We would like to thank the reviewers for their constructive comments that helped to improve the manuscript. We have revised the manuscript accordingly and will be happy to provide a new manuscript. Please find below the reviewer's comments in black and a point-by-point response in blue.

The authors have significantly improved the manuscript, addressing my concerns and those of the other reviewer. They have broadened the scope by adding an analysis of the NEGIS outlet region and by testing a basal power law and various numerical differences, while removing the advection analysis that seemed too limited. They have also clarified some technical and conceptual details and modified the conclusions appropriately.

I have a couple of substantive comments, followed by some suggested minor corrections. Here and below, page and line references are to the version with tracked changes.

The first comment is related to the new Section 5.4. The vertical emergence velocities are large, ~50 m/a. This number is described (l. 390) as "not very extreme", but it exceeds observed height changes by a factor of ~25. What explains these large values? Is there a mismatch between the computed velocities and the observationally derived basal topography? If so, then the emergence velocity is basically a measure of the mismatch. With either stress balance, the model would presumably relax in a few years to a smooth surface slope with a modified velocity field and much smaller emergent velocities.
The large values could be partly explained by the high resolution (up to 100m) employed in this study. At the region where the high emergence velocities develop, the surface slopes (sqrt(nx^2+ny^2)) are exceeding values of 0.1. Assuming a surface velocity of about 1000m/a, the emergence velocity is about 100m/a. As also mentioned in the text (line 390-392), the high values could also be computed (v_x*nx+v_y*ny) when using the observationally derived products, i.e. the GIMP height elevations and the MEASURE velocities.
Certainly, the emergence velocity will reduce when running a transient relaxation run as the setting will reach a steady-state. However, we do not think that the high values are critical, as we are interested in the differences between FS and BP-like.
We added here: *"A transient relaxation run would certainly smooth the surface slopes. However, the high values are not critical as we are interested in the differences between FS and BP-like."*
The details of the relaxation would differ between FS and BP, but I'm not clear on why these details are important for ice sheet models. In particular, I'm not sure how we know that the relaxed surface would be lower for BP than for FS (l. 396).
The surface for BP is expected to be lower as the emergence velocity is lower (see red patches in Figs. 10b and d). This very pronounced in the grounding zone of 79NG. Since the emergence velocity of BP is lower than in FS, it would predict a lower surface than FS within the first time step (assuming SMB=0 or the same SMB in both models). This finding is consistent with the higher discharge in BP and a grounding line position that is located further upstream compared to FS. We have rewritten the sentence in Line 396: *"The emergence velocity differences between FS and BP-like reveals a higher emergence velocity in FS, particularly at the grounding line at 79NG. According to Eq. 14, BP-like would therefore compute a surface elevation that is lower than the FS surface within the first time step."*
Without a better explanation of why the emergent velocity matters, I suggest removing this section.

We think the evaluation of the emergence velocity is an important diagnostic in order to estimate how the model differences could evolve over time. Additionally, the second reviewer recommended this analysis to overcome the lack of time dependence in our study.

Second, the authors now claim (p. 26, l. 469) that based on the new analysis of NEGIS outline regions, "ice flow treatment with FS is essential". This is a strong claim, suggesting that models with BP (or L1L2 or SSA) solvers should switch to FS, regardless of the increased complexity and computational cost. Indeed, the new analysis shows that FS–BP differences are large in the 79N grounding zone. But modelers could compensate for these differences (e.g., by tuning basal drag coefficients), and the differences might matter more or less depending on the science application. I think it would be enough here (as in the Abstract, where the wording is more measured) to say that the BP– FS differences are large, and we should take these differences into account when evaluating results from models with BP or simpler schemes.
We agree with the reviewer that this is a strong claim and needs further research to prove (e.g., transient *simulations). Therefore, we have rewritten the sentence following the reviewer's suggestion: "However, once the outlet regions of NEGIS are included (i.e., 79NG and ZI), the analysis shows that differences between FS and BP-like are large in the 79N grounding zone."*

I'd make a similar comment about the last sentence of the Conclusions (p. 31, l. 610): "a correct representation of the ice dynamics in critical areas is required." Of course, one wants an accurate representation in critical areas, but no model is perfectly correct. It might be more constructive to say that there are regions where FS and BP differ significantly, and where results from simpler models should therefore be viewed with caution.
Done. We followed the reviewer's suggestion hand have rewritten the sentence to: *"Our diagnostic simulations and previous studies indicate that FS and BP differ significantly in regions with a grounding zone and results from simpler models should therefore be viewed with caution."*

Minor corrections:

• P. 1, l. 14: "unveil" -> "show". In general, don't use "unveil" as a synonym for "show".
Done
• P. 3, l. 77: Here and below, I suggest "regions" (a geographic term) instead of "subsets" (a mathematical term).
Done.
• P. 4, l. 100: Delete "is defined"
Done
• P. 4, l. 112: "For floating ice"
Done
• P. 4, l. 113: Delete "used"
Done
• P. 5, l. 127: Delete "up-to-date"
Done
• P. 5, l. 136: Delete "the order"
Done.
• P. 6, l. 164: "circumvented"

Done.

• P. 7, l. 178: Update the last-access date if possible

Done.

• P. 7, l. 191: Add comma after “matrix”. This is one of several places where adding a comma after an initial clause or prepositional phrase would help the reader.

Done.

• P. 7, l. 199: “Dependent” -> “Depending”, comma after “discretization, “run” -> “runs

Done.

• P. 7, l. 200: Not clear what is meant by “The limiting case MUMPS and ASM”

We have rewritten this sentence to: *“For the largest DOF model that could be solved with the direct solver on our cluster system, we also performed a simulation with the iterative linear solver.”*

• P. 7, l. 206: Use quotes for case names consistently (either with or without)

Done. We added quotes to *Ice flow over a bumpy bed*.

• P. 8, l. 211: Add comma after “margins”

Done

• P. 8, l. 218: “using” -> “use”

Done.

• P. 8, l. 234: “upstream of”, “downstream of”.

Done.

• P. 9, l. 242: I’m not clear on what’s meant by “A relaxation run based on e.g., the BP-like scheme and successive simulation runs with BP-like and FS”

We replaced “successive” with “subsequent”.

• P. 10, Fig. 2 caption: The two domains (red and cyan) partly overlap. Here or in the main text, could you explain why this is the case?

There is no specific reason why both regions partly overlap. Both regions were designed to capture slow and fast flowing regions of the NEGIS.

• P. 11, l. 277: “used contained” is awkward. Maybe “span a typical range of resolutions…”

Done

• P. 11, l. 284: Delete “very”

Done.

• P. 11, l. 288: “to” -> “too”

Done.

• P. 12, Table 1: Has formatting issues in this version, but looks OK in the other version

Ok, we will take care of this in the revised version.

• P. 15, Fig. 4: The figure show v_FS – v_BP. Thus where BP is faster, Fig. 4 shows negative values, plotted in blue. This is a bit counterintuitive. If FS is “truth”, then it’s more natural, at least for me, to show the BP error with respect to truth, i.e. red where BP is too fast. If it’s not hard to change the figures, I suggest flipping the sign.

We agree with the reviewer that another choice on how to calculate the differences would be more intuitive (i.e., BP-FS -> BP faster than FS -> positive sign -> ‘red color’). However, to our knowledge there is no rule on how to calculate differences of data (i.e., BP) to ‘truth’/observation (i.e., FS). The calculation of the differences is clearly mentioned in the text and figure captions so there should be no confusion. However, generally we would follow this suggestion but it requires a lot of figure changes (Figs. 4, 5, 7, and 8) and corresponding text changes (flipping sign of mentioned values). So, we do not follow this suggestion.

• Sections 5.2 and 5.3: The new analyses in these sections definitely strengthen the paper.

Thanks. The paper benefits from your suggestions from of the first review.

• P. 16, l. 351: Delete “Equivalent” and “exemplary for”. The word “exemplary” is misused a couple of times; could also delete in the Fig. 6 caption.

Done.

• P. 17, l. 355: “expressed” -> “pronounced”

Done.

• P. 17, l. 359: “stronger” -> “larger”

Done.

• P. 17, l. 367: “Most remarkably” implies that this is a surprising result, but I think it’s to be expected. Maybe “notably” instead.

Done.

• P. 18, Fig. 7: Same comment as for Fig. 4. It would be more intuitive to use red for regions where BP is too fast.

see above

• P. 18, l. 374: The wording here makes it unclear what the numbers (14.9 and 14.2) refer to. Are these the fluxes for 79NG and ZI, respectively, when simulated by FS? Similarly in the Fig. 9 caption. Clearer wording would be “On average, the total simulated flux [for FS?] is 14.7 Gt a^-1 for 79NG and 14.2 Gt a^-1 for ZI.”

You are right the wording is unclear. Yes, it is the average over all settings (resolutions, discretizations, BC implementations, friction exponents) when simulated with FS.

• P. 20, Fig. 9: Similar comment to Figs. 4 and 7. Since BP overestimates fluxes compared to the more correct FS, it would be natural to show the relative Q difference as positive.

see above

• P. 22, l. 425: “stronger pronounced” -> “more pronounced”

Done

• P. 22, l. 434: Add comma after “ratio”

Done

• P. 25, l. 458: Delete “with each other”

Done

• P. 26, l. 464: Delete the second “much”

Done.

• P. 29, l. 565: “less performant” -> “more expensive”

Done.

• P. 30, l. 592: Not sure what “offset” means here.

We have rewritten the sentence to: “In contrast, ten times stiffer ($E=0.1$) ice leads to enhanced ice flow velocities in BP-like compared to FS.”.

• P. 30, l. 595: See the comment above on “essential”

Done. See answer above.