We thank **Referee #1** for their positive feedback and their constructive comments. We considered each comment and revised the manuscript accordingly. Please see below our responses to each comment (in blue, with changes made in *italics*) and the annotated manuscript for the revisions.

These are the main changes we made to the manuscript, all in response to questions from both reviewers:

- To illustrate the evolution of asthenosphere viscosity and velocity to help the discussion of how drag evolves as these two parameters evolve, we added Figure S2-3 (Note that we moved all the appendices to the supplementary material).
- We added further motivation for the used equation for estimating drag, for evaluating drag relative to the main driving force (slab pull), and more information on how parameters for drag estimates in the model were measured (including information marked on the new Figure S2-3).
- We explain better that the correlation between plate size and age is a feature of the Cenozoic plate configuration, not a causal relationship.
- We fixed some copying errors in the table entries of the analytical estimates of Pacific and Cocos plate drag.

1. *Many times in the introduction the authors talk about the viscosity of the asthenosphere and the ratio between lithosphere-asthenosphere viscosity as 'high' or 'low' or 'too low'. Since this is a crucial part of the paper, I think the author should describe this in a more quantitative manner (i.e., mention some estimates with actual numbers suggested by the cited literature). This is true also throughout the rest of the manuscript, even in the results where the viscosity values are easily to extract from the models (e.g., lines 223-224). The same for the conclusions: e.g. line 441-443 "Models with a low-viscosity asthenosphere do reduce the contribution of trench motion to plate convergence to more Earth-like values, as observed in previous studies". How low? Which values range are we talking about?*
   We agree that it is useful to add some more numbers. Many previous models considered ratios of effective lithosphere-asthenosphere viscosities of a factor 100-1000 (with reference asthenospheric viscosities of $10^{21}$ to $10^{20}$ Pa s). Weak asthenosphere models tend to have viscosity ratios of $10^3$ to $10^4$ (for reference asthenospheric viscosities of $10^{20}$ to $10^{19}$ Pa s), which when these are the outcome of non-linear composite rheology still allows for viable subduction. *We now added these numbers plus references to the introduction for context (line 53-55, 71-73 and 85). The values used in the models are discussed and motivated with references in the last paragraph of the 'Methods' section (lines 157-166). We added a new supplementary figure (Fig. S2.3) that displays the evolution of viscosity through time for the long and short reference models along the profile that is used for basal drag calculations. And we include numbers for the viscosity ratios in our models in the conclusions (lines 462 and 468)*

2. *Following the previous point, I would be curious to see the evolution of the viscosity of the asthenosphere in Fig. 4 (or in the Supplementary). The reason for this is that I have a few doubts on (1) the effect of the used methodology and (2) the effect of plate velocity on the viscosity of the asthenosphere and a figure that shows it would likely clear them out. In particular: (1) Regarding the methodology. The viscosity of the asthenosphere is weakened by multiplying the computed viscosity by a factor of 0.5. However, the computed viscosity is a combination of diffusion, dislocation and Pierels creep, thus, it is strain dependent. When lowering the viscosity by a factor 0.5, the stresses in the asthenosphere will be lower too and the viscosity will want to be higher again at the next time step. Is that what happen? Or does it more or less stabilize? (2) Several times the authors say that because the velocity of a plate increases then the basal drag increases too. However, wouldn't the viscosity of the asthenosphere decrease with high plate velocities due to larger deformation? If so, the basal drag would decrease too or is there something else in Eq. 1 that also changes to compensate a decrease of EtaAst?*

(1) Indeed, the viscosity is strain-rate dependent. The lowering of viscosity by a factor 0.5 leads to increased velocities. Increased velocities are associated with increased strain rate, and lead to a feedback that results in further lowering (rather than increasing) of viscosity. The net effect of the scaling is a reduction of viscosity by more than the applied scaling factor. *We write: "The compound effect of applying the 0.5 reduction factor and consequent strain-rate weakening is an order of magnitude decrease in asthenospheric viscosity compared to the reference models."*

(2) The viscosity of the asthenosphere will tend to decrease with higher velocity (if this leads to higher strain rates), reducing the basal drag. However, the direct effect of increased velocity is an increase in drag, which can dominate over the effect of the lowering of viscosity. This happens in the reference case short-plate models, around the time the plates reach 660 (t660), and in the weak asthenosphere long-plate models around t660. The two effect of increased velocity, direct and on viscosity, is also the reason that the plate velocity graph and the basal drag graph do not always show the same trends. *We now discuss this non-linearity in our presentation of the reference models and added a graph showing the evolution of the viscosity and velocity to the supplementary material: "Note that due to the nonlinear feedbacks in the models, there are about 10 Myr right after the plates reach the ULMB where the short plate experiences a higher drag than the long plate. These relatively high drag values for the short plate are a consequence of the high velocities that the short plate attains in response to the initially significantly lower drag (Fig. S2.3)."*

3. *I do not fully understand the reasoning in the Discussion about the correlation between plate size and age at the trench (Fig. 8B). A larger plate would have ridges further away from trenches (e.g., Pacific vs. Cocos) and therefore more likely to have older lithosphere at trenches, could it not be 'that simple' without the need to use the basal drag as explanation?*

The correlation of size and age is an observation. Such a correlation might be expected if spreading rate is relatively constant. However, closer examination of the plate reconstruction reveals a more complex picture of changing spreading rates, subducting ridges, changing trench retreat rates and plate geometry that eventually determine plate size and trench age.

We use this observation to argue that larger plates tend to have a stronger driving force. Increased basal drag with increasing plate size can explain that larger plates are not observed to reach systematically higher velocities. This explanation only applies to the Cenozoic where plate size and age of the subducting plate are correlated.

*We changed some of the phrasing in the revised manuscript to make it clear these are separate observations. In the discussion we write: "A consequence of the observed correlation of average age at the trench and plate size is that plates with a stronger slab pull tend to also have larger surface area and hence a stronger resisting basal drag. This provides a mechanism that offsets the velocity-enhancing effect of larger driving forces of old plates, and could explain the observation that velocities of subducting plates on Earth, today and throughout the Cenozoic, do not increase with age but tend to be mostly stable around 8-10 cm/yr".*

*We also write in the discussion: "Note that the relation between age and size is not a causal relationship, but a feature of the plate configuration that has dominated most of the Cenozoic. Early in the Cenozoic, there are several cases that deviate from the buffered velocity trend. At the start of the Cenozoic, much of the subduction surrounding the Pacific plate consumed relatively young lithosphere, even though the Pacific plate itself was already large in size (horizontally aligned points at the top of the area–age trend in Fig. 8b). Other early Cenozoic deviations include very high velocities of the last remnants of the Izanagi and Kula plates (points with area of about* $2500^2$ *km*$^2$ *and velocities around 17 cm/yr in Fig. 8a) and low velocities of Farallon plate (area about* $6000^2$ *km*$^2$ *in Fig. 8a)".*

*In the conclusions, we reworded the last paragraph to: "Based upon an analysis of a Cenozoic plate motion reconstruction (Müller et al.,, 2016), we suggest that the reason that most plates move at velocities around 8-10 cm/yr is because the plate configuration during this era was such that plate size correlated with plate age at the trench, i.e., both driving and resisting forces* *increased together. Note however, that this correlation between size and age is not causal and may not have existed in other times of Earth history. As a result, during the Cenozoic, the increase in basal drag more or less balanced the increase in plate velocity induced by increased slab pull with increasing age. Such co-variations between plate velocity, age and size should be considered in regional models of subduction systems"*

4. *The authors should discuss how their results and conclusions could be affected by viscous anisotropy? From Becker and Kawakatsu, GRL, 2011: "One of the major limitations of our study is that we only considered a few instantaneous flow examples for which the influence of anisotropy may overall be negligible. This only indirectly addresses more complex, evolving scenarios such as changes in plate motions, or plate boundary dynamics, where mechanical anisotropy may well be relevant." Viscous anisotropy has the potential to have an important effect when looking at plate velocities during the Cenozoic as the authors do here because changes in plate motion direction would change the orientation of the anisotropy and could contribute to change the asthenosphere viscosity (hence, basal drag). I understand that this cannot be included in the calculations, but I think it deserves to be mentioned and discussed.*
   *We now clarify that other factors may contribute to viscosity when we introduce our choice of rheology in the methods: "We use a temperature-, pressure- and strain-rate dependent composite rheology, which naturally results in a lithosphere and asthenosphere. Other factors may contribute to the contrast between lithosphere and asthenosphere viscosity, e.g. hydration, partial melt and anisotropy (e.g., Hirth and Kohlstedt, 1996; Becker and Kawakatsu, 2011). Here, we consider diffusion and dislocation creep mechanisms, combined with a low-pressure yield-stress mechanism to approximate brittle failure and an approximation of Peierls low-temperature plasticity at high pressure (e.g. Čížková et al., 2002; Garel et al., 2014). Equations and parameters for the reference cases are as in Garel et al. 2014 and are given in Table S1.1. We vary parameters, as discussed below, to evaluate different relative lithosphere and asthenosphere viscosities."*

5. *I think the first paragraph of Conclusions belongs to the Discussion. Consider moving it there.*
   The first paragraph of the conclusions is a recap of the motivation for the study previously outlined in the introduction. We think it belongs in the conclusions section rather than the discussion section but hope that with our other clarifications of the introduction, discussion and conclusions this organisation makes more sense.

6. *Eq. 1 and 2. How is the lithosphere defined? Is it defined by the 1100 degC isotherm? Please specify it, since it matters for parameters like hLit, hAst, and Sslab. (I found the answer later on in the lines 144-145, but the authors might want to either repeat it or move it here where the variables of the equations are explained). At what depth is Sslab taken? And is it an horizontal section or perpendicular to the slab? Given the importance of these calculations for the study, I would suggest to have a figure with a schematic cartoon of a model showing where and how all the variables used in Eq. 1 and 2 are taken. It could go in the main manuscript, in the supplementary material, or merged with Fig. 2.*
   The lithosphere is defined as the part of the model that is colder than 1100°C. This does mean the thickness of the lithosphere and asthenosphere evolves with time. It is mentioned in the 'Model set-up' section, on Figure 2 and in the caption for Figure 2. $S_{slab}$ is calculated as the surface of a rectangle with one side the length of the slab (*as explained in lines 204-206*), and the other side the width of the slab, taken as a horizontal line across the slab at 220 km depth. *We now improved the description of the parameters of Eq. 1 and Eq. 2 in the revised manuscript and added a figure in the supplementary material (Figure S2.3) where we also mark where some of the key parameters were measured.*

7. *Table 1. Why is the slab pull force (and the basal drag) for the Cocos plate one order of magnitude larger than for the Pacific? Shouldn't it be the opposite? Please check your calculations. U is the plate velocity, what is deltaU (used in Eq.1)? S should be Splate (as referred to in the text). Is Vslab = Sslab*Ltrench or Sslab*Ltrench*Wslab?*

Thanks for catching this error. There was a copying error in the table (the values of the forces were copied to the wrong row in the table). Vslab is calculated as Lslab*Ltrench*Wslab where Lslab is assumed to be 700 km (lines 364-365). After re-examining the calculation, we found that there was a missing factor 2 for the drag force. *We corrected the mistake, the calculation, the corresponding text and expanded the description of Vslab in the revised manuscript.*

8. *Fig. C1: at the moment this figure is confusing. A better layout could be with each column showing a parameter. And also using more distinct colormaps for age and subduction velocity (now they are very similar). Also the colours of the different subduction zones are different from those used in the other figures (Fig. 1, 8, C2). When possible, I would suggest to keep them the same.*
   *We changed the figure in the revised manuscript according to the comment.*

9. *Technical corrections Line 331: odd sentence, rephrase. Line 338: Other approximations in addition to what? Line 401: "tend to be mostly fall around…" delete 'be' or 'fall'*
   *We changed the sentences in the revised manuscript according to the comments.*