Dear Editors,

We are glad to receive the review report and would like to express our sincere thanks to you and the reviewers. Without the constructive comments from you and the reviewers, the quality of this manuscript cannot be significantly improved. All the comments and suggestions have been carefully considered to revise the manuscript. Detailed reply to all comments and the associate manuscript modifications are given below.

Reply on RC1

1. The language of the manuscript needs to be fully checked and revised by a professional editing service or a native speaker

Reply: *We appreciate your advice and have revised the manuscript carefully to improve the language, with the help of all the co-authors.*

2. In this study, the relation between spreading rate and age of oceanic lithosphere is ignored. Usually, higher spreading rates create younger lithospheres at a constant distance. In this study, the authors assumed that the lithospheric age is constant near the side boundaries (50 Myr). As a result, by imposing some higher velocities near the side boundaries to simulate higher spreading rates, the lithosphere becomes under extension and since the ridge is the weakest point in the system the width of ridge changes (it is clearly seen in e.g., Figs 3e,4e and 5a-b); higher rates lead to wider ridges. Could the authors explain to what extend is this assumption realistic? I think the formation of cracks in the lithosphere is the consequence of this assumption

Reply: *We appreciate this comment which triggers us a lot of thinking. Indeed, the initial temperature distribution of the oceanic plate consists of the half-space cooling model and thermal equilibrium part. The half-space cooling model is used to describe the oceanic plate younger than 50Myr, and the thermal equilibrium structure is used to describe older oceanic parts. The thermal equilibrium thickness of the older lithosphere is constant (i.e., ~100 km; corresponding to a thermal age of 50 Ma). We further* *checked our model results, especially the stress state in the whole lithosphere. The result shows that the lithosphere seems under extension owing to the high internal velocities (Figs. S1, S4). Actually, the "tension cracks" used in the main text may be inaccurate. There is no weakening or plastic deformation in the lithosphere, and there are no normal faults near the surface in the models. We describe the stress distribution in the lithosphere to highlight the stress localization, which occurs in the lithosphere where plume flows to the ridge. As a result, we revised and simplified the description of the "cracks" in the main text (lines 212-213 and lines 249-250).*

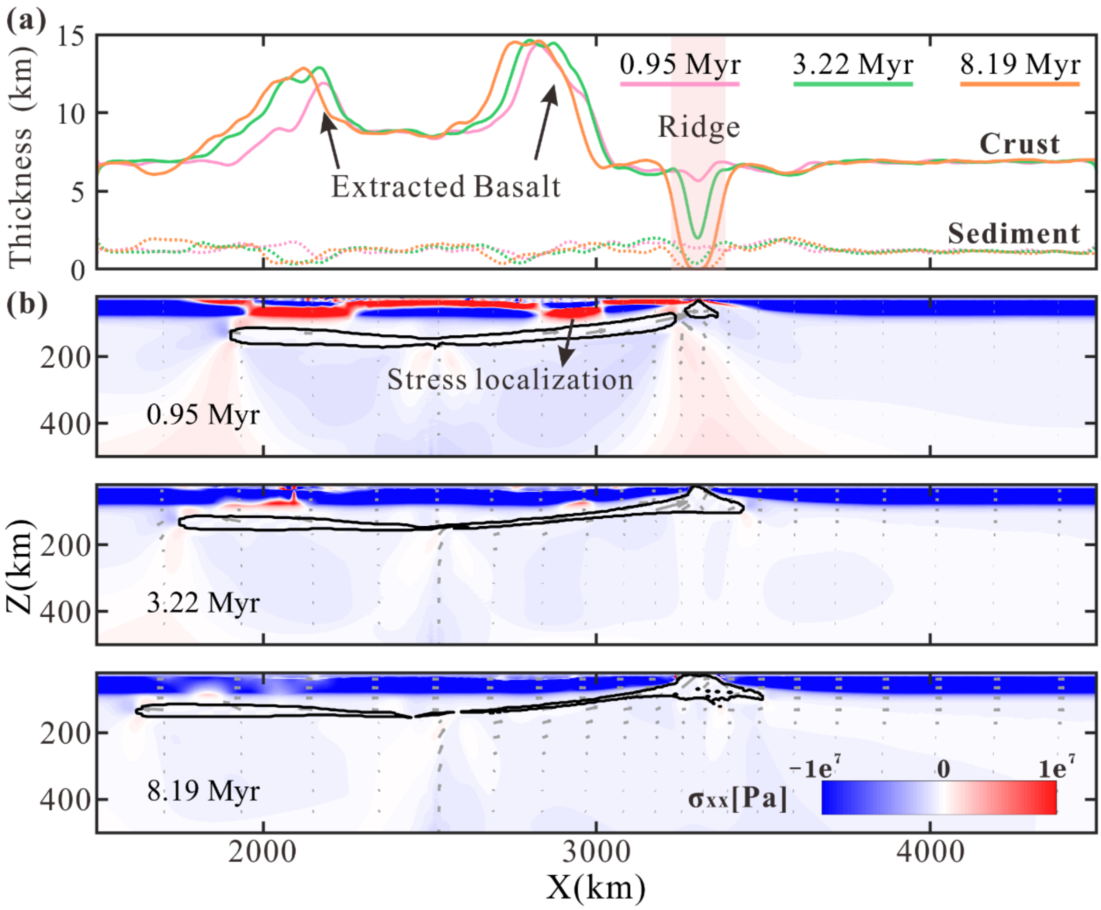

Figure S1. Reference model (M12, see Table S1, same as Fig. 3) evolution of ridge-ward plume flow shown by (a) crust and sediment thickness, (b) normal stress. The mantle plume weakens the overlying oceanic plate and changes the stress state of the overlying oceanic plate. Molten plume material beneath the lithosphere is extracted to the crust.

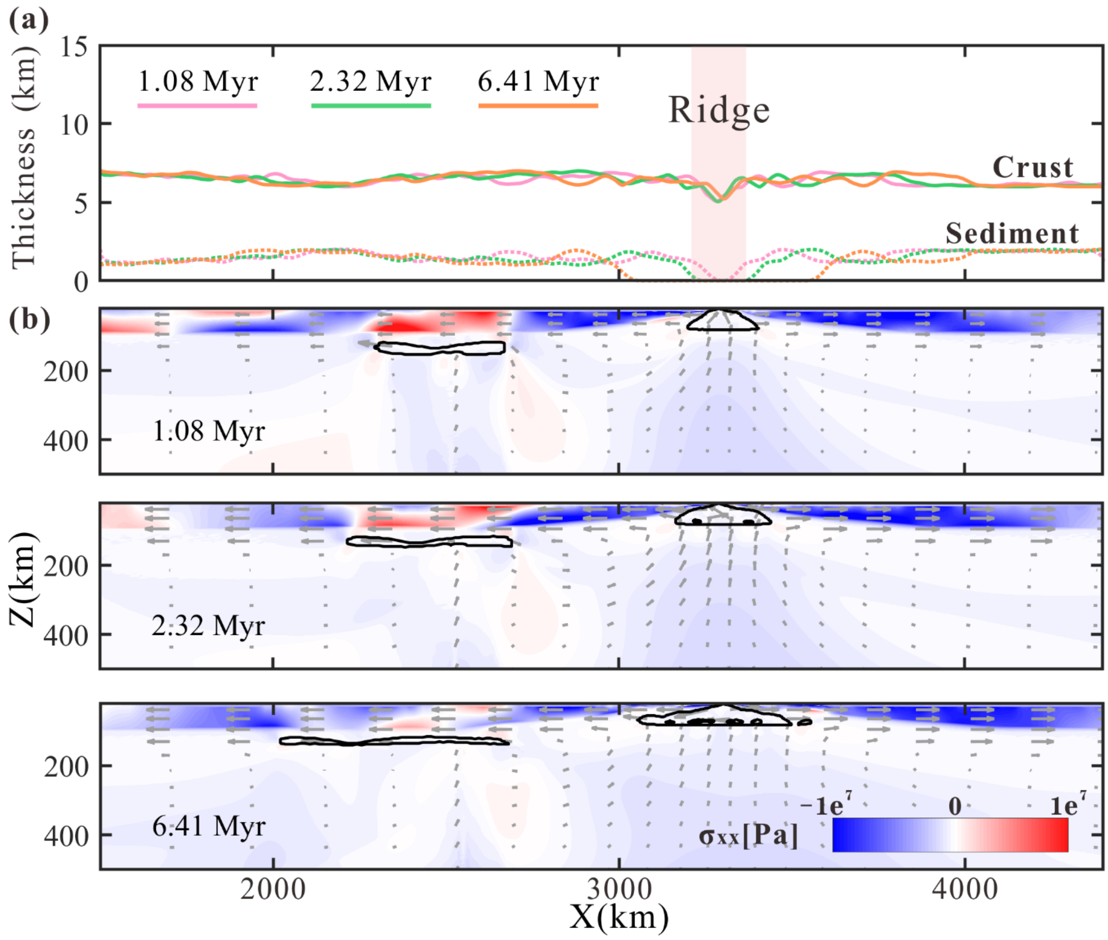

Figure S4. Reference model (M77, see Table S1, same as Fig. 4) evolution of plate-drag plume flow shown by (a) crust and sediment thickness, (b) normal stress. The mantle plume weakens the overlying oceanic plate and changes the stress state of the overlying oceanic plate. Molten plume material beneath the lithosphere is extracted to the crust.

3. In the abstract it is written "plume migration driven by plate drag is promoted by fast-ridge spreading rate." This is true only if the plume radii are small. For large plumes the rate of spreading is irrelevant (Fig. 6). This should be mentioned here and also in discussion and conclusions.

Reply: *We agree with you and rephrased this sentence (lines 19-20). Indeed, plate dragging is most significant when the plume buoyancy is relatively small. When the plume is buoyant enough, plate drag plays a minor role than the plume self-spreading* *on plume-ridge interaction. We have revised this part and the discussion section in the manuscript.*

4. Usually decomposition melting of plume head causes the formation of a plateau above the plume head. Where do plateaus form in the models? I suggest that the authors add information about where plateaus form and how thick the crust is to the manuscript. The temporal evolution of plateaus is also interesting to be investigated.

Reply: *We appreciate this suggestion and add figures to present the temporal evolution of the extracted melt, displaying the crust thickness in the model over time (see figure S1, S4 in the supplementary material). Mantle plumes melt beneath the lithosphere and are then extracted into the oceanic crust, converting into basalt to form a thickened oceanic crust (Specific mechanism described in method section). In the model, oceanic plateaus (thickened crust) are formed directly above the spreading plume head. To describe plateau formation in our models, we amended the text in lines 213-215 and 251-252.*

5. Lines 20-22: "Our results highlight fast-spreading ridges exert strong plate dragging force, rather than suction on plume motion, which sheds new light on the natural observations of plume absence along the fast-spreading ridges, such as the East Pacific Rises." As I indicated above this is true only if plume radii are smaller than 250 km (based on Fig. 6). This conclusion implies that plumes in the Pacific are smaller than those in Atlantic. Are there any observations supporting this? I'm interested in a discussion about this issue in the paper.

Reply: *Thank you for your comments. We revised this sentence (lines 24-25). Observations show approximate plume buoyancy flux distributions in the Pacific and Atlantic (Figure 1 below). There is no obvious correlation between the distribution of buoyancy of mantle plumes in different oceans with their spreading rate. We conclude that plume size is not the deciding factor to explain the difference between the Pacific and Atlantic in terms of plume-ridge interaction mode. We add a figure and discussion about this in the discussion section (lines 470-481). Please also see our reply to* *comment 6.*

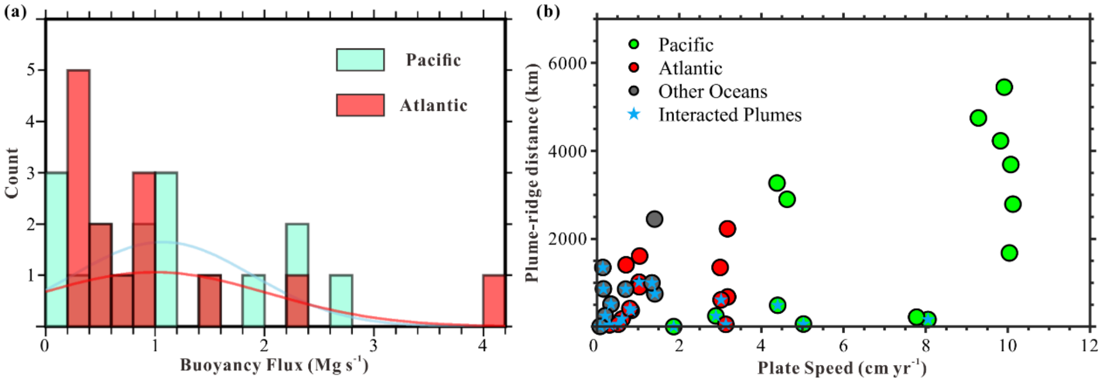

Figure 1. Buoyancy flux, plate speed and plume-ridge distance of mantle plumes in different oceans. (a) Histogram of plume buoyancy flux distributed in the Pacific and Atlantic. The guassian distribution curves are shown in light blue and red lines, respectively. (b) The plot of plate speed at each plume and their off-axis distance. Blue stars mark the plumes shown to be interacted with the nearby ridges.

6. Looking at distribution of plumes and their sizes in Fig.9, one cannot see any correlation between plume size with plate drag and ridge suction. Can authors comment on that? Besides, in conclusion it is written: "The plume size, that is, the plume buoyancy flux, may play a critical role in controlling the connection between the two units, compared with distance and spreading rate." Why does plume size play important role compared to the two other factors? This is not discussed in the main text

Reply: *Thanks for your comments. We suggest that all three factors play an important role (Fig. 8). The predicted effects of plate velocity and plume-ridge distance are fully consistent with observations (Fig. 11b). The effects of plume buoyancy flux are less obvious when compared to observations (Fig. 11a). For discussion, see lines 470-481. We also reconsidered the importance of different influence factors and rephrased the sentence in the conclusion.*

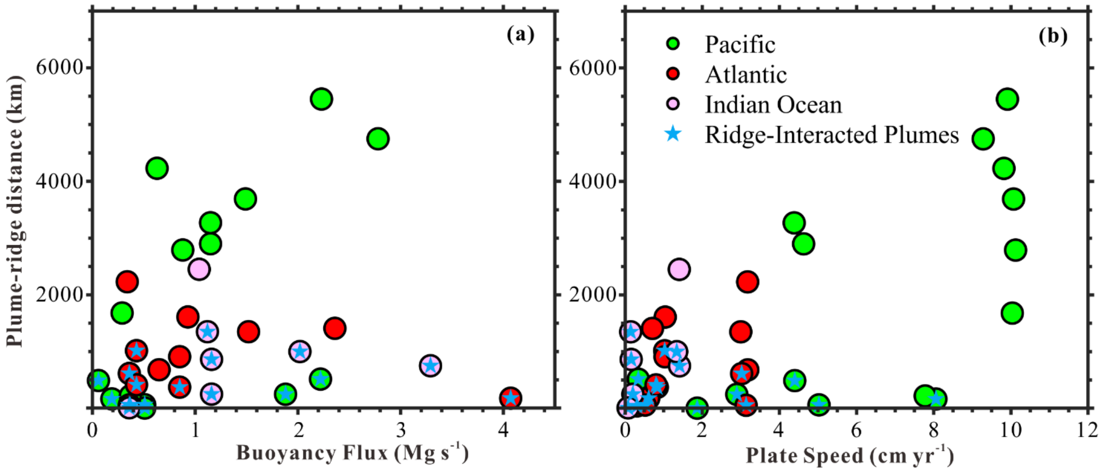

Figure 11. Buoyancy flux, plate speed and plume-ridge distance of mantle plumes in different oceans. Mantle plumes in the Pacific, Atlantic and Indian Ocean are shown in green, red and pink circles, respectively. Blue stars marked the ridge-interacted plumes according to Ito et al. (2003). **(a)** Plot of plume-ridge distance and plume buoyancy flux. Data are from Hoggard et al. (2020). **(b)** Plot of plume-ridge distance and plate speed at the location of plumes. Plume-ridge distance come from GPlates (Müller et al., 2016; Whittaker et al., 2015), and plate speed data come from Becker et al. (2015)

7. Line 413- 415: "Based on a series of numerical modeling as well as geological and geophysical observations, we predict that mantle plumes in the Pacific Ocean are more likely to be dragged away by the spreading ridge." The authors emphasize on the fast spreading rate of Pacific ocean as a main factor for dragging plumes away from the ridges. I think the plume-ridge distance may be a main factor in this case; most of plume tails (shown as blue dots in Fig. 9a) in the Pacific Ocean are located away from the ridges.

Reply: *We appreciate this comment and agree with the reviewer. We reorganized the manuscript discussion (mostly in the final paragraph of section 4) accordingly.*

8. Line 145: Temperature of 2513 K is very high for temperature at 660 km. Considering adiabatic temperature gradient of 0.5 K km-1 and temperature of 1573 K at the base of lithosphere, temperature at the bottom of model should be ~1873 K.

Reply: *The model sizes in our study are set as 6600(width) and 1200(depth). The temperature of 2513K refers to the temperature at the bottom of model, that is, the temperature at 1200 km. Based on adiabatic temperature gradient of* 0.5 K km^{-1}, *the temperature at 660 km in all models is set initial to* 1573 K + (660-120)km*0.5K/km *=1843K.*

9. Fig. 5: I expect the heat flux and melt are initially maximum in the area above the plume head. Then due to underplating of plume and its flowing towards ridge, the location of maximum heat flux and melt changes in time. That would be worth to show the evolution of heat flux and melt in time (similar to what is shown for surface topography in Fig. 3 and 4). For (5e-f): I suggest to show the results of plate drag model from ridge to some distances away from it, similar to what is shown in (c) and (d). I suspect that in plate drag due to imposing higher extension rate, the whole lithosphere is experiencing cracks and becomes extremely weak. Figure 5 shows the results in the early stage of deformation. Can authors provide a figure showing results at later stages?

Reply: *We appreciate this suggestion and add figures to present the temporal evolution of the extracted melt (see figure S1, S4 in the supplementary material). The maximum extracted melt above the plume head, shown as crust thickness in Figure S1, S4, change with flowing of the plume, which is consistent with the topography evolution in Figure 3a and Figure 4a. Besides, considering that the heat flux is not that important to our model result (suggested by another referee), we removed the description of heat flux evolution but reserve the change of extracted melt over time.*

Secondly, we also present the lithosphere stress in a wider perspective (Figures. S1, S4). We agree with you. The whole lithosphere is under extension sightly because of the imposed extension rate. The presence of buoyant mantle plume leads to the stress localization within the above lithosphere. The normal stress in blue represents *extension, while the compression region is plotted in red. Actually, since plastic deformation does not occur in a widespread manner in our models, it indeed does not seem to be appropriate to describe such localized stress areas as "tension cracks". Please also see reply to comment 2. As a result, we removed the description of "cracks" in the main text (lines 212-213 and lines 249-250).*

Other comments

Line 36: What is Amsterdam?

Reply: *The Amsterdam here means the Amsterdam-Saint Paul mantle plume. We have made revision in the main text (line 41).*

*L*ines 136-138: It is not clear what this sentence mean. Please modify this sentence.

Reply: *We rephrased this sentence (lines 170-172) as "To reproduce the oceanic lithosphere, we choose a typical layered model, where the crust is composed of a water level (2 km), a sediment layer (1.5 km), a basalt layer (7.5 km)."*

Lines 147-148: This is not consistence with cooling half space. The temperature of the oceanic lithosphere tends to change linearly with depth when lithosphere is very old (older than ~80 Myr).

Reply: *Thanks for your comments. Indeed, the initial temperature distribution of the oceanic plate is prescribed by the half-space cooling model and thermal equilibrium structure. The half-space cooling model is used to describe the oceanic plate younger than 50Myr, and the thermal equilibrium structure is used to describe older oceanic parts. The thickness of the half-space cooling part is defined by the thermal structure. We set up the models in this way because we consider that the theoretical half-space cooling model has a good match with some geophysical observations when the plate is young, but the fit becomes poor when the age is greater than 60/70 Ma (Turcotte and Schubert, 2014; Stein and Stein, 1994). Therefore, we set a half-space cooling model with a maximum age at 50Ma, and the thermal equilibrium thickness of the older* *lithosphere is constant (i.e., ~100 km; corresponding to a thermal age of 50 Ma). We have made further description of the model initial setup in the main text.*

Line 153: "An additional velocity is imposed on both sides of the ridge to represent the half spreading rate. "Are they internal boundaries? Please explain more about it here; where are they and until which depth they extend.

Reply: *Yes. The velocity boundaries are internal boundaries, which are imposed on 500 km from each side of boundaries in the lithosphere (i.e., from 20 km to 120 km in depth). We made revision in the main text (lines 97-98; lines 187-189).*

Line 181: "The mantle flow vertical velocity profiles" It is a bit confusing. The profiles shown in Fig. 3f are the horizontal component of mantle velocities along two vertical profiles. Please rephrase this part and also explain the depths which were selected for these profiles. Are they from the surface till ~250 km depth?

Reply: *Thanks for your suggestions. The profiles shown in Fig. 3f and Fig. 4f are the horizontal velocities for two vertical profiles (i.e., from 80 km in the lithosphere to 240 km in the asthenosphere). The profiles are located 100 km away from the plume stem. We rephrase these sentences.*

Lines 353-354: "However, without plume further supplies, the overpressure difference from the plume head to the spreading center decreases slowly with time (Fig. 355 8d)." What does it mean?

Reply: *Thanks for your comments. We reworded this paragraph in section 3.4.3 (lines 411-417)*

Lines 186-187: "The overriding plate moves slower than the ponding plume, and hence actually slows down the spreading plume branches." It is not clear what the message of this sentence is. According to model setup, since plume is located on the left side of MOR, the overriding plate motion speeds up the plume flow towards left (since the plume flow and plate motion have the same direction) and slows down the flow in the right plume branch.

Reply: *We disagree with the reviewer. The left plume branch spreads in the same direction but faster than the overriding plate. Hence it is slowed down by plate drag. The right branch is moving in the opposite direction, and hence is also slowed down by plate drag. We improved the description of these processes (lines 235-239).*

Lines 187-188: "Without suction effect from the spreading center, the left plume branch flows out much slower than the right branch." Similar to what I mentioned in my previous comment, I expect faster flow towards left.

Reply: *We improved the sentences (lines 235-239). Please also see our reply to the previous comment. We also refer the reviewer to our added discussion of driving forces for plume spreading in section 3.3.*

Fig. 7: It is a very complicated figure. What do the upper panels of Fig. 7a-c stand for? They show the results at different times. Are the results shown in the lower panels of Fig. 7a-c showing the results at similar times as those shown in the upper panels? The scale of Fig. 7a-c is very small and one can hardly distinguish all the curves shown in the Figure. Please make the figure bigger. I suggest to move the legend of Fig. 7a into the right side of Figure because Fig. 7b-c also shows the results of models with different plume-ridge distances. The colors of curves for different plume-ridge distances are very similar and hard to differentiate them from each other. I suggest to change the colors. What do "plume head stage- positive spreading out" and "plume tail stage- passive flow driven by plate" mean? How are buoyancy fluxes calculated?

Reply: *We appreciate your suggestions and replot the Figure 7 (Fig.6 in the revised manuscript).*

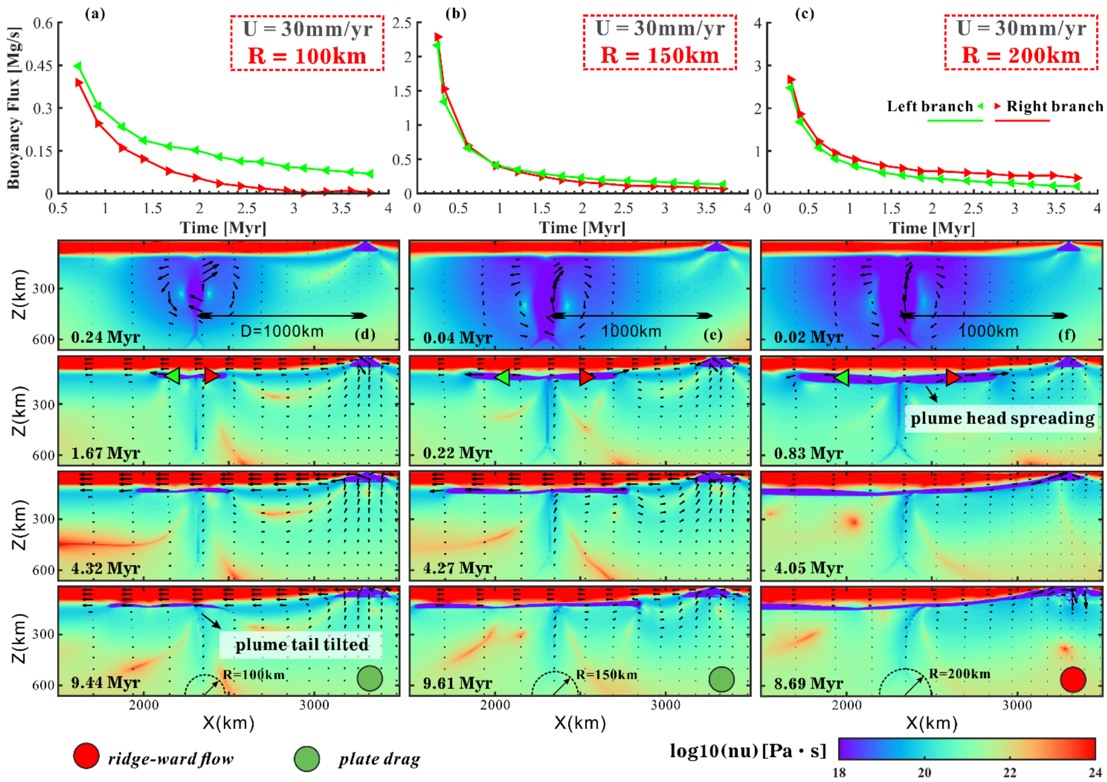

Figure 6 Models varying initial plume head radii (model M53, M58, and M63, Table S1 in supplementary material) shown by buoyancy flux and viscosity. **(a-c)** Buoyancy flux in spreading plume branches over time. Green and red triangles are markers used for buoyancy flux calculation. **(d-f)** Viscosity snapshots of models with different plume head radii. Models with green circle represent plate-drag flow and ridge-ward flow in red.

Line 330: How does Fig. 8a indicate that fast-spreading ridge promotes plume dragging. In this figure, from three models with fast spreading rates two are representing ridge suction mode.

Reply: *We appreciate this comment. We replot the Figure 8 (Fig.9 in the revised manuscript) to demonstrate the effects of spreading rates, choosing different typical models. In fast spreading ridge models, more plume material is dragged away, which make it difficult for plume to interact with the ridge. Therefore, we suggested that fast-*

spreading ridge promotes plume dragging.

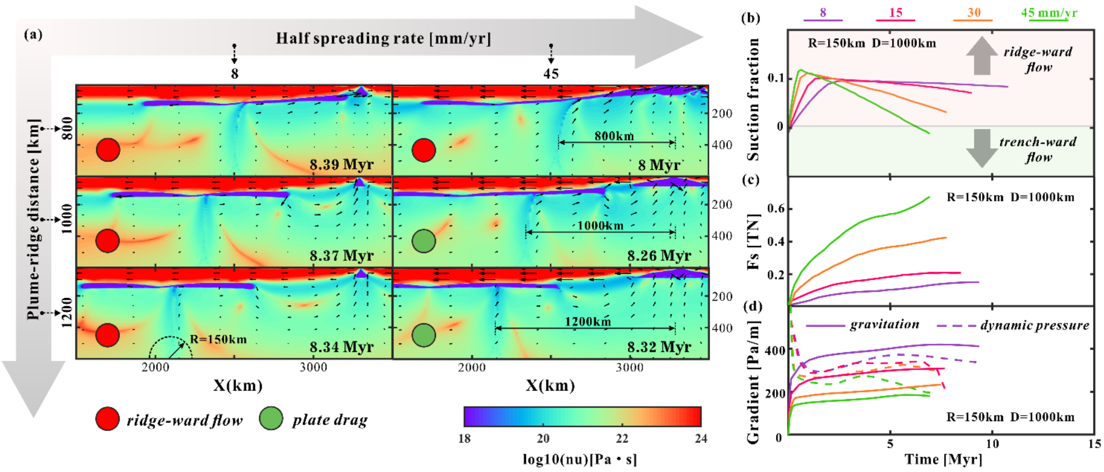

Figure 9. Model results influenced by different half spreading rates. **(a)** Effect of spreading rate on ridge-ward flow verse plate-drag flow. Viscosity snapshots are shown (model M7-M9, M82-M84, Table S1 in supplementary material). Fast-spreading ridge promotes plume material dragged. Models with green circle represent plate-drag flow and ridge-ward flow in red. **(b)** Dynamic evolutions of ridge-ward and plate-drag plume flow, revealed by defined ridge spreading fraction (eq.14). **(c)** Shear force (*Fs*) between moving plate and plume material under different spreading rates. **(d)** Pressure gradient between plume head and ridge center in different half spreading rate models. The solid and dash lines are the plume gravitation and dynamic pressure gradient, respectively.

Lines 340-347: It is not clear how shear force and pressure difference were calculated. Please re-write this part. Was the shear force calculated for the grids in the upper part of plume head or the whole plume head? The box of 50*50 km^2 in Fig. 8a is shown only for the plume head (and not for ridge center).

Reply: *Thanks for your advice. The shear force* F_s *is calculated by integrating the shear stress* σ_{xz} *of the uppermost plume head grids, not the whole plume head. Moreover, we replace the pressure difference with pressure gradient to clarify the* *mechanism of ridge suction. We calculate the plume gravitation and dynamic pressure gradient by tracing the plume markers which record the pressure, density, etc. The method how we compute the shear force and pressure gradient have revised and described in detail (lines 301-308; lines 401-407).*

Fig. 8: What is the distance of plume-ridge in models shown in Fig. 8b-d? What do the dashed color curves in Fig. 8b stand for? Please explain them in the caption. The scale of figures are small. What does “plume head spreading” in Figure 8d mean? What is the effect of plume size on shear force and overpressure difference?

Reply: *We appreciate your suggestions. We replotted the Figure 8(now Fig. 9 in the manuscript) and revised the caption. First, the “plume head spreading” in both figure and main text means the plume head spreads lateral under the plate. We explain it in the revised manuscript. Second, we added figures of shear force and pressure gradient of different plume size models in the supplementary material. The results shows that bigger size plumes are subjected to bigger shear force (Fig. S7). Meanwhile, bigger plume size means a more buoyant strong plume, which creates a larger pressure gradient between the plume and the ridge (Fig. S8). We have discussed this in the section 3.4.1.*

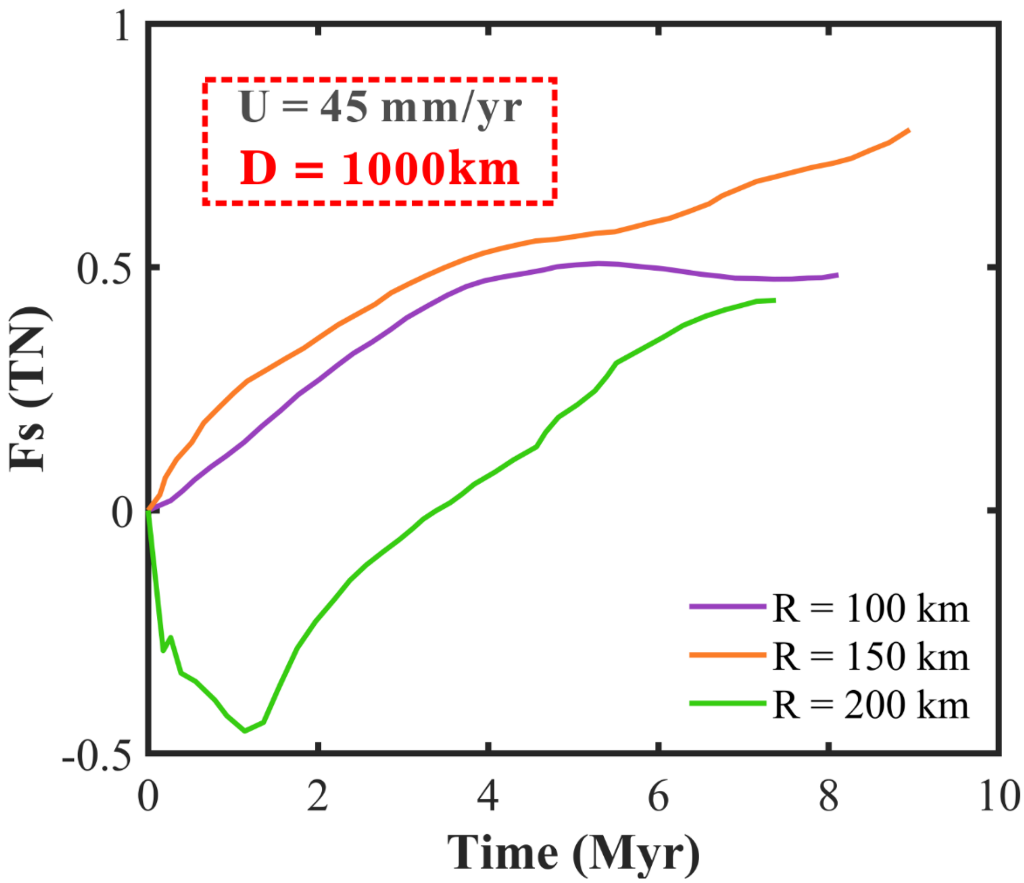

Figure S7. Shear force (*Fs*) between plate and plume in different plume head size models. The shear force imposed on the plume increases with plume size. The negative shear force indicates stronger friction imposed on the ridge-ward flowing (right) plume branch than that on the plate-drag flowing (left) plume branch.

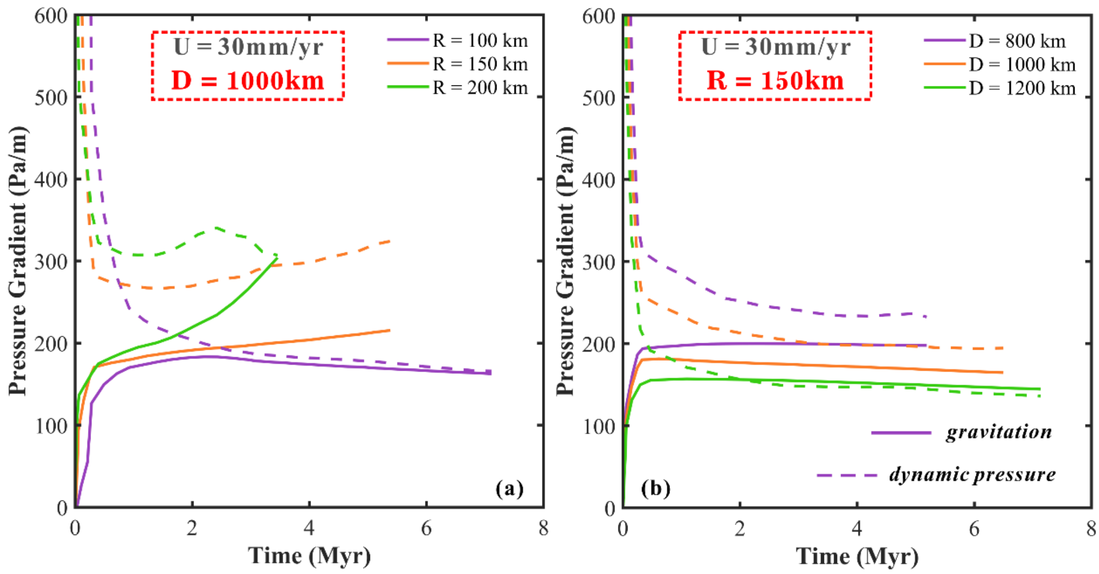

Figure S8. Pressure gradient between plume head and ridge center. The plume gravitation and dynamic pressure gradient of **(a)** different size plumes, **(b)** different plume-ridge distances are shown by solid and dash lines, respectively.

Lines 360-361: "while all models gradually switch from ridge suction in the plume head stage to dominant plate drag in the plume-tail stage" Is it valid for all models or only those representing plate drag regime?

Reply: *We appreciate this comment which triggers us a lot of thinking. The description may be incorrect here. Only those plate drag models shift from ridge suction in the plume head stage to dominant plate drag in the plume-tail stage. We removed this sentence to avoid semantic ambiguity.*

Figure 9: What do "MAR" and "EPR" stand for? Please explain them in the caption. How did the authors obtain the plume buoyancy flux (which indicate plume size) of hotspots shown in Fig. 9?

Reply: *The "MAR" and "EPR" indicate the Mid-Atlantic Ridge and the East Pacific Rise, respectively. The plume buoyancy flux data in Fig.9 come from Hoggard (2020) and are presented in different circle sizes. We modified the Figure 9 (Figure 10 in the revised manuscript) and revised its caption.*

Reply on RC2

1. Please revise the language using a professional editor or native speaker. There are many areas that may cause confusion as they are currently written.

Reply: *We appreciate your advice and have revised the manuscript carefully to improve the language, with the help of all the co-authors.*

2. Definitions: one of the major problems with this work at the moment is the lack of a definition of "ridge suction". Plate drag is reasonably well explained as the frictional force imposed upon the sub-lithospheric plume material, but ridge suction seems to be simply anything that causes plume material to travel toward the ridge. Currently, I have to infer this definition since no clear description is given and the quantitative assessment of ridge suction is a fractional number looking at the volume of plume material flowing toward and away from the ridge. If ridge suction is only assess in this way, then this is inconsistent with the literature and should be retermed as ridgeward flow or something similar. Better, I believe the authors need to reassess their model results with a more consistent definition of ridge suction that can be quantitatively assessed.

Reply:

Thanks for your suggestions. The concept of "ridge suction" refers to Niu (2004), who suggest that the spreading ridge sucks the material from depths due to the pressure gradient between the ridge center and deep hot material. The buoyant mantle plume from deep is overpressured, while the ridge center is in the state of underpressure. Therefore, the ridge suction we termed in the manuscript indicates the dynamic pressure gradient between the plume and ridge, which drives the plume material flowing to the ridge. In any cases, we reworded most of the occurrence of "ridge suction" in the manuscript, because it is only one of the driving mechanisms for ridge-ward plume flow.

On the other hand, the active gravitational spreading of plume also contributes to the ridge-ward plume flow. We discussed and compared the gravitational gradient and ridge-ward dynamic pressure gradient of these two mechanisms in the main text (see section 3.3)

3. In the abstract, the authors claim that plate drag has not been studied very much. However, I think this is perhaps an overly strong statement. There are several studies that incorporate the affect of plate motion (and the consequent drag) on plume spreading including Ribe et al. (1995); Ribe (1996); Ribe and Delattre (1998); Ito et al. (1997), Hall et al., 2003; 2004, etc. Each of these works (and others) incorporates the affects of plate drag on plume spreading in their calculations. I think the authors should be clear about what aspect of their work contributes something these other authors do not.

Reply: *We appreciate this comment. We revised and improved the description of the motivation for our study and its novelty in the abstract and introduction.*

4. Model Comments

4.1 In my opinion, for the scope of this work, the model used is overly complicated and in some respects inaccurate for a mid-ocean ridge setting. For example, the authors are examining the flow of mantle material beneath a lithosphere and have included a 1.5 km thick sediment layer across the model, but near ridge (especially fast ridges) there is little to no sediment. In fact, even along the slow spreading MAR, 1.5 km of sediment does not occur along the ridge axis and, indeed not for a reasonable distance away.

Reply: *We appreciate this comment. In our models, we imposed an average 1.5 km sediment layer vertically following the typical oceanic lithospheric structures. Horizontally, we use a uniform thickness lithosphere. We agree to the reviewer that the sediment near the ridge is negligible. Indeed, the processes of erosion/sedimentation are also considered in our simulations. We plot the thickness of sediment in the reference models (Figs. S1, S4). We are confident that the additional sediment layer thickness does not affect the conclusions of our study.*

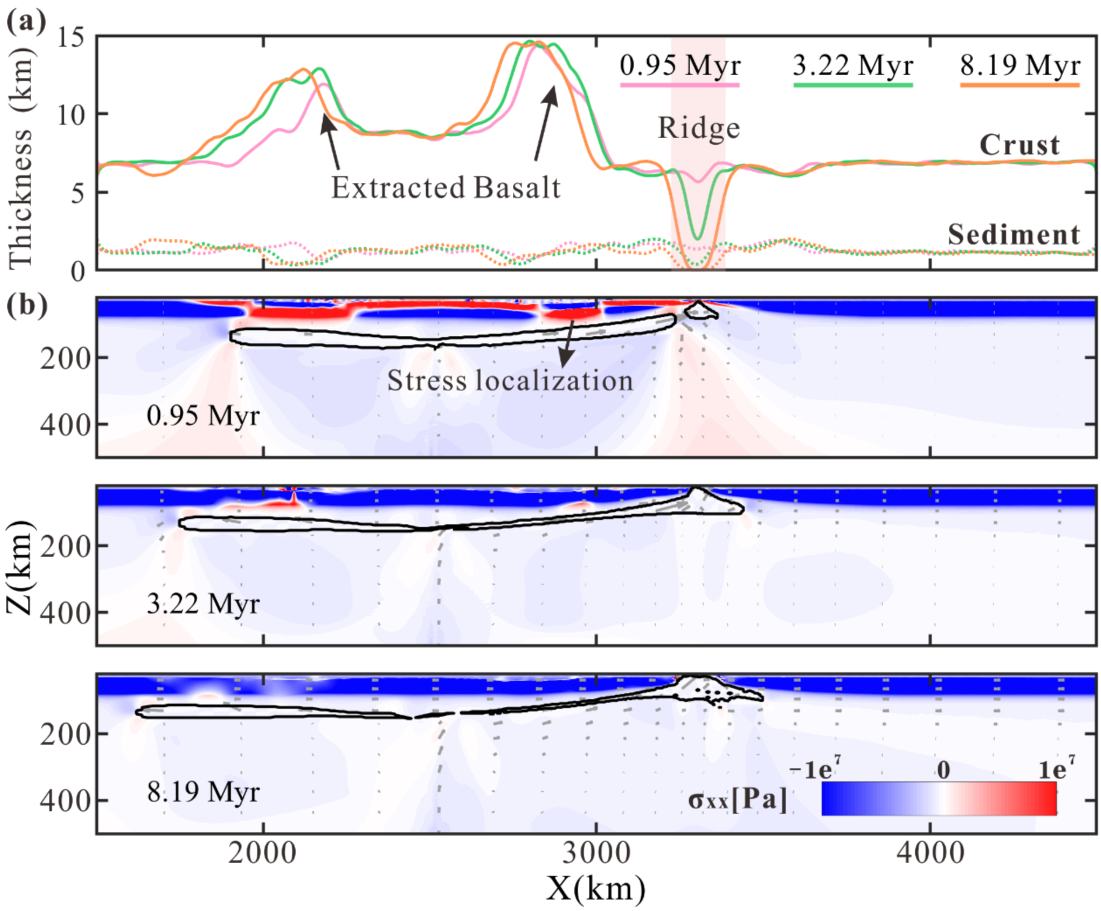

Figure S1. Reference model (M12, see Table S1, same as Fig. 3) evolution of ridge-ward plume flow shown by (a) crust and sediment thickness, (b) normal stress. The mantle plume weakens the overlying oceanic plate and changes the stress state of the overlying oceanic plate. Molten plume material beneath the lithosphere is extracted to the crust.

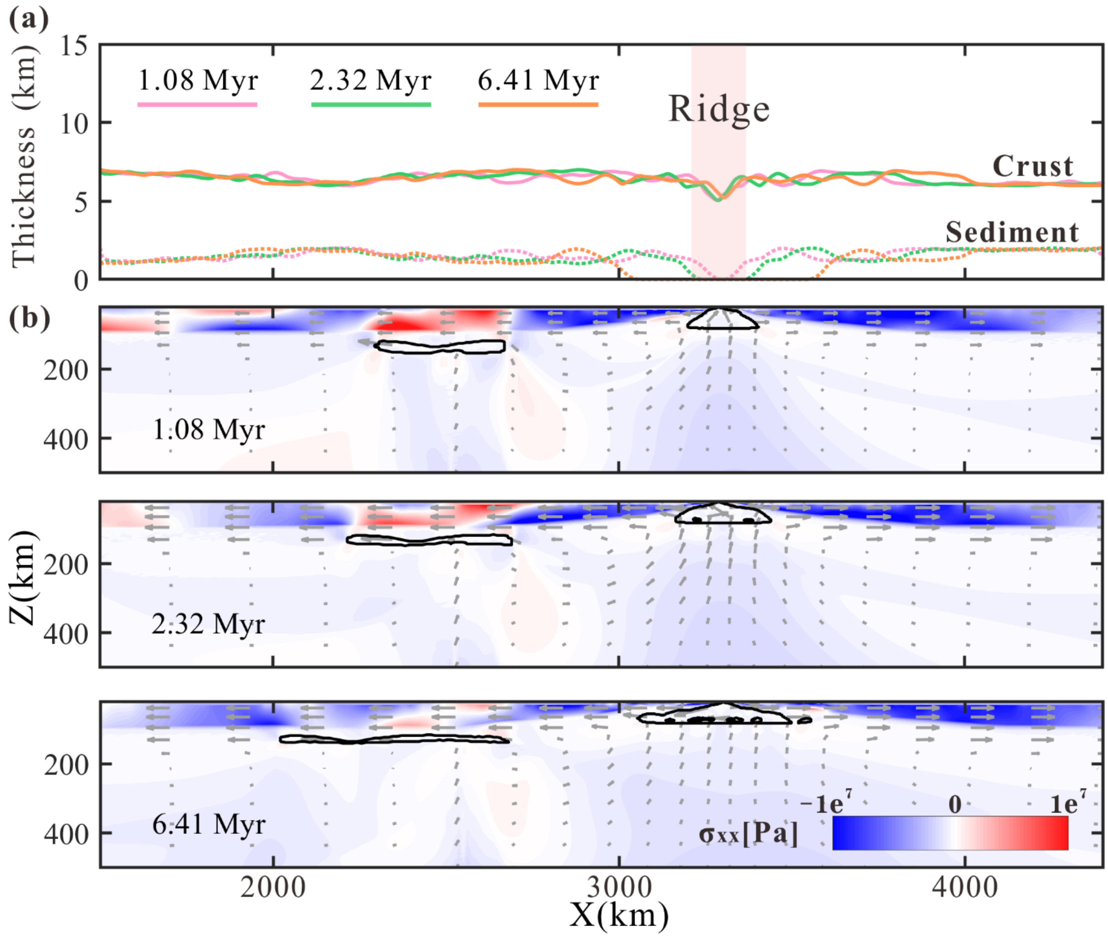

Figure S4. Reference model (M77, see Table S1, same as Fig. 4) evolution of trench-ward plume flow shown by (a) crust and sediment thickness, (b) normal stress. The mantle plume weakens the overlying oceanic plate and changes the stress state of the overlying oceanic plate. Molten plume material beneath the lithosphere is extracted to the crust.

4.2 Next, why is melting and heat flux useful for this study? Given the stated goal of the study to assess plume flow, I do not see (and it was not stated) why melting was useful or necessary. It is also unclear how the movement of melt throughout the system does or does not violate conservation of mass since you are working with an incompressible material and claiming to add material beneath the crust after removing it from another location. Please justify the use of melting and melt movement/accumulation. Also, clearly state any affect this melt has on your model (viscosity? temperature structure? density?, etc.)

Reply: *We appreciate your comments. In the model, melting of plume and asthenosphere are taken into account. The molten plume material is transformed to mafic magma and added to the crust, forming a thickened crust. We discussed the thickened crust between the plume and ridge to indicate the formation of oceanic plateau near the ridge during plume-ridge interaction. Besides, the melt extraction in the model is mass-conservative, as has been applied in previous modeling works* (Gerya et al., 2015; Gülcher et al., 2020). *We have modified the description of the melt extraction and its mechanism in the method section (lines 134-158). We also agree that the heat flux may not be that important for our study, so we remove the discussion of heat flux in the revised manuscript.*

4.3 Why do you need a plastic rheology? Given the scale of the problem you are working on, is the added focusing of the ridge axis to a smaller number of grid cells necessary? I don't believe that the current models can answer this question given the problem I mention next.

Reply: *Thanks for your comments. we employ viscoplastic rheology in the models. However, it is true that the stress and strain of ridge area usually are not strong enough to exceed the yielding criterion. The lithosphere does not exhibit plastic deformation in the model. Since plastic deformation does not occur in a widespread manner in our models, it indeed does not seem to be critical to be included in the model. On the other hand, including it for completeness (as we do here) does not change the main results of our study.*

Secondly, we added smaller grid cells (grid size decreases linearly from 20km at the edges to 2 km at the ridge axis). Indeed, we set denser grids in the middle model *domain in order to better simulate the interaction between plume and ridge. Of course, considering the scale of our research problem, the size of model grid does not have a great influence on the results. According to initial test cases, using smaller grids helps to make the computation more stable and does not fail to converge due to numerical perturbations.*

4.4. Another issue is the lack of adjustment of the lithosphere for plate spreading rate. In other words, the ridge and lithosphere in the "fast spreading, ridge drag dominated" cases do not appear to be in equilibrium before the plume is introduced. Looking that the compositional slices and temperature contours of the model in Figure 4, it appears that the sub-axial lithosphere is flattening out and a new, flatter lithosphere is forming without the initial half-space cooling structure (or with one that is in equilibrium with the faster spreading rate). This will alter the mantle flow field, the upslope topography of the ridge, and, potentially the location of the spreading ridge.

Reply: *We appreciate this comment and further checked our models. According to the model viscosity and temperature structure (Figs. S5, S6), we think that the flattening of the 1300 ℃ isotherm beneath the MOR is caused by the latent heat consumption during asthenospheric melting beneath the ridge. The effect of latent heating is not included in the initial temperature structure. Thus, while it is true that the 1300 ℃ isotherm flattens, the lithosphere does not flatten. It behaves as expected (see plots of the viscosity filed; Figs.S3, S6). Accordingly, we are confident that our predicted temperature and viscosity structure beneath the MOR is realistic.*

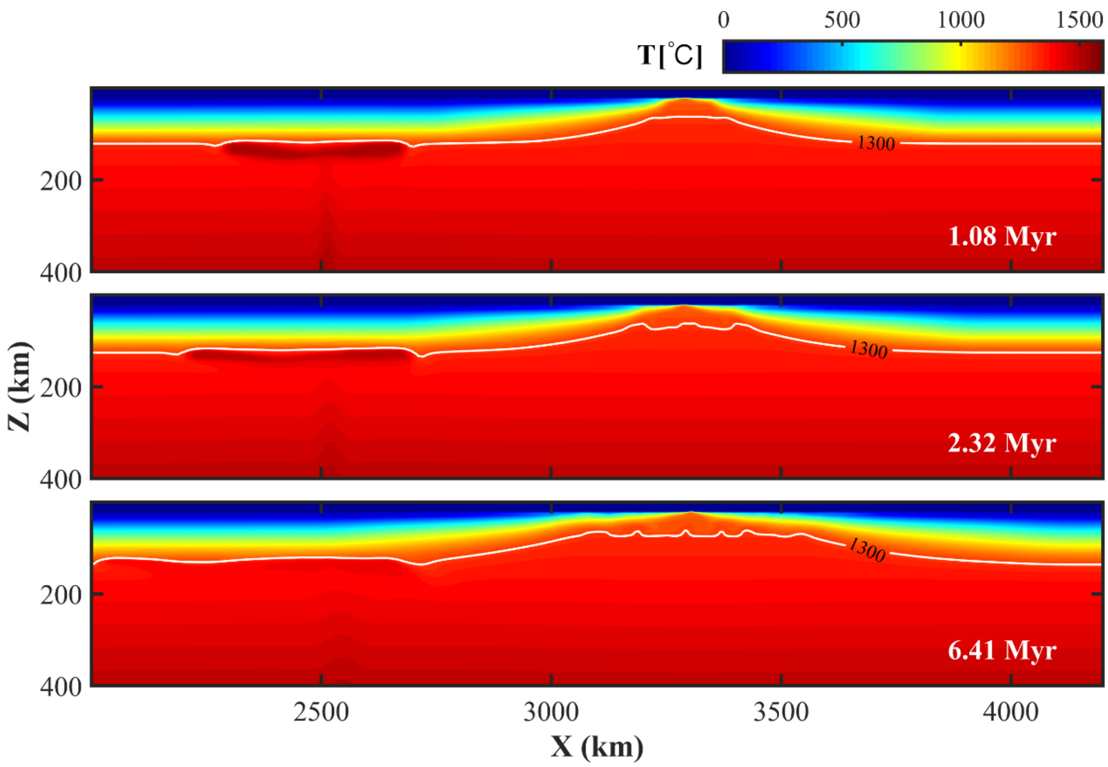

Figure S5. Temperature evolution of reference trench-ward plume flow model (M77, see Table S1, same as Fig. 4).

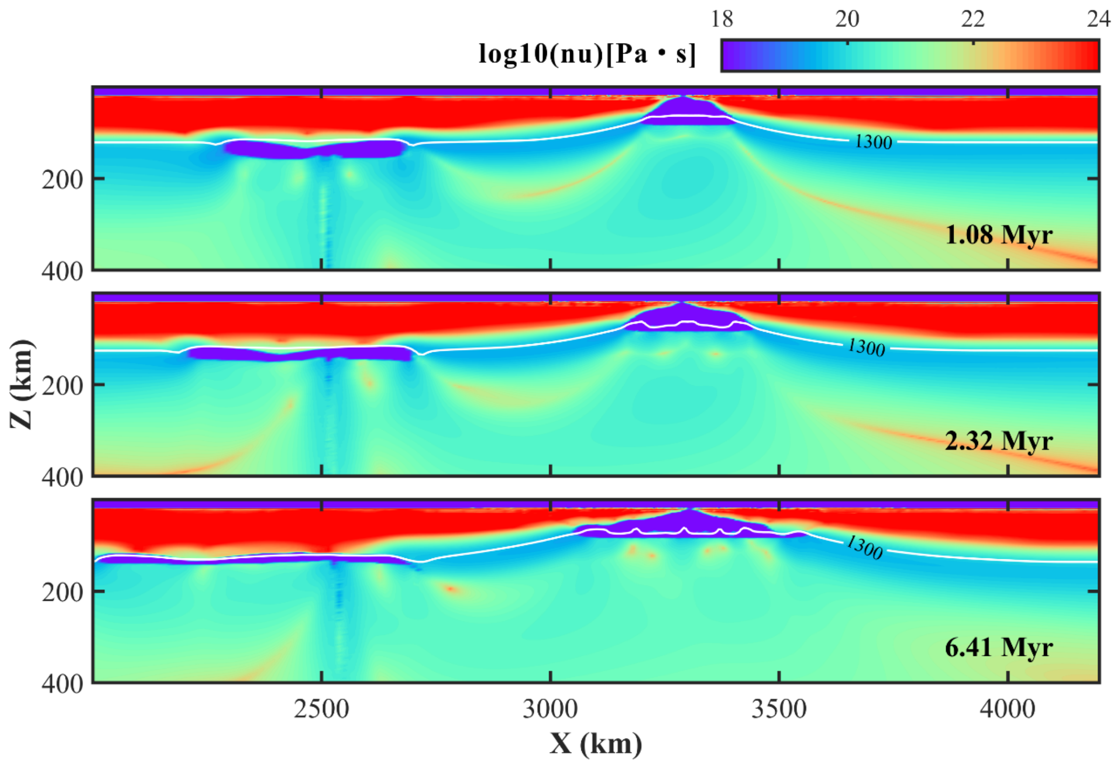

Figure S6. Viscosity evolution of reference trench-ward plume flow model (M77, see Table S1, same as Fig. 4).

4.5. Thermal structure

4.5.1 It is not clear to me how you arrive at a bottom boundary condition of 2513K when the base of the lithosphere has a Tmax = 1573. Since the base of the lithosphere is at ~100 km depth (Figure 1) and there is an imposed 0.5 K/km adiabatic temperature gradient, the max temperature at 660 km depth should be 1573 K + 560km*0.5K/km = 1853K. This is a big discrepancy that might imply a much hotter mantle than is realistic, which would likely have significant impacts on the results of this study.

Reply: *The model sizes in our study are set as 6600(width) and 1200(depth). The temperature of 2513K refers to the temperature at the bottom of model, that is, the temperature at 1200 km. Based on adiabatic temperature gradient of* 0.5 K km^{-1}, *the temperature at 660 km in all models is set initial to 1573 K + (660-120) km*0.5K/km =1843K.*

4.5.2 How is the plume tail maintained? This is not clear or perhaps I missed it

Reply: *Thanks for your comments. The mantle plume imposed in our model is activate by given excess temperature at the beginning. Indeed, this temperature anomaly is only an initial condition, but not a long-term boundary condition. We do not impose a hot patch at the base of model, which means the mantle plume can uprise or erupt only once and does not create a long plume tail self-consistently. Considering that hotspots erupt periodically on earth, at intervals of millions of years. Therefore, in our study, we simplified the formation of mantle plume only considered one pulse of plume uprising. Such plumes are also widely used in similar studies* (Baes et al., 2016; Gerya et al., 2015; Gülcher et al., 2020).

5. Results/Interpretation

5.1 The images in Figure 4 demonstrate a factor that may explain the affect of plume head size on ridgeward flow – the erosion of the lithosphere by the plume head. As pointed out by Kincaid et al., 1995 in their laboratory experiments, the formation of lithospheric levees can act to block plume flow. This appears to be happening here. Small plume heads eat into the lithosphere a bit, effectively create ridges (or levees) that are the same thickness as the plume material and halt its motion. Then, as the plate moves the plume material has no choice but to flow with the plate. In contrast, large plume heads push the lithosphere out of the way all the way to the ridge. Despite the significant ridgeward flow, I would argue that this has nothing to do with ridge suction, but the lithosphere rheology and plume buoyancy forces.

Reply*: We appreciate this comment and give the following reply. As mentioned in our reply to comment 2, we suggest that the dynamic pressure gradient from the high-pressure plume to the low-pressure ridge actually contributes to the ridge-ward plume flow. The pressure gradient from the plume head to spreading center is notable (Fig. 5). Such pressure gradient varies with plume buoyancy force and different off-axis distances. An increase in the plume head size, indeed, enhances the dynamic pressure and promotes the plume flowing ridge-ward (Fig.S8a).*

Primarily, the plume flows to the ridge owing to the pressure gradient at first, which decrease gradually when plume get closer to the ridge. Then, the gravitational spreading of plume starts to drive ridge-ward flow. As the reviewer points out, the gravitational gradient is related to topography at the base of the lithosphere. See section 3.3 for a detailed discussion.

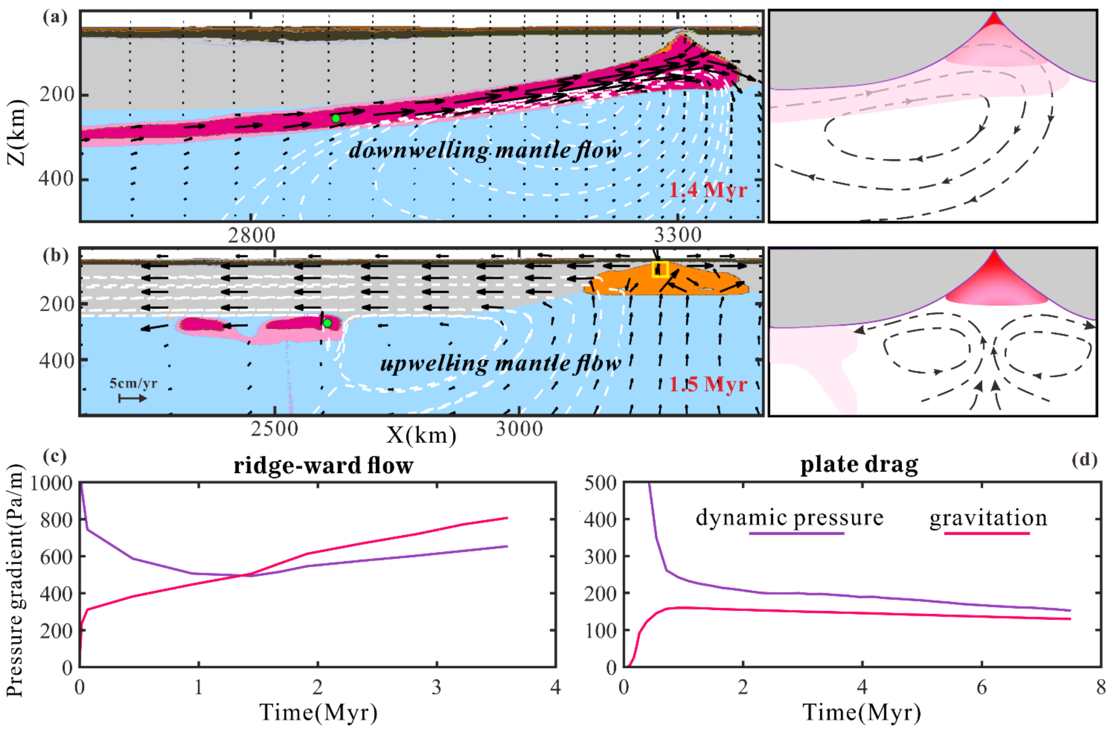

Figure 5. Comparsion between models with ridge-ward vs. plate-drag plume flow. **(a)** Ridge-ward flow with downwelling beneath the MOR (results from case M12 as in Figure 3). White dashed lines are streamlines; black arrows visualize the flow field. Schematic of flow in the sub-panel on the right-hand side. **(b)** Plate-drag flow with upwelling mantle corner flow beneath the MOR (results from case M77 as in Figure 4). **(c)** The dynamic pressure and gravitational gradient of plume marker (i.e. green circle in (a)) over time. The yellow box in (b) marks the location for the computation of average dynamic pressure at the ridge, needed for the calculation of the dynamic pressure gradient (see text). **(d)** The dynamic pressure and gravitational gradient of plume marker (i.e. green circle in (b)) over time.

5.2. The claim of "tension cracks" seems to be based on the stresses in the model. These stresses reach maximums of + or – 3x10^-7 Pa (Figure 5), much too small to actually fracture of rock - especially near the surface, which typically has yield strengths many orders of magnitude larger. Is this a typo? If this should be + or － 3x10^7 Pa (i.e., 30 MPa) that seems very large and so I am left to question how tension cracks are justified here. However, I would note that I don't think these are essential for the results of this paper and fall into the over complication of the model for the state purpose of the modeling.

***Reply**: Thanks for your comments. Yes, the magnitude of normal stress here is a typo. The correct magnitude of these stresses should be ±3×10^7 Pa in the models. Actually, the "tension cracks" used in the main text may be inaccurate. There are no brittle fractures in the lithosphere and normal faults near the surface in the models. While there are horizontal extensional stresses, there is in fact no yielding. As a result, we removed the description of the "cracks" in the main text.*

5.3. The role of ridge suction vs plate drag. I think the authors have glossed over some of the factors likely to contribute to the plume flow including the slope of the lithosphere and its role in guiding plume material up the slope, the buoyancy flux of the plume stem since this is not described by the plume radius definition here (which seems to describe the size of the plume head).

***Reply:** We appreciate this comment which triggers us a lot of thinking. We think the effect of the lithospheric slope is relative to the half spreading rate. The slope of the lithosphere varies with the spreading rate of the mid-ocean ridge. The base of the lithosphere would be flatter at the fast-spreading ridge. Consequently, a slow spreading ridge imposes smaller shear force on plume head and forms a steeper lithosphere base which benefits to the ridge-ward plume flow. Besides, we now evaluate the plume gravitational gradient, considering the local lithospheric slope, and ridge-ward dynamic pressure gradient (Fig. 5). For a detailed analysis, see section 3.3.*

5.4. Much of the interpretation of these results hinge around spreading of the plume head, not the plume after it has established itself beneath the lithosphere. Many plume have been active for 10 Myr or more and the plume head will have greatly diminished or completely spread away by that time. Yet, these plume tails can still interact with ridges since ridges migrate and often approach plumes. How does the long term interaction look - after the plume head has disappeared?

Reply: *According to our results, the plume tail (which follows the plume head at model times 1~8 Myr) bends either toward or bend away from the ridge in ridge-ward flow and plate-drag flow dominated models, respectively. Taking the long-term evolution of the model, the degree of plume tail tilts towards or away from ridge increases with time, and ends up maintaining a relative steady state. Please also see our previous reply to 4.5.2.*

5.5. Related to 4., I don't think the authors should be claiming to assess plume radius, as this commonly is used to refer to the radius of the plume stem. Instead, I think the manuscript would be much clearer if the authors would state that they were varying the plume head radius.

Reply: *We appreciate this comment and made revisions. In this study, we varied the plume buoyancy flux by changing its radius. The plume radius in our model only refers to the initial size of mantle plume. The width of plume stem decreases dramatically when the plume rising to the plate and plume head spreading out, which is different from the columnar plume stem radius detected by the geophysical tomography. Thus, as you suggested, we use the "initial plume head radius" to replace "plume radius" in the main text.*

Other minor specific comments:

•Line 47-48 – I'm not clear as to what this statement has to do with the EPR sucking in plumes so that they do not appear near the ridge.

Reply: *Thanks for your comments. We removed this sentence and reword this paragraph in introduction section.*

•Line 52 – the use of the work "push" is inappropriate here and should be changed to “drag” or similar

Reply: *Thanks for your advice. We replace the “push” with “convey” in the text (line 54-55).*

•Line 58 - The authors should reconsider how they phrase things – for example, “slow spreading rate, short distance (small plume-ridge distances??), and large plume radii promote ridge suction,…” is an inaccurate statement – really, I think what the authors are trying to say is that these factors favor plumes being pulled toward ridges by ridge suction

Reply: *Thanks for your advice. We rephrased all these inaccurate phrases in the main text.*

•Line 59 – maybe try a more careful wording – it is the fast plate motions associated with fast-spreading ridges that exert strong drag forces on plumes

Reply: *Thanks for your advice. We revised this paragraph to clarify our goal of this study in the main text.*

Figure 2 – this does not look like a half-space cooling model. Is this a plate cooling model or some modified half-space model? The half-space cooling model does not flatten like this.

Reply: *We appreciate this comment and made revisions in figure 2 and the main text. Indeed, the initial temperature distribution of the oceanic plate is prescribed by the half-space cooling model and thermal equilibrium structure. The half-space cooling model is used to describe the oceanic plate younger than 50Myr, and the thermal equilibrium structure is used to describe older oceanic parts. We set up the models in this way because we consider that the theoretical half-space cooling model has a good match with some geophysical observations when the plate is young, but the fit becomes poor when the age is greater than 60/70 Ma (Turcotte and Schubert, 2014; Stein and Stein, 1994). Therefore, we set a half-space cooling model with a maximum age at 50Ma, and the thermal equilibrium thickness of the older lithosphere is constant (i.e., ~100* *km; corresponding to a thermal age of 50 Ma).*

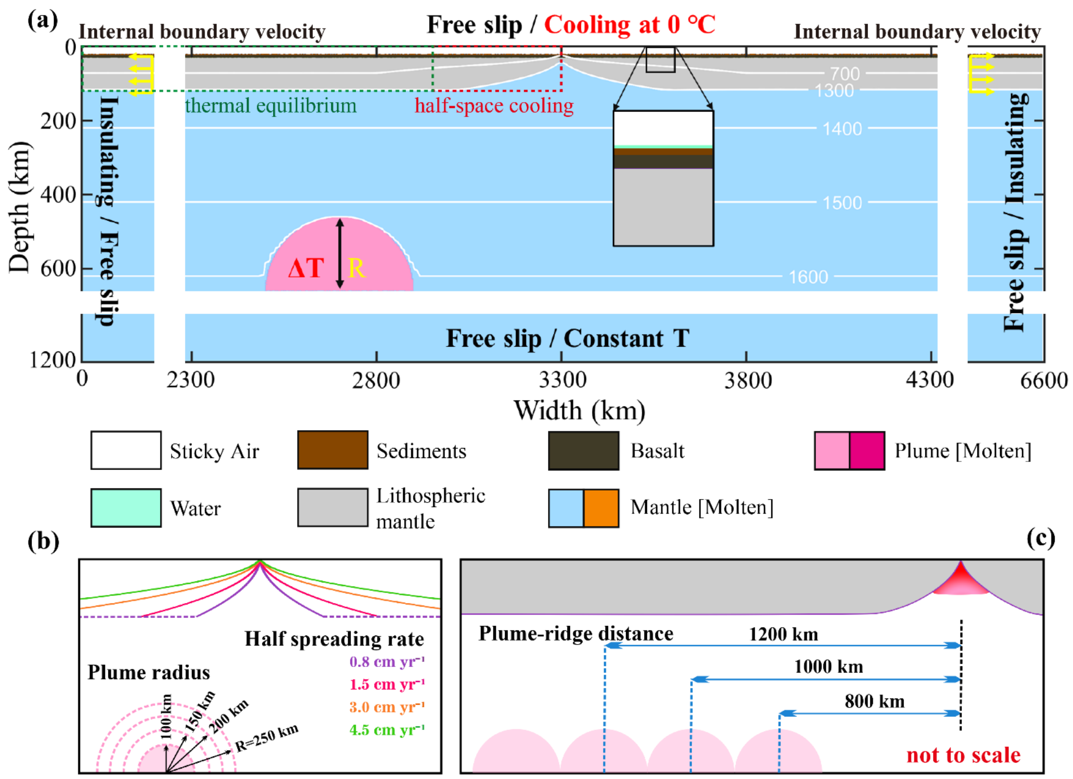

Figure 2. Model setup. **(a)** Initial composition and boundary conditions. The oceanic plate consists of half-space cooling part and the thermal equilibrium part. A 50-Myrs-old mid-ocean ridge sets in the middle of the model based on half-space cooling temperature structure. A thermal and chemical anormal mantle plume locates at 660 km. Different colors indicate the initial rock types and corresponding newly formed molten rock types. Yellow arrows are the half-spreading rates imposed internal in the lithosphere (i.e., from 20 km to 120 km in depth) to simulate ridge spreading. **(b)** Initial tested ridge and plume configurations. (**c)** Initial tested plume-ridge distances.