Response to Reviewer RC1 – Prof. Ed Bueler

Egusphere-2024-1052 Article

Author Responses in Red

I thank Prof. Bueler for a detailed and helpful review!

Summary: This paper rewrites the standard glaciological (Glen law) Stokes model in a form which resembles a shallow approximation, the Blatter-Pattyn (BP) model. This expresses the saddle-point structure of the Stokes problem in a form close to the unconstrained-optimization form of the BP model. The stability and finite element (FE) analysis of the new form is addressed, and new mixed FE pairs for vertically-extruded meshes are propsed. Small-scale experiments are presented, and then prospective applications at larger scale are discussed. The resulting essentially-theoretical paper is both frustrating and promising. The manuscript's current form is notably inefficient, with 1500 lines of text. The presentation is likely to be hard to read for those who have not already done battle with BP equations and related technical matters. Despite doing numerical experiments, the author provides no open-source code basis for further development by readers, a clear demerit in 2024. The manuscript avoids the function-space understanding of the Stokes and BP problems---this is the viewpoint from which these problems are known to be well-posed and by which they are solved by mainstream finite element libraries---but then it labors to build a fragmented substitute for this viewpoint. Despite these flaws, the paper illuminates important matters. It shows how the (transformed) Stokes equations are close to an "extended Blatter-Pattyn" (EBP) form, and thereby how the solvability conditions of the Stokes model work in practice over vertically-extruded meshes. The EBP model has similar numerical and stability issues as the Stokes problem, which is actually clarifying because the numerical and FE character of the standard BP and Stokes models otherwise appear very different. The inf-sup stability of the mixed Stokes problem is recognized here, when the mesh is extruded and when one simultaneously wants the EBP model to be solvable on the same mesh, as the requirement of unique solvability of the continuity (incompressibility) equation for the vertical velocity from the horizontal velocity. A necessary condition for this to work is that the number of vertical velocity and pressure unknowns must be exactly the same, or rather that a particular matrix in the blockwise form of the discrete equations must be invertible.

Recommendation: A manuscript which made the same points in half the length, and which provided open source code in a widely-used language, facilitating further development, would be an excellent paper. Of course it is not realistic to expect re-coding at that level. However, significant revisions should be attempted. A much-shortened abstract is offered below, along with several other suggestions for trimming.

An effort has been made to tighten and shorten the manuscript while preserving the content. The line count has been reduced to 1340 while preserving most of the content. Unfortunately, it is not possible to provide open source code in a widely used language because of the piecemeal way that the work was carried out using the Mathematica program, as pointed out in the paper.

Specific Comments on Manuscript

lines 9-35: This long abstract could be halved without losing meaning, by removing the sales pitches and by other simple edits. However, changes are also needed to clearly identify the models (systems) under consideration. The following is a guess/suggestion for an abstract which meets these objectives. It has 191 words vs 371 in the original: """We introduce a novel transformation of the Stokes equations into a form that resembles the shallow Blatter-Pattyn (BP) equations. The two forms only differ by a few additional terms, and the variational formulations differ only by a single term in each horizontal direction, but the BP form also lacks the vertical velocity in the second invariant of the strain rate tensor. The transformed Stokes model has the same type of gravity forcing as the BP model, determined by the ice surface slope. An apparently intermediate "extended Blatter-Pattyn" (EBP) form is identified, which is actually the same as the standard BP model although it retains a pressure variable. The role played by the vertical velocity in the transformed Stokes and EBP forms, reflected in the block-wise structure of their discrete equations, motivates the construction of new finite element velocity/pressure pairs for vertically-extruded meshes. With these new pairs, examples of which are demonstrated in 2D and 3D, the discrete continuity equation can be uniquely and stably inverted for the vertical velocity. We describe how to incorporate the new forms into codes that adaptively switch between Stokes and BP models, where the latter would lose accuracy."""

I have rewritten the abstract using many of these suggestions. Thank you.

line 41: "full" is unnecessary.
Removed

line 52-72: The style of glaciology, used at unnecessary length in these lines, says some models are shallow and some are higher order. It is more accurate to say all are shallow, and to not claim some are "higher-order" because the order depends on which scaling argument is use.
I have used the term “shallow” only as part of the accepted names of some simple approximations. The term “higher-order” is commonly applied to the Blatter-Pattyn and other more accurate approximations.

line 99: "THE LOWER BOUNDARY OF an ice sheet ...". (A 3D ice sheet can't be divided the way the text says.) (1)
Don’t quite understand what the problem is. This is an idealized situation of course. I will be glad to make whatever change is required.

lines 103-105: This "vertical line of sight" phrase appears here and later. Surely one can just say: "We assume the glacier's geometry is described by an upper surface function z_s(x,y) and a lower surface function z_b(x,y)."
This was intended to mean that there should not be various indentations so that various multiple upper and lower surfaces would exist along a vertical line. Although unlikely, these could be handled but would complicate things considerably. I have changed this to say that there should be just one upper and one lower surface.

lines 105-106: There is nothing about the rest of the paper, in my reading, that excludes the techniques being used for floating ice. (Put f_i=0 in equation (11)?) It is true that there must be sufficient drag--see the inequality in Schoof (2006)--*somewhere at the base* so that the velocity field is unique, but the techniques apply across grounding lines.
I have modified the sentence to say that ice shelves can be handled.

lines 112--126 Briefer notation is surely possible.
I have simplified by removing superscripts on unit normal vectors. Not sure what else can be done.

line 149: "positive-definite" --> "nonnegative"
Changed to “a positive quantity”

line 178-180: Whether or not the surface kinematical equations can be added "easily", the way this is said here is silly. The whole paper assumes fixed ice geometry.
Yes, fixed geometry is assumed. What this says is that flux inflows or outflows are allowed through a fixed geometry (which may be a crude representation of melting or refreezing at the bed).

lines 192-195: I don't know what this means. "There are some stress boundary conditions and it is easier for the author to think about them in the variational formulation."? No need for this?
This means that evaluating derivatives at boundaries is less accurate or more complicated because one-sided formulas have to be used due to the absence of information from across the boundary. I have changed the wording to make this clearer.

lines 197-200: No need for this.
I think this needs to be pointed out because most people use the weak formulation method and may not be familiar with the variational method.

lines 204-209: Is this option ever used later in the paper? (Line 233 suggests not.) If not, it can be removed and replaced with a simple declaration that the boundary conditions can be weakly imposed if desired.
I have indeed used it but most computations were done using direct substitution, as stated on Line 233. Of course, there is no difference in the results. However, it is a useful option and some people may prefer it. There are some consequences when Lagrange multipliers are used. For example, the “solvability condition” must be modified (see Line 626 in the originally submitted paper). For this reason, I prefer to leave this section as is.

lines 238-252: This is a valuable observation, namely form (17) which shows ~P solves a trivialized problem. If this observation is original, then great. Otherwise cite it more clearly; did it appear in DPL 2010? (The nearby citations to DPL do not refer to this main idea as far as I can tell.)

It is original but only insofar as it refers to the transformed pressure ($\tilde{P}$ not P) in the Blatter-Pattyn approximation. It does not appear in DPL, 2010, since the new transformation was not invented yet.

Figure 2: This basic point is greatly appreciated: The deviation from hydrostatic is relatively small. However, in this and almost all figures, the fonts are too small! (Also these figures are bad on a monochrome printer, but I suppose that train has left ...)
Changing all figures would be difficult. Should be OK for young eyes ....

line 282: I don't think (22) is actually used *here*.
Yes, it is used in the strain rate tensor (6) and in the second invariant (7). See (26) and (28).

around line 282: Warn the reader that "dummy variables" ("flag variables"?) are about to be used. As the text is written, they are finally explained on the next page.
Done

lines 286 and onward: I find "modified" really unpleasant here. For (25) the tensor ~\tau_ij is actually modified; it is not equal to the original. But in (26) the tensor is merely rewritten; neither "modified" nor the tilde have the same meaning as they do in the equation above. Similarly (27) and (28) are not "modified" but merely rewritten, as far as I can tell. I therefore would not say "modified" or add a tilde; just write out the new form. Equality means equality.
I must disagree here. Equations (26), (27), (28) are indeed modified because $\partial w/\partial z$ is replaced by $-\left(\partial u/\partial x+\partial v/\partial y\right)$ according to (22). They may have the same numerical value at convergence but they are discretized differently, so they are “modified”. It is also important to distinguish quantities in the transformed Stokes equations from the standard or traditional Stokes to avoid confusion.

line 325: "implies the use of" --> "uses"
Done

lines 327-336: This is a rambling paragraph that can be shortened to something like "As noted earlier we require the upper and lower surfaces of the glacier to be functions of the

horizontal coordinates x,y. That is, as expected in glacier modeling, overhangs are not permitted."
Thank you, this is better. Text has been changed.

line 344-348: Repetitive. Say *once* (earlier, presumably) that one could impose boundary conditions weakly, and that you won't do that.
Shortened, but did mention can use Lagrange multipliers, if desired.

line 360: Help the reader by referencing/comparing (23).
I have referenced (23) and (25) following (37).

lines 361 and 404: Separate these into 2 displays. (Or better, just be more efficient. Use vector notation?)
I have done it this way in an effort to be more compact (long paper!) I think it's quite clear that I have combined equations and boundary condition. Vector notation would not be good because the rest of the paper uses Cartesian tensors.

lines 437-439: This use of the continuity equation is completely mainstream in glaciology. It applies in all shallow theories including BP. (And the current manuscript illuminates it!) Please say this some other way.
This has been reworded.

lines 459-460: Again, deriving FE discretizations from variational principles is the normal way to do business. Why "except"?
My understanding is that the normal way to do FE business is by means of the weak formulation.

ine 475: There is no reason to use capital "U" here, and it is a source of confusion because capital U is used shortly in subscripts with a different meaning.
I have changed U to V.

line 495: "u, w, AND M_{UP}, M_{WP}"
Section 4.3: This section needs editing most. The main point of the entire paper is made in subsection 4.3.3, I believe. Roughly-speaking the main point is that, for the

transformed Stokes or EBP equations, the block M_{WP} must be invertible, thus square, when an extruded mesh with z-aligned cells is used. This point is buried after laborious and repetitive text. The main point of the paper *does* require a block-wise presentation of the Newton step equations, so the text will necessarily be somewhat technical, but it doesn't have to bury the main idea. There would seem to be no reason not to start a section with (47) and (48); the notation here is obvious. In any case, this reader had to get 600 lines into the document before getting to the key lines (roughly starting at line 596), and only then have an "oh ... that is what he is trying to say ..." moment.

lines 596-600: The main point of the paper, right? Which this reader appreciates! The blockwise form of the EBP model is therefore the central object of the paper, and could be put much earlier and more prominently.

Section 4.3 has been completely rewritten. I believe it may now address these comments.

lines 616-618: I would not permit my undergrad linear algebra students to say what is said here. The necessary condition is that *M_{WP} must be non-singular*, from which it *follows logically* that it must be square. The text literally says that non-singularity is "in addition" to squareness, thereby asserting that square matrices are invertible! (Line 1521 is worse.) Equation (56) could instead say "M_{WP} is non-singular"; one is allowed to put text in displayed LaTeX equations.

I have been careless here. In Section 4.3.2 it now says: "matrix M_{WP}^T must be invertible and so it must be square and full rank. Since in general M_{WP}^T is an $n_p \times n_w$ matrix, for solvability this requires that $n_p = n_w$".

Section 5: I think the paper would be improved by removing this section. I understand that the transformed Stokes model is the same as the Stokes model, and the EBP model is the same as the BP model. So recapitulating the ISMIP-HOM purpose, which is (I suppose) to examine how close BP results are to Stokes results, should not come out any differently here, and thus it is not worth doing. Of course it is true that different numerical approaches generate different results in detail. But what exactly should the reader know about this numerical comparison? Can this be summarized in a sentence or two?

I have shortened this section considerably, keeping the figures and only a minimum amount of text to describe them.

lines 778-780: For efficiency I assume that BP is first used everywhere, then some criteria is applied, and then Stokes is used where the criteria applies. But do you want to demonstrate that the Stokes calculation everywhere gives the nearly same criteria-satisfying region?
I think this is done visually in Fig. 8. It is quite obvious that the Adaptive (AH) and Stokes (TS) calculations are quite close while the Blatter-Pattyn calculation is not very accurate in the details up through the column in the vicinity of the obstacle.

line 785: Is the "counterintuitive" aspect of this explained by noting that the effective viscosity is often actually largest in the top of the ice column, which implies the greatest longitudinal and bridging stress transmission up there? I often find that visualizing the effective viscosity, in these shear-thinning flows, illuminates where stresses de-localize the problem.
It is counterintuitive because I would have expected the Stokes calculation to be needed just in the vicinity of the obstacle and not far away at the top of the domain. Your explanation is probably correct but it would need a more detailed analysis to verify than is justified in this paper.

line 811-813: It is not the personal computer etc. which stops an analysis of the cost savings, but rather the lack of a performance model for the solver. This could be added, but it requires a bit of thinking.
Yes, but a more realistic calculation on representative computer hardware would be able to provide believable information on cost savings.

Subsection 6.2 and Section 7: This seems like tedious overkill. If a reader gets the main points of the paper then they can probably imagine lagging the Newton iteration and/or dual grids and/or higher order. In any case, another 300 lines are burned before the summary. If these are important enough then they could be a separate paper? Otherwise most readers won't have the endurance; really I don't.
In introducing the new transformation I stated that I wanted to bring out two of its applications (although there may be more): Adaptive switching and improved approximations that are more accurate than BP. I think both are equally important. Breaking it up into two papers is possible but it would lose some continuity. Honestly, I would not have the stamina to do that. Readers can always skip over parts that don't interest them.

Section 8 (Summary): Too long.
Substantially shortened.

Appendix A-C: On and on.
Appendix D: The manipulations shown in (79) and (80) are again very close to the main novel point of the paper. I see no reason why they can't be written into a new and prominent form which makes subsubsection 4.3.3 into the central material.
line 1521: Again, please don't say that all square matrices are invertible. (Literally the text says "the solvability condition [n_u=n_p] implies the invertibility of M_{WP}". Just no.)
Appendices A and D eliminated. Material from Appendix D shortened and transferred to subsection 4.3.2. Sloppiness re matrix invertibility has been corrected.

Response to Reviewer RC2 – Prof. Christian Schoof

Egusphere-2024-1052 Article

Author Responses in Blue

I thank Prof. Schoof for a detailed and insightful review!

Overview

The paper builds a unified variational formulation for Stokes flow and the somewhat misnamed Blatter-Pattyn model for ice flow. To get the trivia out of the way, I will call this the Herterich-Blatter-Pattyn model from this point onwards. History may be written by the winners, and Herterich clearly hasn't been one of them, *but* the first instance of the Herterich-Blatter-Pattyn model being formulated was (to my knowledge) the following paper:

- Herterich, K. 1987. On the flow within the transition zone between ice sheet and ice shelf, *in* Dynamics of the West Antarctic Ice Sheet. Proceedings of a workshop held in Utrecht, May 6–8, 1985, pp.!185–202. D. Reidel, Dordrecht,

This predates the commonly used Blatter (1995) and Pattyn (2003) references by eight and sixteen years, respectively. If we're going to fall into the trap of naming things after people, we probably owe it to ourselves to do that accurately. Not that I've been able to convince anyone of this point so far.

I was not familiar with this reference (I doubt many people are) so I looked into this model to compare it to the BP model. <u>It turns out that the Herterich model is NOT equivalent to the Blatter-Pattyn model. Instead, it is the same as the BP+ approximation that's described in §6.2.1</u>. I've attached a derivation of this result at the end after these responses, in an Appendix, and I've also added a short discussion in §6.2.1 regarding this. Needless to say, I will not be changing the name of the Blatter-Pattyn model. Thank you for pointing out this reference.

Back to the present. It took me a bit of effort in stripping away detail and side notes to understand that the primary advance of the paper is a numerical formulation in which one can dynamically switch between the simpler Herterich-Blatter-Pattyn and the more complete Stokes flow model. The latter applies to flows with arbitrary aspect ratios and abrupt changes in boundary conditions, while the former requires a "shallow" flow.

At face value, such a "switchable" model makes a lot of sense, since ice sheet flow is shallow in most places, but often contains boundary layers (at ice divides, ice stream margins, grounding lines etc) where the Stokes equations must be solved. The present paper uses the variational structure of both models to create a unified formulation, by re-writing the Lagrangian for the Stokes equations so that it takes the form of the Herterich-Blatter-Pattyn Lagrangian with a few extra terms retained. The unified formulation then consists of introducing a flag that activates or deactivates these Stokes correction terms. This is an intriguing idea and deserves to be published. I am not entirely convinced that

You are right, this is an important advance introduced in the paper. However, I also want to point out the additional innovations, such as the new approximations that improve on Blatter-Pattyn and the introduction and use of grids, such as P1-E0, that permit the solution of the continuity equation for $w(u,v)$. To this effect, I have rewritten Section 8 (I hope you don't mind that I used some of your words to do this).

This is an intriguing idea and deserves to be published. I am not entirely convinced that *The Cryopshere* is the right vehicle as there are a number of technical issues that deserve more thorough scrutiny; *Geoscientific Model Development* would have seemed a more appropriate journal in the EGU stable to use, but really I would have been inclined to go with something like *J. Comp. Phys.*.

I did consider Geosci. Model Dev. and J. Comp. Phys. Unfortunately, I was not able to comply with GMD's code availability requirement. JCP did not seem appropriate because the paper is too specialized to ice sheet modeling.

As far as I can tell, the modified Stokes Lagrangian does not alter the saddle point strucutre of the Stokes flow problem (in the sense that the solution maximizes the modified Stokes functional with respect to $\tilde{p}$ and minimizes with respect to $\boldsymbol{u}$). Still, the modified Stokes Lagrangian is a fairly ugly object that obscures the natural symmetries of the original Stokes flow Lagrangian and introduces a much larger null space to the elliptic operator. For the usual Stokes flow Lagrangian, that null space consists of rigid body rotations modulo any such motions that are precluded by the boundary conditions (by way of restrictions on the space of admissible functions). In the modified Stokes Lagrangian of the present paper, the vertical velocity w can be changed by adding an arbitrary function of the vertical coordinate z only while leaving the elliptic part of the Lagrangian uchanged; such functions of z are then penalized through the incompressibility constraint.

That ugliness appears to be an unavoidable part of a unified formulation. It does however mean that you would probably not choose to use the modified Lagrangian for the analysis of general Stokes flow problems, except to unify the Stokes and Herterich-Blatter-Pattyn problem computationally. The paper can probably streamlined by focusing on that aspect at the exclusion of some of the more peripheral commentary, and the title of the paper could probably be made more informative.

You are quite right. I do not advocate the transformed Stokes formulation for the solution of general Stokes problems. As presented in the paper I view it primarily as a means to (a) "unify the Stokes and Blatter-Pattyn problem computationally", §6.1, and (b) as a means of developing new approximations that improve on BP, §6.2. On the other hand, the transformed Stokes model can perform better computationally than the standard Stokes model. Section 5 compares some computational results for the standard and transformed Stokes formulation. Fig. 3 shows that the transformed model is more accurate than the standard model for Test B (no-slip) calculations at all resolutions. However, this is probably a fortuitous result because the two cases in the Test D* (frictional sliding) calculations show similar accuracy.

In terms of pursuing that unification of Stokes and Herterich-Blatter-Pattyn models, there are two significant issues that I can see:

1. While the variational formulation of the Stokes flow problem *is* a saddle point problem, the same is not true of the Herterich-Blatter-Pattyn problem. The latter naturally wants to be solved as an unconstrained minimization problem, with the incompressibilty condition solved a posteriori for the vertical velocity component w. Enforcing incompressibility through a Lagrange multiplier as part of the variational formulation leads to problems that occupy most of the technical material in the paper: unlike the modified Stokes Lagrangian, the elliptic part of the Herterich-Blatter-Pattyn Lagrangian does not contain w at all, which must instead be solved for through the (hyperbolic) incompressibility condition alone — for which the standard function spaces used in Stokes flow solvers are unsuitable. This is perhaps unsurprising, as the standard function spaces used in finite elements are problematic for hyperbolic equations, and discontinuous basis functions (as used in discontinuous Galerkin methods) might be preferable for w in the Herterich-Blatter-Pattyn problem; they might at least alleviate issues such as $\nabla \cdot \boldsymbol{u} = 0$ being an overdetermined problem if we assume that $\boldsymbol{u}$ is represented by P1 basis functions, which work well for the force balance part of the Herterich-Blatter-Pattyn problem.[1]

   The main challenge identified in the paper (which motivates the choice of those P1–E0 soaces, and the others discussed in sections 6,7 and appendix C) is to make sure that the choice of basis functions for the "Blatter-Pattyn system" (that is, the Blatter-Pattyn equations and the incompressibility condition) can also be used for the modified Stokes system. Unfortunately, I think that there are some issues with how this has been addressed in the paper, and I suspect that there is a misunderstanding of what the inf-sup ("LBB") condition really does. More on this below.

[1]By contrast, the "P1-E0" spaces advocated here go in a different direction: w is still represented by piecewise linear, and therefore continuous, basis functions, but the incompressibility constraint is weakened through the choice of a "coarser" basis function for the Lagrange multiplier that enforces incompressibility, averaging over two adjacent elements.

I don't have a quarrel with what is said here so I will postpone a response to where the inf-sup condition is discussed later.

2. More practically, I am not convinced that the method developed here will be adopted widely. Consider this: unlike its depth-integrated variants developed by Richard Hindmarsh, the Herterich-Blatter-Pattyn requires a three-dimensional domain to be resolved. This leads to a large number of computational degrees of freedom. In fact, the only advantage relative to a Stokes flow model is that there are fewer variables to be solved for in *the same* computational domain, as you can solve for (u, v) using the elliptic solver, and the compute w and p *a posteriori* if required. As far as I can tell, the Seroussi et al (2012, cited in the manuscript) tiling method makes use of that. The method proposed here, of solving for (u, v, w, p) using the same basis functions for both models without explicitly using the simpler structure of the Herterich-Blatter-Pattyn model, seems to get rid of that last advantage. Which begs the question, why bpther, given that the Stokes model is preferable in terms of the physics it represents. Unfortunately, the last paragraph of the conclusion puts paid to any hope that the paper might address whether the new approach is actually going to be computationally competitive.

I think this comment refers to the use of the transformed model in a "unified formulation". There is no question that the BP model is much cheaper, although less accurate, than a full-Stokes model. This is the reason why BP rather than Stokes is used in several production code packages (e.g., ISSM, MALI, CISM). Depth-integrated models are still cheaper but even less accurate. The approximations presented in §6.2 are at the high end of this scale; they are more accurate and more expensive than BP but still cheaper than Stokes. There is a tradeoff between accuracy and computational cost so a choice has to be made depending on the application.

I would not say that the standard Stokes formulation is preferable in terms of the physics it represents. The two formulations represent exactly the same physics and provide the same solutions to a given problem. They are, after all, mathematically related by a transformation. They differ, if at all, by their numerical behavior.

The adaptive switching example of §6.1 makes use of the "Blatter-Pattyn system", i.e., solving for (u,v,w,p), for simplicity and just to illustrate the method, even though this is more expensive than it needs to be. In a practical application it should be possible to switch to using the Blatter-Pattyn model instead, i.e., solving for (u,v), and coupling to the Stokes model using $p=0$ and $w=w(u,v)$ at the interface, even though this would involve more complicated programming. I have added some sentences to this effect in the paper.

To follow up on the first point above, I am concerned that appendix D is misleading. Apologies if this gets a little long below; I am not as much of an expert at this as I'd like to be, so it took me a bit more explanation. My understanding of the inf-sup or Ladyzhenskaya-Babuška-Brezzi condition for Stokes flow problems is the following. Take the Hessian matrix in (55) of the present paper

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & \mathbf{0} \end{pmatrix} \tag{1}$$

and we have to solve $\mathbf{Au} + \mathbf{Bp} = \mathbf{f}$, $\mathbf{B}^{\mathrm{T}}\mathbf{u} = \mathbf{0}$ where $\mathbf{f}$ is the relevant residual of the Stokes equations. $\mathbf{A}$ is positive semi-definite, which makes the solution of this problem (if it exists) equivalent to a saddle point: to find

–

–

–

Long discussion of inf-sup condition

–

–

–

I hope I have been clear to say that I may be wrong. If I am, the text of the paper should be more explicit about the technical issues that I have raised; it's currently fairly vague in spelling out what the "LBB" (inf-sup) condition actually is, what it does, and whether the basis functions described in sections 6–7 and appendix C satisfy the inf-sup condition. (Also, if you're going to introduce a concept like that, please spell out what the acronym "LBB" stands for before using it, and maybe cite the original places where it comes from.)

[3]since I would otherwise have $B(\boldsymbol{v}_h, q_h) = 0$ for all $\boldsymbol{v}_h$ and some q_h

[4]There is another issue in play here: the inf-sup condition provides sufficient conditions for a convergence, but not necessary ones. You might also decide that you don't care about convergence of the discretized pressure solution $\tilde{p}_h$, but only of $\boldsymbol{u}_h$, which may conceivably relax the conditions you need to impose. I am not in a position to comment on that, however — you need a real expert reviewer.

I have eliminated Appendix D and rewritten Section 4 to bypass the difficult issue of the inf-sup condition. The inf-sup condition is relevant only so as to point out that the standard and transformed Stokes models are subject to it and that one must use one of the many inf-sup-stable elements available in the literature in the discretization.

I hope I have made it clear that the inf-sup condition does not apply to problems using elements satisfying the "solvability condition" because they are no longer constrained problems since incompressibility is built-in when using $w(u,v)$.

My other main points would be the following (partially elaborated under "specific points")

1. The notation in the paper is fairly idiosyncratic, which made it more time-consuming for me to follow various pieces. There is a mix of standard subscript notation, a very much nonstandard "round bracket around subscript means projection onto the horizontal plane" variant to subscript notation, and the use of explicit component

notation as in equations (7), (28) and (35) to write scalar quantities that should really be denoted by contractions over subscripts. The paper makes enough assumption about the reader's level of mathematics that I would recommend streamlining this as I don't imagine much of an audience who will be helped rather than hindered by the nonstandard notation. In particular, I would discourage the round bracket notation in favour of a more standard projection operator: if P is the projection onto the horizontal plane, $P(v_1, v_2, v_3) = (v_1, v_2, 0)$, then you can simply replace a_i by $P_i(\boldsymbol{a})$ and retain summation over indices from 1 to 3. I say this even though the author has used it previously: the use of nonstandard notation it may make the paper harder to decipher for numerical analysts, who should really be encouraged to delve deeper into the theory relevant to the paper.
The use of superscripts $^{(s)}$, $^{(b_1)}$ and $^{(b_2)}$ on surface normals is also redundant, as you specify the parts of the domain boundary to which the stated boundary conditions apply. Keep this as simple as possible, because it it certainly isn't simple.

The second invariant was written in subscript notation, $\dot{\varepsilon}^2 = \dot{\varepsilon}_{ij}\dot{\varepsilon}_{ij}/2$, just before equations (7) and (28) but then expanded for clarity.

I prefer using the nonstandard horizontal-index notation, i.e., $u_{(i)}$ vs $P_i(u)$, because it is more compact. However, I've added a sentence to clarify this in the paragraph following Fig. 1.

I have removed superscripts on surface normals as per your suggestion. Thank you.

2. The paper is quite ungenerous to the prior literature. I can see the appeal of referencing only your own paper ("DPL" in the present case) as the default reference because you know its content extremely well, but many of the introductory concepts in this manuscript have been developed in other places, which remain uncited. That may put off other practicioners who should read this paper (and who might in fact get some sort of notification, alerting them to your paper if you were to cite them!). It's also unhelpful to any reader who wants to make sense of the field: the may conclude that, really, only DPL is relevant as a prior publication. In particular, I see almost no reference to the extensive numerical analysis literature on Herterich-Blatter-Pattyn and Stokes flow models in glaciology (except my ca. 2010 own effort in that direction, which does get a citation somewhere).

I am happy to add additional or more appropriate references. Any suggestions?

3. The "Improved Blatter-Pattyn or BP+ Approximation" in section 6.2.1. I am unconvinced that it makes sense to include this in the present paper, unles you want to "patent" the idea for eternity. (Who knows, maybe someone will in future refer this approximation by someone else's name, the same fate that befell Herterich?) The reason why I am unconvinced is that this material further breaks the flow of the paper, without being robust. The supposedly improved approximation remains an *ad hoc*, partial retention of higher order terms in the Stokes flow model, higher order being in sense of the aspect ratio. There is no theoretical justification for doing that, as a partial retention of higher order terms in a model comes with no guarantee of a reduction in model error. In fact, it can make the model error worse. We do see a reduction in model error for the "BP+" model ("HBP+" / "D"?) in the single test that the "improved" model has been subjected to (ISMIP-HOM Test B, figure 9 of the manuscript). However, that is not really a robust demonstration of an "improved" model. It's also unclear whether the supposedly improved model is really competitive relative to solving the full Stokes model in terms of the tradeoff between accuracy and computational effort. (See my second major point above.)

I wonder if this comment is relevant now that we understand that the BP+ approximation is the same as the Herterich model? I doubt that it's necessary to have a full scale analysis when introducing a new approximation. For example, the Blatter-Pattyn model did not have a scale analysis for 15 years until Schoof and Hindmarsh (2010).

Specific points

A number of specific points, some of which will replicate elements of what I've written above.

- p3, line 81: "finite element grid" — "grid" might be seen to imply regularity. "finite element basis functions"? I mean, closer scrutiny does indicate that the basis functions used here are restricted to quite regular meshes, so perhaps the terminology is not wrong here, but that required regularity actually needs to be discussed somewhere.

Changed "grids" to "discretizations"

- p3 line 84 "these two elements are so-named because they employ edge-based pressures" — I actually found that confusing. The basis function for pressure is not really defined on edges (since there are plenty of edges in the mesh that don't have a pressure defined on them). They are really P0 basis functions in which two adjacent triangles of the triangulation are assigned the same pressure.
  Also, to make it clear that you are inventing these basis functions, avoid the passive voice here. "I have named these E0 to indicate that pressure is defined on select element edges" or similar. It took me a while to realize that I would look in vain for references to these elements in the literature.

I have changed the wording as follows "these two elements are novel and are so-named because they employ pressures located on vertical grid edges"

- p 3 line 89 onwards " A conventional ice sheet Stokes model discretized on such a grid is numerically equivalent to an inherently stable positive-definite minimization (i.e., optimization) problem, as demonstrated in Appendix D. This is in contrast to the ubiquitous Stokes finite element practice of needing to use elements that satisfy the "inf-sup" or "LBB" condition for stability (see Elman et al., 2014, and the brief discussion in §4.3.1)." — This honestly confused me. If you're saying that current practice in solving the Stokes equations in ice sheet dynamics ("A conventional ice sheet Stokes model") is to use unconstrained optimization, then please provide a reference. It's very hard to do unconstrained optimization for Stokes flow as you need divergence-free basis functions. If you mean that *your new approach* is equivalent to constructing such a divergence-free basis, then say that instead. However, you really also need to show that the reduction of the single Newton iteration step to a problem for $(u_1, \ldots, u_N, w_1, \ldots, u_N)$ only in appendix D (taking the dimensionality argument for $\mathbf{M}_{WP}$ at face value) really is equivalent to a convergent solution (under mesh refinement) of the unconstrained Stokes optimization problem (that is, the problem in which the incompressibility constraint is directly imposed on admissible $\boldsymbol{u}$). I don't think this is entirely trivial, see also my discussion of the inf-sup criterion above. (Put this another way, for any set of basis functions that allows the problem to be solved, I must be able in principle to eliminate the discretized pressure variable in favour of retaining on the nodal values of velocity, even if I would not choose to do the matrix manipulations involved. By your logic, any such scheme should therefore be robust. Is that true?) I may be mistaken about what you're trying to say here, but it may therefore be worth finding a different way of saying it.

As mentioned on Page 5 of this response, I have eliminated Appendix D and rewritten Section 4 and I hope that this now addresses the issue. No, my approach is not equivalent to constructing divergence-free basis functions. In fact, I doubt that such a basis exists (Boffi et al. (2008), a new reference added in the paper). What I am trying to say is that given an arbitrary discrete horizontal velocity field (u) on a grid that satisfies the solvability condition (i.e., square, full rank matrix M_{WP}^T), I can always find an "incompressible" velocity field $(u, w(u))$. This, when substituted into a constrained functional (Jacobian?) for variables (u, w, p) converts it into an equivalent unconstrained functional for (u) alone. This is what I'm trying to say in the new Section 4.3.2. I have also reworded the end part of the Introduction to express this.

- p 4 line 116. $S = S_{B1} \cup S_{B2}$?

Changed

- p 4 line 121–123. See above re: notation.
- eqs (8), (10, (11), (13)–(15). The superscripts on your unit normals seem to be redundant. There is only one (outward-pointing) unit normal to each of these parts of the boundary.

Corrected, see Point 1 above.

- p 6 line 174, "f_i is a specified frictional sliding force vector." — f_i has the wrong units for a force. It's a traction, but why not just call it an interfacial shear stress? τ_i would probably be a more common symbol.

Changed throughout the paper except that I used the symbol to τ_i^S to avoid confusion with the stress tensor τ_{ij} .

- p6, line 170 "the simplest representation" — I beg to differ. There is a standard way of writing shear stress, as the traction $\Sigma_i = \sigma_{ij} n_j$ minus the normal component of traction $n_i \Sigma_j n_j$, or

$$(\delta_{ij} - n_i n_j)\sigma_{jk} n_k.$$

That is equation (71) in appendix A, and I think you'd be well advised to just use that (standard) form. (The point about weak solutions is surely that you never need to actually evaluate the shear stress itself from the stress tensor and the normal to the surface; you just need to know the constitutive relation, which here is just

$$(\delta_{ij} - n_i n_j)\sigma_{jk} n_{k}. = \beta u_i.$$

I think what the paper says isn't wrong, but manipulations don't seem they will help the unwitting reader, who is just wrapping their head around the basic model formulation. Especially at this point, where you haven't motivated your use of the bracketed indices at all yet, in terms of the mathematics you're doing.

- page 13, equation (32). As above, that is pretty ugly. I can't tell if there is a less confusing form, but might be worth trying.

You are right! The form I was using is computationally easier because it eliminates the complicated quantity $\tau_n = n_i \tau_{ij} n_j$, the normal component of the stress force. But if the discretization is based on a variational principle (or on a weak formulation) this is irrelevant because this quantity never needs to be explicitly calculated. One might as well eliminate Appendix A (shortens the paper!) and obtain the tangential frictional shear stress as in DPL (2010) or your expression above, except that I prefer to use the deviatoric stress tensor as in the rest of the paper. Equations (11) and (32) have been updated.

- p 7 line 180 "... can easily be added" → "... can esaily be added to equation (10)"?

Done

- p 8 line 218 " Here we use z_b as a shorthand notation for $z_b(x, y)$' — this is surely redundant? Same on line 259.

Removed in both places.

- p 7 "(see DPL, 2010, for a fuller description of the variational principle applied to ice sheet modeling)" — you should refer to
  Chen, Q, M. Gunzburger and M. Perego. 2013. Well-posedness results for a nonlinear Stokes problem arising in glaciology. *SIAM J. Math. Anal.*, 45(5), 2710-2733. This is probably the most comprehensive analysis of Stokes flows with sliding in glaciology (and you'll find the relevant functional there, too, naturally).

Reference added

- p 7 line 200 "... (see Schoof, 2010, in connection with the Blatter-Pattyn model)" — I feel a little unfairly singled out here. There are earlier papers by Coling and Rappaz (MSAN, 1999), Glowinski and Rappaz (2003), Chow et al 92004) Rappaz and Reist (M3AS, 2005) that deserve an equal mention as having contributed to the analysis of the Herterich-Blatter-Pattyn model

I've inserted "as well as earlier references contained therein"

- p8 line 208 " As in DPL (2010), arguments enclosed in square brackets, here u_i , P, λ_i ,Λ , indicate those variables that are used in the variation of the functional." — this is an odd way of putting it. A functional is simply a mapping (or function) from some vector space (or perhaps an affine space) into the real numbers. So why not just say that you're enclosing the arguments of the functional (which is a function and therefore takes arguemnts like any other function![5]) in square brackets.

[5]Unless you insist that a function have to take arguments in $\mathbb{R}^n$ for some finite n.

I have rephrased it as "As in DPL (2010), arguments enclosed in square brackets, here $u_i, P, \lambda_i, \Lambda$, indicates those functions that are subject to variation as arguments of the functional ". Hope this helps.

- page 9 line 254"... as is done in the Blatter-Pattyn approximation (see DPL, 2010)."— is the apropriate reference not one of Herterich (1987), Blatter (1995) or Pattyn (2003(?

Yes, replaced by Pattyn (2003). I couldn't actually find it in Blatter (1995).

- page 12 line 296. It would be a good idea to explain the role of the flag parameters ξ and $\hat{\xi}$ earlier. I spent a page half guessing what they were.

The dummy variables are now explained directly below where they are introduced, i.e., right after equations (23)-(28).

- p 8 line 210 onwards. I don't think this correct, or at least, it seems misleading. In my understanding, there are two ways of imposing Dirichlet conditions: by Lagrange multiplier, or by restricting the space of admissible functions. The latter is not the same as the kind of explicit substitution you're doing here. If I take the functional defined in equation (15) and I *don't* separately impose the constraint $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ at $\S_b$ on the arguments u_i, then I don;t see that taking the first variation of the functional will recover that constraint. If you impose the constraint $\boldsymbol{u} \cdot \boldsymbol{n} = 0$, then you do not need to make the substitution in equation (14) in the friction term of the Stokes Lagrangian
  The substitution here is again quite nonstandard, with little advance warning or real motivation. I think you're really trying to lay the groundwork for the modified Stokes variational principle later (by eliminating w where that's going to be necessary) but here is an awkward place to do it, unless you explain why you're eliminating w.

Again an oversight on my part! I should have pointed out that direct substitution is possible only in the discrete formulation of the functional where boundary variables are directly accessible (except of course in the surface integral terns where surface values are accessible even in the continuous formulation). I have tried to clarify this in the text.

- page 9 line 232 "...the specified values of velocity are then obtainable a posteriori from (9) or (14)" — this is a strange way of putting it. What is a specified value? One that is prescribed? I don't think that's what you mean. I think you mean that the taking the first variation does not recover the boundary condition $\boldsymbol{u} \cdot \boldsymbol{n} = 0$, see previous point. However, you can't impose that "a posteriori" since you cannot solve the Euler-Lagrange equations for the functional in (15) without using the Dirichlet condition on velocity. The standard way of putting this (I believe) is to restrict the functional to the vector space of suitable smooth velocities (in $[W^{1,1+1/n}(\Omega)]^3$) *that satisfy the Dirichlet condition.*

Yes, I do mean the prescribed values. I have changed the wording in the text to hopefully make this clearer.

- page 10, line 265 "The standard Stokes pressure P is some three orders of magnitude larger than the transformed pressure $\tilde{P}$" — asymptotic analysis of the problem will show that, more generally, the "Stokes correction" $\tilde{P}$ scales as $O(\varepsilon)$ when there is signficant sliding, where ε is aspect ratio, and of $O(\varepsilon^{1+1/n})$ in the absence of sliding (see the Schoof and Hindmarsh 2010 reference).

My main objective here was to highlight the absence of the large hydrostatic pressure component in the "Stokes correction".

- page 15 line 376 Remark # 1: seems like splitting hairs, especially as you wouldn't bother with this if you wrote down the weak form.

Maybe, but I think it's useful to point out that the presented BP equations differ, if only slightly, from the original.

- p 13 line 353 "The standard (or traditional) Blatter-Pattyn approximation (originally introduced by Blatter, 1995; Pattyn, 2003; later by DPL, 2010; Schoof and Hewitt, 2013)" — as above, Herterich (1987) was the person who introduced this. I don't think the Schoof and Hewitt review paper did much more than describe the theory, as oppposed to contributing to it. If I did anything here, then Schoof and Hindmarsh (2010) provided the first self-consistent asymptotic analysis of the model. If you want to list DPL for its description of the variatoinal formulation, you probably ought to cite the numerical analysis papers I have listed above (Coling and Rappaz, Glowinski and Rappaz, Chow et al, Rappaz and Reist; full citations in my 2010 paper that is in the reference list), who previously dealt with the same variational formulation.

Herterich is not pertinent since he did not introduce BP, just BP+. I've added "and references therein" to Schoof and Hewitt, 2013. Hope this is adequate. I'm reluctant to add a lot of extra references.

- page 15 line 419 "In summary, the extended Blatter-Pattyn model, (40)-(42), is equivalent to the standard Blatter-Pattyn model, (36), for the horizontal velocities, u,v , except that it also includes two additional equations that determine the pressure P ! and the vertical velocity w , which are usually ignored in the standard Blatter-Pattyn approximation when only the horizontal velocity is of interest. Because of this, we distinguish between the Blatter- Pattyn model that solves for just the two horizontal velocities (i.e., the standard Blatter- Pattyn approximation, (36)), and the Blatter-Pattyn system that solves for all the variables (i.e., the extended Blatter-Pattyn approximation, (40)-(42))." — again, this seems like splitting hairs. Anyone who needs to solve for temperature in an ice sheet (which is a standard part of any ice sheet simulation code) has to solve for the vertical velocity component, so I don't think insisting on a difference between the Blatter-Pattyn model and system is helpful. Solving for $\tilde{P}$ is not particularly relevant as we

  know $\tilde{P} = 0$ for the Blatter-Pattyn model. For the sake of not going down rabbit holes, I would omit this.

- p 17 line 437 "The use of the continuity equation to solve for the vertical velocity w is a novel feature of the Blatter-Pattyn approximation since the continuity equation is not normally used for this purpose." — as per my previous comment, I don't think this is true or novel. The vertical velocity component is a pretty important quantity to be able to compute in ice sheet codes.

- p 17 lie 439 onwards. The use of a depth-integrated mass conservation equation is a boilerplate approach for thin film models, so a lengthy discussion without citation seems unnecessary, especially as this doesn't really lead anywhere in developing the novel material in the paper.

I have rewritten the relevant sections and shortened them considerably. Hope this resolves the issue.

- p 17 line 482 "Additional details about the grid and the associated discretization…" — I flagged this above (grid versus mesh), but again: I think it's worth pointing out somewhere that some of the basis functions developed here really do require a regular mesh (grid?) rather than an unstructured triangulation. For the P1-E0 basis functions, I have to be able to match every triangle with precisely one neighbour, not leaving any neighbourless triangles. I don't think that's possible with every fully unstructured mesh. (?)

I think this is already done in Appendix B and especially in Fig. B2, which shows that the P1-E0 element is quite flexible and can be used with quite general triangulations. (Note: Appendix A has been eliminated so all remaining appendices have been renumbered.)

- page 19 line 504. Italicize the variables u and w.

Done

- page 21 line 526, "…or else they are "saddle point" problems since the Hessian matrix $M(u,w)$ is symmetric but indefinite, with both positive and negative eigenvalues…" — you can be more definitive about this I think: the "saddle point" terminology refers to the fact that there really is a saddle point, where you mimiize the quadratuc form associated with the matrix with respect to the velocity variables and maximize with respect to pressure.

I have added a sentence to this effect in the text.

- p 22 line 565 Please define the acronym "LBB" before using it.

Done on p. 22 in Section 4.3.1

- p 23 line 601 Taylor-Hood elements have gone from being an illustration of an element that satisfies the inf-sup condition (line 569) to becoming the standard reference for stable elements. Is it true in general that all elements that satisfy the LBB condition for the Stokes problem will leave M_{WP}^{T} non-invertible on dimensional grounds alone? If so, that is important to point out here.

Pointed out in Section 4.3.1

- Sections 5.1–5.3 You're focusing on velocity solutions here. If there are stability issues with your choice of basis functions, I'd expect these to show up in the pressure solution. Can we see resultss for some of those? Obviously not relevant for Herterich-Blatter-Pattyn, but for the transformed and original Stokes flow problems. Especially pressure along the bed would be useful, but also convergence or lack thereof.

I have added the following at the end of Section 5: "Pressure results are not shown because pressure, particularly in the transformed case, has little or no physical significance. However, pressure calculated on the P1-E0 grid is particularly smooth and well behaved."

- Figure 7: this is fine, but more convincing would be a grounding line, where you have to worry about a normal stress constraint determining the grounding line location (and that normal stress is in general not cryostatic as assumed by Herterich-Blatter-Pattyn)

That's true, but it's too late to make such a major change

Appendix - Identification of the Herterich model

From §3.4.1 of the present paper the 2D Blatter-Pattyn model is given by

$$\frac{\partial}{\partial x}\left(4\mu_{BP}\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial z}\left(\mu_{BP}\frac{\partial u}{\partial z}\right)-\rho g\frac{\partial z_s}{\partial x}=0\,, \tag{1}$$

where the effective viscosity is

$$\mu_{BP}=\eta_0\left(\dot{\varepsilon}_{BP}^2\right)^{(1-n)/2n}\quad , \tag{2}$$

and the second invariant is

$$\dot{\varepsilon}_{BP}^2=\left(\frac{\partial u}{\partial x}\right)^2+\frac{1}{4}\frac{\partial u}{\partial z}^2\;. \tag{3}$$

After allowing for minor changes in notation, this is the same as found in Pattyn (2003), Schoof (2010), and DPL (2010).

The Herterich (1987) system, corresponding to (1)-(3), consists of his equations (2.11) and (2.12) which, using the present notation (i.e., x, u in the horizontal and z, w in the vertical), may be written as follows

$$\begin{gathered}\frac{\partial}{\partial x}\left(4\mu_H\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial z}\left(\mu_H\left(\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}\right)\right)-\rho g\frac{\partial z_s}{\partial x}=0,\\ \frac{\partial u}{\partial x}+\frac{\partial w}{\partial z}=0,\end{gathered} \tag{4}$$

where

$$\mu_H=\frac{f}{2A^{1/n}}=\frac{1}{2A^{1/n}}\left(\dot{\varepsilon}_H^2\right)^{(1-n)/2n}, \tag{5}$$

and

$$\dot{\varepsilon}_H^2=\left(\frac{\partial u}{\partial x}\right)^2+\frac{1}{4}\left(\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}\right)^2. \tag{6}$$

Comparing (5) to (2), we can identify η_0 with $1/2A^{1/n}$. Thus, comparing Herterich's model, (4) and (5), to the Blatter-Pattyn model, (1)-(3), we see that the two models are

NOT the same because of the presence of the $\partial w/\partial x$ terms in the Herterich model. (Note: There is a typo in Eq. (2.10) of the Herterich paper, i.e., $\partial v/\partial z$ should be $\partial v/\partial x$.)

Now, consider the BP+ approximation as given by the two functionals (62)-(64) from §6.2.1 in the paper. Ignoring the boundary condition terms since they don't matter for the present purpose, the functionals may be written as

$$\begin{gathered}\tilde{\mathcal{A}}_{PS1}[u] = \int_V dV \left[\frac{4n}{n+1} \eta_0 \left(\tilde{\dot{\varepsilon}}^2 \right)^{(1+n)/2n} + \rho g u \frac{\partial z_s}{\partial x} \right], \\ \tilde{\mathcal{A}}_{PS2}[p] = \int_V dV \; p \left(\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} \right),\end{gathered} \tag{7}$$

where

$$\tilde{\dot{\varepsilon}}^2 = \left(\frac{\partial u}{\partial x} \right)^2 + \frac{1}{4} \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right)^2 . \tag{8}$$

Since $\eta_0 = 1/2 A^{1/n}$, the variation of $\tilde{\mathcal{A}}_{PS1}[u]$ with respect to u , and of $\tilde{\mathcal{A}}_{PS2}[p]$ with respect to p , yields precisely the Herterich system (4)-(6)! Together with the fact that $\tilde{\dot{\varepsilon}}^2 = \dot{\varepsilon}_H^2$, we conclude that Herterich's model is identically the same as the BP+ approximation.

A paragraph has been added to Section 6.2.1 pointing this out.