Reply to Comments by G. Stadler

The paper contains benchmark results obtained for the HYTEG code framework, which solves the typical equations governing mantle flows. Such comparisons and benchmarks are valuable, even if they do not add real new methods/algorithms, or new geophysics results. Generally the paper is well written, but I've a list of comments and suggestions below.

Main comments:

1. Line 108: I would not claim that the surface velocity components are typically taken from plate-tectonic observations. This is one option to force the system, but it should arguably be the other way round, i.e., the forces in the mantle together with the physics should result in plate-like surface velocities. Imposing the plate motion as Dirichlet bdry conditions can result in unphysical forces in the system; I suggest that you at least not claim that this is the usual/only thing done in mantle flow. It is again pointed out as the normal case in line 308/309–I dont think this is the only choice.

   Authors' reply: You are of course correct here in both respects. Using plate velocities is not the only option and doing that incorrectly can introduce unphysical forcing. We usually use (appropriately scaled) paleo-plate velocities in our simulations for retrodicting mantle flow in the past for the stabilising effect they have in the inversion. We will stress in the paper that this is, but only one of the possible options.

2. Eq 30, line 396 – something does not seem to be correct in the nonlinear viscosity here, there might be a missing square root. I would think that the highest power the denominator should be 1 in terms of the strain rate, corresponding to plastic yielding.

   Authors' reply: You are correct, apologies, the text will be corrected.

3. Regarding nonlinear Stokes solves: I assume all nonlinearity is treated with Picard fixed points iterations, which has limitations for stronger nonlinearities. Can you give some information about how many orders of viscosity variations occur in these benchmarks? Since Picard just solves a sequence of variable viscosity Stokes problems, I dont understand the comment made on line 274, namely that harmonic averages work better for nonlinear problems. These averages only affect the multigrid coarse grids, which is a purely linear issue–or am I missing something here?

   Authors' reply: From Fig. 3, we can see that the final non-dimensionalised viscosity changes by nearly 4 orders of magnitude across the domain. With respect to the harmonic averaging, you are fully correct. It only affects the multigrid solver. Our observation was that in this setting the multigrid solver worked better with the harmonic averaging. We will express these two points in text more clearly.

4. Rigid body modes, Sec 3.4: Makes sense to penalize rigid body modes as you discuss and thus save an MPI reduction since you avoid projecting out the rotation. It is surprising that, as discussed in Sec 4.1.2 there is sensitivity to the penalty parameter–if systems are solved reasonably accurate, one would expect that there is no sensitivity as these modes are in the null space of the PDE operator, so they should be pretty easy to control (and the solution should not degenerate if the penalty parameter is large). Does the sensitivity (and thus the need to tweak this parameter) have to do with the iterative solver?

   Authors' reply: Thanks for the affirmation. No, it does not have to do with the iterative solver, indeed we get the same behaviour with the direct solver as well. As this was a first test with this penalty approach, we had only presented the results that we observed.

5. Regarding scalability in Sec 5: Since you focus on the Stokes solves, the task should be the scalability of those solves, and not (only) at the scalability of a fixed number of iterations towards that solve. What you study is sometimes called the parallel scalability and it's a purely implementation issue; but it could be that more and more iterations are needed for finer discretization (in weak or even strong scaling), degrading the overall scalability of the complete solve (that second component is sometimes called algebraic scalability or simply mesh independence/robustness). For problems with severely varying coefficients (or even nonlinearities!) it's challenging and sometimes impossible to achieve that, and it requires an effective and scalable smoother, often with more than a few smoothing steps etc. This issue that two components contribute to the scalability does not come up with explicit time stepping which typically just amounts to matrix-vector applications, but it matters for implicit systems (like Stokes) and makes them much harder to scale. Please at least add a discussion of that issue to emphasize that just being able to do Krylov iterations in a scalable fashion does not necessarily imply a scalabe solver! For the larger scalability test you even reduce the number of smoothing steps and the GMRES history to save memory; I'm not convinced that you get that second part needed for overall scalability of the solver for reasonably challenging problems.

Authors' reply: The idea of this scalability test is indeed purely to test the implementation of the Stokes solver and how well it can cope with the increasing resolution of the models. Although we tried to mention that the solver can only 'handle' the system, it is definitely not very clear. We will add an extra discussion on nailing this point down even more in the final version. A parallel study from Burkhart et al. (2025) was also conducted which analyses the Stokes solver considering the residual reduction with large viscosity variations in the HyTeG framework.

6. It's great that you can scale to 1e11 unknowns, but it would be interesting to see settings where such a resolution is actually needed. I would guess that it only matters for extremely nonlinear rheologies, extremely varying coefficients, very fine geometric structures etc. Can you comment on your solver being used (or plans to being used) for such problems? My concern is that you vastly overresolve problems and thus do not gain any significant accurarcy by going to so many DOFs.

Authors' reply: As you have pointed out, for extremely nonlinear rheologies/varying coefficients, higher resolution is necessary. Although, from our tests, we see that to achieve mesh convergence even at Rayleigh numbers $\sim 10^6$, we require $\sim$ 25 km resolution to capture the small scale features. A baseline case in our groups' interest is a model with a radially varying viscosity with the inclusion of an asthenosphere which introduces $\sim 10^4$ orders of viscosity changes within $\sim$ 150 km. To capture the small scale dynamics and features at Earth-like Rayleigh numbers $\sim 10^7$, an ideal resolution would be with 10^{11} DoFs which corresponds to $\sim$ 6.25 km resolution of the FE element ($\sim$ 3 km resolution between grid points). A trade-off would be not to do global refinement but rather an adaptive mesh refinement. This is also an ongoing study in our framework in doing this efficiently under the matrix-free setting (Mann, B., & Rüde, U. (2025)).

Minor:

1. line 83: "A fact to which ..." This isn't a sentence.
2. Equation 9: The dot denotes the inner product between two vectors, but you also seem to use a dot for the matrix-vector multiplication between tau and $\hat{r}$ – is that intentional?
3. line 167: ...a brief *overview* of some...
4. line 185: a Stokes -> Stokes
5. line 308: setup more closer -> setup closer
6. line 344: in same as -> in the same as
7. line 483: upto -> up to
8. Missing commas make reading some sentences tricky–please add commas in the following spots (and probably some more spots):
   (a) line 100 after mantle
   (b) line 110 after no-outflow
   (c) line 148 after Here
   (d) line 197 after factor
   (e) line 216 after optimisations
   (f) line 226 after enough
   (g) line 310 after Next
   (h) line 479 after First

Authors' reply: Thanks for pointing these out, we will correct them in the revised version.