit collocation via Lobatto or Legendre points. Please clarify this section and suggest removing text that does not apply to what you are using in the paper. I also don't agree that the horizontal directions decouple when using collocation. In DG, the fluxes always couple edge-neighbor elements.

Answer: this first might be a problem with our nomenclature: we mean 'nodal' as opposite to an orthogonal 'modal' base, but we don't necessarily use 'nodal' in connection with only Gauss-Lobatto points (or generally using interpolation points at the cell edges). As you say, we construct the Lagrange base on triangles via PKD orthogonal polynomials. We completely agree about your statement that a *Gauss-Lobatto* like quadrature rule on triangles does not maintain the correct approximation order (and this is in our understanding one main statement in Helenbrook (2009)). However, throughout the whole paper we use the standard Gauss-Legendre quadrature in the vertical and the Williams, Shunn, Jameson (2014) quadrature rules on the triangle. This paper has been cited in B20, but it is probably reasonable to cite it here again. The latter work gives quadrature points only in the interior of a triangle and can therefore be denoted as the equivalent of Gauss-Legendre quadrature (in contrast to Gauss-Lobatto) on triangles. We agree that the optimal point set of interpolation points for high approximation order is not identical with a quadrature point set. However, for the relatively low order that we use (4th order in the horizontal) we wouldn't expect bigger numerical problems. Taking all this together, we think that using collocation with the above quadrature point set is reasonable also on triangles/prisms.

About the 'coupling in horizontal directions': we meant that the *coefficient matrices* of the LSE decouple, but of course not the complete DG scheme. We have now explicitly written this and together with the information of these matrix sizes we hope that this point is clearer now.

L. 343: I would also add "strong" scaling to a good property of DG methods.

Answer: We have included this in the text. Indeed, DG methods are often praised for both, but especially for strong scaling in high-performance computing.

L. 370: stating what is done in collocated methods is a moot point since you are not using these methods.

Answer: The implementation in the BRIDGE code is sufficiently general to support both collocated methods and general nodal and modal basis functions. However, for efficiency reasons, the

vertically implicit scheme in particular exploits the collocation property. Therefore, all numerical experiments presented in the manuscript employ collocation points in the horizontal direction.

L. 402: Again, it matters whether you are using modal or nodal bases. If you use collocation in a nodal method then it is diagonal but at the cost of under-integration of the mass matrix. If you use modal, you are only diagonal if the elements are straight-edged/flat. This will not be the case on the sphere as your Jacobians are not constant. So, representing the mass matrix accurately will prevent the mass matrix from being diagonal.

Answer: In a nodal DG method, the unknowns are represented by their values at interpolation nodes. As you have pointed out, this can in general lead either to a non-diagonal mass matrix or, if the exact integral is replaced by an insufficiently accurate quadrature rule, to an under-integrated mass matrix. However, this caveat commonly applies to Gauss–Lobatto collocation. In our case, the nodal DG basis is defined at Gauss quadrature nodes, and we employ the corresponding Gauss quadrature. Then the mass matrix is diagonal and due to non-polynomial metrics it is in general under-integrated (has been added to the text)

$$M_{ij} = \sum_{q=1}^N w_q \ell_i(\xi_q) \ell_j(\xi_q) \sqrt{g}(\xi_q) = \sum_{q=1}^N w_q \delta_{iq} \delta_{jq} \sqrt{g}(\xi_q) = w_i \sqrt{g}(\xi_i) \delta_{ij},$$

with the covariant metric tensor determinant $g(x)$.

L. 420: See "Adaptive hurricane simulations" for h-refinement in a CG model (Tissaoui et al. JAS 2025). Conceptually, it's no different (perhaps even harder with h-refinement than p-refinement). I don't understand the issue you are describing here.

Answer: Horizontal mesh refinement could likewise be used to increase the resolution. However, as noted by the reviewer, h-refinement is technically challenging, particularly with respect to parallel load balancing. In this context, p-refinement appears to be an elegant adaptation strategy, as it leaves the computational mesh unchanged. It is indeed possible that the order of accuracy varies between different columns of prism elements, and we have revised the text accordingly.

The discussion in this paragraph, however, addresses a slightly different issue, namely whether different variables can be approximated by finite element components of different polynomial order. What makes the DG-HEVI approach computationally feasible in the BRIDGE code is the decoupling of the implicit systems at each horizontal point. Since the finite element contributions are evaluated at these points in the form of coupled terms, each term in the equation generally involves a mixture of different solution components. This approach therefore remains efficient only if no intermediate interpolation is required and all components share the same horizontal points. In the vertical factor of the tensor-product finite element space, however, p-refinement remains possible.

L. 440: Good point. This is why this type of RK method is preferred.

Answer: :-)

L. 455: Not sure this statement is necessary but if you want to include this, you should add that the value of a weak BC is that the BC will be satisfied to the same order of accuracy as the interior solution.

Answer: Thank you for this hint! We think that this is indeed worth mentioning.

L. 521: Is this reference state constant? This does not work very well for real weather problems. This is the reason why using, e.g., the previous solution is typically the method of choice in HEVI methods. See, e.g., Giraldo et al JCP 2024.

Answer: Yes, it is constant, but not used for the linearisation. We agree, see our answer above to 'l. 143'.

L. 550: For completeness, should mention Vandeven's theorem. See, e.g., Vandeven JSC 1991.

Answer: A citation has now been added to the manuscript; see the remark below.

L. 571: I agree. Filtering is something we have moved away from. One other problem that you didn't mention is that most filters proposed are not idempotent. Therefore, changing the time-step changes the filtering - unlike using hyperdiffusion which is time-stepped. This, I think, is one of the main issues with filtering.

Answer: We thank the reviewer for highlighting the importance of a time-step-independent interpretation of the method and for raising the implications of using a spectral filter. The exponential filter employed in the BRIDGE code is not a spectral projector. It preserves the mean mode and smoothly damps the higher modes, rather than acting idempotently. We have added a short remark to the manuscript to clarify this point and included a citation to the classical reference by Vandeven on filters in spectral approximations. We have put this remark directly before the paragraph discussing convergence order because it touches related aspects. At the same time, a direct one-to-one comparison with the ICON model is difficult for several reasons. Moreover, in our geophysical setting, mass and energy conservation is of greater importance than approximate idempotence. For this reason, we chose to stabilize the method using this simple spectral filter, which we consider less intrusive than a generic hyper-diffusion term. More sophisticated stabilization strategies are certainly worth investigating in future studies.

L. 585: Is there a reference for this case? I don't understand how you use the reference state.

Answer: Now this isothermal reference state is detailed in sec. 2.2.

L. 587: do you mean w here? It's 1D in space but not in the equations?

Answer: No, it is in fact u meant here (we now added the term 'horizontal base' velocity). However, we agree that the description as such is misleading. Though we consider here a 1D 'column' model, we have the full Euler equations not only using w but also u and v. The relevant point is to initialize with w=0. Now we start this part with the hint to use a steady atmosphere *at rest* for the initialization. (The hint to possibly set u/=0 is now put as a short side remark into brackets; As said, choosing u with any finite value is not relevant for this test, it just demonstrates that there isn't anything wrong with the free slip boundary conditions.) We also slightly changed the denotation of the test: instead of 'only 1D vertical Euler equations' we now write 'Euler equations in an only 1D vertical column setup'.

L. 610: Convergence is difficult to see unless you plot the theoretical convergence slopes in the plots.

Answer: yes, our formulation can lead to misunderstandings. This short section only should clarify that proper convergence rates are achieved if stability occurs (according to the Lax Richtmeyer theorem), independent on the well-balancing issue. But we don't think that it is worth to show an own convergence plot, here. For this we have the much more demanding and more general convergence test in sec. 6.5. To clarify this we have slightly reformulated this paragraph, indicating that we don't show convergence results here.

L. 625: I am surprised about this result. I have not seen this behavior in our models.

Answer: Thank you for pointing this out. We had looked only at time steps for which the output times are (simple) integer multiples. Therefore, this result appears a bit too harsh. Now, we have sampled the possible time steps with a higher accuracy and in fact we can run equation

set 'rho*Theta' with 0.23s whereas equations set 'E' allows a slightly higher time steps of 0.26s. This is now mentioned.

L. 643: Why don't you expect a similar behavior between the solution of the two equation sets?

Answer: Right, the term 'unstationary' is not clear enough. We simply wanted to express the fact that flow over such high and steep mountains becomes highly irregular or even chaotic and therefore we cannot expect closer similarity. This is the reason why we have plotted two forecast times (23h and 24h) to give an impression of this temporal behaviour. However, apart from this fundamental source of irregularities, one can recognize different flow pattern for both equation sets that probably come from the different conservation properties. We replaced the term 'highly unstationary' by 'chaotic flow'.

L. 648: Please state the resulting Courant numbers.

Answer: Done

L. 653: Would be good to show a plot of mass conservation for the simulation. Probably best for the longer simulation cases though like the Baroclinic Instability or Held-Suarez.

Answer: Thank you for this suggestion. Indeed, a time series plot is more instructive and also lead us to consider this issue a bit more in detail (by the way, we formerly noted the energy conservation error erroneously by a factor of 10 too high). We decided to show such plots (new Fig. 9) for the baroclinic instability case because here we expect not only mass but also energy conservation (in contrast to the Held Suarez test that uses energy producing artificial source terms).

L. 659: please elaborate on the plotting range comment - I don't understand.

Answer: Thank you for this question. Our procedure was in fact too complicated by trying to exactly follow Baldauf et al. (2014), who calculated the errors on a latitude-z slice. To do this we wrote an external interpolator script that produced artifacts at the borders of this slice near the poles and near the upper and lower boundaries. To avoid these interpolator artifacts we defined those a bit odd limits. Now we simply calculate the error measures in the full 3D spherical shell, i.e. for *all* output points (=quadrature points in every cell volume) of the BRIDGE code (the Python script doing these calculations/integrations is now much, much simpler!). Of course, the error numbers changed a bit (in particular for the coarser resolutions), but the big picture about the convergence properties remains the same. Therefore, we reproduced figure 6 and rewrote (i.e. simplified) the explanation of how to calculate the errors.

L. 673: Spatial error will dominate temporal error for a sufficiently small enough time-step regardless of the order of the time-integrator.

Answer: This is right. Although we cannot be sure that the time step is small enough for this argumentation, because it is chosen merely by stability requirements, it is obviously the right explanation. We have extended this sentence by mentioning 'the time step is small enough' and 'that spatial convergence order dominates over temporal convergence order.'

L. 676-681: unclear

Answer: We have completely reformulated this paragraph.

L. 685: Why use this case? Why not the Ullrich BI test case which is designed for z-coordinate models and can handle nonhydrostatic deep-planet equations?

Answer: We just think that the Jablonowski, Williamson (2006) test setup is still used by researchers and has a long tradition in dycore developments. However, we followed your advice and added the Ullrich et al. (2014) setup. We agree that it is much easier to implement in a dynamical core using the fully compressible, non-hydrostatic Euler equations and a height based

coordinate (took about half a day compared to several days for JW2006). However, we don't expect larger fundamentally different results since the initial state has quite similar properties concerning the strength and shape of the two jets and the overall temperature distribution. There is now a new sec. 6.6.2 with this test and a new figure 11.

L. 710: But without filtering it is better and for other tests, you showed that the theta version requires extra stabilization. If you run the Ullrich baroclinic instability with hyperdiffusion it should stay stable forever. See Giraldo et al. JCP 2024 where simulations are shown for over a 100 days. It would be useful to do this and avoid filters.

Answer: We agree that hyper-diffusion is an interesting stabilization procedure, since it seems to be relatively cheap and acts like a physical process (in contrast to filtering). Due to the relative short time that we had for the corrections of the review we only could do a quick and dirty implementation of the Giraldo et al. solution and we didn't succeed. However, this is something that we will try with more care in the near future. However, a general detrimental property of hyper-diffusion (at least if one does some simplifications to implement it in a computational efficient way for DG schemes) is the violation of conservation. Now, we mention the option of hyper-diffusion in the conclusions.

L. 713: This is not a conservative method. In 10 days you are at 10^{-15} mass conservation and at 100 days at 10^{-14} . Not sure why your model does not conserve mass but this could be a concern. How do you compute your metric terms? Giraldo et al JCP 2024 discuss conservation and metric terms.

Answer: We agree. Although the conservation error is extremely small and practically irrelevant (in the new version of the paper we illustrate it with a 1000 year climate run) it is an annoying property. We don't do anything special about the metrics that would go beyond what is described in sec. 2.3. In particular, we calculate metric properties (metric tensor, Christoffel symbols, and transformation matrices) just by the transformation rules (given after eq. (20)) in the sequence of the described coordinate systems. We also have added a plot (new figure 9) for a run without filtering and here no linear trend occurs any more, neither for mass nor for energy. This is another argument to get rid of filtering in a future perspective of our developments.

L. 715: How many degrees of freedom are you computing in these simulations?

Answer: An ICON R2B3 grid uses 5120 grid cells horizontally; with 10 vertical levels we end up at 51200 grid cells. With 4th order horizontally and 5th order vertically we use $10 \times 5 = 50$ base functions for each grid cell (and variable). This means 50 quadrature points in the cell and $3 \times 4 \times 5 + 2 \times 10 = 80$ quadrature points at all prism faces. For the 5 Euler variables this means in total $51200 \times 50 \times 5 = 12.8 \text{e}6$ degrees of freedom, calculated on $51200 \times (50 + 80) = 6.656 \text{e}6$ quadrature points. A slightly reduced version of this text is now added after l. 715.

L. 720: What is the time-to-solution and how does this compare to the current ICON model? Why not use hyper-diffusion for stabilization to make more apples-to-apples comparisons with the current operational model (or whatever stabilization ICON uses).

Answer: we have added a work-precision diagram for the Baldauf et al. (2014) test case (which possess an analytic solution) in sec. 6.5 and compare this with ICON (using the same computer). Of course, for the well resolved fields of this test, high order methods must by far outperform a lower order model like ICON as can be seen in the new figure 7. This is discussed in brevity. Additionally, we give a runtime comparison between BRIDGE and ICON for the baroclinic instability test at the end of sec. 6.6.1.

ICON uses a whole bunch of stabilization methods: 4th order hyper-diffusion is surely one of the most important ones, but off-centering in the implicit solver, and 2D and 3D divergence damping

are also applied.

L. 725: what's the Courant number?

Answer: This question goes into a similar direction as a question by reviewer 2. In the baroclinic test case, sec. 6.6, we didn't use vertical grid stretching, so even larger vertical velocities are tolerated with $\Delta t=80s$. Therefore, the time step is mainly determined by the horizontal grid spacing of $\Delta x=315km$, this is now noted in the new paper version. In the Held-Suarez test we used a vertically stretched grid (to make this test a bit closer to more realistic setups used e.g. in numerical weather prediction). Thus mainly the (expected) vertical velocity w limits the stability of the explicit part of the HEVI scheme, so that here the time step is not any more limited by the horizontal grid spacing but mainly by w . This is demonstrated by (partly estimated) Courant numbers now given for both test cases.

L. 735: Perhaps BRIDGES is not expected to be used for climate simulations but isn't this loss in mass a concern?

Answer: It is (at least for double precision runs as we do) indeed of minor concern even for most climate applications simply due to its relatively tiny trend. However, please see our answer above for l. 713.

L. 743: you state that an important goal is to achieve local conservation of the variables. However, I don't believe you have satisfied this with respect to mass. Is that a concern?

Answer: We indeed think that *local* conservation of the prognostic variables should be an important property of modern dynamical cores simply to avoid unphysical production or destruction just by transport and this is one reason why we are interested in DG methods. Due to the tininess of the linear trend in the global mass and energy change we would say that this property of our DG solver is still given. Global conservation (although related to local conservation) is a different issue and in principle could be achieved also by other means. However, as we have now figured out (see our answer above to l. 713), our DG solver itself does not have a conservation problem but rather the filtering method.

L. 755: Understandable. However, you used a stabilization technique that is non-standard (non-idempotent, etc.). Since you DID need additional stabilization why then not use a standard method (i.e., hyper-diffusion) which is well understood and used by most operational models?

Answer: Please see our answer to l. 710. We now also mention the textbook of Giraldo (2020), who gives a larger overview of possible stabilization mechanisms for DG schemes.

L. 760: The method proposed in Souza et al is kinetic energy preserving - not entropy conserving/dissipating like the work in Waruszewski.

Answer: Thank you for this clarification; now we have removed this reference here. Indeed, also due to a question of reviewer 2 to this paragraph, we have reformulated some parts of this paragraph and can be slightly more precise in describing previous work.

L. 760-764: you ran the Jablonowski-Williamson BI, Waruszewski et al ran the Ullrich BI. These are different cases so not sure I would conclude anything from different tests.

Answer: As mentioned above in our answer to l. 685, both setups are not fundamentally different concerning the initial state. Therefore, both should work comparably and this is essentially what we got for both tests.

L. 762: Entropy stable methods have already been formulated for triangles by Jesse Chan - likely extendable to triangular prisms. See Jesse Chan, On discretely entropy conservative and entropy stable discontinuous Galerkin methods, Journal of Computational Physics, Volume 362, 2018, Pages 346-37.

Answer: Thank you very much for this hint. We have added a sentence with this reference in the Conclusions.

L. 764: The work of Montoya et al 2025 won't help much since it uses collocation. You need to read the work by Chan.

Answer: The work of Montoya et al. is a first step to treat equations in *covariant* form (here the shallow water equations) by these provably stable approaches. It is planned by these colleagues to extend their work towards the Euler equations in our chosen covariant form. However, you are right that Chan's work is a crucial step.

L. 785: But is this the true setup you used in the experiments? I don't think so since you mention a horizontal velocity. How then does this apply?

Answer: We see that our remark about the horizontal velocity setting in sec. 6.1 is more misleading than helpful. See our answer to l. 587 ... therefore, this is in fact the analogous setup to sec. 6.1.

L. 785: This normal model analysis shows that the theta form is unstable whereas the energy form is not. Or rather that the energy form is stable in the stable regime and unstable in the unstable regime. Is this an issue with the PDE, numerics or the linearization? The theta equation set is standard in many models (MPAS-A) and have not heard about such issues before. Which equation set does ICON use?

Answer: From the pure PDE's, one can derive the stability criteria of a steady, stratified atmosphere, and these are of course the same for both equation sets 'E' and 'rho*Theta'. It is in fact the numerical discretization by a DG scheme that makes the difference (the linearizations used here are the same as described in sec. 3.1.1 and 3.1.2, otherwise they are quite trivial in this simple 1D case). We also have cross checked that our BRIDGE code, applied to just one grid cell, produces just the behaviour predicted by this numerical normal mode analysis.

Answers to reviewer 2

We would like to thank the reviewer for his thorough review and his encouraging words. These hints helped us to improve our paper significantly at several places. Below please find our point by point answers. [In the revised .pdf version our changes are marked in blue color \(green color for changes with respect to both reviewers\).](#)

1. The most important issue is general. The authors compare here two different formulations of the Euler equations to be discretized with a classical DG approach with HEVI time integration, one using weighted potential temperature as prognostic variable for the energy equation, the second using total energy instead. The authors choice of studying the total energy formulation is mainly motivated by local conservation requests and the results shows a general better well-balancing properties of the total energy formulation over the potential temperature ones. Another important aspect is the fact that the total energy-based formulation does not require any STF filter to be applied in any of the tests presented in the article, differently from the potential temperature-based formulation. However, it is also shown that the total energy formulation suffers from nonlinear instabilities which are much stronger than in the potential temperature-based formulation. I especially refer to the results of the baroclinic instability test. What is missing here is a clear plan about how to address these stronger nonlinear instabilities of the total energy-based formulation to make it the option of choice. This also affects the main conclusions of the paper: it is not clear to me at this point if the best choice for the BRIDGE project will be the total energy-based formulation or if the

nonlinear instabilities found with it will preclude this conclusion. A clear message on this would be needed.

Answer: Right, this should be clearer formulated. We now have partly reformulated our 'outlook' part of the 'Conclusions' section. The main message is that we want to follow further on equation set 'E' due to its otherwise promising properties. We tried to express in a clearer way our next research steps in inspecting the entropy stable schemes. Also induced by reviewer 1, we have extended the discussion about stabilization schemes in the 'Conclusions'. Additionally, we shortly explained that for the near future we can work with the stabilization that turbulence schemes offer in more realistic model setups (this is at least the experience that we could get until now with a TKE scheme, but couldn't mention it in this report).

2. Talking again about the baroclinic instability test, I was wondering why the authors use a grid spacing which is 5 times larger than by the benchmark model: is this to run a simulation at the same "equivalent" resolution of the benchmark model, in the sense that it takes care of the fact that internally to each grid cell there is the $p=4$ polynomial representation with the 5 points? If this is the case using a 5 times larger resolution would be still correspond to coarser resolution than in the benchmark model considering the repeated degrees of freedom at the element boundaries. Please could you explain more on this?

Answer: Right, we haven't motivated this well enough. We have added a general motivation sentence about the required properties of new dycore developments (accuracy, efficiency, ...) in sec. 6.6.1. This leads to the requirement of inspecting as coarse as possible grids in the comparison with benchmark simulation results. Since we use in horizontal direction polynomials of degree $p=3$, i.e. 4th order basis functions, one would roughly expect an equivalent grid spacing of $dx/4$. If BRIDGE would need a finer grid than this for this standard test case, something would definitely be wrong. That we can use even slightly coarser grids than this estimation might be an indication of additional accuracy benefits of a higher order scheme (even in this non-diffusion-limited case). To be honest, this factor 5 is also a rather technical outcome, since the standard grids in ICON are R2Bn or R3Bn grids and the chosen R2B3 grid is just closest to the targeted ' $dx/4$ ' grid spacing (but this is of course not a hard limitation, in principle the ICON grid generator could produce also other (less common) resolutions).

We also have added one sentence in this respect in the section with the density current (which is, in contrast to the baroclinic instability, a diffusion-limited test setup).

3. Another clarification needed is about timestep size and wallclock time. In baroclinic wave test the authors report a timestep size of just 30 seconds for a grid resolution of 630 km. This leads to a wallclock time around 623s for simulated/model day measured on 32 processors: this is sensibly higher than the wallclock times reported on Jablonowsky and Williamson QJRM 2006 reference paper (cited by the authors) in its table 4. From a quick comparison this might be related to the very small timestep size used here (30s) in comparison to the timestep sizes used in the Jablonowsky and Williamson paper as reported in their table 3. This could indicate a potential computational overhead to take into consideration when examining the overall efficiency of this approach. Please expand this point with some further assessment about the efficiency of the current time stepping.

Answer: This question is in the same line as posed by reviewer 1 (about used Courant numbers). In the baroclinic test case, sec. 6.6, we didn't use vertical grid stretching, so even larger vertical velocities are tolerated with $dt=80s$. Therefore, the time step is mainly determined by the horizontal grid spacing of $dx \approx 315km$, this is now noted in the new paper version. In the Held-Suarez test we used a vertically stretched grid (to make this test a bit closer to more realistic setups used e.g. in numerical weather prediction). Thus mainly the (expected) vertical velocity w limits the stability of the explicit part of the HEVI scheme, so that here the time step is not any more limited by

the horizontal grid spacing but mainly by w . This is demonstrated by (partly estimated) Courant numbers now given for both test cases.

Concerning the wall clock time: our measured 623s on 32 processors is quite close to the value for the finite-volume scheme at a similar resolution in table 4 of JW2006. Of course, this article is 20 years old. However, the IBM Power 4 processors used there, are not so slow compared to today's processors. We now shortly mention this comparison in sec. 6.6.1.

In sec. 6.1 we now have given a work-precision diagram (which is possible due to the availability of an analytic solution) and compared it with ICON. In this well-resolved case with smooth fields BRIDGE is by far superior. However, in more realistic tests like the baroclinic instability we have to admit that the BRIDGE code is still much slower than ICON on the same computer and we now give a runtime comparison with ICON at the end of sec. 6.6.1.

Additionally we have extended the 'Conclusions' by a remark about possible time step limitations by too excessive vertical velocities and how they could possibly be reduced by the provably stable schemes. We now also mention possible time step increase by alternative IMEX-RK schemes.

4. About the polynomial order used: it is said that the choice $p=4$ in the horizontal vs $p=5$ in the vertical is made. Is there any reason for this choice? Have you found some results suggesting this is an optimal choice in some sense? That would be also interesting to know.

Answer: First for clarification: we use polynomial degree 3 horizontally and 4 vertically, leading to approximation order 4 horizontally and order 5 vertically. The horizontal order 4 was mainly a result from Baldauf (2020) where it was demonstrated that this is the lowest order where grid irregularities are not any more visible in the simulation for shallow water equation tests. The vertical order 5 is mainly a result from the simple well-balancing test in sec. 6.1: from Fig. 2 it can be seen that the overall well-balancing violation is small and acceptable mainly for order 5. These motivations now are given at the end of section 6.1. One may think that this would further be improved when going to even higher order. However, since a DG scheme only is locally conserving on a grid cell base but (at least for the classical DG method) not inside of a grid cell, one does not want to go to too high order (we shortly mentioned this point in the introduction).

5. In the baroclinic instability test section it is written "the update frequency of the implicit solver is every 10th time step". This sounds to me very unclear. As it is written, it seems to be suggesting that the implicit solver is called every 10 timesteps which of course cannot be the case especially because of stability reasons. Being obvious that the implicit system is solved every timestep, I guess that what is meant here by this sentence is that the coefficients of the implicit solver are recomputed every 10 timesteps, in other words the linearized solution on which the implicit solver is based is computed every 10 time steps. If this is the case, then this should be expressed in a bit clearer way.

Answer: Yes, our implicit solution strategy seems to be not clearly enough described (there are similar questions by reviewer 1). As you say, it is the calculation of the linearisation coefficients that is done every 10 timesteps. In contrast, the implicit solver (i.e. the matrix-vector multiplication of the so-called back-substitution step of an implicit scheme) is of course called in every RK stage. We extended the last paragraph of sec. 3.1 to explain this a bit better (e.g. introduce there already the term 'update', ...). Furthermore, we have extended most occurrences of the term 'update' by 'update of linearisation coefficients' (or similar).