General comments:

In this paper authors describe the effort associated with making the GRIST dynamical core computationally cheaper through adopting reduced precision to selected parts of the code. The paper is really well written, logically well structured which helps its readability. Like that it is a joy to be followed for a reader. The thorough evaluation part is quite impressive as it represents a lot of hard work. It is also done with great care to cover all possible aspects potentially impacted by reduced precision.

Reply:

We would like to express our gratitude to Dr. Filip Váňa for detailed review and constructive feedback on our manuscript. We appreciate your insightful comments and suggestions. Each issue you highlighted has been thoroughly addressed in our revised manuscript. Your expertise has been essential in enhancing both the quality and clarity of our work.

The only complaint is that the paper doesn't really attempt to modify the original code in order to make it more profitable for reduced precision. By that I mean that authors were only trying to identify precision sensitive parts requiring to be exclusively evaluated with double precision in the original code. There is no discussion trying to explain this sensitivity neither an attempt to eventually propose a modification or new method allowing the ussage of single precision also there. But that is perhaps subject to another paper.

Reply:

Indeed, the focus of this work has been on identifying precision-sensitive parts of the original code, without delving into further modifications of complex or "tricky" code segments. This work could serve as an initial step towards the broader application of reduced precision within the GRIST dynamical core. Future developments could include rewriting suboptimal code implementations and exploring the use of half-precision computations, which could further enhance computational efficiency and resource utilization.

Small points:

1. In 110 some error norm is computed based on two model variables Ps and VOR. How those norms are evaluated and is there any scaling applied to one of them to make the two norms roughly comparable? The way it is described here is too generic to be followed.

Reply:

In this manuscript, Eq. (2) evaluates the error norms for two model variables: surface pressure (Ps) and relative vorticity (VOR). The error norms are defined as follows:

$$L_1 = \frac{I(|\mathcal{H} - \mathcal{H}_T|)}{I(|\mathcal{H}_T|)} \tag{1}$$

$$L_2 = \sqrt{\frac{I[(\mathcal{H} - \mathcal{H}_T)^2]}{I[(\mathcal{H}_T)^2]}} \tag{2}$$

$$L_\infty = \frac{max\forall|\mathcal{H} - \mathcal{H}_T|}{max\forall|\mathcal{H}_T|} \tag{3}$$

$$I(\mathcal{H}) = \int_{z=z_{surface}}^{z=z_{top}} \oiint \mathcal{H}\, dAdz \qquad (4)$$

where $\mathcal{H}$ and $\mathcal{H}_T$ are the computational solution and true solution, respectively. Max $\forall$ means select the maximum value from the field. $I(\mathcal{H})$ denotes global 3D integration for an arbitrary quantity $\mathcal{H}$, A denotes cell area and z denotes height. Vertical integral will be omitted for a 2D integration.

Your observation is accurate. As you noted, the error magnitude associated with pressure (Ps) is smaller compared to vorticity (Vor). Consequently, according to our current criteria (defined beforehand), only "Vor" significantly influences our optimization outcome. This point has been clearly stated in the revised manuscript.

For the baroclinic wave test, "Vor" demonstrates a higher sensitivity to small perturbations than other physical variables. This heightened sensitivity makes it a good metric in our optimization procedure.

2. I found bit unintuitive to digest the results of splitting supercell thunderstorms in section 4.2. Especially, the text belonging to 245 part describing results presented on figure 4. I am bit surprised by the great similarities between double precision and mixed solution until the 5400s with almost bifurcation behaviour afterwards. It feels like something strange happens at that time range.

Reply:

The revised manuscript's Figure 5 provides a detailed depiction of the temporal evolution of the domain maximum vertical speed and area-integrated rainfall rate in the supercell thunderstorms simulation. During the initial 5400 seconds of the simulation, the results from the mixed-precision and double-precision simulations are quite comparable. This similarity can be attributed to the relatively weak vertical motions present before the supercell reaches maturity, which are not highly sensitive to the level of precision used in the computations.

After 5400 seconds, as the supercell matures and develops more complexity, including the generation of small-scale features (now Fig. 4). As the simulation progresses, these small-scale perturbations amplify. Therefore, as the storm develops and becomes more dynamic, the sensitivity to the precision level becomes more pronounced, but still limited to relatively small scales.

I am also quite surprised by finding the highest resolution runs continue to remain similar across the two precisions while lower resolution runs show difference. From our experience it was rather opposite: higher resolution runs exhibited higher sensitivity to used numerical precision. Could this be somehow explained?

Reply:

In our revised manuscript, we have added tests at G4 (~4 km) grid. We then show the differences between mixed-precision and double-precision simulations at heights of 2.5 km, 5 km, and 10 km, respectively. It is found that at 4 km, the differences between mixed-precision and double-precision simulations are much smaller than at other resolutions, for all examined heights (Figs. 4b-c). As the resolution increases from 4 km to 1 km, the differences in the supercell become more pronounced (Figs. 4b-d, f-h, j-l). The contrasting behaviours between 4km and 1km are consistent with earlier experience and research studies.

As you noted, the differences between mixed-precision and double-precision simulations do not increase monotonically as resolution moves from 1 km to 0.5 km and to 0.25 km. Instead, the differences diminish. The design of the DCMIP2016 supercell test is intended to ensure converged solutions as resolution increases. At 0.25 km, the mixed-precision simulations are closer to the double-precision results at all heights (Figs. 4n-p). This is likely related to the solution convergence.

Please note that in the original manuscript, the horizontal section shown in Figure 3 was not actually interpolated to a height of 5 km. We have corrected this mistake in the revised manuscript.

Despite my general comment and the two rather questions than really complain I would suggest the paper is accepted for publication. If author wish they could eventually address my points, but it could be published straight away the way as it was submitted. Bravo!

Filip Vana (ECWMF)

Reply:

Thank you! Dr. Váňa.