Review of Gassmöller et al. *Benchmarking the accuracy of higher order particle methods in geodynamic models of transient flow*

This is a well-written article that provides valuable insights into the source and time-evolution of the errors introduced in forward models by commonly used advection methods in geodynamic modeling software.

The study illustrates the higher accuracy of higher order particle methods for transient flow, and how this can potentially lead to significantly different results in geodynamic modeling applications compared to lower order methods. The higher order method proposed in this paper is also highly appealing as it requires minimal code modifications for those codes already using Particles-in-Cell methods to advect material properties.

This article also showcases the importance of benchmarking in numerical modeling and presents two new very interesting and well-thought benchmarks from which the existing software can benefit, as they can easily be implemented and reproduced in other existing codes.

For the reasons stated above, the article certainly deserves to be published in GMD. Please find below a set of minor general and in-line comments for the authors consideration.

General comments

(1) The integration method presented here as RK2 and RK2FOT is a particular case of the general two-stage Runge Kutta method

$x \leftarrow x + \Delta t((1 - (1/2\alpha))\ u(t\ ,x) + (1/2\alpha)\, u(t + \alpha\Delta t,\ x + \alpha\Delta t\, u(t\ ,x)))$

where $\alpha = 0.5$, known as the *midpoint method*. This should be noted somewhere when the different integration methods are described and discussed.

In my numerical codes I typically use $\alpha = 2/3$ (Ralston's method), which in my tests seemed to produce (slightly) better results and lower particle injection rates, compared to the midpoint method. Have the authors tested other values of α different from 0.5?

As the existing code would require minimal modifications to accommodate the general RK2 method, it would be worth it to test (and quantify) whether other RK2 variations produce more accurate results or not during the model time evolution, or whether we should avoid some values of α. Additionally, the extra computational time should also be negligible when compared to the midpoint method.

(2) Many of the conclusions discussed in the last paragraphs of Subsections 5.1 and 5.2 are very similar. Perhaps it would be a good idea to merge them in a small Subsection 5.3, to avoid redundancy.

As later mentioned in one of my comments below, I believe it's important to quantify the performance of RK2 vs RK2FOT, instead of vaguely stating that RK2 is somewhat slower.

This will give a better idea of the trade-off between accuracy and speed, and perhaps help other code users in deciding which method to use based on the application of their models and their computational resources. Plots of performance vs time could be presented and discussed also in this new subsection. It is also stated that higher accuracy particle methods yield a faster solver convergence, if so, a comparison of the Stokes solver runtime for both cases can also be included here.

(3) I really like the two benchmarks in this paper. However, adding some more details about them in code may be beneficial. For example, the manuscript does not mention whether these benchmarks and models do inject or delete particles. If this is the case, it is worth briefly discussing whether RK2 and RK2FOT yield a similar or different number of particles at the end of the models, as well as the time-history of particle injection and deletion rates. The dimensions used in this article for the domain of these benchmarks should also be stated in the text.

Line-by-line comments

L26: It's clearly stated that the papers cited here are just some out of many. Nonetheless, some of the examples have a large collection of previous works that I think it would be appropriate to at least include a second reference. And also add (e.g. …) to the citation to emphasize the fact that they are just some of the many examples.

L34-35: Give some examples/references for every stated method. For example, ASPECT and Moresi et al. 2014 for FEM, Gerya and Yuen 2003 and Kaus et al. 2016 for FDM, or Tackley 2008 for FVM, among others.

L39: I think it is better to mention it is "application agnostic" a bit below, where it is described that the particle infrastructure is implemented in a general purpose FEM library such as deal.II, rather than in ASPECT itself. Is this particle architecture specific to FEM models? For example, can it deal with a finite difference code with a staggered grid?

L39: most modern HPC centers offer a considerable amount of GPU resources. Does the software here presented run on a (multi)GPU environment? If not, please clarify in the text.

L54: *"...due to interpolating properties from particles to **fields**"*
I think using "grid" or "mesh" may be more clear than "fields"

L80: Since you write the vector fields in bold, mention this fact in the lines below.

L82: strain rate tensor

L88: *N* is not explicitly defined.

L108: ~~value~~ values

L109: ~~results~~ result

L109: ~~(2))~~ (2)

L111: Why use *j* as a sub-index? *i* is not used elsewhere for any other purpose as a sub-index, and it seems a more obvious choice since the particle information is likely to be stored in 1D arrays.

L119: *"the equation the differential equation solver really tries to solve is"*
I find this phrasing a bit weird, I suggest something in the line of "the equation being solved is actually", or similar.

L130 I find somewhat random the use of inverted commas and italic font for the word exact.

L145 *"although many other methods are available and have been used in geodynamic applications."*
Provide some examples/references.

Eq 7. Later on in equation (29), the total time derivative d/dt is not in italic. For consistency, denote d/dt everywhere either in italic or normal font.

L223: *"properties: First"*
"properties: first" or "properties. First"

Section 4.1: Specify, either in the main text or in the caption of Figure 1, what are the inner and outer radii of the spherical shell.

L235: ~~In other words~~ In other words,

L248: ~~appendix~~ Appendix A

L266: Shouldn't it be $x \leftarrow$ for the coordinate transformation?

Equations 20, 21, 22, 23, 25 : all of these equations are missing a \right) bracket to close the first trigonometric function. E.g, eq. (20) should be $\sin(\pi(x - \tau(t)))$ instead of $\sin(\pi(x - \tau(t))$

L287: Backward Differentiation Formula (BDF2)

L295: Perhaps add a brief comment on the Q2 x Q1 element, as you do for Q2 and DGQ2. I believe Q1 is not mentioned elsewhere.

L296: Clarify here whether there is injection and removal of particles. And if this is the case, please mention what are the minimum and maximum number of particles per element.

L311: ~~Analyzing the~~ The analysis

L314: What is the real difference in % of the error of RK2 between the first and last time step?

L314: *"While RK2 and RK2FOT start off at the same error value, and RK2 almost maintains this error over the evolution of the model, the error of RK2FOT increases significantly over time."*

May benefit from some rephrasing. Along the lines of "*Both RK2 and RK2FOT start with the same error value and, while RK2 error remains near-constant over the evolution of the model, the error of RK2FOT increases significantly with time."*

L329-330: The extra computation that RK2 is doing $u(t + dt/2) = (u^{n} + u^{n+1}) * 0.5$ which is $(2\, mem\, read + 1\, add + 1\, multiply) * ndim$. Where the slow part is the extra mem reads, and the extra flops are likely to be virtually for free. Thus, the price of doing $u^{n} * dt * 0.5$ in RK2FOT indeed doubles, give or take. However, unless I am missing something or there is something else going on, this still appears to be far from doing 2x operations during the entire integration algorithm, compared to RK2FOT. Could you please elaborate on this point?

Additionally, it would be beneficial to quantify the overhead of RK2 in comparison to RK2FOT. A Plot could be added, to compare and discuss the time evolution of the performance of both methods (shown as [runtime RK2] / [runtime RK2FOT]) for different grid sizes. This will also show whether the performance of RK2 deteriorates with time more or less than in RK2FOT.

Last, doesn't RK2 have a higher memory footprint since it needs the storage of two velocity solutions?

Section 6: Some information is missing in the description of model setup. What is the resolution of this model? is the mesh regular or spatially refined? is AMR active? I found this information in the .prm files, but it would be useful to add it in the manuscript as well, since not every reader will check the input files.

L369: ~~Perhaps~~ perhaps

L370: ~~sources and therefore~~ sources, and therefore

L377: *"…, and right (outflow) boundary and linearly decreases with depth starting at 100 km towards 0 cm/yr at the bottom of the model;..."*
I find this sentence a bit confusing. I guess the authors mean that there is a prescribed outward velocity at the right hand side boundary, which is constant for depth < 100km, and then decreases linearly to 0 at the bottom. However, the arrows in the right boundary of the model in the top panel of Figure 5 appear to be of the same length, top to bottom. It is worth rephrasing it in a clearer manner.

L369: ~~boundary~~ boundaries

L383: what is the magnitude of the anomalies? is it a Gaussian shape or constant value

L402: is “field” really meant to be in italic?

L407: *“For a detailed discussion of these terms and all parameter values we refer to (Dannberg et al., 2017).”*
It may make sense to move this sentence at the end of the very same paragraph.

L441: model ~~run~~ runs

Figure 2:
- Looks like the aspect ratio of the model is 1:2 but this is not written anywhere. You could just add the corner x and y coordinates in one (or all) of the plots. Or just write it down in the caption
- The ticks of both colorbars in the upper plots display some integers as proper integers (e.g. “**2”**) and others as floats (e.g. “**1.**”). For consistency, I suggest displaying all the integers as proper integers, as in Figure 1.

Figure 5:
- Would be nice add a video of these two models to the Zenodo repository
- Better to write “Plate age” in the titles of the middle and bottom panels, instead of just “Age”
- Light gray text and arrows are not the easiest to read on top of a white background. I would suggest the use of a different (darker) color

Code availability:
- I haven’t tried to install ASPECT and run these benchmarks and models myself, but it looks like all the necessary files to reproduce the results in sections 5.1, 5.2 and 6 are indeed included in the repository, along with a good and much appreciated description of where to find these files and (potential installation problems aside) how to run them.
- Perhaps add the link to the Zenodo repo to the manuscript to make it easy to the readers => https://zenodo.org/records/10161412
- I may be looking in the wrong place, but I don’t see a 2.6.0-pre branch. And the given git hash seems to link to a unrelated commit (https://github.com/geodynamics/aspect/commit/299a6456385b1fde6564fc079f3aa01cac075f24)

Other minor suggestions

Here is a list of some minor style and writing suggestions that may improve the readability of some sentences. It is up to the authors of the manuscript to decide which suggestions they believe will improve the quality of the text:

L21: scalable, and efficient
L22: investigate ~~more~~ complex
L42: general purpose open source finite element ~~software~~ library

L42: to model a wide range of geoscientific applications
L61: "and therefore fast changes over time"
Is this missing a verb?
L74: ~~that are too specific to be relevant for the main text~~
L89: , for example,
L124 The particle positions contain error contributions from the…
L138: ~~itself~~
L146: ~~then~~
L149: , in practice
L150: number of around
L154: ~~in the following~~
L199: (1), (2), and (4)
L230: geometry, and
L234: … for the density, not allowing to…
L235: in other words,
L280: in equations (15) and (23)
L299: In the following subsections,
L312: , as expected
Fig.3 L3: for ~~an~~ the exact
L371: …method may be sufficient

References

Moresi, Louis, Zhong, Shijie, Han, Lijie, Conrad, Clint, Tan, Eh, Gurnis, Michael, Choi, Eunseo, Thoutireddy, Pururav, Manea, Vlad, McNamara, Allen, Becker, Thorsten, Leng, Wei, & Armendariz, Luis. (2014). CitcomS v3.3.1 (v3.3.1). Zenodo. https://doi.org/10.5281/zenodo.7271920

Gerya, T.V. and Yuen, D.A., 2003. Characteristics-based marker-in-cell method with conservative finite-differences schemes for modeling geological flows with strongly variable transport properties. *Physics of the Earth and Planetary Interiors*, *140*(4), pp.293-318.

Kaus, B.J., Popov, A.A., Baumann, T., Pusok, A., Bauville, A., Fernandez, N. and Collignon, M., 2016, February. Forward and inverse modelling of lithospheric deformation on geological timescales. In *Proceedings of nic symposium* (Vol. 48, pp. 978-983). John von Neumann Institute for Computing (NIC), NIC Series.

Tackley, P.J., 2008. Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid. *Physics of the Earth and Planetary Interiors*, *171*(1-4), pp.7-18.