101 Geodynamic Modelling: How to design, carry out, and interpret numerical studies

Review by *Laurent Montesi*, University of Maryland.

This manuscript serves as a well-designed guide for modeling mantle and crustal-scale processes. As is necessarily the case with an exercise of the sort, it does in place reflect the personal experience and opinions of the authors, but it overall does a very nice job of remaining neutral, and even the more engaged sections are full of important information that will be useful for moving the field forward. I particularly appreciated the discussions of figure accessibility and Section 9. There are also some very important discussions of the objectives of modeling, in particular the difference between "specific" and "generic" modeling. Both are presented as equally valuable, which is an important message to pass to both modeling and non-modeling communities.

Although the paper is currently well organized and well written, I do have a few suggestions that could lead to significant rewriting. They are organized here as topics and followed by several more isolated comments. Most suggestions can be seen as a matter of personal preference and should not stand in the way of the publication of the paper. Maybe they are best seen as a discussion of the material presented. I apologize that as my review is already much overdue, I did not study the appendix or glossary in detail.

1. Although the title sets the scene for "Geodynamic modelling" in general, the authors focus on the dynamics of the mantle and the crust. In place, they contrast their discussion with the core, especially in their neglect of inertia and Coriolis forces. There are other geodynamical settings where these are also neglected: the hydrosphere (especially the oceans) and the atmosphere, to say nothing of giant planets. The effects of fluids (including but not restricted to magma) are noted in passing, even though they are a growing segment of geodynamical studies. I do not think that presenting specific equations would be necessary to discuss modeling strategies and philosophies but as the authors have chosen to focus on a specific system, I believe that the title must reflect that choice. Maybe specify "Geodynamic Modelling of the Mantle and Crust" (lithosphere instead of crust would be OK too) and make this focus (including the neglect of two-phase flow) clear at the outset. The occasional mention of core studies can be seen as choosing a related science but neglecting other applications. Minor points related to this topic:
    - Line 20: The mention of "spherical shells" implies a global focus. Local and even regional studies do not need to consider shells.
    - Line 37: "magma dynamics and grain dynamics" The citation to Solomatov and Reese shows that you have global mantle convection in mind. This is a good paper but please also mention studies about magma dynamics, and also other fluids (CO_2, water) that have been shown to matter in volcanic and seismogenic systems, and maybe also serpentinization (important in mid-ocean ridges).
    - Line 80: I don't have the statistics to evaluate if applications to Earth's crust and mantle are indeed "the most common applications". I expect that people studying climate and atmosphere may have something to say about that. The concepts presented (except for the specific equations) are also applicable to their disciplines. It may be better to just say that these are the disciplines you worked on and therefore are most familiar with.
    - I hate making titles longer (believe it or not) but it strikes me that the important communication steps that are in the paper are not included in the title.
2. Figure 2 is key to the entire paper. It is well-thought-out and I am sure the authors have spent much time discussing it. I do have several issues with it as it stands.
    - It is easy to assume, based on Figure 2, that geodynamic modeling is a linear process. I believe this is an oversimplification. This paper would be a great place to emphasize instead the need to constantly reevaluate foundational hypotheses. Results and analysis, especially following evaluation against observations (necessary whether the model is "specific" or "generic") can

lead to updates in any of the previous steps. "Nature" should feed directly into the validation (maybe) and analysis (certainly) steps.

- What is missing in Figure 2 and section 1.2 is the mention of a hypothesis. I strongly believe that the most important models are those that were designed at the outset to answer a specific scientific hypothesis. As the paper is currently organized, it would be possible to write a code and verifying it before a hypothesis is identified (see line 127). To me, this is backward: we should design codes that enable us to answer a question, therefore the question needs to come first, not in the "setup" stage.
- This may be a matter of semantic, but it bothers me that "physical models" are described only using equations (also Line 87). I believe that physical models can be discussed in terms of concepts (e.g. buoyancy, inertia, diffusion, or the rheologies that are illustrated without equations in Figure 3) and that by the time you have introduced an equation, you have already moved on to the next level of abstraction: a mathematical model. That model can be solved analytically or numerically. Therefore I believe that an intermediate step ("mathematical model"?) is missing in Figure 2.
- Somewhat related to the previous, don't forget that chemical processes and increasingly considered in geodynamics models. You might argue that they are included in "physical processes" as thermodynamics can be seen as either, but issues of trace element partitioning (important for evaluating the origin of magma or fluid interaction) are typically discussed only in geochemistry. Like physical models, chemical models need a mathematical description to be included in the later numerical modeling efforts.
- This is not an issue, but maybe consider distributing your figure as an open workflow (e.g. on https://workflowhub.eu/)

3. Model simplification
   - You often use "simplified" to describe your model. This is perhaps a matter of preference but I favor "idealized". Complexity is not necessarily your enemy (and you do a good job discussing the pitfalls over oversimplification) but whenever you settle on mathematical relations, especially in the constitutive equations that are needed to close your balance equations, you are looking at an idealized but not necessarily simpler view of reality.
   - Line 789: the concept of over-simplification is not shown in Figure. Another issue is that what is "too much" simplification depends really on the hypothesis to be tested (that's a reason why the hypothesis needs to be identified first). The "simple" convection model in Fig.7 is also quite complex to me. The triangle heart may be sufficiently realistic for some purpose. The link between oversimplification and hypothesis (line 791) should a major motivation for motivating the identification of the hypothesis before designing the model (see my comments about Figure 2) Here again, I would favor idealized over simple, and maybe realistic over complex.
   - Figure 8 could mention constitutive equations that coupled vary from non-linear to linear to fixed parameters. The degree of coupling in Multiphysics problems could also be mentioned.
4. Line 247 etc. High thermal conductivity does not mean that the curvature $\nabla \cdot \nabla T$ is steep. Just consider a simple steady-state solution with fixed temperatures at both ends of a 1D solid. You know the temperature varies linearly between the two ends. Curvature is 0, regardless of the conductivity. What the high conductivity does is allow for large heat flow through that material. It also makes it faster (everything else being the same) to obtain this solution, so that in general, the curvature of the temperature field is less in the case with high k, at least in this specific example.
5. The description of physical concepts emphasizes the momentum equations. While I can appreciate their importance, most models also need to include constitutive equations. These appear in section 2.2 but should be at least mentioned earlier, as part of the model design. The fact that these evolutions, including their parameters, can change over time should also be described in general with an evolution equation, and that aspect is buried in "more complex processes" (section 2.3). It goes beyond the idea of fault gouge development (line 422). I don't think this is any more complex

than the other kinds of equations described. Many constitutive parameters evolve following an ODE, so in a way, this is simpler.

- Figure 3: As much as I like the emphasis on rheology, it strikes me that it is only one of the many constitutive equations that can enter the model. Perhaps it is more fundamental because, without it, there would be no way to link stress and velocity. In that case, it has to make the distinction clear but highlight in the caption that other constitutive and evolution equations can (should?) be considered. Minor point: the diagram for "plastic" is brittle.
- Line 224: Stokes equations need not only eq. 2 and 5 but also viscous rheology. If you use elasticity you end up with a different set of equations.
- Line 229-231: Seismic waves do need inertia but they do not follow the viscous relations (Stokes or Navier-Stokes equations) that you describe here.
- In detail, Equation (10) doesn't work as it takes a tensor to a power that is not necessarily an integer. This equation should be written with a scalar measure of the strain rate and the stress tensor. We often use it in models with an effective viscosity approach. I think it would be important to explain this in the text. By the way, H2O should be styled as subscript (and not italicized) in $f_{\mathrm{H_2O}}$.
- Line 374: In fact, grain size dependence is fairly well constrained. Grain size evolution is not (but you do not consider evolution at this stage). Also, dependence on porosity ($\alpha^* \phi$) is even less well constrained.
- I find the discussion of elastic deformation to be a distraction: it cannot be included as is with the other relations and the framework of Stokes flow. We can see that in line 471, it is written that "velocity … enter the stress tensor". That is incompatible with elasticity. I recommended either skipping section 2.2.3 or describing implementations of viscoelastic materials.
- Line 480: Why the specific focus on diffusive quantities? You can have ODEs, e.g. grain size evolution.
- Line 505: the addition of Eq. 15 is inconsistent with the statement of introducing "the general concepts of geodynamics". I would argue that constitutive relations are more fundamental and general.

6. Section 2.2.4 needs more discussion of the way brittle failure can be implemented.
    - Line 392: brittle deformation takes place when rocks break, but this is not necessarily (in fact rarely) a rupture of the crystal lattice.
    - Line 394: Actually, it is possible to represent discrete faults using split nodes (Melosh and Raefsky, 1981, A simple and efficient method for introducing faults into finite element computations, Bulletin of the Seismological Society of America 71 (5): 1391–1400), slippery nodes (Melosh and Williams, 1989, DOI 10.1029/JB094iB10p13961) and augmented FEM or X-FEM approaches (Ortiz et al., 1987, DOI 10.1016/0045-7825(87)90004-1; Dolbow et al., 2001, DOI 10.1016/S0045-7825(01)00260-2)
    - Line 398: there is a difficulty, here again, of comparing and tensor (stress) and a scalar (yield stress). This difficulty should be mentioned. It is particularly important as the difference between criteria that are later mentioned (e.g. Tresca vs. Von Mises) stands in the way that this comparison is made (stress invariant for Tresca and Drucker-Prager, which leads to a smooth yield envelope, vs. resolved shear stress for Tresca and Mohr-Coulomb, which leads to a segmented yield envelope).
    - Line 410: Byerlee's law has two branches. You describe here the friction coefficient of the high-pressure branch but that branch also has a cohesion term. The intermediate branch (important in the crust) does not have cohesion but its friction coefficient is 0.85.
    - Line 415: Drucker-Prager and Mohr-Coulomb cannot be equivalent. One is smooth and depends only on the first and second stress invariants, the second is segmented. For an elastic-plastic deformation, localization arises spontaneously at the corners of the yield envelope, so the difference is significant.

7. Section 3 misses some of the fundamental differences between FEM and FDM. As the authors are well aware, FEM uses the original equations (in their weak form) while FDM modifies these equations. Conversely, FEM restricts the solution to a predefined space. Therefore, the statement in line 510 that we are considering approximate solutions to Eq. 2, 5, and 6 is only strictly valid for FEM (FDM approximates the equations, not necessarily the solution)
   - Line 520: there are even approaches where one equation (energy?) is solved with FDM and the others by FEM.
   - Line 560 etc. The notation of $Q_1 \times P_0$ etc. the concept of "bi/trilinear velocity" must all be described here or in a glossary. I don't think it's reasonable for non-initiated readers to know what this means.
   - Line 576: ALE is only one of several semi-Lagrangian methods.
8. Modeling philosophy
   - The analogy between car/model vs. engine/code did not work for me. To me, the code is more general than the model. The model takes the code and restricts it if using the setup. The code may have additional capabilities that are not used in a particular model. The car also uses many other components that are not the engine (steering wheel, headlamp, seats, doors). I am sure that every modeler can come up with their preferred analogy (and you may argue that your analogy best fits your vision) but I might suggest that the code is like a road network and the model is like an itinerary that uses these roads to read a destination.
   - It would be useful in Section 5.2 to emphasize the importance of "failed" models. It is often impossible to conclude that a certain phenomenon is responsible for observations but it can be shown which situation does not explain observations (this is valid for both specific and generic models). Similarly, these failed models would not verify the criterion in Line 1085 of being "consistent with our understanding of geodynamic processes" but are important to probe the limits of our understanding (this may not be relevant for Phanerozoic Earth, but what about other circumstances?)
   - Line 847: add "the PROPOSED control mechanism(s)". I am not convinced that assumptions and simplification are necessary for ALL control mechanisms. Some models may even be designed to find the necessary level of simplification.
   - Line 851: Other examples of regime diagrams can be found in Citron et al. (2020, DOI 10.1029/2019GC008895) and Gülcher et al. (2020, DOI https://doi.org/10.1038/s41561-020-0606-1)
   - Line 856. Spiegelman focuses more on magma transport than on magma dynamics (often linked to eruptions). It would be good to also cite the newer work by Sim et al. (2020, https://doi.org/10.1016/j.pepi.2020.106486).
9. Boundary and Initial conditions
   - Line 898: prescribed stress can also be used to mimic topographic loads.
   - Line 910, 924: periodic BC are also commonplace in mantle convection dynamics, not just lithosphere dynamics (for example Gurnis, 1988; 10.1038/332695a0; Lowman et al., 2001 10.1046/j.1365-246X.2001.00471.x).
   - Line 925: why restricting the criticism to specific model setups? Generic models of long-term mantle convection are more realistic if the core temperature is allowed to vary according to its thermal balance.
   - Line 926: I think you need a citation to support your statement that models of the other core are better off prescribing heat flux boundary conditions.
   - Line 955: Maybe highlight the issue of initial strain rate for models with non-linear viscosity (there is no solution if velocity is 0 everywhere).
   - Line 965-6: Why specify the mantle when mentioning chemical heterogeneities. The crust is also highly heterogeneous.

- Line 979 paragraph. Note that weak sees can also be implemented as a random field of initial values, which is less constraining than a single seed (e.g. Jammes and Lavier, 2016, 10.1002/2016GC006399).

10. Unphysical behavior vs. numerical problems (e.g. Line 1005 paragraph).
    - I am not certain what "nonphysical artefacts" you have in mind. An incorrect IC or BC is not sufficient to have nonphysical behavior but the resulting result could be irrelevant for the geological problem at hand (still being physical). Maybe you mean that the IC or BC is incorrectly implemented (the verification step should have taken care of that). In some cases, there could be issues of convergence, but not necessarily. I would imagine that the issues you identify (resolution, drunken sailor effect) would be identified at the verification stage (are you solving the problem correctly) not at the validation stage (are you solving the problem you think you are solving)
    - Lines 1035-1038. Stabilization by diffusion is helpful, but one should also remember that it potentially changes the set of physics included in the model (you are changing the equations). Same thing for mass scaling in FLAC codes.
    - Lines 1045: smoothing is indeed important for $Q_1 \times P_0$ elements but I think the recommended solution should be to use stable elements instead. I always refer to the list in the Bathe (2014). *Finite Element Procedures* textbook (Table 4.8 in the second edition). Certainly, there are other references)
11. Modelling manuscript. I found section 8.1 to be overly prescriptive. It presents one possible manuscript organization, which is indeed very common and effective. However, this is far from the only solution (this manuscript does not follow the prescribed structure, for example). I also notice that the general papers about manuscript structure and methods come from the clinical literature, which could have different conventions than geodynamics. At a minimum, the wording must be toned down (should or may instead of must). I would go so far as recommending you delete Figure 11 and much of section 8.1. I always prefer flexibility over rigidity when planning a publication. Scientific manuscripts DO NOT have a rigidly defined structure (Line 1316) although I can recognize that a structure can be useful especially for the first papers one might write. More specific comments follow.
    - Modern publications tend to deemphasize the method (especially in general journals) and emphasize results and insight instead. In my opinion, it is a good thing and helps broaden the reach of our papers. The IMRAD structure that you describe is appropriate for highly specialized publications but not for generic ones. Non-specialist will gloss over the technical aspects and focus on the take-home message of the paper. I would certainly emphasize more the message (you have just one sentence on telling a story in line 1317, yet that is what controls the impact of the paper).
    - As the editor of a journal, I strongly disagree that "it is always good practice to write a complete methods section for every manuscript" (line 1271). If the focus of the paper is on the method, that's true. Otherwise, it is better to refer to other papers that have developed the method and maybe summarize that method in supporting information documents. This separation can strengthen the take-home message of the paper (and remember we study geodynamics to gain new insight). Also, it avoids having identical (or nearly-identical) method sections in different papers, which violates dual-publication policies. Finally, modern open science strategies require sharing codes in FAIR-enabling platforms (as you describe very well), in which case it should be possible to include a citation to the code and its version. It is to be expected that code publications will have documented verification and maybe validation steps. The paper can then focus on the hypothesis, setup, and results.
    - On a related topic, line 1231 "The methods section is considered one of the most important parts of any scientific manuscript" is more true of social and clinical sciences (the Kallet reference you include), which struggle with reproducibility due to reliance on human subjects

and survey methods, than it is in geodynamics modeling. Note also that "who performed the experiment" (line 1241) should be irrelevant: the models should be reproducible by anyone. I think again that the context of the cited paper (here Annesley, 2010) makes it irrelevant for our discipline.
- My other problem with IMRAD as described is that it links discussion and conclusions. In our disciplines, these sections have different purposes: one puts the results in a broader perspective, and may even speculate on future hypotheses, while the other summarizes the paper.
- Line 1239-1240 I don't know what journal specifies "how many words can be used" to describe methods, but certainly not the one I edit.
- Line 1288: The results section should describe the model results. Answering the central question or hypothesis of the paper should only happen after these results are analyzed and evaluated, which is best done as a discussion. This is in fact what you prescribe in Line 1297.
- A couple of pitfalls should be mentioned in the Line 1306 paragraph: the abstract should give a preview of the results as well as the work. Too often I see abstracts that say what is done without saying what is learned (which is what will inspire a broader audience to read the paper). Second: the best titles remain succinct (I was told to limit titles to ten major words). I know I suggested additions to your title but in my opinion, the title should also stop at the column. Two-part titles are often cumbersome. You may also want to mention the use of Plain Language Summary and Graphical Abstracts, for publishers that allow them, as alternative ways to engage a broad readership (note that the purpose and mechanics of Plain Language Summaries are not the same as those of abstracts).

12. Not surprisingly considering who the authors are, the section on visualization is very strong. I have a couple of points to make, though
    - Line 1323: Bar plot should have a well-justified baseline, but it does not have to be 0. Imagine that the quantity reported varied over several orders of magnitude. I may best to use a log scale, and 0 may not be the best reference (e.g. grain size varying from micron to millimeters).
    - Line 1341: there are circumstances when it is necessary to change the range of a color scale. I agree it typically should be avoided and if the range changes, that needs to be emphasized in the caption. However, saying the range should "always" be the same is overly prescriptive.
    - Perceptually uniform color maps are certainly to be preferred for an unbiased reporting of results. However, figures should also inform the readership, and it may be useful to take advantage of non-uniformity to highlight a result (10.1109/TVCG.2018.2855742). This is in a sense what is done with a multi-sequential color map (oleron, highlighting the sea level) and there are also tools for interactively creating colormaps (https://sciviscolor.org/colormoves/overview/)
13. Section 9 is also very strong and useful. It contains many important resources.
    - I wonder if the underlined words should be links, though. In that case, they would best included as URL citations (or better permanent citations if a DOI is available).
    - You may also want to point to CIG's best practice documents at https://geodynamics.org/cig/dev/best-practices/.
    - Line 1418. Another argument for storing and sharing numerical data is to enable the R of FAIR (Reusability). Other scientists may be interested in new analyses of the model runs for different purposes or quantitative comparison with other studies. Sharing results saves on the time and computational cost needed to reproduce the results.
    - Line 1430: Mention the Planetary Data System?
    - Line 1442: Mention Earth and Space Science journal (https://agupubs.onlinelibrary.wiley.com/journal/23335084)?
    - Line 1443: here I think you are not strong enough: research not only can but must create and use persistent identifiers whenever possible (and you do include great pointers on how to do this).

- Line 1446: note that ORCID iD is required by some journals.

14. Isolated points
    - Line 24: replace hundreds by thousands of km (the scale of the largest plates, or the "penetration" of plate boundary deformation in Asia or North America).
    - Line 35: "THEY take place"? (subject surface processes)
    - Figure 1: 1) make the axes labels darker? 2) what about aseismic transients and creep processes?
    - Line 55: While I am proud that you included my paper here, I would recommending mentioning the seminal papers of Hager and O'Connell (1981, https://doi.org/10.1029/JB086iB06p04843), McKenzie (1969 https://doi.org/10.1111/j.1365-246X.1969.tb00259.x). There has been action between 1879 and 2006…
    - Line 63: What do you mean by "the physical properties of the variables" (variables can have value and they can represent physical properties, but they don't have physical properties).
    - Line 97: I would regard boundary conditions as a part of the model setup, not simulations.
    - Line 112 etc. It would be good to mention data assimilation.
    - Line 130: I would think that it's the model results, not the setup, that need to be compared to observations.
    - Line 133: add "openly" to "clearly and reproducibly" (BTW: reproducibly, not reproducible)
    - Line 136: elasticity can be important at long time scales too. See plate bending. Its role in long-term tectonics is still debated.
    - Line 141: 450 years assumes a certain viscosity. I am quite certain it is longer when you enter the lithosphere (still in the mantle). Do you assume the asthenospheric mantle?
    - Line 162: I would focus on the physical processes that are of interest (not just relevant).
    - Line 169 misses a space in "viscousFluid".
    - Line 255: Definitions may be different for different people, but I believe that shear heating is the more general term (it does not imply a shear mechanism) and that it can be divided into frictional and viscous dissipation. This is opposite to the relation you include here.
    - Line 270: Melting is much more complex than shown here, as the heat can be transported by melt flow. You also have the issue that the fraction that has melted (X) may be different for the (retained) melt fraction.
    - Line 283: density OF the …profile.
    - Line 297-309: It may be worth mentioning that many studies have used the EBA but called it Boussinesq.
    - Line 458: why do you assume the relation between density and temperature is linear?
    - Line 584: typo: 10^{24}. I think the shear forces should be negligible, but it's not true that there are no shear forces.
    - Figure 4 gives the impression that triangular meshes are necessary for conforming to an interface. You can have quad meshes that do the same. Maybe include one on the bottom right panel? Anyway, as currently designed, the panel with the two fluids appears in the "mesh" row, so that's a little confusing. Define the fluids in the Field Method panel. I also find the grey fonts too light (I honestly didn't see the words "mesh" and "method" when I first looked at the figure).
    - Line 640: the viscosity "MAY DEPEND" on the velocity… (it's not a requirement)
    - Line 719: Explain what you mean by an "as simple hello world" test.
    - Line 725: I know it's impossible to be exhaustive here, but I would like to see mentioned example of corner flow, viscous folding, and half-space cooling. The latter is particularly important as otherwise there would be no example using the energy equation. Folding also presents interesting numerical effects (see Schmalholz and Podlachikov, 1999, DOI 10.1029/1999GL900412)

- Line 744: missing words “It is important…”
- Line 759: Laboratory experiments do not have “infinite resolution” especially in tectonic applications with granular media. Also, even if all possible physics is indeed included, the constitutive relations are not always fully understood (Katz et al., 2005 Tectonic microplates in a wax model of sea-floor spreading, New J. Phys. 7 37, Di Giuseppe et al., 2012. https://doi.org/10.1007/s00397-011-0611-9). These limitations are important to mention as it highlights the complementary between numerical and analog models (in numerical studies, only selected sets of physics are included, but that means they are well controlled).
- I would have preferred to see the non-dimensional numbers of lines 1130-1 introduced with the concept of the regime diagram (Line 851).