We thank Ömer Bodur (**Referee #2**) for the detailed feedback and comments. We considered the comments carefully and revised the manuscript accordingly. Please find our replies to the comments in the following section (in blue, with changes made in *italics*) and see the annotated manuscript for changes to the text.

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. *I do find the arguments about linking slab strength to trench advance/retreat convincing and important. However, I do have concern that the analysis of the numerical models might have been done with partial ignorance of the forces (mostly resistive) acting on the subducting plate (Solomon and Sleep, 1974; Forsyth and Uyeda, 1975), henceforth limiting our understanding of how basal drag solely affects the whole plate subduction system. For example, the drag around the sinking slab (i.e. slab resistance) must also be acting along with basal drag and this has not been considered in the analysis so that the readers cannot separate the contribution of each. Therefore, it's not clear if it's the basal drag component itself driving the changes in plate velocity and slab morphology.*
   Similarly, once the slab dip is small and there is longer slab sinking into the mantle, analysis of relative significance of basal drag to the slab pull can be misleading because slab resistance is not taken into account with the increased slab area (i.e. slab interface in 2D) although it was considered for slab pull. The arguments of force balance cannot be reduced only to basal drag and slab pull as other forces also act and vary in time in the numerical model.
   We agree that the force balance cannot be reduced to basal drag and slab pull. However, we make a different argument in the text. We show, throughout the evolution of the model, how much of the slab pull is resisted by basal drag. We do not make the point that the evolution of velocity and slab morphology is controlled solely by these two forces, nor do we try to show the full force balance acting on the plate and slab throughout the evolution. By referring to the ratio of basal drag to slab pull, we can compare the significance of basal drag in the subduction process to other forces, such as viscous dissipation and ridge push, which have been examined against slab pull in previous studies. *We discuss the range of forces that contribute in the second paragraph of the introduction and have now worded this more clearly when we introduce the equations used to estimate basal drag and slab pull: "To evaluate the relative importance of basal drag in the overall force balance, we estimate basal drag below the subducting plate for each time step in the models, and compare it to the main driving force, slab pull."*

2. *Equally important, some of the model results contradict with what the paper proposes throughout the text. For example, in the numerical experiments, a decrease in asthenosphere viscosity results in an increase in relative basal drag right after (for about 10 Myr) the plate has sunk to 660 km depth*

(comparing light red line with green line in Fig 6d). This is contrary to what is expected and needs clarification well before addressing other effects of basal drag (e.g. slab morphology).

Indeed, the non-linear feedbacks in the models sometimes cause unexpected results. In this case, the lower viscosity of the asthenosphere leads to high velocities which by the time the plate reaches 660, mean that the long-plate model with lower viscosity asthenosphere, briefly (~ 10 Ma) experiences a higher drag than the reference case with higher asthenospheric viscosity at this same stage of the model. However, in spite of the complexities in how basal drag evolves, all results are consistent with the conclusion that basal drag is important.

We added a short discussion of this to the description of the Reference models in the Results section in the revised manuscript: "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. *The authors use a derived equation to estimate the slab pull and basal drag forces over time. These equations are (Eqn. 1, 2), most likely, valid only for iso-viscous plate and asthenosphere, and therefore may not be suitable to apply on the numerical modelling results. The authors also need to be cautious about necking of the subducting plate which results in significantly lower viscosities at the subducting plate as can be seen in Figure 3. This means, the slab pull force cannot act efficiently on the unsubducted part of the plate, hence an interpretation of the relative strength of basal drag and slab pull may become misinforming.*

   Indeed Eq. 1 and 2 approximate the forces. In theory one could integrate over velocity, viscosity and density fields for more detailed measurements. However, this too would involve assumptions in choosing the boundary of the driving lithosphere and resisting asthenosphere, and a more non-linear dependence on those choices. We found the approximate solution illustrates model behaviour well enough. Due to our adaptive mesh, we are able to resolve the strong core of the plate even if it is thin. Note that the minimum element size is 0.4 km, as noted in the text, which allows us to resolve a thin strong core, which may not be clearly reflected on the colour scale in Fig. 3 (but illustrated clearly in figures in Garel et al., 2014). This means that stress is always quite effectively transmitted from the surface plate to the slab and from the slab to the plate. In all of the results we analysed, we found that the viscosity in the bulk of the slab was at least an order of magnitude higher than the surrounding mantle

   We added a mention of this where the equations are introduced in the revised manuscript. "Note that the expression for F_{SP} assumes stresses are effectively transmitted from the slab to the plate, which is appropriate because the adaptive grids of our models ensure we always resolve the strong slab core, even when it is quite thin (Garel et al., 2014)".

4. *Abstract: It's worth briefly indicating why the force balance in subduction dynamics is incompletely understood (e.g. fails to explain plate velocity?) I think the force balance, or the method itself, is not to blame, but the contributions to the force balance by different forces are quite uncertain. This should be made clear.*

   We reworded this sentence to "Subducting slabs are an important driver of plate motions, yet the relative importance of different forces in governing subduction motions and styles remains incompletely understood."

5. *[Lines: 140-150]: The lithosphere has a number of definitions and one of which is by viscosity profile (e.g. (Conrad and Molnar, 1997). In numerical models, the viscosities will vary and the effective*

lithospheric thickness that you have used in the estimation of basal drag will also change accordingly (Bodur and Rey, 2019).

That is indeed the case in our models. We use the 1100 ˚C isotherm as the base of the lithosphere. The depth and shape of the isotherm changes with time and the thickness of the lithosphere changes with it. To account for that in the drag calculation, we measure the thickness at a given from the trench at each time step. *The process is detailed in lines 200-204, and new supplementary Figure S2.3 illustrates the evolution of various parameters, including the thickness of the lithosphere, through time, for the reference model.*

6. *[Lines 158-160]: Please justify why avoiding any slab detachment during subduction is favoured. If the model results are only applicable to plates not showing slab detachment, then this should be mentioned early in the text or in the abstract. It's important to acknowledge that slab detachment has been used to explain important features of the Earth (Göğüş and Psyklywec, 2008; Duretz and Gerya, 2013; Hacker and Gerya, 2013)*

   Our aim in this paper is to model basic continuous subduction behaviour. Slab detachment is the terminal stage of subduction, which has been observed in different settings than ocean-ocean subduction (e.g. continental collision, subduction of continental fragments or seamounts, etc. See examples in Hacker and Gerya, 2013). We therefore aim for a rheology that allows continued subduction rather than slab detachment, or subduction freeze.

   We added this explanation to the Model Set-Up section in the revised manuscript: "With this range of lithospheric and asthenospheric viscosities, we can generate continuous Earth-like oceanic style subduction while avoiding immediate slab detachments which result from to lithospheric weakening due to high strain rates, or the stalling of subduction due to unattainable forces required for the bending of very strong lithosphere".

7. *[Eqn1]: Please provide the derivation of the equation for basal drag and/or the page # of the citation you provided.*

   The equation is based on Eq. (6.24) in section 6.3, Page 419, Turcotte, D. L. and Schubert, G.: Geodynamics, 2nd ed., Cambridge University Press, Cambridge. Available from: http://dx.doi.org/10.1017/CBO9780511807442, 2002. *We added the section number to the reference in the text and include the doi link in the reference list.*

8. *Section 4.3 [Lines 364-365]: The slab dip you consider here is quite higher than numerical models show. Why? It's also unclear what sort of data you have used to calculate the slab pull and basal drag estimates in Table 1. Please be more specific so that one can derive the same results individually for further reference. Also, for different plates, the asthenosphere viscosity ($\eta_{asth}=10^{19}$ Pa×s) is not necessarily the same, so you may need to consider different viscosities, at least mentioning about it.*

   This is partly the result of copying mistake. The dip used for the calculation was 70° rather than 80°. This dip angle was chosen as it is representative for dip angles observed on Earth (Lallemand et al. 2005). It should be noted that:

   - Considering a single value of slab dip angle, as well as single value for the slab width and other parameters, for a plate as varied as the Pacific cannot provide more than a first order estimation.
   - The difference between considering an angle of 70° and 50° is a factor of 1.2. This is a minor uncertainty considering the estimation of slab width and other values.
   - As first order estimation of forces, we think it is justified to use a single value of asthenospheric viscosity.

We corrected the error and added a statement that the calculation is used for first order approximation in the revised manuscript. And we explain that the dip used is a representative dip from a global compilation of subduction parameters.

9. *[Line 391]: Although they can be correlated, Fig. 8b doesn't show subduction zone length, but the plate size vs. plate age at trench. It's better to be more specific.*
   Fig. 8b presents the square root of the surface area of the plate vs. trench age. The square root has units of length. In Fig. 1, plate size is displayed in terms of a typical length calculated in two different ways. The way typical length is calculated is explained in the captions. *We changed the labels of Fig. 1 and 8b to "Plate size [typical length in km]". To avoid any confusion, we removed references to length when discussing 3D plates. It is only used when describing the results of the 2D models.*

10. *[Lines: 399-401]: The correlation is weak already (based on error bars and scattered points in Fig. 8a), and the argument on explaining an already weak correlation "at least in part" is making this sentence more confusing for readers. I recommend restructuring those lines.*
    We rephrased this part in the revised manuscript to make our intention clearer. We explain that old plates tend to be large, and as a result their movement is resisted more. This provides a possible mechanism (which probably works in tandem with other mechanisms) to lower the potentially high subduction velocities which are expected in plates with old and cold slabs. *We removed the phrase "(at least in part)"*

11. *[Line 156]: 1024 Pa×s – 1025 Pa×s should be changed to 10^{24} Pa×s – 10^{25} Pa×s.*
    [Line 383]: Fig. B1 needs to be changed to Fig. C1.
    We revised the manuscript according to these comments