**List of relevant changes:**

- The abstract has been shortened and reformulated on request of the reviewers
- An additional figure has been added on request of the reviewers
- A description of how the elliptical inclusions in the mantle are emplaced has been added to Appendix B
- Appendix C has been reformulated for clarity
- Conclusions have been reformulated for clarity
- The comments of both reviewers have been implemented

The authors thank Jianfeng Yang for his constructive comments. Implementing his suggestions improved the manuscript significantly. In the following, the referee's comments are denoted *RC* and the author's reply is denoted *AR*.

*AR:* The abstract has been shortened and rephrased for clarity.

*RC*: Line 23: "the increase of horizontal driving force", why results in subduction halting but not vice versa?

*AR:* We suggest that when the value of the horizontal force, required to drive the collision, exceeds the magnitude necessary to initiate subduction, a new subduction zone could be initiated in a nearby region. This would then cause a decrease in plate driving force and halt, or significantly slow down, the currently active subduction.

*RC:* Line 62: remove "lithosphere"

*AR:* This has been changed in the revised version of the manuscript.

*RC:* Line 82: "buoyancy forces" -> "density contrast" because buoyancy forces might result in misunderstanding (to negative buoyancy force due to the slab).

*AR:* We have changed "buoyancy forces" to "upward-directed, positive buoyancy forces" for clarity.

*RC:* Line 84: "at the topo of the mantle", a bit vague, the strength is only changed at the surface? You may rephrase it.

*AR:* This sentence has been rephrased during the review process.

*RC:* Line 97: the "thermal softening" = shear heating? If so, you may bracket with shear heating, otherwise additional explanation is needed or at least a reference is needed.

*AR:* The term "shear heating" has been added in the revised version of the manuscript.

*RC:* Line 112-123: as for the boundary conditions, it's better to show the details in this paper rather than referring to another paper. For example, I feel abrupt to see 1 cm/yr of total absolute extension in the later part, and the occurrence of calcite, mica is also very abrupt. But I do understand them only after checking your previous paper.

*AR:* The models presented here build upon models presented in a previous study. In the previous study, the large-scale and long-term modelling approach we apply here has been established and tested rigorously. For keeping the manuscript as concise as possible, we only mention the most important modelling aspects here and refer to our previous study for more detailed explanations.

*RC:* Line 146: “except that …” could be rephrased to, for example “except that a feldspar-dominated upper crust is replaced with a weaker rheology of quartz-dominated upper crust”

*AR:* This sentence has been rephrased in the revised version of the manuscript.

*RC:* Line 342-343: it’s not easy to understand, please rephrase

*AR:* This sentence has been rephrased in the revised version of the manuscript.

*RC:* Line 350: which model shows the first-order feature? Put (Fig. xx) behind “Pyrenees”

*AR:* This has been addressed during the review process.

*RC:* Line 406: as for the overestimated topography, what’s the erosion rate used (and observed from literature)? The overestimated topography is probably mainly due to the large subducted crustal volume, especially the upper crust. BTW, what’s the justification for the upper and lower crust thickness with 25 and 8 km, respectively?

*AR:* We used a constant erosion rate of 0.5 mm/yr above an elevation of 2 km. This value agrees with averaged exhumation rate estimates in the European Alps since 35 Ma (see [1]). Values for the initial crustal thickness are also consistent with average estimates (see [3]). We list the erosion and sedimentation rates in the revised version of the manuscript.

*RC:* In the model configuration part, it’s better to explicitly describe the correspondence of flow law (dry, wet olivine) vs. density (bulk DMM, hydrate peridotite). For example, feldspar (table 1) uses wet anorthite (table A1)? Quartz uses ‘wet quartzite’?

*AR:* This has been changed in the revised version of the manuscript.

*RC:* Line 418-423, just a comment for future study: the resolution used here is fairly high, the large amount of crustal material subducting to mantle depth might be attributed to the boundary condition, if the inflow is only imposed on the lithospheric domain, while outflow for the asthenosphere, I suspect that the slab is probably free to advance or rollback, in this case, much less crustal material could be entrained to the mantle depth due to its buoyancy.

*AR:* Indeed, the type of boundary condition might also impact on the force evolution. Exploration of different types of boundary conditions was not the goal of this study but is indeed a potentially interesting topic of future studies.

*RC:* Eq. (A9): isn’t it ½ rather than 1/3?

*AR:* We here used the diffusion creep law derived in laboratory experiments conducted under uniaxial conditions and therefore 1/3 is the correct conversion factor (see [2], p. 76).

*RC:* Line 491, as for the second invariant strain rate, do you miss ½?

*AR:* Indeed, this mistake has been corrected in the revised version of the manuscript.

*RC:* Fig. 3: do you correctly show the phase diagrams for the MORB, Hydr Peridotite, Andesite, Serpentinite, pelite? You may have to reshape the matrix

*AR:* The density tables are shown correctly, the results are consistent with the prediction of large density at high pressures (e.g. significant densification at the coesite field). The consistency of our results can be checked also by observing the almost-vertical (T-dependent) boundaries caused by dehydration of serpentinite minerals (for hydrous peridotites).

*RC:* Fig. 4: though the effective viscosity is shown, the most important part: the viscosity around the channel is not shown which inhibits better understanding, moreover, the relative strength of feldspar and quartz used (for example, no such description at Line 146) is not well known for readers (I suggest to plot either of the viscosity around the channel, or strength profile for different flow laws).

*AR:* We have added a new figure (new Fig. 7) to the main text, showing the viscosity in the channel and the crust. This clearly shows the low viscosity at the plate interface for weak serpentinites and a relatively stronger plate interface for strong serpentinites.

*RC:* Fig. 10: using different shading zones separating the timeline for those 5 stages may look better.

*AR:* We have tested various visualization techniques for this figure. However, the different stages are transient and partly overlap each other in time (f.e. subduction initiation and basin closure). Therefore, we chose the points 1-5 to only mark the onset of a stage. For clarification we have updated the caption of Fig. 10 during the review process.

## REFERENCES

[1] Bernet, M., Zattin, M., Garver, J. I., Brandon, M. T., & Vance, J. A. (2001). Steady-state exhumation of the European Alps. *Geology*, *29*(1), 35-38.

[2] Gerya, T. (2019). *Introduction to numerical geodynamic modelling*. Cambridge University Press.

[3] Huang, Y., Chubakov, V., Mantovani, F., Rudnick, R. L., & McDonough, W. F. (2013). A reference Earth model for the heat-producing elements and associated geoneutrino flux. *Geochemistry, Geophysics, Geosystems*, *14*(6), 2003-2029.

The authors thank Jonas Ruh for the very constructive criticism. Implementing his suggestions significantly improved the manuscript. In the following, the referee's comments are denoted *RC* and the author's reply is denoted *AR*.

*RC 1)* The model involves the mantle and therefore claims to have an advantage to crustal wedge models. That is a valid argument. But one point I want to address is the mantle rheology. Obviously, the mantle rheology should strongly affect the outcome of the models, and therefore, its implementation should be clearly introduced. The mantle rheology depends on various types of flow laws, including dislocation and diffusion creep. I spotted in the appendix table annotations that all mantle grain size is constant at 1 mm. This has a huge affect on mantle flow, as it results in close to 100% diffusion creep and much lower related viscosities than for pure dislocation creep. Grain sizes of 1 cm would already result in dislocation creep and actual grain sizes in the mantle are thought to be even larger based on a wealth of tomography papers and exhumed xenoliths. Now, I wonder why such a small grain size was applied, because it immensly affects the entire model dynamics. What happens if diffusion creep is switched off? A great example of the effect of diffusion creep is Fig. 5, where it can be observed that the lithosphere is 50 km thick everywhere and not restricted to less than 1300°C, which may be the temperature to enhance dislocation creep. One can also observe lithospheric dropping-type delamination or thermal erosion of the continental lithosphere, mainly driven by very low viscosities due to (unnatural) diffusion creep forced by very small grain sizes. I don't say the authors should redo all models, but just explain better what they implemented and what the effects are of the implemented grain size, maybe refering to Jiao et al. (2017; JGR).

*AR:* Indeed, diffusion creep flow laws require knowledge of the grain size. The natural grain size in the mantle and its temporal variation remains still little constraint in our opinion. Nevertheless, recent studies of whole mantle convection including state-of-the-art grain size evolution models predict grain sizes in the order of 1 millimeter between the lithosphere and 660 km depth (Danneberg et al. 2017, Schierjott et al. 2020). Based on these results we applied a grain size of 1 mm. Including in addition a grain size evolution law in our models was simply beyond the scope of our study. Most important, our previous work has demonstrated that the applied flow law parameters, with a constant average grain size of 1 mm, results in a viscosity structure that is consistent with geophysical constraints for the convection dynamics of the mantle and the thermal thickness of the overlying lithosphere (Candioti et al. 2020). In particular, the thickness of the lithosphere in the models presented here is ca. 120—150 km (see the depth of purely horizontal glyphs in for example Fig. 5). In addition, the up- and downwelling at the base of the lithosphere are rather the boundaries of convection cells that establish in the upper mantle and not lithosphere drops. Thus, in all the presented models, the lithosphere remains stable with more or less constant thickness over large time scales. The consistency of our predicted lithosphere thickness and mantle dynamics with geophysical constraints (heat flow, Rayleigh number etc., see Candioti et al., 2020) indicate that our viscosity structure is reasonable and applicable to natural mantle deformation and convection. We have added a paragraph in the model section of the revised manuscript for clarification.

*RC 2)* Also related to diffusion creep: Table A1 shows weird values. In the first part of the table it is stated that rheologies are taken from H+K2003, but the values are not correct

(wet diffusion for const. OH: Q = 335, not 375 for example). Also, all r = 0.0 should be changed to r = -. Because r = 0 would change the A according to your equation for A(Pa^-n-r.....)

*AR:* All chosen values are within the error range of published flow laws. As mentioned above, we have calibrated the flow law parameters in Candioti et al. (2020) to obtain a viscosity structure which is consistent with geophysical constraints and natural Rayleigh and Nusselt numbers. By using r = 0.0, the water fugacity term becomes 1 in the flow law and the parameter r is "removed" from the units of A. Using r = - would be an alternative, but we think using r = 0 is equally suitable.

*RC 3)* A general comment on using frictional / brittle / plastic. I also often write brittle-plastic, but I like it less every time I use this term. Of course, theoretically brittle processes are described by a viscous process in the numerical model and some authors use "plastic" for everyting that has a yield strength. However, brittle and plastic is still something different, where the latter describes a process acting at the molecular framework of crystals. I know that there is no exact definition on how to use those terms and I keep having arguments with many people from different fields, but anyway. Maybe my comment serves as the initiation of an interesting discussion. I'd write that the code is mimicking brittle deformation with a viscous implementation by reducing the viscosity based on a yield strength, or so.

*AR:* Indeed, some further elaboration on the nomenclature can be done here. In this study, the term brittle-plastic refers to rate independent (instantaneous) plastic deformation using a yield function that is motivated by experiments (Byerlee 1978) and mimics friction due to shear fractures in rocks at low temperature. The term viscous refers to all rate dependent (crystal) plastic deformation (dislocation, diffusion and Peierls creep). This has been clarified in the revised version of the manuscript.

Minor suggestions:
*RC:*- The abstract is very long and very wordy. I had difficulties to follow it. You can easily delete some unnecessary information to better attract the reader to go on and read the entire manuscript.

*AR:* The abstract has been shortened in the revised version of the manuscript.

*RC:*- L34: I'd delete "-plastic"

*AR:* "Plastic" has been deleted in the revised version of the manuscript.

*RC:*- L39 and 41: Twice: "crustal wedge models have also been" in a row, sounds weird

*AR:* These sentences have been rephrased in the revised version of the manuscript.

*RC:*- L44: May cite Jammes et al., 2012 after ... actual collision

*AR:* Jammes et al. 2012 do not model the opening of wide basins and, therefore, this reference is not well suited here. However, we have added this reference to the discussion where we apply our models to the Pyrenean orogeny.

*RC:*- L47/48: indicate that the overriding lithosphere is meant. Of course, without this information is much stronger taking into consideration the second part-sentence. Of course, also your wedges have a plane décollement, but with the unclear sentence the reader might understand that lower plate lithosphere is detached and involved in the wedge. As is, the overridig lithosphere just forms a backstop...

*AR:* This has been changed in the revised version of the manuscript.

*RC:*- L50: likely significantly – rephrase

*AR:* This sentence has been rephrased during the review process.

*RC:*- L55: maybe mention "body forces" in contrast to "surface forces"

*AR:* "Body force" is the general term that includes forces due to f.e. gravity, electric or magnetic fields. In this context, we explicitly refer to forces due to gravity and, therefore, kept the term "gravitational forces".

*RC:*- L60/61: That is not true, of course they do, they have gravity and density. But it is not very important as in mantle-scale models

*AR:* Of course, the crustal wedge models consider gravity and density which modulate the frictional behavior and the gradients of topography. However, most wedge models employ a singularity point at the base of the backstop that inhibits deep subduction of crustal material. Consequently, these models cannot predict exhumation of subducted crustal units driven by upward-directed buoyancy forces arising from density contrasts between the subducted crust and the surrounding mantle. For clarity, this sentence has been reformulated in the revised version of the manuscript.

*RC:*- L62/63: I don't understand. Tey don't consider shear forces at the interface? And what happens along the décollement? Shear strength along the interfaces defines these models, very similar as in the presented models.

*AR:* This sentence was indeed unclear and we have removed it from the revised version of the manuscript.

*RC:*- L63: what is lithosphere-upper mantle? lithospheric mantle? or lithosphere and non-lithosphere mantle? I'd use crust in constrast to mantle, and asthenosphere in contrast to lithosphere.

*AR:* The term upper mantle refers to the asthenosphere and the transition zone, whereas the term lithosphere comprises the crust and the mantle lithosphere. This has been clarified in the revised version of the manuscript.

*RC:*- L65ff: maybe refer to Fig. 1 in this paragraph

*AR:* A reference to Fig. 1 has been added in the revised version of the manuscript.

*RC:*- L95: Brittle deformation
*RC:*- L96: simplify instead of repeating. We do not apply any frictional nor viscous strain weakening

*AR:* These two sentences have been reformulated in the revised version of the manuscript.

*RC:*- L102: I couldn't find mica and calcite in the text, although they are introduced in the figures. Maybe also explain that mica is a weak inclusion at shallow levels and a strong inclusion at deeper levels (in contrast to quartz for example).

*AR:* This has been explained in the revised version of the manuscript.

*RC:*- L143: how did you come up with 18? I always found it very weird just to multiply the pre-exponent, but ok. You may just call it a stronger rheology. I don't know how much it still has to do with serpentine. But why 18? Any explanation in the text would be appreciated.

*AR:* We increased the prefactor gradually and the most significant change in dynamics was observed for a prefactor value of 18. An explanation has been added to the revised version of the manuscript.

*RC:*- Table 1: I would write in the annotations what 1 and 2 are, not only refer to another table in the appendix

*AR:* This has been changed in the revised version of the manuscript.

*RC:*- L165: 120-130 km??

*AR:* The depth of 20-30 km is correct and refers to the region below the rift center (see Fig. 4d). This has been clarified in the revised version of the manuscript.

*RC:*- L166: the width of the left margin

*AR:* This has been implemented in the revised version of the manuscript.

*RC:*- L183: sheared off

*AR:* This mistake has been corrected in the revised version of the manuscript.

*RC:*- L303: refer to Figure after "diagram"

*AR:* A reference has been added during the review process.

*RC:*- L338: The natural examples need a better introduction. I for example would be interested in the average shortening. Is there less exhumed material in the Pyrenees because there is less shortening? How does it compare to the model results?

*AR:* We did not vary the convergence velocity or the width of the marine basin for this study. However, for constant shortening we observe exhumation in some of our simulations and in others we do not. This indicates that the plate interface and crustal strength are probably more important parameters controlling the exhumation than the amount of shortening. A detailed parametric investigation on the shortening was beyond the scope of this study. We have introduced the plate kinematics of the Pyrenean orogeny in a bit more detail during the review process.

*RC:*- L346: refer to figure after "convergence"

*AR:* A reference to a figure has been added to the revised version of the manuscript.

*RC:*- L353: delete "model", as in L339

*AR:* This has been changed during the review process.

*RC:*- L356: maybe worth citing Cristina Malatesta's paper from 2012 in Lithos that investigates subduction of serpentinized oceanic mantle

*AR:* While Malatesta et al. indeed investigate subduction of narrow oceanic basins, they consider an oceanic crust that includes also basalts and gabbros (i.e. a mature oceanic crust) along the entire basin. Whether the Piemont-Liguria ocean was mainly a mature or embryonic ocean is still actively debated. Malatesta et al. also model intraoceanic subduction without continental subduction and subsequent collision, which is key for the Alpine orogeny. We, therefore, decided not to refer to this study when discussing implications for the Alps.

*RC:*- L373/374: because the lower crust is stronger? Then explain that it is because there is no decoupling at the Moho but a weak lower upper crust.

*AR:* This has been explained in more detail in the revised version of the manuscript.

*RC:*- L393/394: that is weird. The lithosphere just acts as a backstop could be argued. It does not involve the entire lithosphere in wedging (there are also crustal models with elastic beams depending on bouyancy: e.g., Stockmal et al., 2007; Fillon et al., 2012; Ruh, 2020).

*AR:* Indeed, in a lithospheric wedge model the mantle forms the backstop whereas in crustal wedge models the crust forms the backstop. The important difference is that in lithospheric wedge models the continental crust may subduct and induce upward-directed buoyancy forces that impact on the collision dynamics. Such upward-directed forces cannot be observed in crustal wedge models, because deep (>50 km) subduction of continental crust is not modelled.

*RC:*- L395: the model by Platt is rather crustal. He appied it and compared it mantle-scale orogens

*AR:* We have reformulated this sentence in the revised version of the manuscript.

*RC:*- L398-401: I see that the entire downgoing lithospheric mantle remains undeformed (except bending) beneath the crustal wedge. Hence, there is no strong variation in plate-parallel velocity. This is similar to apply a lower boundary condition to a crustal model...

*AR:* Indeed, the predicted velocity in the subducting plate of our models might be similar to an applied lower boundary velocity condition in crustal models. However, our models (1) avoid kinematic singularity points, (2) allow for continental subduction and (3) predict the subducting plate velocity far away from the applied boundary which is the advantage of our models. This sentence has been reformulated during the review process.

*RC:*- L411/412: as written, it sounds like delamination is also called roll-back..

*AR:* This sentence has been rephrased during the review process.

*RC:*- L522: introduce before that particles are applied. And state how many per cell etc.

*AR:* This information has been added to the modelling approach section.

*RC:*- Table A1: Kohlstedf should be Kohlstedt

*AR:* This mistake has been corrected in the revised version of the manuscript.

## REFERENCES


[1] Byerlee, J. (1978). Friction of rocks. In *Rock friction and earthquake prediction* (pp. 615-626). Birkhäuser, Basel.

[2] Candioti, L. G., Schmalholz, S. M., & Duretz, T. (2020). Impact of upper mantle convection on lithosphere hyperextension and subsequent horizontally forced subduction initiation. *Solid Earth*, *11*(6), 2327-2357.

[3] Dannberg, J., Eilon, Z., Faul, U., Gassmöller, R., Moulik, P., & Myhill, R. (2017). The importance of grain size to mantle dynamics and seismological observations. *Geochemistry, Geophysics, Geosystems*, *18*(8), 3034-3061.

[4] Jammes, S., & Huismans, R. S. (2012). Structural styles of mountain building: Controls of lithospheric rheologic stratification and extensional inheritance. *Journal of Geophysical Research: Solid Earth*, *117*(B10).

[5] Malatesta, C., Gerya, T., Scambelluri, M., Federico, L., Crispini, L., & Capponi, G. (2012). Intraoceanic subduction of “heterogeneous” oceanic lithosphere in narrow basins: 2D numerical modeling. *Lithos*, *140*, 234-251.

[6] Schierjott, J., Rozel, A., & Tackley, P. (2020). On the self-regulating effect of grain size evolution in mantle convection models: application to thermochemical piles. *Solid Earth*, *11*(3), 959-982.