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.

Changes in response to Prof. Bueler in Red.
Changes in response to Prof. Schoof in Blue.

A Novel Transformation of the Ice Sheet Stokes Equations and Some of its Properties and Applications

John K. Dukowicz
Guest Scientist, Group T-3, Los Alamos National Laboratory,
Los Alamos, New Mexico, 87545, USA
Correspondence to: John K. Dukowicz (jn.dk@outlook.com)

Abstract. We introduce a novel transformation of the Stokes equations into a form that closely resembles the shallow Blatter-Pattyn equations. The two forms differ by only a few additional terms, while their variational formulations differ only by a single term in each horizontal direction. Specifically, the variational formulation of the Blatter-Pattyn model drops the vertical velocity in the second invariant of the strain rate tensor. Here we make use of the new transformation in two different ways. First, we consider incorporating the transformed equations into a code that can be very easily converted from a Stokes to a Blatter-Pattyn model, and vice-versa, by switching these terms on or off. This may be generalized so that the Stokes model is switched on adaptively only where the Blatter-Pattyn model loses accuracy. Second, the key role played by the vertical velocity in the Blatter-Pattyn approximation motivates new approximations that improve on the Blatter-Pattyn model. These applications require a grid that enables the discrete continuity equation to be invertible for the vertical velocity in terms of the horizontal velocity components. Examples of such grids, such as the first order P1-E0 grid and the second order P2-E1 grid are given in both 2D and 3D. It should be noted, however, that the transformed Stokes model has the same type of gravity forcing as the Blatter-Pattyn model, determined by the ice surface slope, thereby forgoing some of the grid-generality of the traditional formulation of the Stokes model.

1 Introduction

Concern and uncertainty about the magnitude of sea level rise due to melting of the Greenland and Antarctic ice sheets have led to increased interest in improved ice sheet and glacier modeling. The gold standard is a Stokes model (i.e., a model that solves the nonlinear, non-Newtonian Stokes system of equations for incompressible ice sheet dynamics) because it is applicable to all geometries and flow regimes. However, the Stokes model is computationally demanding and expensive to solve. It is a nonlinear, three-dimensional model involving four variables, namely, the three velocity components and pressure. In addition, pressure is a Lagrange multiplier enforcing incompressibility and this creates a more difficult indefinite "saddle point" problem. As a result, full-Stokes models exist but are not commonly used in practice (examples are FELIX-S, Leng et al., 2012; Elmer/Ice, Gagliardini et al., 2013).

Because of these difficulties there is much interest in simpler and cheaper approximate models. There is a hierarchy of very simple models such as the shallow ice (SIA) and shallow-shelf (SSA) models, and there are also various more accurate higher-order approximations. These culminate in the Blatter-Pattyn (BP) approximation (Blatter, 1995; Pattyn, 2003), which is currently used in production code packages such as ISSM (Larour et al., 2012), MALI (Hoffman et al., 2018; Tezaur et al., 2015) and CISM (Lipscomb et al., 2019). This approximation is based on the assumption of a small ice sheet aspect ratio, i.e., $\varepsilon = H/L \ll 1$, where H, L are the vertical and horizontal length scales, and consequently it eliminates certain stress terms and implicitly assumes small basal slopes. Both the Stokes and Blatter-Pattyn models are described in detail in Dukowicz et al. (2010), hereafter referred to as DPL (2010). Although the Blatter-Pattyn model is reasonably accurate for large-scale motions, accuracy deteriorates for small horizontal scales, less than about five ice thicknesses in the ISMIP–HOM model intercomparison (Pattyn et al., 2008; Perego et al., 2012), or below a 1 km resolution as found in a detailed comparison with full Stokes calculations (Rückamp et al, 2022). This can become particularly important for calculations involving details near the grounding line where the full accuracy of the Stokes model is needed (Nowicki and Wingham, 2008). Attempts to address the problem while avoiding the use of full Stokes solvers include variable grid resolution coupled with a Blatter-Pattyn solver (Hoffman et al., 2018) and variable model complexity, where a Stokes solver is embedded locally in a lower order model (Seroussi et al., 2012). Better approximations, more accurate than Blatter-Pattyn but cheaper than Stokes, are currently not available.

The present paper introduces two innovations that may begin to address some of these issues. The first is a novel transformation of the Stokes model, described in §3, which puts it into a form closely resembling the Blatter-Pattyn model and differing only by the presence of a few extra terms. This allows a code to be switched over from Stokes to Blatter-Pattyn, and vice-versa, globally or locally, by the use of a single parameter that turns off these extra terms. As a result, variable model complexity can be very simply implemented, as described in §6.1. The second innovation is the introduction of new finite element discretizations that decouple the discrete continuity equation and allow it to be solved for the vertical velocity in terms of the horizontal velocity components. Several elements used to construct such grids are described in Appendix C in both 2D and 3D, primarily the first order P1-E0 and second order P2-E1 elements (these two elements are novel and are so-named because they employ pressures located on vertical grid edges). Within the framework of the transformed Stokes model these grids facilitate new approximations that improve on the Blatter-Pattyn approximation so that it is no longer strictly limited to a small ice sheet aspect ratio. We describe two such approximations in §6.2. There is another very significant benefit. An ice sheet Stokes model is conventionally discretized as a constrained minimization problem requiring special "stable" finite elements for solution. However, the same model on these new grids can be formulated as an inherently stable and numerically equivalent unconstrained minimization problem, as demonstrated in §4.3.2.

2 The Standard Formulation of the Stokes Ice Sheet Model

2.1 The Assumed Ice Sheet Configuration

An ice sheet may be divided into two parts, a part in contact with the bed and a floating ice shelf located beyond the grounding line. The Stokes ice sheet model is capable of describing the flow of an arbitrarily shaped ice sheet, including a floating ice shelf as illustrated in Fig. 1, given appropriate boundary conditions (e.g., Cheng et al., 2020). One limitation of the methods proposed here, in common with the Blatter-Pattyn model, will be that there should be just one upper and one basal surface, as is the case in Fig. 1. Here we will only consider a fully grounded ice sheet with periodic lateral boundary conditions, i.e., no ice shelf, although in general ice shelves can be handled.

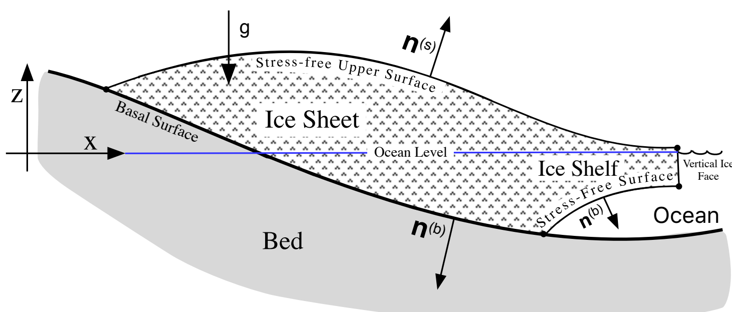

Figure 1 A simplified illustration of the admissible ice sheet configuration.

Referring to Fig. 1, the entire surface of the ice sheet is denoted by S. An upper surface, labeled S_S and specified by $\varsigma_s(x,y,z) = z - z_s(x,y) = 0$, is exposed to the atmosphere and thus experiences stress-free boundary conditions. The bottom or basal surface, denoted by S_B and specified by $\varsigma_b(x,y,z) = z - z_b(x,y) = 0$, is in contact with the bed. The basal surface may be subdivided into two sections, $S_B = S_{B1} \cup S_{B2}$, where S_{B1}, specified by $z = z_{b1}(x,y)$, is the part where ice is frozen to the bed (a no-slip boundary condition), and S_{B2}, specified by $z = z_{b2}(x,y)$, is where frictional sliding occurs. We assume Cartesian coordinates such that $x_i = (x,y,z)$ are position coordinates with $z = 0$ at the ocean surface, and the index $i \in \{x,y,z\}$ represents the three Cartesian indices. Later we shall have occasion to introduce the restricted index $(i) \in \{x,y\}$ to represent just the two horizontal indices. Note that this is equivalent to applying a projection operator but is more compact, i.e., $u_{(i)} = P_i(u) = (u,v,0)$. Unit normal vectors appropriate for the ice sheet configuration of Fig. 1 are given by

$$\begin{aligned} n_i = (n_x, n_y, n_z) &= \frac{\partial \varsigma_s(x,y,z)/\partial x_i}{\left|\partial \varsigma_s(x,y,z)/\partial x_i\right|} = \frac{(-\partial z_s/\partial x, -\partial z_s/\partial y, 1)}{\sqrt{1+(\partial z_s/\partial x)^2 + (\partial z_s/\partial y)^2}} \quad \text{at surface } S_S, \\ n_i = (n_x, n_y, n_z) &= -\frac{\partial \varsigma_b(x,y,z)/\partial x_i}{\left|\partial \varsigma_b(x,y,z)/\partial x_i\right|} = \frac{(\partial z_b/\partial x, \partial z_b/\partial y, -1)}{\sqrt{1+(\partial z_b/\partial x)^2 + (\partial z_b/\partial y)^2}} \quad \text{at surface } S_B. \end{aligned} \tag{1}$$

2.2 The Stokes Equations

The Stokes model is a system of nonlinear partial differential equations and associated boundary conditions (Greve and Blatter, 2009; DPL, 2010). In a Cartesian coordinate system the Stokes equations, the three momentum equations and the continuity equation, for the three velocity components $u_i = (u,v,w)$ and the pressure P are given by

$$\frac{\partial \tau_{ij}}{\partial x_j} - \frac{\partial P}{\partial x_i} + \rho g_i = 0, \tag{2}$$

$$\frac{\partial u_i}{\partial x_i} = 0 \,, \tag{3}$$

where ρ is the density, and g_i is the acceleration vector due to gravity, arbitrarily oriented in general but here taken to be in the negative z-direction, $g_i = (0,0,-g)$. Repeated indices imply summation (the Einstein notation). The deviatoric stress tensor τ_{ij} is given by

$$\tau_{ij} = 2\mu_n \dot{\varepsilon}_{ij} \,, \tag{4}$$

where the strain rate tensor is

$$\dot{\varepsilon}_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \tag{5}$$

the nonlinear ice viscosity μ_n is a defined by

$$\mu_n = \eta_0 \left(\dot{\varepsilon}^2 \right)^{(1-n)/2n} \,, \tag{6}$$

and $\dot{\varepsilon}^2 = \dot{\varepsilon}_{ij}\dot{\varepsilon}_{ij}/2$ is the second invariant of the strain rate tensor that may be written out in full as follows

$$\dot{\varepsilon}^2 = \frac{1}{2}\left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial w}{\partial z}\right)^2 \right] + \frac{1}{4}\left[\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right)^2 \right]. \tag{7}$$

Note that the second invariant is a positive quantity, i.e., $\dot{\varepsilon}^2 \geq 0$. As usual, ice is assumed to obey Glen's flow law, where n is the Glen's law exponent ($n = 1$ for a linear Newtonian fluid. Typically $n = 3$ in ice sheet modeling, resulting in a nonlinear non-Newtonian fluid). The coefficient η_0 is defined by $\eta_0 = A^{-1/n}/2$, where A is an ice flow factor, here taken to be a constant but in general depending on temperature and other variables (see Schoof and Hewitt, 2013). The three-dimensional Stokes system requires a set of boundary conditions at every bounding surface, each set being composed of three components. Aside from the periodic lateral boundary conditions used in our test problems, the relevant boundary conditions are given as follows

(1) Stress-free boundary conditions on surfaces S_S not in contact with the bed, such as the upper surface S_S:

$$\tau_{ij} n_j - P n_t = 0 \,. \tag{8}$$

(2) No-slip or frozen to the bed conditions on surface segment S_{B1}:

$$u_i = 0 \tag{9}$$

(3) Frictional tangential sliding conditions on surface segment S_{B2} in two parts:

(3a) A single condition enforcing tangential flow at the basal surface:

$$u_i n_i = 0\,. \tag{10}$$

(3b) Two conditions specifying the horizontal components of the tangential frictional stress force vector, as follows

$$\tau_{(i)j} n_j - \tau_n n_{(i)} + \tau^S_{(i)} = 0\,, \tag{11}$$

where $\tau_n = n_i \tau_{ij} n_j$ is the normal component of the shear stress, and τ^S_i is a specified interfacial shear stress, tangential to the bed $\left(n_i \tau^S_i = 0\right)$. The tangential shear stress or traction is obtained as in DPL (2010) by subtracting out the normal component from the shear stress. However, the three components of the tangential shear stress are not independent because they already satisfy the tangency condition at the basal surface and therefore we retain only the horizontal components. The interfacial shear stress τ^S_i is potentially a complicated function of position and velocity (e.g., Schoof, 2010). However, here we assume only simple linear frictional sliding,

$$\tau^S_i = \beta\left(x\right) u_i\,, \tag{12}$$

where $\beta\left(x\right) > 0$ is a position-dependent drag law coefficient. For simplicity we assume there is no melting or refreezing at the bed resulting in vertical inflows or outflows. If needed, these can be easily added to (10) (Dukowicz et al., 2010; Heinlein et al., 2022).

2.3 The Stokes Variational Principle

A variational principle, if available, is usually the most compact way of representing a particular problem. The Stokes model possesses a variational principle that is particularly useful for discretization purposes and for the specification of boundary conditions (see DPL, 2010, and Chen et al., 2013, for a fuller description of the variational principle applied to ice sheet modeling). There are a number of significant advantages. For example, all boundary conditions are conveniently incorporated in the variational formulation, all terms in the variational functional, including boundary condition terms, contain lower order derivatives than in the momentum equations, and the resulting discretization automatically involves a symmetric matrix. In discretizing the momentum equations, stress terms at boundaries involve derivatives that would normally have to be evaluated using less accurate one-sided approximations. This problem does not arise in the variational formulation since all terms are evaluated in the interior. Finally, stress-free boundary conditions, as at the upper surface for example, need not be specified at all since they are automatically incorporated in the functional as natural boundary conditions. In discrete applications, the variational method presented here is closely related to the Galerkin finite element method, a subset of the weak formulation method in which the test and trial functions are the same (see Schoof, 2010, and earlier references contained therein in connection with the Blatter-Pattyn model).

The variational functional for the standard Stokes model may be written in two alternative forms:

(1) Basal boundary conditions imposed using Lagrange multipliers:

$$\begin{aligned}\mathcal{A}[u_i,P,\lambda_i,\Lambda]=&\int_V dV\left[\frac{4n}{n+1}\eta_0\left(\dot{\varepsilon}^2\right)^{(1+n)/2n}-P\frac{\partial u_i}{\partial x_i}+\rho g w\right]\\&+\int_{S_{B1}}dS\,\lambda_i u_i+\int_{S_{B2}}dS\left[\Lambda u_i n_i+\frac{1}{2}\beta\left(x\right)u_i u_i\right],\end{aligned}\qquad(13)$$

where λ_i and Λ are Lagrange multipliers used to enforce the no-slip condition and frictional tangential sliding, respectively. 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.

(2) Basal boundary conditions imposed by direct substitution:

In this case, the two conditions (9), (10) are used directly in the functional to specify all three velocity components u_i in the first case, and the vertical velocity w in terms of the horizontal velocity components in the second case, along the entire basal boundary in both the volume and surface integrals in (13). However, this can only be done in the discrete formulation of the functional since only then are boundary values of velocity accessible (except in the surface integral terms where they are always accessible). In particular, the tangential flow condition (10) is used in the following form,

$$w=-\frac{u_{(i)}n_{(i)}}{n_z}=u_{(i)}\frac{\partial z_b}{\partial x_{(i)}},\qquad(14)$$

to eliminate w on the basal boundary segment S_{B2} of the variational functional, to obtain

$$\mathcal{A}[u_i,P]=\int_V dV\left[\frac{4n}{n+1}\eta_0\left(\dot{\varepsilon}^2\right)^{(1+n)/2n}-P\frac{\partial u_i}{\partial x_i}+\rho g w\right]$$
$$+\frac{1}{2}\int_{S_{B2}} dS\,\beta(x)\left(u_{(i)}u_{(i)}+\left(u_{(i)}n_{(i)}/n_z^{(b2)}\right)^2\right). \tag{15}$$

It is important to emphasize again that boundary conditions (9) and (14) must also be applied in the volume integral part of the discretized functional (15) as part of direct substitution to replace velocity variables that lie on the basal boundary. In the case of (14), horizontal velocity variables remain undisturbed while w is eliminated, thus implementing the tangential sliding boundary condition.

As described in DPL (2010), a variational procedure yields the full set of Euler-Lagrange equations and boundary conditions that specify the standard Stokes model, equivalent to (2)-(11). In the case of (13), the system determines all the discrete variables specified on the mesh: the velocity components and the pressure, u_i, P, as well as the Lagrange multipliers, λ_i, Λ. In the direct substitution case, (15), the numerical solution determines only the pressure P and those velocity variables u_i that were not directly prescribed as boundary conditions in (9) or (14). These prescribed (known) values of boundary velocities are then added a posteriori. As a result, the direct substitution method is smaller and simpler, and therefore is the one primarily used in the paper.

3. A Transformation of the Stokes Model

3.1 Origin of the Transformation

The transformation is motivated by the Blatter-Pattyn approximation. Consider the vertical component of the momentum equation and the corresponding stress-free upper surface boundary condition in the Blatter-Pattyn approximation (from DPL, 2010, for example), which are given by

$$\frac{\partial}{\partial z}\left(2\mu_n\frac{\partial w}{\partial z}\right)-\frac{\partial P}{\partial z}-\rho g=0,$$
$$\left(2\mu_n\frac{\partial w}{\partial z}-P\right)n_z=0 \quad \text{at} \quad z=z_s(x,y). \tag{16}$$

These equations may be rewritten in the form

$$\frac{\partial}{\partial z}\left(P-2\mu_n\frac{\partial w}{\partial z}+\rho g\left(z-z_s\left(x,y\right)\right)\right)=0,$$
$$\left(P-2\mu_n\frac{\partial w}{\partial z}+\rho g\left(z-z_s\left(x,y\right)\right)\right)n_z=0 \quad \text{at} \quad z=z_s\left(x,y\right), \tag{17}$$

suggesting a new variable $\tilde{P}$, to be called the transformed pressure, as follows

$$\tilde{P}=P-2\mu_n\frac{\partial w}{\partial z}+\rho g\left(z-z_s\left(x,y\right)\right), \tag{18}$$

which simplifies system (17) to give

$$\frac{\partial \tilde{P}}{\partial z}=0,$$
$$\tilde{P}n_z=0 \quad \text{at} \quad z=z_s\left(x,y\right). \tag{19}$$

This is a complete one-dimensional partial differential system, that, when integrated from the top surface down yields

$$\tilde{P}=0\,. \tag{20}$$

Thus, the transformed pressure vanishes in the Blatter-Pattyn case. The definition (18) forms the basis of the present transformation but we also use the continuity equation to eliminate $\partial w/\partial z$ as is done in the Blatter-Pattyn approximation (e.g., Pattyn, 2003). Therefore, the transformation consists of eliminating P and $\partial w/\partial z$ in the Stokes system (2), (4)-(11) (i.e., everywhere except in the continuity equation (3) itself) by means of

$$P=\tilde{P}-2\mu_n\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+\rho g\left(z_s-z\right), \tag{21}$$

$$\frac{\partial w}{\partial z}=-\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right). \tag{22}$$

The pressure P in the standard Stokes system is primarily a Lagrange multiplier enforcing incompressibility, but with a very large hydrostatic component. The transformation eliminates the hydrostatic pressure from $\tilde{P}$, as illustrated in Fig. 2 where the two pressures are compared. The transformed pressure $\tilde{P}$ is some three orders of magnitude smaller than the standard Stokes pressure P primarily because of the absence of hydrostatic pressure.

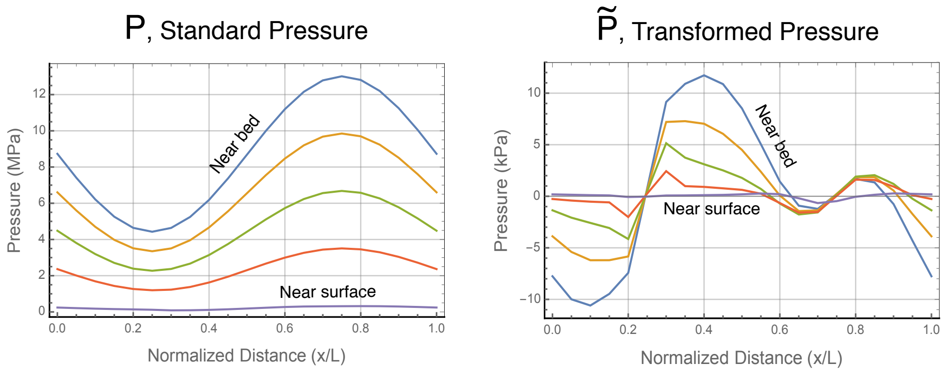

Figure 2. Standard pressure P compared to the transformed pressure $\tilde{P}$ in Exp. B from the ISMIP–HOM model intercomparison (Pattyn et al., 2008) at L = 10 km. Note that P is in MPa while $\tilde{P}$ is in kPa.

The transformed pressure $\tilde{P}$ is again a Lagrange multiplier enforcing incompressibility. Alternatively, since $\tilde{P} = 0$ in the Blatter-Pattyn approximation, the transformed pressure may be written as $\tilde{P} = P - P_{BP}$, where

$$P_{BP} = -2\mu_n \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + \rho g \left(z_s - z \right)$$

is the effective Blatter-Pattyn pressure (Tezaur et al., 2015). As a result, $P = P_{BP} + \tilde{P}$ and therefore $\tilde{P}$ is actually the "Stokes" correction to the Blatter-Pattyn pressure.

3.2 The Transformed Stokes Equations

Introducing (21), (22) into the Stokes system of equations (2)-(11) results in the following transformed Stokes system:

$$\frac{\partial \tilde{\tau}_{ij}}{\partial x_j} - \hat{\xi} \frac{\partial \tilde{P}}{\partial x_i} - \rho g \frac{\partial z_s}{\partial x_{(i)}} = 0 \, , \qquad (23)$$

$$\hat{\xi} \frac{\partial u_i}{\partial x_i} = 0 \, , \qquad (24)$$

where quantities that are modified in the transformation are indicated by a tilde, e.g., $\tilde{P}$.

Here and in the following we will be using dummy variables $\xi, \hat{\xi}$ to indicate terms that are absent in the Blatter-Blatter approximation, as explained below. Corresponding to (4) the modified Stokes deviatoric stress tensor $\tilde{\tau}_{ij}$ is given by

$$\tilde{\tau}_{ij} = 2\tilde{\mu}_n \left(\tilde{\dot{\varepsilon}}_{ij} + \frac{\partial u_{(i)}}{\partial x_{(i)}} \delta_{ij} \right), \tag{25}$$

where δ_{ij} is the Kronecker delta, the modified strain rate tensor $\tilde{\dot{\varepsilon}}_{ij}$, corresponding to (5), is given by

$$\tilde{\dot{\varepsilon}}_{ij} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial u}{\partial z} + \xi\frac{\partial w}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) & \frac{\partial v}{\partial y} & \frac{1}{2}\left(\frac{\partial v}{\partial z} + \xi\frac{\partial w}{\partial y}\right) \\ \frac{1}{2}\left(\frac{\partial u}{\partial z} + \xi\frac{\partial w}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial v}{\partial z} + \xi\frac{\partial w}{\partial y}\right) & -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \end{bmatrix} \tag{26}$$

and, corresponding to (6), the modified viscosity,

$$\tilde{\mu}_n = \eta_0 \left(\tilde{\dot{\varepsilon}}^2 \right)^{(1-n)/2n}, \tag{27}$$

is given in terms of the second invariant, $\tilde{\dot{\varepsilon}}^2 = \tilde{\dot{\varepsilon}}_{ij}\tilde{\dot{\varepsilon}}_{ij}/2$, that in expanded form becomes

$$\tilde{\dot{\varepsilon}}^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \frac{\partial u}{\partial x}\frac{\partial v}{\partial y} + \left(\frac{\partial v}{\partial y}\right)^2 + \frac{1}{4}\left[\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial z} + \xi\frac{\partial w}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial z} + \xi\frac{\partial w}{\partial y}\right)^2 \right]. \tag{28}$$

Since (28) differs from (7) only by the use of the continuity equation (22), the transformation will leave the second invariant $\tilde{\dot{\varepsilon}}^2$ and viscosity $\tilde{\mu}_n$ unchanged, i.e., $\tilde{\dot{\varepsilon}}^2 = \dot{\varepsilon}^2$ and $\tilde{\mu}_n = \mu_n$, and the transformed second invariant remains positive, i.e., $\tilde{\dot{\varepsilon}}^2 \geq 0$.

The dummy variables $\xi, \hat{\xi}$ in (23)-(25) and (26)-(29) are used to identify terms that are neglected in the two types of the Blatter-Pattyn approximation discussed in §3.4. These are (a) the standard Blatter-Pattyn approximation, $\xi = 0, \hat{\xi} = 0$, as originally derived (Blatter, 1995; Pattyn, 2003; DPL, 2010), which solves for just the horizontal velocity components u, v, and (b) the extended Blatter-Pattyn approximation, $\xi = 0, \hat{\xi} = 1$, described more fully later, that contains the standard approximation and also contains additional equations that determine the vertical velocity w and the pressure $\tilde{P}$. Keeping all terms, i.e., $\xi = 1, \hat{\xi} = 1$, specifies the full transformed Stokes model.

Boundary conditions for the transformed equations, corresponding to (8)-(11), are given by

BCs on S_S: $$\tilde{\tau}_{ij} n_j - \hat{\xi} \tilde{P} n_i = 0 \, , \tag{29}$$

BCs on S_{B1}: $$u_i = 0 \, , \tag{30}$$

BCs on S_{B2}: $$u_i n_i = 0 \, , \tag{31}$$

$$\tilde{\tau}_{(i)j} n_j - \tilde{\tau}_n n_{(i)} + \beta(x) u_{(i)} = 0 \, , \tag{32}$$

where $\tilde{\tau}_n = n_i \tilde{\tau}_{ij} n_j$ as before. Equations (31), (32) constitute the three required boundary conditions for frictional sliding.

The transformed system, (25)-(32), and the standard Stokes system, (2)-(11), yield exactly the same solution. However, in common with the Blatter-Pattyn approximation, transformation (21) needs to use a gravity-oriented coordinate system because of the particular form of the gravitational forcing term, while the standard Stokes model does not have this restriction. This is not a major limitation. A somewhat more restrictive limitation is the appearance of $z_s(x, y)$, an implicitly single valued function, to describe the vertical position of the upper surface of the ice sheet. This means that care must be taken in case of reentrant upper surfaces (i.e., S-shaped in 2D) and sloping cliffs at the ice edge, a restriction not present in the standard Stokes model. For simplicity, as noted before we assume that there is just one upper and one basal surface, i.e., as is usual in ice sheet modeling we do not permit overhangs.

3.3 The Transformed Stokes Variational Principle

It is easy to verify that the transformed Stokes system (23)-(32) results from the variation with respect to $u_i, \tilde{P}$ of the following functional:

$$\tilde{\mathcal{A}}[u_i,\tilde{P}] = \int_V dV \left[\frac{4n}{n+1}\eta_0 \left(\tilde{\dot{\varepsilon}}^2\right)^{(1+n)/2n} - \hat{\xi}\tilde{P}\frac{\partial u_i}{\partial x_i} + \rho g u_{(i)} \frac{\partial z_s}{\partial x_{(i)}} \right] + \frac{1}{2}\int_{S_{B2}} dS\, \beta\left(x\right)\left(u_{(i)}u_{(i)} + \left(u_{(i)}n_{(i)}/n_z\right)^2 \right), \tag{33}$$

where $\tilde{\dot{\varepsilon}}^2$ is the transformed second invariant from (28). Basal boundary conditions are imposed by direct substitution, as in (15). Alternatively, one could also impose boundary conditions using Lagrange multipliers as in (13), if desired.

3.4 Two Blatter-Pattyn Approximations

3.4.1 The Standard Blatter-Pattyn Approximation

The standard (or traditional) Blatter-Pattyn approximation (originally introduced by Blatter, 1995; Pattyn, 2003; later by DPL, 2010; Schoof and Hewitt, 2013, and references therein) is obtained by setting $\xi = 0, \hat{\xi} = 0$ in the transformed system. This yields the following Blatter-Pattyn variational functional,

$$\mathcal{A}_{BP}[u_{(i)}] = \int_V dV \left[\frac{4n}{n+1}\eta_0 \left(\dot{\varepsilon}_{BP}^2\right)^{(1+n)/2n} + \rho g u_{(i)} \frac{\partial z_s}{\partial x_{(i)}} \right] + \frac{1}{2}\int_{S_{B2}} dS\, \beta\left(x\right)\left(u_{(i)}u_{(i)} + \varsigma\left(u_{(i)}n_{(i)}/n_z\right)^2 \right), \tag{34}$$

in terms of horizontal velocity components only, where the second invariant $\dot{\varepsilon}_{BP}^2$ follows from (28) with $\xi = 0$,

$$\dot{\varepsilon}_{BP}^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \frac{\partial u}{\partial x}\frac{\partial v}{\partial y} + \left(\frac{\partial v}{\partial y}\right)^2 + \frac{1}{4}\left[\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2 + \frac{\partial u}{\partial z}^2 + \frac{\partial v}{\partial z}^2\right], \tag{35}$$

and therefore the Euler-Lagrange equations and boundary conditions become

$$\frac{\partial \tau_{(i)j}^{BP}}{\partial x_j} - \rho g \frac{\partial z_s}{\partial x_{(i)}} = 0; \quad \begin{cases} \tau_{(i)j}^{BP} n_j + \beta\left(x\right)\left(u_{(i)} + \zeta\left(u_{(j)}n_{(j)}/n_z\right)n_{(i)}/n_z \right) = 0 \\ \text{on } S_{B2}, \quad \tau_{(i)j}^{BP} n_j = 0 \text{ on } S_S, \quad u_{(i)} = 0 \text{ on } S_{B1}, \end{cases} \tag{36}$$

where the Blatter-Pattyn stress tensor $\tau_{(i)j}^{BP}$ is

$$\tau_{(i)j}^{BP}=\eta_0\left(\dot{\varepsilon}_{BP}^2\right)^{(1-n)/2n}\begin{bmatrix} 2\left(2\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right) & \left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right) & \frac{\partial u}{\partial z} \\ \left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right) & 2\left(\frac{\partial u}{\partial x}+2\frac{\partial v}{\partial y}\right) & \frac{\partial v}{\partial z} \end{bmatrix}. \quad (37)$$

These last two equations correspond to (23) and (25) in the transformed Stokes system. There is a new dummy variable ζ in (34) introduced to identify the basal boundary term normally dropped $(\zeta=0)$ in the standard Blatter-Pattyn approximation but restored $(\zeta=1)$ in Dukowicz et al. (2011) to better deal with arbitrary basal topography.

The Blatter-Pattyn model is a well-behaved nonlinear approximate system for the horizontal velocity components u,v because in this case the variational formulation is a convex optimization problem whose solution minimizes the functional. As noted in the Introduction, the Blatter-Pattyn approximation is widely used in practice as an economical and relatively accurate ice sheet model. If desired, the vertical velocity component w may be computed a posteriori by means of the continuity equation.

Remark #1: The original formulation (e.g., Pattyn, 2003) approximates the normal unit vectors n_i on the frictional part of the basal boundary S_{B2} by making the small slope approximation. However, this additional approximation is unnecessary since any computational savings are negligible (Dukowicz et al., 2011; Perego et al., 2012).

3.4.2 The Extended Blatter-Pattyn Approximation

A second form of the Blatter-Pattyn approximation is obtained from the transformed variational principle (33) by making the assumption,

$$\left\|\frac{\partial w}{\partial x}\right\| \ll \left\|\frac{\partial u}{\partial z}\right\|, \quad \left\|\frac{\partial w}{\partial y}\right\| \ll \left\|\frac{\partial v}{\partial z}\right\|, \quad (38)$$

and therefore neglecting $\partial w/\partial x, \partial w/\partial y$ in the transformed second invariant $\tilde{\dot{\varepsilon}}^2$, or equivalently, in the strain rate tensor $\tilde{\dot{\varepsilon}}_{ij}$ from (26), consistent with the original small aspect ratio approximation (Blatter, 1995; Pattyn, 2003; DPL, 2010; Schoof and Hindmarsh, 2008). This corresponds to setting $\xi=0, \hat{\xi}=1$ in the transformed Stokes model. In other words, we neglect vertical velocity gradients but keep the pressure term.

This will be called the extended Blatter-Pattyn approximation (EBP) because, in contrast to the standard Blatter-Pattyn approximation, all the variables, i.e., $u, v, w, \tilde{P}$, are retained. Notably, assumption (38) is equivalent to just setting $w = 0$ in the second invariant $\tilde{\dot{\varepsilon}}^2$ in the full transformed Stokes model. That is, the extended BP approximation is obtained by neglecting vertical velocities everywhere in (33) except where they occurs in the velocity divergence term. This aspect of the transformed Stokes model will be exploited later to obtain approximations that improve on Blatter-Pattyn. Thus, the extended Blatter-Pattyn functional is given by

$$\begin{aligned} \mathcal{A}_{EBP}[u_i, \tilde{P}] &= \int_V dV \left[\frac{4n}{n+1} \eta_0 \left(\dot{\varepsilon}_{BP}^2 \right)^{(1+n)/2n} - \tilde{P} \frac{\partial u_i}{\partial x_i} + \rho g u_{(i)} \frac{\partial z_s}{\partial x_{(i)}} \right] \\ &+ \frac{1}{2} \int_{S_{B2}} dS \, \beta(x) \left(u_{(i)} u_{(i)} + \varsigma \left(u_{(i)} n_{(i)} / n_z \right)^2 \right), \end{aligned} \tag{39}$$

and the Blatter-Pattyn second invariant $\dot{\varepsilon}_{BP}^2$ is given by (35). Taking the variation of the functional, the system of extended Blatter-Pattyn Euler-Lagrange equations and their boundary conditions is given by

(1) Variation with respect to $u_{(i)}$ yields the horizontal momentum equation:

$$\frac{\partial \tau_{(i)j}^{BP}}{\partial x_j} - \frac{\partial \tilde{P}}{\partial x_{(i)}} - \rho g \frac{\partial z_s}{\partial x_{(i)}} = 0; \quad \begin{cases} \tau_{(i)j}^{BP} n_j - \tilde{P} n_{(i)} = 0 \text{ on } S_S, \quad u_{(i)} = 0 \text{ on } S_{B1}, \\ \tau_{(i)j}^{BP} n_j + \beta(x) \left(u_{(i)} + \zeta \left(u_{(k)} n_{(k)} / n_z \right) n_{(i)} / n_z \right) = 0 \\ \qquad \text{on } S_{B2}, \end{cases} \tag{40}$$

where $\tau_{(i)j}^{BP}$ is given by (37).

(2) Variation with respect to w yields the vertical momentum equation:

$$-\frac{\partial \tilde{P}}{\partial z} = 0; \qquad \tilde{P} n_z = 0 \text{ on } S_S, \tag{41}$$

(3) Variation with respect to $\tilde{P}$ yields the continuity equation:

$$\frac{\partial w}{\partial z} + \frac{\partial u_{(i)}}{\partial x_{(i)}} = 0; \qquad w = 0 \text{ on } S_{B1}, \text{ or } w = -u_{(i)} n_{(i)} / n_z \text{ on } S_{B2}. \tag{42}$$

It is apparent that the vertical momentum equation system (41) is decoupled, yielding $\tilde{P} = 0$, as was already shown in §3.1. This eliminates pressure from the horizontal momentum equation (40), making it a decoupled equation for the horizontal velocities

$u_{(i)}$, identical to the standard Blatter-Pattyn system (36). In addition, having obtained the horizontal velocities from (40), the continuity equation (42) may now be solved for the vertical velocity w (but see the comments regarding the discrete case that follow (43)).

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 $\tilde{P}$ and the vertical velocity w that are usually ignored in the standard Blatter-Pattyn approximation where only the horizontal velocity is calculated. 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)). Perhaps the main distinction between the two, which may be imporant in some applications, is that the Blatter-Pattyn system obtains the vertical velocity on the same grid as the horizontal velocities, while in the Blatter-Pattyn model the calculation of vertical velocity is completely decoupled and may be done on an unrelated grid. These models must obtain the vertical velocity w from the continuity equation (42) once horizontal velocities u,v are available. In the continuous case this can be done using the Leibniz's theorem, as follows

$$w(u,v) = w_{z=z_b} - \int_{z_b}^{z} \frac{\partial u_{(i)}}{\partial x_{(i)}} dz' = u_{(i)} \frac{\partial z_b}{\partial x_{(i)}} - \int_{z_b}^{z} \frac{\partial u_{(i)}}{\partial x_{(i)}} dz' = -\frac{\partial}{\partial x_{(i)}} \int_{z_b}^{z} u_{(i)} dz' . \qquad (43)$$

In the discrete case one may consider discretizing (43) directly. However, later we consider special finite element grids where the continuity equation is stably solved for w.

So far we have only considered continuum properties of Stokes-type systems. However, a discrete finite element formulation may not be well behaved. The solution of discretized Stokes models and Blatter-Pattyn approximations, and the solution for vertical velocity from the continuity equation will depend on the choices made for the grids and the finite elements that are to be used. These issues will be discussed next.

4. Finite Element Discretization

4.1 Standard and Transformed Stokes Discretizations

In practice, both traditional Stokes and Blatter-Pattyn models are discretized using finite element methods (e.g., Gagliardini et al., 2013; Perego et al., 2012). We follow this practice except that here the discretization originates from a variational principle. This has a number of advantages (see §2.3 and DPL, 2010). The following is a brief outline of the finite element discretization. Additional details about the grid and the associated discretization are provided in Appendix B. For simplicity, we confine ourselves to two dimensions with coordinates (x,z) and velocities (u,w). Generalization to three dimensions is possible (an example of a three-dimensional grid appropriate for our purpose is discussed in Appendix B). Further, we discuss only the case of direct substitution for basal boundary conditions in the variational functional, i.e., (15) or (33). The remarks in this Section will apply to both the standard and transformed Stokes models; for example, the discrete pressure variable p may refer to either the standard pressure P or the transformed pressure $\tilde{P}$.

Consider an arbitrary grid with a total of $N = n_u + n_w + n_p$ unknown discrete variables at appropriate nodal locations $1 \le i \le N$, with n_u horizontal velocity variables, n_w vertical velocity variables, and n_p pressure variables, so that

$$V = \{V_1, V_2, \cdots, V_N\}^T = \left\{\{u_1, u_2, \cdots, u_{n_u}\}, \{w_1, w_2, \cdots, w_{n_w}\}, \{p_1, p_2, \cdots, p_{n_w}\}\right\}^T = \{u, w, p\}^T \quad (44)$$

is the vector of all the unknown discrete variables that are the degrees of freedom of the model. If using Lagrange multipliers for basal boundary conditions then discrete variables corresponding to λ_z, Λ must be added. Variables are expanded in terms of shape functions $N_i^k(\mathbf{x})$ associated with each nodal variable i in each element k, so that $V^k(\mathbf{x}) = \sum_i V_i N_i^k(\mathbf{x})$ is the spatial variation of all variables in element k, summed over all variable nodes located in element k. Shape functions associated with a given node may differ depending on the variable (i.e., $u, w,$ or p). Substituting into the functional, (15) or (33), integrating and assembling the contributions of all elements, we obtain a discretized variational functional in terms of the nodal variable vectors u, w, p, as follows

$$\mathcal{A}(u,w,p) = \sum_k \mathcal{A}^k(u,w,p), \quad (45)$$

where $\mathcal{A}^k(u,w,p)$ is the local functional evaluated by integrating over element k. Since the term in the functional involving the product of pressure and divergence of velocity is linear in pressure and velocity, and the term responsible for gravity forcing is linear in velocity, the functional (45) may be written in matrix form as follows

$$\mathcal{A}\left(u,w,p\right)=\mathcal{M}\left(u,w\right)+p^{T}\left(M_{UP}^{T}u+M_{WP}^{T}w\right)+u^{T}F_{U}+w^{T}F_{W}\ , \qquad (46)$$

where the notation from (44) has been used, i.e., $u=\left\{u_1,u_2,\cdots,u_{n_u}\right\}^T$, etc. Parentheses indicate a functional dependence on the indicated variables. Comparison with (15) and (33) indicates that $\mathcal{M}\left(u,w\right)$ is a nonlinear positive-definite function of the velocity components u,w, M_{UP}^T, M_{WP}^T are constant $n_p \times n_u$ and $n_p \times n_w$ matrices, respectively, arising from the incompressibility constraint, and F_U, F_W are constant gravity forcing vectors, of dimension n_u and n_w, respectively. Note that $F_U=0, F_W\neq 0$ specifies the standard Stokes model, and $F_U\neq 0, F_W=0$ the transformed Stokes model. The discrete functional $\mathcal{M}\left(u,w\right)$ differs but it remains positive-definite in both.

Discrete variation of the functional corresponds to partial differentiation with respect to each of the discrete variables in V. Thus, the discrete Euler-Lagrange equations that correspond to the u-momentum, w-momentum, and continuity equations, respectively, are given by

$$R\left(u,w,p\right)=\begin{bmatrix} R_U\left(u,w,p\right) \\ R_W\left(u,w,p\right) \\ R_P\left(u,w\right) \end{bmatrix}=\begin{bmatrix} \mathcal{M}_U\left(u,w\right)+M_{UP}p+F_U \\ \mathcal{M}_W\left(u,w\right)+M_{WP}p+F_W \\ M_{UP}^T u+M_{WP}^T w \end{bmatrix}=0\ , \qquad (47)$$

where $R\left(u,w,p\right)$ is the residual vector with components $R_U\left(u,w,p\right)=\partial\mathcal{A}/\partial u$, $R_W\left(u,w,p\right)=\partial\mathcal{A}/\partial w$, and $R_P\left(u,w\right)=\partial\mathcal{A}/\partial p$. The functionals $\mathcal{M}_U\left(u,w\right)=\partial\mathcal{M}/\partial u$, $\mathcal{M}_W\left(u,w\right)=\partial\mathcal{M}/\partial w$ are nonlinear vectors of dimension n_u and n_w, respectively. Altogether, (47) is a set of N equations for the N unknown discrete variables V_i. As explained previously, all boundary conditions are already included in functional (46), and therefore are also incorporated into the discrete Euler-Lagrange equations (47).

Since the overall system (47) is nonlinear, it is typically solved using Newton-Raphson iteration, expressed in matrix notation as follows

$$M\left(u^K,w^K\right)\Delta V+R\left(u^K,w^K,p^K\right)=0, \tag{48}$$

where K is the iteration index, $M\left(u,v\right)=\partial^2\mathcal{A}\left(V\right)/\partial V_i\partial V_j$ is a symmetric $N\times N$ Hessian matrix, and ΔV is the column vector given by

$$\Delta V=\left\{\Delta u,\Delta w,\Delta p\right\}^T=\left\{u^{K+1}-u^K,w^{K+1}-w^K,p^{K+1}-p^K\right\}^T.$$

Given V^K from the previous iteration, (48) is a linear matrix equation that is solved at each iteration for the N new variables V^{K+1}. In view of (46) and (47), the Hessian matrix $M\left(u,w\right)$ may be decomposed into several submatrices, as follows

$$M\left(u,w\right)=\begin{bmatrix} M_{UU}\left(u,w\right) & M_{UW}\left(u,w\right) & M_{UP} \\ M_{UW}^T\left(u,w\right) & M_{WW}\left(u,w\right) & M_{WP} \\ M_{UP}^T & M_{WP}^T & 0 \end{bmatrix}. \tag{49}$$

Submatrices $M_{UW}\left(u,w\right)=\partial^2\mathcal{M}/\partial u\partial w$, etc., depend nonlinearly on u,w. Thus, $M_{UU}\left(u,w\right)$, $M_{WW}\left(u,w\right)$ are square $n_u\times n_u$, $n_w\times n_w$ symmetric matrices, respectively, while $M_{UW}\left(u,w\right)$ is a rectangular $n_u\times n_w$ matrix since n_u, n_w may not be equal. As noted earlier, M_{WP}^T is an $n_p\times n_w$ matrix and therefore not square unless $n_p=n_w$.

4.2 Blatter-Pattyn Discretizations

For completeness, we express the Blatter-Pattyn approximations from §3.4 in matrix form, as follows

(1) The standard Blatter-Pattyn model from §3.4.1 takes the simple form

$$R^{BP}\left(u\right)=\mathcal{M}_U\left(u,0\right)+F_U=0, \tag{50}$$

whose Newton-Raphson iteration is given by

$$M^{BP}\left(u^K\right)\Delta u+R^{BP}\left(u^K\right)=0, \tag{51}$$

and therefore the Blatter-Pattyn Hessian matrix is given by $M^{BP}\left(u\right)=M_{UU}\left(u,0\right)$.

(2) The extended Blatter-Pattyn approximation from §3.4.2 becomes

$$R^{EBP}\left(u,w,p\right)=\begin{bmatrix} \mathcal{M}_U\left(u,0\right)+M_{UP}p+F_U \\ M_{WP}p \\ M_{UP}^T u+M_{WP}^T w \end{bmatrix}=0\,, \tag{52}$$

with the Newton-Raphson iteration given by

$$M^{EBP}\left(u^K\right)\Delta V+R^{EBP}\left(u^K,w^K,p^K\right)=0\,, \tag{53}$$

and the associated Hessian matrix is

$$M^{EBP}\left(u\right)=\begin{bmatrix} M_{UU}\left(u,0\right) & 0 & M_{UP} \\ 0 & 0 & M_{WP} \\ M_{UP}^T & M_{WP}^T & 0 \end{bmatrix}. \tag{54}$$

4.3 Solvability Issues

4.3.1 Solvability of Stokes and Blatter-Pattyn Models

We now consider the solution of the three linear matrix problems (48), (51), (53) associated with the Stokes and the corresponding Blatter-Pattyn approximate models. While there are no issues in the continuous case, this is not so in the discrete case depending on the choice of the grid and the finite elements, as noted earlier. The discrete system to be solved has the general form

$$\mathbf{M}\begin{Bmatrix} \mathbf{u} \\ p \end{Bmatrix}=\begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix}\begin{Bmatrix} \mathbf{u} \\ p \end{Bmatrix}=\begin{Bmatrix} \mathbf{f} \\ g \end{Bmatrix}, \tag{55}$$

where $\mathbf{u}=\left\{u,w\right\}^T$ in the linear case or $\left\{\Delta u,\Delta w\right\}^T$ in the nonlinear case, and similarly for the vector of pressures or pressure increments p. The form (55) is characteristic of Stokes-type problems, or more generally of constrained minimization problems using Lagrange multipliers. In finite element terminology these are called "mixed" or "saddle point" problems, meaning that velocity components and the pressure occupy different finite element spaces, and that the solution of (55) is actually at the saddle point with respect to the velocity and pressure variables of the quadratic form associated with (55). The matrix $\mathbf{M}$ is symmetric but indefinite, with both positive and negative eigenvalues. As a result, the matrix inverse may not be bounded and may lack stability.

There are three cases to consider:

(1) The standard Blatter-Pattyn model, (51). In this case only the matrix A exists, it is elliptic, and $B = B^T = 0$. As a result, the standard Blatter-Pattyn model is a well-behaved and stable unconstrained minimization problem. The model (51) is self-contained and is solved for u while the vertical velocity w is potentially available a posteriori from a separately obtained continuity equation.

(2) The extended Blatter-Pattyn model, (53), (54). The middle row of the Hessian (54) indicates that the solution for the pressure will be zero. Using this in the top row of the Hessian, one obtains the standard Blatter-Pattyn system and therefore the same well-behaved horizontal velocity u as above, with the result that the bottom row of the Hessian, the continuity equation, is the only way to obtain a solution for the vertical velocity w. However, this is possible only if matrix M_{WP}^T is invertible, which at minimum requires a square matrix, i.e., $n_p = n_w$, and this depends on the finite element grid chosen for the discretization. For example, the popular second-order Taylor-Hood (P2-P1) element with piecewise quadratic velocity and linear pressure (Hood and Taylor, 1973) typically has $n_p \ll n_w$. As a result, the linear system for w is greatly underdetermined and cannot be solved for w. In fact, this is a problem for all inf-sup stable elements with $n_p \neq n_w$, such as the Taylor-Hood element, for example.

(3) The standard and transformed Stokes models, (48), (49). These models require the use of pressure as a Lagrange multiplier to enforce incompressibility and therefore these are mixed or saddle point problems, as mentioned previously. To avoid problems with the solution these finite elements must satisfy a certain condition, the Ladyzhenskaya–Babuška–Brezzi (LBB, or inf-sup) condition. There is a very large literature on the subject, e.g., Boffi et al. (2008), Elman et al. (2014), Auricchio et al. (2017). Both the standard and transformed Stokes models are subject to this problem and in general must use inf-sup stable finite elements. Testing for stability is not trivial. However, collections of inf-sup stable elements for the Stokes equations may be found in many papers and books on mixed methods, e.g., Boffi et al. (2008). The popular second-order Taylor-Hood P2-P1 element (Hood and Taylor, 1973) is an example of an inf-sup stable element. Some results involving this element are shown in Fig. 13 for Test B, one of the test problems described in Appendix A.

4.3.2 A Special Case: Invertible Continuity Equation

In the continuous case, the Blatter-Pattyn approximation (§3.4.1) implies that vertical velocity w is obtainable from the continuity equation after having solved for the horizontal velocities u,v. As mentioned previously, this is possible to do in the continuum but not necessarily so in the discrete case. The 2D discrete continuity equation from (47) or (52) is given by

$$M_{UP}^T u + M_{WP}^T w = 0 \,. \tag{56}$$

For this to be solvable for w in terms of the horizontal velocity, i.e., $w(u)$, means that matrix M_{WP}^T must be invertible and thus it must be square and full rank. Since in general M_{WP}^T is an $n_p \times n_w$ matrix, for solvability at minimum this requires that

$$n_p = n_w \,. \tag{57}$$

Assuming that we are dealing with reasonable discretizations, we shall presume for our discussion that matrix M_{WP}^T is always full rank. If Lagrange multipliers were to be used, this means that the number of unknown pressures n_p would have to be augmented by the number of Lagrange multipliers so (57) would become $n_p + \lambda_z + \Lambda = n_w$ (See Appendix B, §B2, for more details). We shall refer to (57) (together with the assumption of full rank) as the solvability condition. In Appendix B we present several grids and elements that satisfy this condition, including one variant in particular, the P1-E0 element, that will be used in most of the 2D test problems featured in this paper. Thus, if the solvability condition is satisfied, the discrete continuity equation (56) may be inverted for the vertical velocity, to obtain

$$w(u) = -M_{WP}^{-T} M_{UP}^T u \,, \tag{58}$$

where matrix M_{WP}^{-T} is defined by

$$M_{WP}^{-T} = \left(M_{WP}^T\right)^{-1} = \left(M_{WP}^{-1}\right)^T . \tag{59}$$

Here wee have used the fact that if matrix M_{WP}^T is invertible then so is its transpose M_{WP}. Note that (58) is one discrete form of equation (43).

Invertibility of the continuity equation has several important applications. First, it is a necessary requirement for the new Stokes approximations that are discussed in §6.2.

Since these approximations are based on approximating the vertical velocity in the transformed second invariant, (28), it is necessary to obtain the vertical velocity independently of solving the entire coupled Stokes problem. Second, we noticed earlier that the extended Blatter-Pattyn model does not work with a Taylor-Hood P2-P1 grid because the solvability condition is not satisfied. However, this model does work with a variant of the Taylor-Hood grid, the P2-E1 grid, illustrated in Fig. 13A, which does satisfy the solvability condition and this therefore allows for a successful calculation of the vertical velocity.

Perhaps the main reason for the importance of the solvability condition is that it implies that the Stokes variational principle, (15) or (33), may be transformed into and therefore that it is equivalent to an optimization or minimization problem. Consider the discrete form of the variational functional given by (46). Working with a grid that satisfies the solvability condition, we may substitute the vertical velocity given by (58) into the functional (46). This immediately eliminates the term responsible for the continuity equation, including the pressure, and one obtains a functional in terms of horizontal velocity u only, as follows

$$\mathcal{A}\left(u\right)=\mathcal{M}\left(u,w\left(u\right)\right)+u^{T}F_{U}+w\left(u\right)^{T}F_{W}\,. \tag{60}$$

Since the functional $\mathcal{M}\left(u,w\left(u\right)\right)$ is positive semi-definite, this is now an unconstrained minimization problem, entirely analogous to the standard Blatter-Pattyn problem of §3.4.1 except that here it represents the full Stokes problem for either the standard (F_W) or the transformed (F_U) formulation. This result suggests that a conventional Stokes problem, when solved on a grid satisfying the solvability condition, is equivalent to an unconstrained minimization problem and therefore is well behaved. This is because any problem will give the same answer whether formulated as (46) or (60) on a grid that satisfies the solvability condition.

Note that functional (60) is actually the discrete version of a pressure-free formulation that was attempted analytically by Dukowicz (2012). It is possible to consider solving problems in practice using the pressure-free formulation (60) instead of a standard saddle point formulation such as (46) or (47). However, this produces a dense Hessian matrix that makes a solution using Newton-Raphson iteration very costly and therefore impractical, particularly for large problems.

5. Comparison of the Standard and Transformed Stokes Models

The standard and transformed Stokes models are expected to converge to the same solution. To verify that this is indeed the case we do a number of calculations for some 2D test problems based on the ISMIP-HOM benchmark (Pattyn et al, 2008). These tests are described in Appendix A where they are referred to as Test B and Test D*. Test B involves no-slip boundary conditions on a sinusoidal bed, and Test D* evaluates sliding of the ice sheet along a flat bed in the presence of sinusoidal friction. The tests are discretized using P1-E0 elements on a regular grid composed of n quadrilaterals in the x-direction and m quadrilaterals in the z-direction, illustrated in Fig. B1, with each quadrilateral divided into two triangles. Results are presented for two domain lengths, $L = 5$ km and 10 km, to test the aspect ratio range where the Blatter-Pattyn model begins to fail, and using a relatively coarse grid, i.e., $m = n = 40$, except when we consider the convergence of the models with grid refinement in Fig. 3.

Fig. 3 evaluates the convergence of the two Stokes models as a function of grid resolution r, where r is the number of quadrilaterals in either direction. The models do converge to the same solution and convergence is second order as expected from the use of linear elements. Interestingly, the transformed Stokes model displays considerably smaller error at all resolutions in Test B. As a result, we observe that standard Stokes calculations are not fully converged even at the 40x40 resolution.

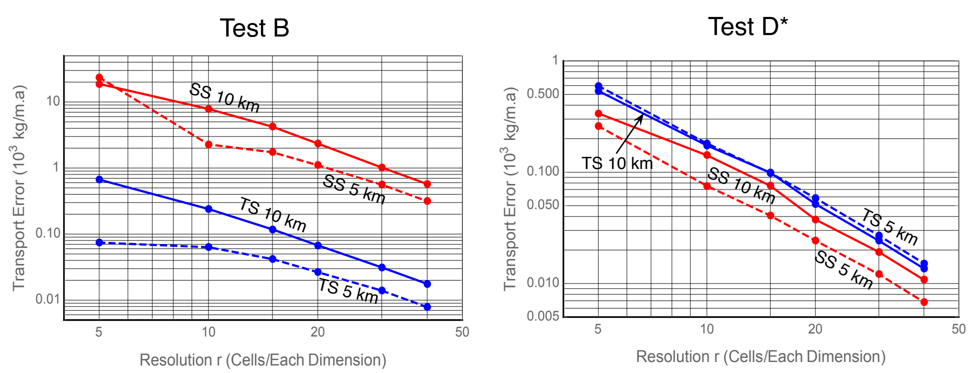

Figure 3. Convergence of ice transport in Tests B and D* with grid refinement. Transformed Stokes (TS) plots are in blue and standard Stokes plots (SS) are in red.

Fig. 4 shows the vertical profiles of the horizontal velocity u at outflow, $x = L$. We plot results from the no-slip Test B problem and the two frictional sliding problems, Tests D and D*. The Test D profile from the ISMIP-HOM benchmark is almost vertically constant, indicating that the value for basal friction originally chosen is too small. This is what motivated the change from Test D to Test D* in Appendix A.

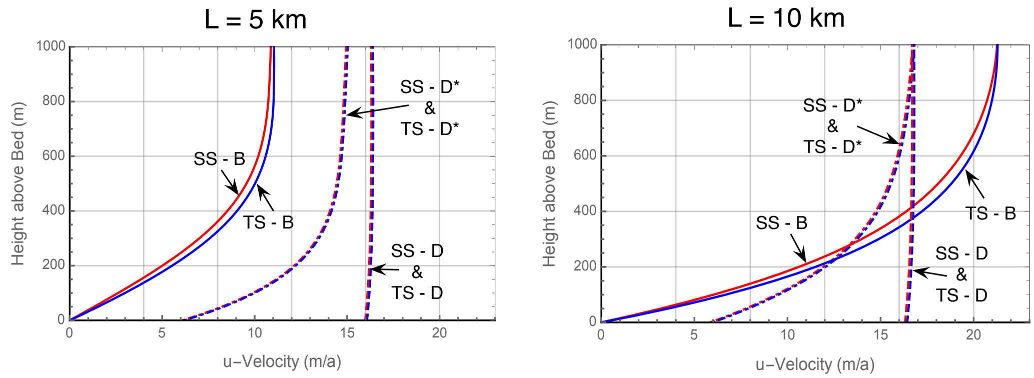

Figure 4. The u-velocity profile at location $x = L$ as a function of height from the bed.

Figs. 5 and 6 show the u-velocity at the ice sheet upper surface for Tests B and D*. This is the benchmark used in ISMIP-HOM to compare the different ice sheet models. Here we compare four cases: the standard Stokes model (SS), the transformed Stokes model (TS), the Blatter-Pattyn (BP) model, and for reference, a very high resolution full-Stokes Test B calculation "oga1" (SS-HR), available from the ISMIP-HOM paper and also independently available in Gagliardini and Zwinger (2008). The TS and the SS-HR plots lie on top of one another (they have been offset slightly for clarity), indicating that the TS model is already fully converged. We again observe that the SS model is not yet converged in Test B, particularly at $L = 5$ km. As also seen in the ISMIP-HOM paper, the Blatter-Pattyn calculation (BP) shows large deviations from the Stokes results, especially so at $L = 5$ km where the surface velocity is entirely out of phase with the Stokes results. Test D* results in Fig. 6 for the SS and TS models are very similar (the SS plot has been slightly offset upward for visibility). As expected, the error in the Blatter-Pattyn results is noticeable at $L = 10$ km and very large at $L = 5$ km.

Pressure results are not shown because, particularly in the transformed case, pressure has little or no physical significance. However, pressures calculated on the P1-E0 grid are particularly smooth and well behaved.

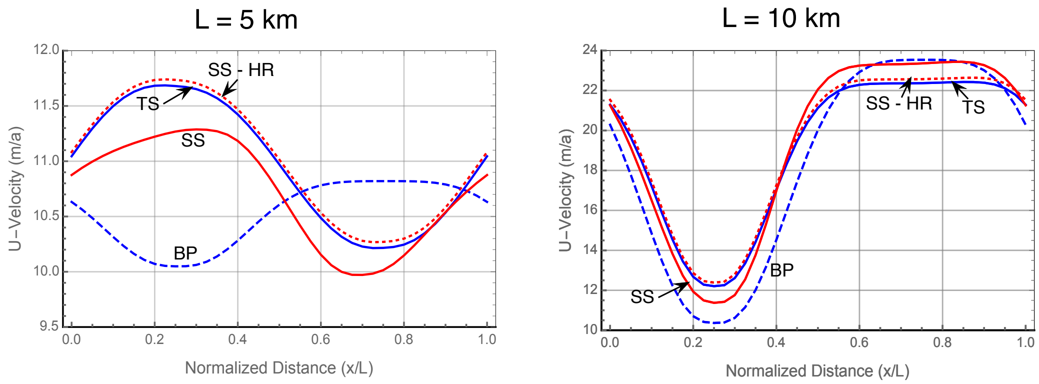

Figure 5. Upper surface u-velocity, $u(x,z_s)$ - Test B, No-slip boundary conditions.

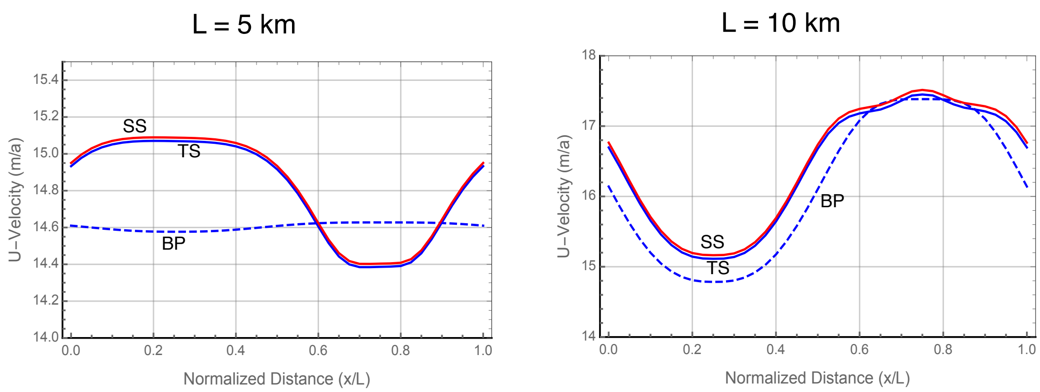

Figure 6. Upper surface u-velocity, $u(x,z_s)$ - Test D*, Modified frictional sliding case.

6. Applications of the Transformed Stokes Model

6.1 Adaptive Switching between Stokes and Blatter-Pattyn Models

One way of reducing the cost of a full Stokes calculation is to use it adaptively with a cheaper approximate model. That is, one may use the cheaper model in those parts of a problem where it is accurate, and the more expensive full Stokes model where the approximate model loses accuracy. One example of such an adaptive approach is the tiling method by Seroussi et al. (2012). However, there are drawbacks to such methods, such as the difficulty of incorporating two or more presumably quite different models into a single model, and the additional complexity of a necessary transition zone to couple the disparate models.

The transformed Stokes model used in such an adaptive role is attractive because it may be switched between the Stokes and Blatter-Pattyn cases simply by switching the parameter $\xi \in \{0,1\}$ between its two values. For simplicity the extended Blatter-Pattyn approximation ($\xi = 0, \hat{\xi} = 1$) is being used since both the Stokes and the Blatter-Pattyn parts of the code have the same number of discrete variables. The extended Blatter-Pattyn model requires the use of a grid that satisfies the solvability condition as explained in §4.3. We therefore use the P1-E0 element. However, it would be computationally cheaper to use the standard Blatter-Pattyn approximation ($\xi, \hat{\xi} = 0$) instead, solving only for the horizontal variables and coupling to the Stokes model with $p = 0$ and $w = w(u,v)$ at the interface. This, however, implies much more complicated programming.

To demonstrate the idea of adaptive switching with a transformed Stokes model, we introduce a new test problem, Test O, described in Appendix A and illustrated in Fig. A1. This consists of an inclined ice slab whose movement is obstructed by a thin obstacle protruding 20% of the ice depth up from the bed. No-slip boundary conditions are applied along the bed and on the obstacle itself. Because of the localized nature of the obstacle, the Blatter-Pattyn approximation conditions, (38), must fail near the obstacle and therefore the full Stokes model is needed for good accuracy, at least locally.

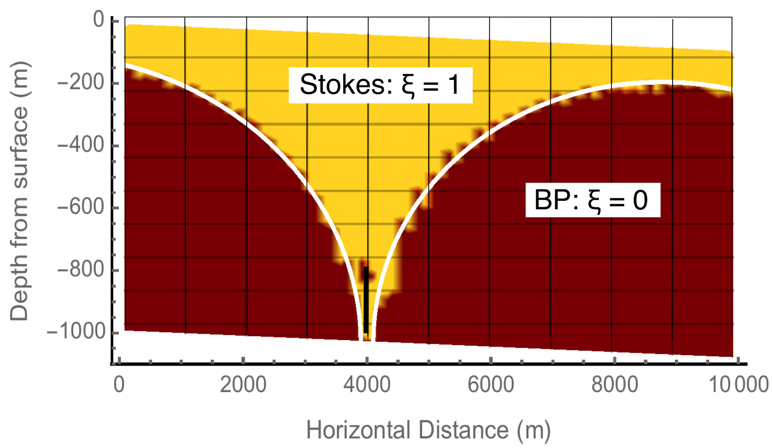

Figure 7. Mask function (white curve, $z = F_M(x)$) to indicate where the Stokes and BP models are activated in the 20% obstacle test problem. The dark brown region delineates the region where $|\partial w/\partial x| \le 0.1|\partial u/\partial z|$ in a Blatter-Pattyn calculation.

To implement this, we first use a Blatter-Pattyn calculation to outline regions where $|\partial w/\partial x| \le 0.1|\partial u/\partial z|$, approximately localizing where the Blatter-Pattyn approximation is valid. This determines a mask function $z = F_M(x)$, illustrated in Fig. 7 by the white curves, that specifies where the two models must be used. Defining the centroid of a triangular element by (x_C, z_C), the code makes a selection in each element:

$$
\begin{aligned}
z_C \le F_M(x_C) \quad &\Rightarrow \quad \text{Set } \xi = 0, \quad \text{i.e., the Blatter-Pattyn region,} \\
z_C > F_M(x_C) \quad &\Rightarrow \quad \text{Set } \xi = 1, \quad \text{i.e., the Stokes region.}
\end{aligned}
$$

Somewhat counterintuitively, the Stokes region occupies the upper part of the domain in Fig. 7 and includes the obstacle, while the Blatter-Pattyn region occupies much of the bottom part of the domain. A transition zone, e.g., $0 \le \xi(x,z) \le 1$, is possible but was not used in the present calculation.

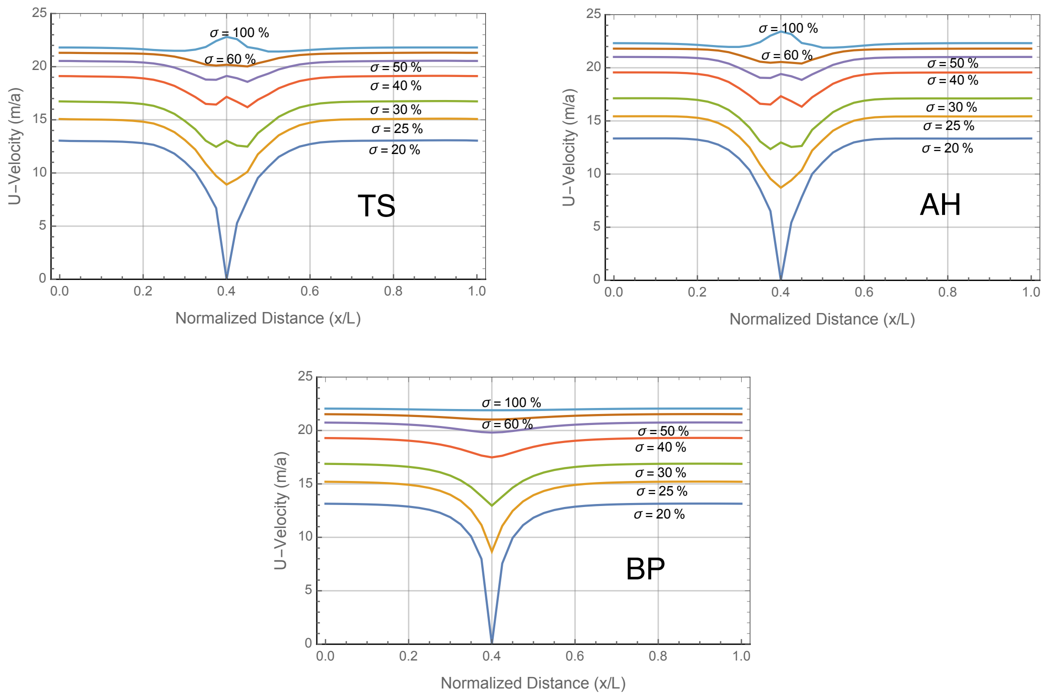

Figure 8. Comparing results for the Transformed Stokes (TS), the Adaptive-Hybrid (AH), and the Blatter-Pattyn (BP) models in Test O.

The Adaptive-Hybrid results are shown in Fig. 8, which shows curves of the horizontal velocity u at seven different vertical positions specified as a percentage of the distance between top and bottom, with $\sigma = 100\%$ at the top surface. The top right panel shows the results for the adaptive-hybrid model (AH). For comparison, the top left panel and the bottom panel show results for the Stokes (TS) and the Blatter-Pattyn (BP) calculations, respectively. All calculations are at the 40x40 resolution. The Adaptive-Hybrid results are very similar to the full Stokes results, reproducing most features of the velocity profiles, including the velocity bump at the top surface, indicating that even the top surface feels the presence of the obstacle. The Blatter-Pattyn results are much less accurate; they completely miss the details of the flow near the obstacle. We also measure the RMS error in the u-velocity relative to the Stokes results. The RMS error in the Blatter-Pattyn case is 0.493 m/a and 0.440 m/a in the Adaptive-Hybrid case, smaller in the Blatter-Pattyn case as expected, but the difference is not as large and striking as the visual difference in Fig. 8. Nevertheless, the adaptive-hybrid method is successful judged by the Fig. 8 results alone. Unfortunately, an estimate of the computational cost savings will have to wait a more realistic implementation.

6.2. Two Stokes Approximations Beyond Blatter-Pattyn

As shown in §3.4, simply setting $w = 0$ in the second invariant $\tilde{\dot{\varepsilon}}^2$ in the transformed functional $\tilde{\mathcal{A}}$ results in the standard Blatter-Pattyn approximation. This suggests that approximating the vertical velocity w in the functional would be a good way to create approximations that improve on the Blatter-Pattyn approximation since $w = 0$ already produces an excellent approximation. We will look at two such methods in this Section although others are possible. The first method, to be called the BP+ approximation, is implemented using a combination of Newton and Picard iterations such that at each Newton iteration the pressure-free variational functional is evaluated using a lagged vertical velocity w^K from the previous iteration. The pressure is used in a subordinate role as a "test function" to obtain a decoupled invertible continuity equation to obtain $w\left(u^K\right)$. Although this method improves on the accuracy of the Blatter-Pattyn approximation, its overall accuracy is limited because it uses only the horizontal momentum equation and neglects the vertical momentum equation. The second method, to be called the Dual-Grid approximation, keeps the pressure and vertical velocity as in the transformed Stokes model but approximates it by discretizing the continuity equation on a coarser grid. Since vertical velocity w is determined by inverting the continuity equation, this has the effect of approximating the vertical velocity while reducing the number of pressure and vertical velocity variables. This preserves the structure of the Stokes model, while the degree of approximation is determined by the amount of coarsening of the continuity grid.

6.2.1 An Improved Blatter-Pattyn or BP+ Approximation

To prepare, we introduce a pair of 2D variational quasi-functionals, $\tilde{\mathcal{A}}_{PS1}[u]$ and $\tilde{\mathcal{A}}_{PS2}[\tilde{P}]$. Noting that $\tilde{P}=0$ in the Blatter-Pattyn approximation, we drop the pressure term from the transformed functional (33) and define a new pressure-free functional,

$$\tilde{\mathcal{A}}_{PS1}[u]=\int_V dV\left[\frac{4n}{n+1}\eta_0\left(\tilde{\dot{\varepsilon}}^2\right)^{(1+n)/2n}+\rho g u\frac{\partial z_s}{\partial x}\right] + \frac{1}{2}\int_{S_{B2}} dS\,\beta(x)\Big(u^2+\zeta\left(u\,n_x/n_z\right)^2\Big), \tag{61}$$

where

$$\tilde{\dot{\varepsilon}}^2=\left(\frac{\partial u}{\partial x}\right)^2+\frac{1}{4}\left(\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}\right)^2. \tag{62}$$

Since the continuity equation has been eliminated from (61), incompressibility is introduced separately by defining a second functional,

$$\tilde{\mathcal{A}}_{PS2}[p]=\int_V dV\; p\left(\frac{\partial u}{\partial x}+\frac{\partial w}{\partial z}\right). \tag{63}$$

Note that functional (61) is a function of u,w but variation is to be taken only with respect to u, and similarly, functional (63) is a function of u,w,p but variation is taken only with respect to p. Direct substitution is used for boundary conditions, as before. Here we are effectively viewing the pressure p as a “test function” in the finite element sense. This gives us great flexibility to create elements that satisfy the solvability condition (57) as desired. In a triangulation, for example, pressures may be assigned to every two triangles, as in a P1-E0 grid, while others may be assigned to a single triangle so as to achieve an equal number of pressure and vertical velocity variables.

The variation of $\tilde{\mathcal{A}}_{PS1}[u]$ with respect to u, results in a set of n_u discrete Euler-Lagrange equations,

$$\hat{R}_U\left(u,w\right)=\frac{\partial\tilde{\mathcal{A}}_{PS1}\left(u,w\right)}{\partial u}=M_U\left(u,w\right)+F_U=0. \tag{64}$$

This may be recognized as the Blatter-Pattyn model, (50), when w is set to zero. The discrete variation of $\tilde{\mathcal{A}}_{PS2}[p]$ with respect to p, results in the continuity equation, (56),

$$\hat{R}_P(u,w) = \frac{\partial \tilde{\mathcal{A}}_{PS2}(u,w,p)}{\partial p} = M_{UP}^T u + M_{WP}^T w = 0 \,. \tag{65}$$

These two systems are now combined to form the BP+ approximation, as follows

$$\hat{R}(u,w) = \left[\hat{R}_U(u,w), \hat{R}_P(u,w) \right]^T = 0 \,. \tag{66}$$

This is a single system of $n_u + n_p$ equations to determine the $n_u + n_w$ discrete variables u, w, implying that (66) is viable only on grids satisfying the solvability condition, $n_p = n_w$. Just as in the standard Blatter-Pattyn approximation in §3.4.1, the vertical momentum equation is missing, but instead of neglecting w, the vertical velocity is now consistently obtained from the continuity equation.

The continuum version of the discrete Euler-Lagrange system (64), (65) may be written as follows

$$\begin{gathered} \frac{\partial}{\partial x}\left(4\tilde{\mu}\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial z}\left(\tilde{\mu}\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right)\right) - \rho g \frac{\partial z_s}{\partial x} = 0, \\ \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0, \end{gathered} \tag{67}$$

whose the boundary conditions are

$$\begin{gathered} 4\tilde{\mu}\frac{\partial u}{\partial x} n_x + \tilde{\mu}\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right) n_z = 0 \;\text{ on }\; S_S, \quad u = w = 0 \;\text{ on }\; S_{B1}, \\ \left.\begin{aligned} 4\tilde{\mu}\frac{\partial u}{\partial x} n_x + \tilde{\mu}\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right) n_z + \beta(x)\left(u + \zeta\, u\left(n_x/n_z\right)^2\right) = 0, \\ w = -u\, n_x/n_z, \end{aligned}\right\} \;\text{ on }\; S_{B2}, \end{gathered} \tag{68}$$

where $\tilde{\mu} = \eta_0\left(\tilde{\dot{\varepsilon}}^2\right)^{(1-n)/2n}$ and the second invariant $\tilde{\dot{\varepsilon}}^2$ is given by (62). Remarkably, a model exactly equivalent to (67), i.e., the BP+ approximation, was introduced by Herterich (1987) to study the transition zone between an ice sheet and an ice shelf[1]. This predates the less accurate, widely used Blatter-Pattyn model by some eight years. Unfortunately, this anticipatory work seems to have faded into obscurity.

There are two ways of solving the BP+ system (66), as follows

[1] Reference pointed out to me by C. Schoof.

(1) BP+, Quasi-variational, Newton iteration version:

Although a single variational principle does not exist in this case, it is still possible to make use of Newton-Raphson iteration to obtain second order convergence. To do this, we treat (66) as a single multidimensional nonlinear system and solve it using Newton-Raphson iteration, as follows

$$\begin{bmatrix} M_{UU}\left(u^K,w^K\right) & M_{UW}\left(u^K,w^K\right) \\ M_{UP}^T & M_{WP}^T \end{bmatrix}\begin{bmatrix} \Delta u \\ \Delta w \end{bmatrix}+\begin{bmatrix} \hat{R}_U\left(u^K,w^K\right) \\ \hat{R}_P\left(u^K,w^K\right) \end{bmatrix}=0\,, \qquad (69)$$

where $M_{UU}\left(u,w\right)=\partial\hat{R}_U\left(u,w\right)/\partial u$ and $M_{UW}\left(u,w\right)=\partial\hat{R}_U\left(u,w\right)/\partial w$ are the same matrices as appear in (49). Convergence is rapid (quadratic) once in the basin of attraction but each step is more expensive than the Picard iteration described next.

(2) BP+, Newton/Picard iteration version:

A single step of the Newton-Raphson system (69) may be written as follows

$$\begin{gathered} M_{UU}\left(u^K,w^K\right)\Delta u+M_{UW}\left(u^K,w^K\right)\Delta w+\hat{R}_U\left(u^K,w^K\right)=0, \\ M_{UP}^T u^{K+1}+M_{WP}^T w^{K+1}=0. \end{gathered} \qquad (70)$$

If we lag the vertical velocity, i.e., $w^{K+1}=w^K \Rightarrow \Delta w=0$ in the first equation, we obtain a Picard iteration algorithm as follows

$$\begin{aligned} &\text{Starting from } K=0,\ \text{choose an initial guess, } u^0\neq 0, w\left(u^0\right), \\ &\quad \text{Solve:} \\ &\qquad M_{UU}\left(u^K,w^K\right)\Delta u+\hat{R}_U\left(u^K,w^K\right)=0, \\ &\qquad u^{K+1}=u^K+\Delta u, \\ &\qquad w^{K+1}=w\left(u^{K+1}\right)=-M_{WP}^{-T}M_{UP}^T u^{K+1}, \\ &\qquad K=K+1, \\ &\quad \text{Repeat until convergence.} \end{aligned} \qquad (71)$$

Each step of this iteration is inexpensive since it is equivalent to a step of the standard Blatter-Pattyn model, (36). On the other hand, Picard iterations typically converge only linearly. It remains to be seen which version is preferable in practice.

Both BP+ versions converge to the same solution. Fig. 9 compares the upper surface u-velocity from the improved Blatter-Pattyn (BP+) approximation to the standard Blatter-Pattyn approximation and to a reference exact Stokes calculation. The RMS u-Error of the BP+ approximation relative to the exact Stokes case is shown in Fig. 12. The

BP+ approximation is noticeably more accurate than the BP approximation, particularly in the $L = 5$ km case where the Blatter-Pattyn solution bears no resemblance to the correct solution while the BP+ approximation shows excellent accuracy. This is confirmed by the RMS u-Error results in Fig. 12 where BP+ is two to three times as accurate as BP.

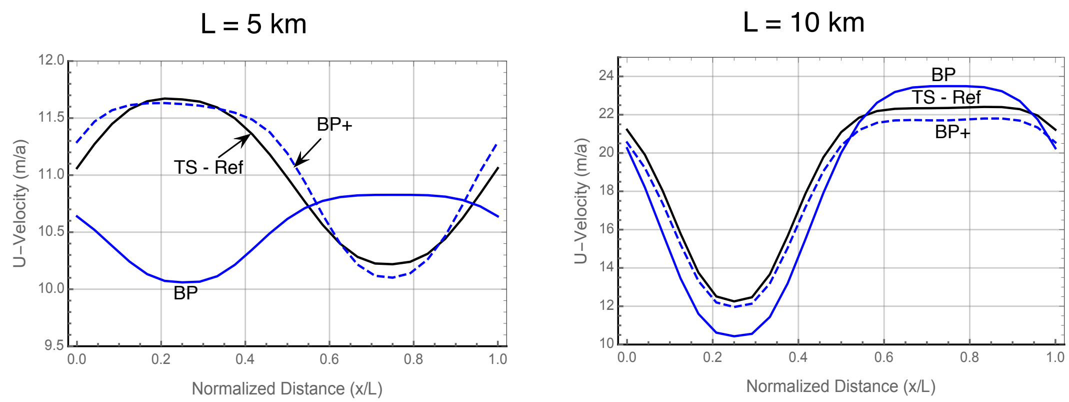

Figure 9. Comparing Approximations. Test B, Upper surface u-velocity. TS-Ref: Transformed Stokes; BP: Blatter-Pattyn; BP+: Improved Blatter-Pattyn. Resolution: 24x24.

These two versions depend on an invertible continuity equation to obtain $w = w(u)$. However, vertical velocity w may already be available for the purpose of temperature advection in production code packages that either incorporate or are based on the Blatter-Pattyn approximation. Thus, the BP+ approximation, and particularly the Newton/Picard version, may be attractive for use in such codes since they improve the accuracy of the basic Blatter-Pattyn model, as seen in Fig. 9, at little additional cost.

6.2.2 A "Dual-Grid" Transformed Stokes Approximation

Here we take a different approach and approximate the continuity equation in the transformed Stokes model, which indirectly approximates w. Thus, the continuity equation is discretized on a grid coarser than the one used for the momentum equations, and then interpolate the vertical velocity to appropriate locations on the finer grid. This reduces the number of unknown variables in the problem, making it cheaper to solve but hopefully without much loss of accuracy. As described in Appendix A, our test problem grids are logically rectangular, divided into n cells horizontally and m cells vertically. The coarse grid is constructed by dividing the fine grid into s equal segments in each direction. This assumes that the integers n and m are each divisible by s, such that there are nm/s^2 coarse cells in total, with each coarse cell containing s^2 fine cells. The primary grid (i.e., the fine grid) was chosen to have $n = m = 24$, resulting in a reference 24×24 fine grid, so as to maximize the number of different coarse grids that may be used for this test. Coarse grids were constructed using $s = 2,3,4,6$, and this resulted in fine/coarse grid combinations labeled by $24 \times 12, 24 \times 8, 24 \times 6, 24 \times 4$, respectively. Similar to a P1-E0 fine grid, coarse grid vertical velocities w are located at vertices and pressures at vertical edges. Fig. 10 illustrates the case of a single coarse and four fine quadrilateral cells for a grid fragment with $n = m = 2$ and $s = 1$. For the Test B problem, using direct substitution for basal boundary conditions, there will be nm u-variables and nm/s^2 w- and p-variables each, for a total of $nm\left(1+2/s^2\right)$ unknown variables, considerably fewer than the $3nm$ variables in the full resolution (i.e., fine grid) case, depending on the value of s. The coarse grid terms in the functional that are affected, $\tilde{P}\left(\partial u/\partial x + \partial w/\partial z\right)$ and $\partial w/\partial x$, are computed using coarse grid variables and interpolated to the fine grid. We will consider two versions depending on how the coarse grid terms are calculated and distributed on the fine grid.

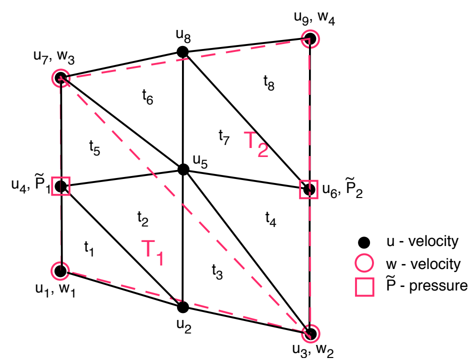

Figure 10. A Sample of a Coarse/Fine P1-E0 Grid for the Dual-Grid Approximation. Resolution: $n = m = 2, s = 1$. Coarse grid is in red, fine grid in black.

(1) <u>Approximation A, Bilinear interpolation</u>:

Referring to Fig. 10, the four velocities at the vertices of the coarse grid quadrilateral, i.e., u_1, u_3, u_7, u_9 and w_1, w_2, w_3, w_4, are used to obtain u, w at the remaining five vertices of the fine grid by means of bilinear interpolation. Thus, the five velocities u_2, u_4, u_5, u_6, u_8 are obtained in terms of vertex velocities u_1, u_3, u_7, u_9, and similarly for the w velocities. The resulting complete set of fine grid variables, interpolated from coarse grid variables, are used calculate the divergence $D=\left(\partial u/\partial x+\partial w/\partial z\right)$ and the quantity $\partial w/\partial x$ in each of the eight triangular elements $t_1,t_2,\cdots,t_8$ of the fine grid. Coarse grid pressures $\tilde{P}_1,\tilde{P}_2$ are associated with the coarse grid triangles T_1,T_2. The products $\tilde{P}_1 D$ in elements t_1,t_2,t_3,t_5 and $\tilde{P}_2 D$ in elements t_4,t_6,t_7,t_8 are then accumulated over the entire grid to obtain $\tilde{P}\left(\partial u/\partial x+\partial w/\partial z\right)$ for use in the transformed functional $\tilde{\mathcal{A}}$. Similarly, the quantity $\partial w/\partial x$ is computed in the fine grid elements from coarse grid variables for use in the second invariant $\tilde{\varepsilon}^2$.

(2) Approximation B, Linear interpolation:

In this version, the three velocities at the vertices of the two coarse grid triangles T_1 and T_2, i.e., u_1,u_3,u_7 and w_1,w_2,w_3 in T_1, and u_7,u_3,u_9 and w_3,w_2,w_4 in T_2, approximate the divergence $D=\left(\partial u/\partial x+\partial w/\partial z\right)$ and the quantity $\partial w/\partial x$ as constant values in the two coarse triangles. The constant quantities $\tilde{P}_1 D$, $\tilde{P}_2 D$ are then accumulated over the entire grid. The constant quantity $\partial w/\partial x$ in each coarse triangle is then distributed to each of the eight fine grid elements $t_1,t_2,\cdots,t_8$ depending on whether the centroid of the fine triangular element is in T_1 or T_2. As in the previous case, this is then used in the second invariant $\tilde{\varepsilon}^2$ when evaluating the transformed functional $\tilde{\mathcal{A}}$.

While the number and type of unknown variables is the same in the two versions, they differ considerably in accuracy, as is seen in Figs. 11 and 12. Fig. 11 compares the upper surface u-velocity in both version, Approximations A and B, for the four coarse grid combinations and the reference 24x24 fine grid calculation. Fig. 12 compares the overall accuracy the same way by means of the RMS u-Error. As might be expected, the accuracy of Approx. A is better than the accuracy of Approx. B, particularly in the case when $L=10$ km. Both versions are more accurate than the Blatter-Pattyn and BP+ approximations, except at the lowest 24x4 resolution when only the Approx. A version retains that distinction.

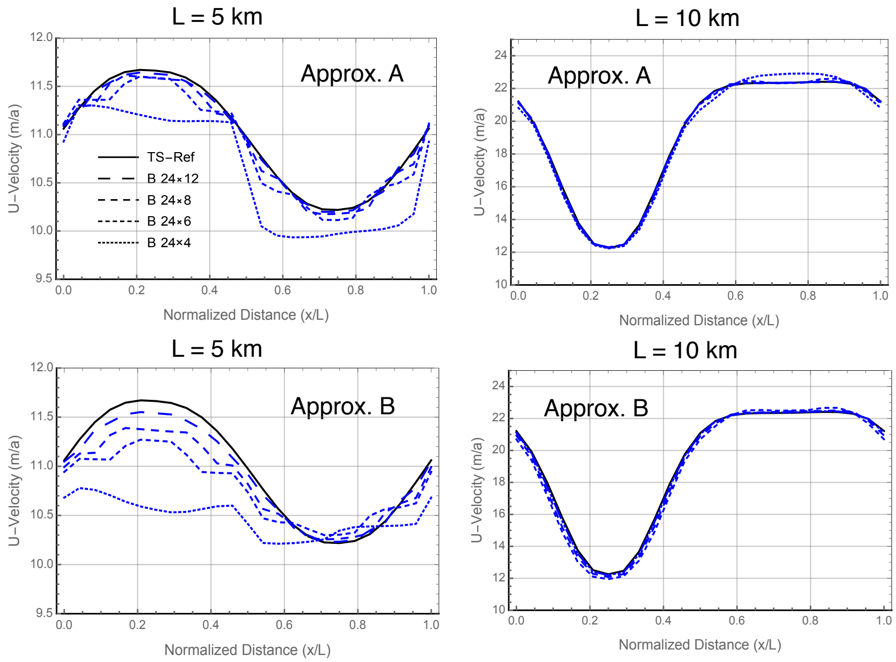

Figure 11. Comparing Approximations A and B. Test B. Upper surface u-velocity. TS-Ref: Reference Stokes 24x24; Fine/Coarse resolutions (r x R): 24xR, R=12, 8, 6, 4.

In summary, the dual-grid approximation improves on the Blatter-Pattyn approximation in both versions and at all resolutions, as seen in Fig. 12. Compared to the BP+ approximations, here the vertical momentum equation is retained, although in approximated form. In fact, here the solution procedure is very similar to that of the unapproximated Stokes model except that the dimensions of the pressure and the vertical velocity variables are reduced.

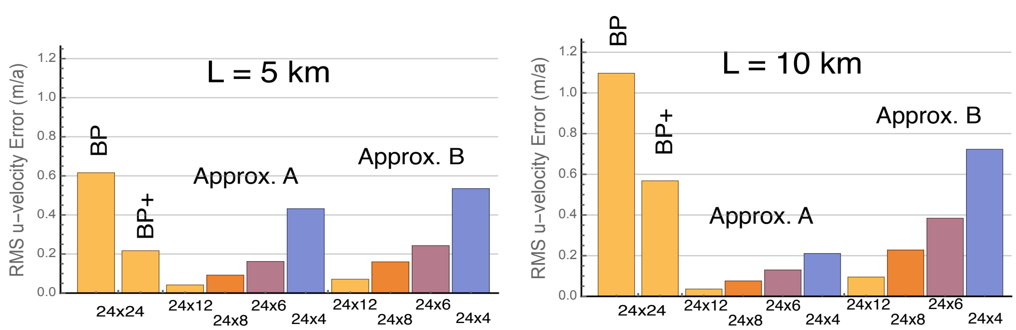

Figure 12. Comparing RMS u-Error in Different Approximations, Test B, Resolutions (r x R): Approx. BP, BP+: 24x24; Approx. A, B: 24xR, R=12, 8, 6, 4.

7. Second-Order Discretizations

So far we have been using first-order elements, primarily P1-E0. However, in current practice Stokes models are more often based on second-order elements such as the popular Taylor-Hood P2-P1 element (Leng et al., 2012; Gagliardini et al., 2013). In 2D the P2-P1 element, illustrated in Fig. 13A, has velocities on element vertices and edge midpoints and pressures on element vertices, resulting in a quadratic velocity and linear pressure distributions. The element satisfies the conventional inf-sup stability condition (e.g., Elman et al., 2014) but not the solvability condition (57). For example, in Test B with direct substitution for basal boundary conditions, the number of vertical velocity variables is $n_w = 4nm$, much larger than the number of pressure variables, $n_p = n(m+1)$.

Stokes models work well with a Taylor-Hood grid, as illustrated in Fig. 13, where both P2-P1 and P1-E0 models converge to a common Test B solution, but models that require the solvability condition (57) will not work on a P2-P1 grid, as is the case with the extended Blatter-Pattyn approximation. However, a second-order element can be constructed that is consistent with an invertible continuity equation. This is called the P2-E1 element and it is illustrated in Fig. 13A. It is second-order for velocities and linear for pressure, just like the P2-P1 element, but the pressure is edge-based, as in the P1-E0 element. Pressure is located midway between the velocities on the vertical cell edges, including an "imaginary" vertical edge joining the velocities in the middle of the vertical column. Since pressures are collinear with vertical velocities along vertical grid edges, as in the P1-E0 element, the same analysis in Appendix B demonstrates that this element also satisfies the solvability condition (57). As explained in Appendix B, this grid should be constructed using vertical columns of quadrilaterals. A 3D analog exists as explained in Appendix B.

Fig. 13B shows the approximate error of the ice transport as a function of grid refinement for the second-order P2-P1 and P2-E1 grids in transformed Stokes Test B calculations, together with similar results for the first-order P1-E0 grid from Fig. 3, for comparison. We note that both second order models show approximately the same error at resolution $r = 16$ as the first order P1-E0 model at resolution $r = 40$, and similarly for coarser resolutions such as $r = 8$ and $r = 20$, respectively. However, it is safe to say that these second-order calculations are considerably more expensive than the first-order calculations at comparable resolution or accuracy.

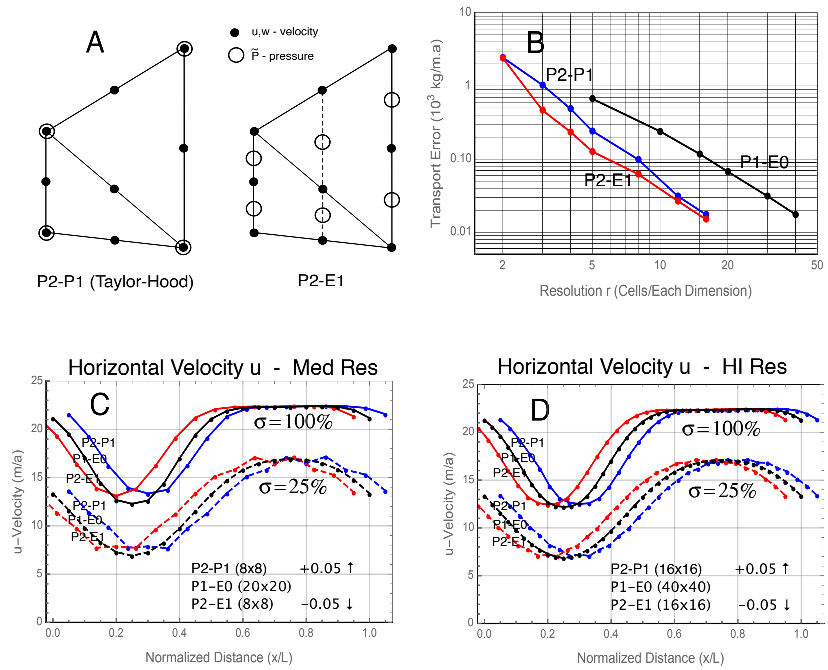

Figure 13. Comparing second-order discretizations based on the P2-P1 and P2-E1 elements from panel A to first-order discretizations using the P1-E0 element running Test B with L=10 km. Only transformed Stokes calculations are compared; standard Stokes results behave similarly. Panel B compares the convergence and accuracy of the various schemes with increasing resolution, while panels C, D compare the horizontal velocities at medium and maximum resolutions.

Panels C, D in Fig. 13 compare the u-velocities from several Test B calculations using the two second-order models in comparison with first-order P1-E0 model results from Fig. 3. Each panel shows results from the upper surface ($\sigma = 100\%$) in solid lines and results from a surface a quarter of the way up from the bottom ($\sigma = 25\%$) in dashed lines. Panel C shows results from medium resolution calculations ($r = 8, 20$ in the second-order and first-order calculations, respectively) and panel D shows the corresponding results from the higher resolution calculations ($r = 16, 40$). At these resolutions the accuracy of the first- and second-order calculations is very similar so for clarity the second-order results are displaced horizontally from the first-order results by 0.05 nondimensional units. The P2-E1 results in magenta are displaced to the left and the P2-P1 results in blue are displaced to the right. In general, models satisfying the solvability condition, namely the P1-E0 and P2-E1 models, are better behaved than the P2-P1 model. This is possibly related to the well-known “weak” mass conservation of the Taylor-Hood element. This problem is greatly improved by “enriching” the pressure space with constant pressures in each triangular element (Boffi et al., 2012). In the 2D Test B problem this increases the number of pressure variables from $n_p = n(m+1)$ in the basic Taylor-Hood element to $n(3m+1)$, much closer to the $4nm$ needed to satisfy the solvability condition. On the other hand, the pressure in the P2-E1 case is highly oscillatory but well behaved in the P2-P1 case. However, this is not at all concerning since the transformed pressure, a Lagrange multiplier, has no physical significance.

8. A Summary and Discussion

In summary, this paper presents two innovations in ice sheet modeling. The first involves a transformation of the ice sheet Stokes equations into a form that differs from the approximate Blatter-Pattyn system by a small number of terms. In particular, the variational formulations differ only by the absence of terms involving the vertical velocity w in the second invariant of the strain rate tensor in the Blatter-Pattyn system.

We focus on two applications of the new transformation. The first is that these extra terms in the transformed Stokes equations may be “switched” on or off to convert the code from a full-Stokes model to a Blatter-Pattyn model, if desired. Ice sheet flow is generally shallow but often contains limited areas where Stokes equations must be solved. Thus, the switch from Blatter-Pattyn to Stokes may be done locally and adaptively only where the extra accuracy is required.

The fact that neglecting the vertical velocity in only one localized place creates the Blatter-Pattyn approximation suggests that approximating the vertical velocity instead will create improved approximations. We present two such approximations. The first approximation, called the BP+ approximation, solves the pressure-free horizontal momentum equation with the vertical velocity obtained from the continuity equation. Remarkably, this approximation turns out to be the same as a model originally proposed by Herterich (1987). An intriguing idea would be to replace the BP with the BP+ approximation in the adaptive switching method. The second approximation simply approximates the vertical velocity by discretizing the continuity equation on a coarser grid than the rest of the model.

The second innovation involves the introduction and use of finite element discretizations that feature a decoupled invertible continuity equation permitting the numerical solution for the vertical velocity in terms of the horizontal velocity components, i.e., $w = w(u,v)$. Some examples of such grids for use in 2D and 3D are given in Appendix B. An important example is the P1-E0 grid that is used in most of the test problems in this paper. However, one can alternatively obtain $w = w(u,v)$ by other means, as for example by discretizing (43). For example, this is done in MALI (Hoffman et al., 2018), a code based on the Blatter-Pattyn approximation, to obtain the vertical velocity w for the advection of ice temperature (Mauro Perego, private communication).

Finally, no cost comparisons have been presented because the present calculations are only proof of concept, made on a personal computer using the program Mathematica. This is not at all representative of the computer hardware or the methods used in practical ice sheet modeling. Furthermore, no effort was made to optimize the calculations or to take advantage of parallelization. As a result, cost comparisons would be inaccurate and possibly misleading.

Code Availability

All calculations were made using the Wolfram Research, Inc. program Mathematica in a development environment. No production code is available.

Competing Interests

The author has acknowledged that there are no competing interests.

Acknowledgements

I am grateful to Steve Price for support, and especially to Mauro Perego, William (Bill) Lipscomb, and reviewers Ed Bueler and Christian Schoof for many helpful comments and suggestions that helped to improve the paper.

References

Auricchio, F., da Veiga, L.B., Brezzi, F., and Lovadina, C.: Mixed Finite Element Methods, In Encyclopedia of Computational Mechanics Second Edition (Eds E. Stein, R. de Borst, and T.J.R. Hughes), John Wiley & Sons, Ltd., 2017.

Blatter, H.: Velocity and Stress Fields in Grounded Glaciers: A Simple Algorithm for Including Deviatoric Stress Gradients, J. Glaciol., 41, 333-344, 1995.

Boffi, D., Brezzi, F., Fortin, M.: Finite Elements for the Stokes Problem. In: Boffi D., Gastaldi L. (eds) Mixed Finite Elements, Compatibility Conditions, and Applications. Lecture Notes in Mathematics, vol 1939. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78319-0_2, 2008.

Boffi, D., Cavallini, N., Gardini, F., and Gastaldi, L.: Local Mass Conservation of Stokes Finite Elements, J. Sci. Comput., 52, 383–400, 2012.

Chen, Q., Gunzburger, M., and Perego, M.: Well-Posedness Results for a Nonlinear Stokes Problem Arising in Glaciology, SIAM J. Math. Anal., 45(5), 2710-2733, 2013.

Cheng, G., Lötstedt, P., and von Sydow, L.: A Full Stokes Subgrid Scheme in Two Dimensions for Simulation of Grounding Line Migration in Ice Sheets Using Elmer/ICE (v8.3), Geosci. Model Dev., 13, 2245-2258, 2020.

Dukowicz, J.K., Price, S.F., and Lipscomb, W.H.: Consistent Approximations and Boundary Condition for Ice Sheet Dynamics from a Principle of Least Action, J. Glaciol., 56, 480-496, 2010.

Dukowicz, J.K., Price, S.F., and Lipscomb, W.H.: Incorporating Arbitrary Basal Topography in the Variational Formulation of Ice Sheet Models, J. Glaciol., 57, 461-467, 2011.

Dukowicz, J.K.: Refolmulating the Full-Stokes Ice Sheet Model for a More Efficient Computational Solution, The Cryosphere, 6, 21-34, 2012.

Elman, H.C., D.J. Silvester, and A.J. Wathen, 2014: *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*, 2nd Ed., Oxford University Press, 494 pp.

Gagliardini, O., and Zwinger, T.: The ISMIP-HOM Benchmark Experiments Performed Using the Finite-Element Code Elmer, The Cryosphere, 2, 67–76, 2008.

Gagliardini, O., Zwinger, T., Gillet-Chaulet. F., Durand, G., Favier, L., de Fleurian, B., Greve, R., Malinen, M., Martín, C., Råback, P., Ruokolainen, J., Sacchettini, M., Schäfer, M., Seddik, H., and Thies, J.: Capabilities and Performance of Elmer/Ice, a New-Generation Ice Sheet Model, Geosci. Model Dev., 6, 1299–1318, doi:10.5194/gmd-6-1299-2013, 2013.

Greve, R. and Blatter, H.: Dynamics of Ice Sheets and Glaciers, Springer-Verlag, Berlin Heidelberg, 2009.

Heinlein, A., Perego, M., and Rajamanickam, S.: FROSch Preconditioners for Land Ice Simulations of Greenland and Antarctica, SIAM J. Sci. Comput., 44, V339-B367, doi: 10.1137/21M1395260, 2022.

Herterich, K.: 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, (Eds. C. J. Van Der Veen and J. Oerlemans), D. Reidel, Dordrecht, 185-202, 1987.

Hoffman, M. J., Perego, M., Price, S. F., Lipscomb, W. H., Zhang, T., Jacobsen, D., Tezaur, I., Salinger, A. G., Tuminaro, R., and Bertagna, L.: MPAS-Albany Land Ice (MALI): A Variable-Resolution Ice Sheet Model for Earth System Modeling Using Voronoi Grids, Geosci. Model Dev., 11, 3747–3780, doi:10.5194/gmd-11-3747-2018, 2018.

Hood, P. and Taylor, C.: Numerical Solution of the Navier-Stokes Equations Using the Finite Element Technique, Comput. Fluids, 1, 1-28, 1973.

Larour, E., Seroussi, H., Morlighem, M., and Rignot, E.: Continental scale, high order, high spatial resolution, ice sheet modeling using the Ice Sheet System Model (ISSM), J. Geophys. Res., 117, 1–20, doi:10.1029/2011JF002140, 2012.

Leng, W., Ju, L., Gunzburger, M., Price, S., and Ringler, T.: A Parallel High-Order Accurate Finite Element Nonlinear Stokes Ice Sheet Model and Benchmark Experiments, J. Geophys. Res., 117, 2156–2202, doi:10.1029/2011JF001962, 2012.

Lipscomb, W.H., Price, S.F., Hoffman, M.J., Leguy, G.R., Bennett, A.R., Bradley, S.L., Evans, K.J., Fyke, J.G., Kennedy, J.H., Perego, M., Ranken, D.M., Sacks, W.J., Salinger, A.G., Vargo, L.J., and Worley, P.H.: Description and Evaluation of the Community Ice Sheet Model (CISM) v. 2.1, Geosci. Model Dev., 12, 387-424, 2019.

Nowicki, S.M.J. and Wingham, D.J.: Conditions for a Steady Ice Sheet-Ice Shelf Junction, *Earth Planet. Sci. Lett.,* **265**(1-2), 246-255, 2008.

Pattyn, F.: A New Three-Dimensional Higher-Order Thermomechanical Ice Sheet Model: Basic Sensitivity, Ice Stream Development, and Ice Flow across Subglacial Lakes, J. Geophys. Res., 108(B8), 2382, 2003.

Pattyn, F., Perichon, L., Aschwanden. A., Breuer, B., de Smedt, B., Gagliardini, O., Gudmundsson, G.H., Hindmarsh, R.C.A., Hubbard, A., Johnson, J.V., Kleiner, T., Konovalov, Y., Martin, C., Payne, A.J., Pollard, D., Price, S., Ruckamp, M., Saito, F., Soucek, O., Sugiyama, S., and Zwinger, T.: Benchmark Experiments for Higher-Order and Full-Stokes Ice Sheet Models (ISMIP–HOM*),* The Cryosphere, 2, 95-108, 2008.

Perego, M., Gunzburger. M., and Burkardt, J.: Parallel Finite-Element Implementation for Higher-Order Ice-Sheet Models, J. Glaciol., 58, 76-88, 2012.

Rückamp, M., Kleiner, T., and Humbert, A.: Comparison of ice dynamics using full-Stokes and Blatter-Pattyn approximation: application to the Northeast Greenland Ice Stream, The Cryosphere, 16, 1675-1696, 2022.

Schoof, C.: Coulomb friction and other sliding laws in a higher order glacier flow model, Math. Models. Meth. Appl. Sci., 20(1), 157–189, 2010.

Schoof, C. and Hewitt, I.: Ice-Sheet Dynamics, Annu. Rev. Fluid Mech., 45, 217–239, 2013.

Schoof, C. and Hindmarsh, R.C.A.: Thin-Film Flows with Wall Slip: An Asymptotic Analysis of Higher Order Glacier Flow Models, Quart. J. Mech. Appl. Math, 63, *73-114*, 2010.

Seroussi, H., Ben Dhia, H., Morlighem, M., Latour, E., Rignot, E., and Aubry, D.: Coupling Ice Flow Models of Varying Orders of Complexity with the Tiling Method, J. Glaciol., 58, 776-786, 2012.

Tezaur, I. K, Perego, M., Salinger, A. G., Tuminaro, R. S., and Price, S. F.: Albany/FELIZ: A Parallel, Scalable and Robust, Finite Element, First-Order Stokes Approximation Ice Sheet Solver Built for Advanced Analysis, Geosci. Model Dev., 8, 1197-1220, 2015.

Appendix A: Test Problems

We will use three 2D test problems to demonstrate the new methods. The geometrical configuration of the three problem grids is illustrated in Fig. A1. The first problem, Test B, is actually Exp. B from the ISMIP-HOM benchmark suite (Pattyn et al., 2008); it features a no-slip condition (infinite friction) on a sinusoidal basal surface. The second problem, Test D*, incorporating sinusoidal friction along a uniformly sloped plane basal surface, is a replacement with modified parameters for Exp. D from the benchmark suite since ice flow in Exp. D is nearly vertically uniform (see Fig. 4), more characteristic of a shallow-shelf approximation, and this is rectified by increased basal friction.

A third problem, Test O (for "Obstacle") is introduced to the illustrate adaptive switching discussed in §6.1. Test O has a unique feature, namely, a thin no-slip obstacle, located at $x = 4\ km$ and extending vertically $200\ m$ from the bed (20 % of the ice sheet thickness), as illustrated in Fig. A1, which forces the ice flow near the obstacle to adjust abruptly. Because of the no-slip boundary conditions along the obstacle surface, a triangular element in the lee of the obstacle, with one vertical edge and one edge along the bed, would have all zero vertex velocities. This implies zero stress and therefore a local singularity in ice viscosity. To avoid this, all elements at the back of the obstacle are "reversed" as compared to the ones at the front of the obstacle, as shown in Fig. A1.

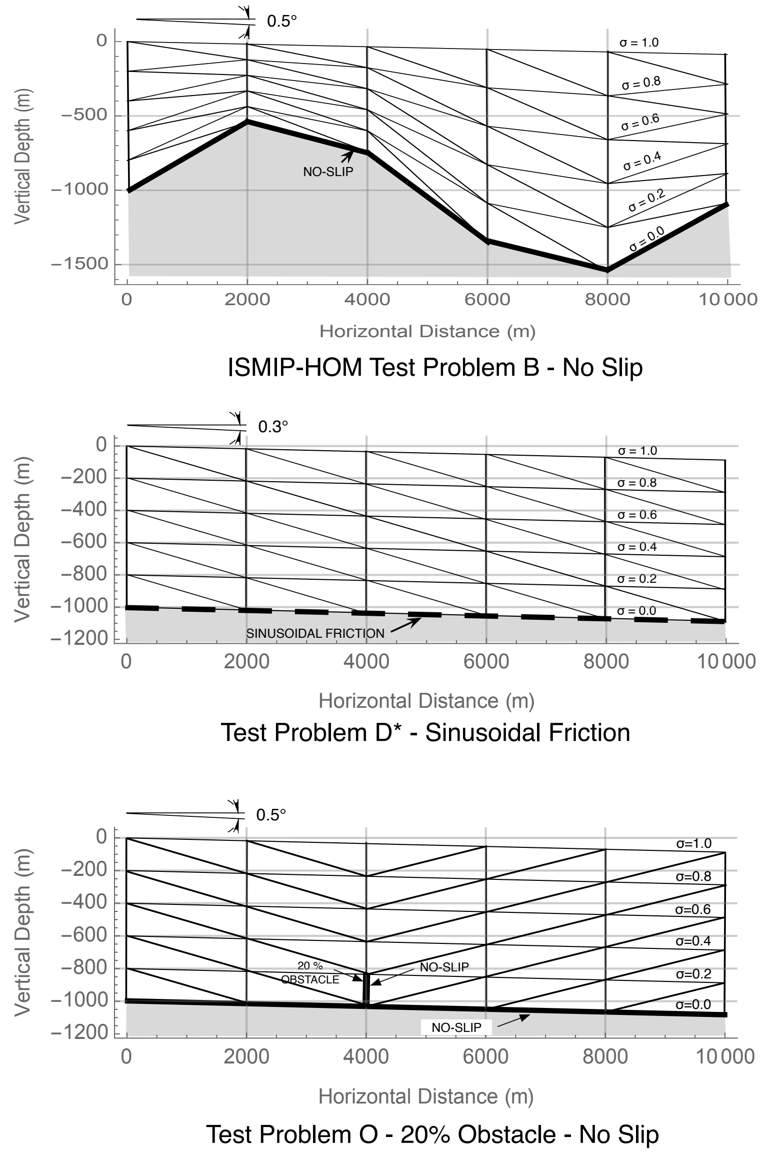

Figure A1. Test problem grids. For clarity, a coarse 5x5 configuration is shown.

All tests feature a sloping flat upper surface, given by $z_s(x) = -x\,\mathrm{Tan}(\theta)$, where $\theta = 0.5^\circ$ for Tests B and O, and $\theta = 0.3^\circ$ for Test D* (this differs from the 0.1° slope in Test D), with a free-stress upper boundary condition in all cases. The sinusoidal bottom surface elevation for Test B is specified by $z_b(x) = z_s(x) - H_0 + H_1 \mathrm{Sin}(\omega x)$, where the depth $H_0 = 1000\ m$, $H_1 = 500\ m$, $\omega = 2\pi/L$, and L is the perturbation wavelength, which is also the domain length. The bottom surface elevation in Tests D* and O is $z_b(x) = z_s(x) - H_0$, parallel to the upper surface. The length L in the ISMIP-HOM suite ranges from $5\ km$ to $160\ km$, but here we consider only the two cases at the high end of the aspect ratio H_0/L range, namely, $L = 5\ km$ and $L = 10\ km$, where the inaccuracy of the Blatter-Pattyn approximation becomes noticeable. In all cases the lateral boundary conditions are periodic. The spatially varying friction coefficient for Test D* is given by $\beta(x) = \beta_0 + \beta_1 \mathrm{Sin}(\omega x)$, where $\beta_0 = \beta_1 = 10^4\ Pa\ a\ m^{-1}$ (these are an order of magnitude higher than in Test D). The physical parameters are the same as in ISMIP-HOM, namely, ice-flow parameter $A = 10^{-16}\ Pa^{-3} a^{-1}$, ice density $\rho = 910\ kg\, m^{-3}$, and gravitational constant $g = 9.81\ ms^2$. In general, units are MKS, except where time is given per annum, convertible to per second by the factor $3.1557 \times 10^7\ s\ a^{-1}$.

Appendix B: Grids Satisfying the Solvability Condition

B1 An Invertible Continuity Equation

As discussed in §4, the invertibility of the discrete continuity equation requires special grids that satisfy the solvability condition. Here we discuss several such grids and their properties. Fig. B1 shows three 2D elements on triangles or quadrilaterals that satisfy the solvability condition (57) in certain circumstances. The P1-E0 element is quite general, as demonstrated in §B2. It has velocities located at triangle vertices, resulting in a linear velocity distribution within each triangle (P1), and pressure located on the vertical edge of each triangle, providing a constant pressure over the two triangles that share that edge (E0). A second order version of this element, the P2-E1 element, is shown in Fig. 13A. The other two elements, the P1-Q0 and Q1-Q0 elements, satisfy the solvability condition when used in Tests B and D* but may not do so in other problems. The P1-Q0 element also has velocities on triangle vertices for a linear velocity distribution within each triangle (P1) but the pressure is constant within the quadrilateral (Q0) formed by the two adjoining triangles. The Q1-Q0 element has velocities located at quadrilateral vertices and pressure centered in the quadrilateral, resulting in a bi-quadratic velocity distribution (Q1) and a constant pressure within the quadrilateral (Q0).

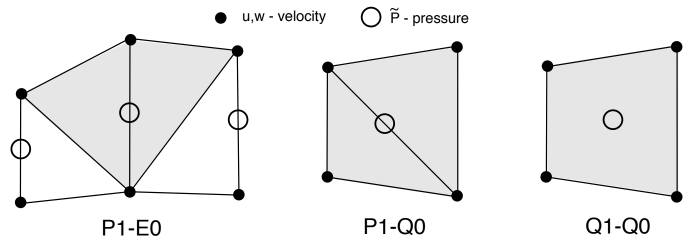

Figure B1. Three first-order 2D elements that may be used to satisfy the solvability condition, (57), in Tests B and D*.

The solutions are stable, as expected, and they all converge with the same value for ice transport. The pressure distribution is smooth in the P1-E0 case, but contains small fluctuations near the upper surface in the P1-Q0 and Q1-Q0 cases that tend to disappear as resolution is increased. The Q1-Q0 element is unstable in conventional applications because it contains a checkerboard pressure null space and is only used in a stabilized form (see Elman et al., 2014, where the element is called Q1-P0). Here, however, the Q1-Q0 grid does behave well, presumably because it satisfies the solvability condition. Overall, this confirms our expectation of stability when the solvability condition is satisfied. As we discuss next, the P1-E0 element is special because the solvability condition is satisfied along each vertical edge, as opposed to being satisfied over the entire grid as in the other two elements.

B2 The Solvability Condition in the P1-E0 Element

Fig. B2 illustrates the P1-E0 element used in a representative grid. We assume that the grid is composed of vertical columns subdivided into triangular elements. Consider a single vertical edge from bottom to the top. Assuming there are m edge segments in the vertical direction, there will be $m+1$ discrete w variables and m discrete $\tilde{P}$ variables since each $\tilde{P}$ variable is located between a pair of w variables. However, since the w variable at the bed is specified as a boundary condition, either directly as a no-slip condition or as part of a no-penetration condition, there will be only m unknown w variables. As a result we have $n_w = n_p$ along each vertical grid edge, and therefore over the entire grid, satisfying the solvability condition. In case Lagrange multipliers are used, there will be $m+1$ unknown discrete w variables (since now the basal vertical velocity w is also an unknown). However, this is matched by m unknown $\tilde{P}$ variables, supplemented by one λ_z or one Λ unknown Lagrange multiplier variable, depending on the type of boundary condition. Thus, again the number of unknown variables equals the number of equations along every vertical edge, thereby satisfying the solvability condition whether Lagrange multipliers are used or not. This means that the P1-E0 element can be used to satisfy the solvability condition irrespective of the boundary conditions on quite arbitrary grids, as illustrated in Fig. B2. These arguments apply for other versions of the P1-E0 element as well, such as the second order version P2-E1 in Fig. 13A or the 3D version in Fig. B3.

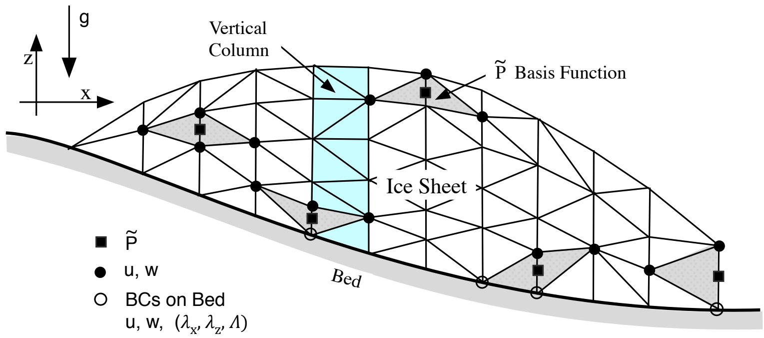

Figure B2. An illustration of a 2D edge-based P1-E0 grid, composed of vertical columns randomly subdivided into triangles. Pressures are located on the vertical edges. The triangulation and the configuration of the associated pressure basis functions (shown in gray) is quite general, allowing for a flexible triangulation of the domain.

B3 Two- and Three-Dimensional Meshes Based on the P1-E0 Element

The P1-E0 element has been used on the test problem grids in Fig. A1 and performs well. Moreover, the element has great geometric generality, as in Fig. B2, so it may be used for quite complicated grids. Generally, there are two triangles associated with a pressure variable, one on each side of a vertical edge, except in situations where the ice sheet ends at a vertical face, as in Fig. B2. However, there is no problem since the pressure is simply associated with the single triangle on one side of the vertical face.

Meshes composed of P1-E0 elements have another useful property. Since pressure and vertical velocity variables alternate along vertical grid lines, the matrix-vector products $M_{WP}p$, $M_{WP}^{T}w$ in (47), corresponding to $\partial\tilde{P}/\partial z$ and $\partial w/\partial z$ in the vertical momentum and continuity equations, respectively, consist of simple decoupled bi-diagonal one-dimensional difference equations along each vertical grid line for determining pressure and vertical velocity. This should be particularly advantageous for parallelization.

The two-dimensional P1-E0 element has a relatively simple three-dimensional counterpart, shown in Fig. B3. The mesh again consists of vertical columns, this time composed of hexahedra. Each hexahedron is subdivided into six tetrahedra such that each vertical edge is surrounded by as few as four to as many as eight tetrahedra. As in the 2D case, velocity components are collocated at vertices, yielding a piecewise-linear velocity distribution in each tetrahedral element, and pressures are located in the middle of each vertical edge so that pressure is constant in the tetrahedra that surround that edge. Lagrange multipliers, if used, are located at the vertices on the basal surface, yielding a piecewise linear distribution on the basal triangular facet. Since pressures and vertical velocities are again intermingled along a single line of vertical edges from top to bottom, we see that this satisfies the solvability condition (57) since the argument used in the 2D case applies here also.

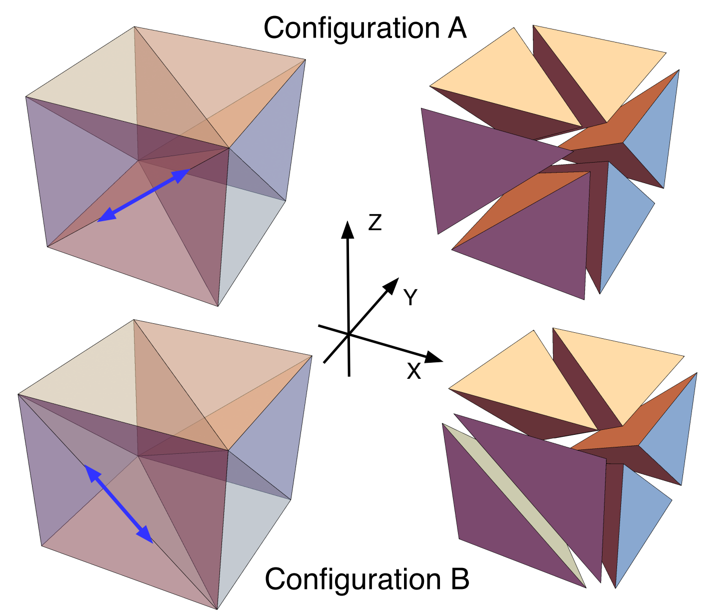

Figure B3. Three-dimensional P1-E0 tetrahedral elements that generalize the 2D P1-E0 element of Fig. C1. Configurations A and B differ by having an internal triangular face rotated, as indicated by the blue arrows. Both configurations satisfy the solvability condition.

Fig. B3 shows two of the several possible configurations of a typical hexahedron, including an exploded view of each configuration for clarity. The two configurations differ in having the internal face of the two forward-facing tetrahedra rotated, creating two different forward facing tetrahedra. The remaining six tetrahedra are undisturbed.

Since edges must align when hexahedra (or tetrahedra) are connected, this showa that the 3D mesh can be flexibly reconnected and rearranged, just as in the 2D case of Fig. B2.

Remark #2: A closely related but perhaps even simpler three-dimensional P1-E0 element is one based on the P2-P1 prismatic tetrahedral element that is used in Leng et al. (2012). A grid of these elements is composed of vertical columns of triangular prisms, with triangular faces at the top and bottom, which are then each subdivided into three tetrahedra. As in Fig. B3, pressures are located on the vertical prism edges so this again satisfies the solvability condition.

Just as the 2D second-order P2-E1 element in Fig. 13A is a generalization of the P1-E0 element, a 3D second-order P2-E1 element may be constructed as a generalization of the P1-E0 element illustrated in Fig. B3. Velocities would be located at the vertices and at midpoints of the tetrahedral edges, and pressures halfway between the velocities on vertical edges, including the imaginary vertical edges through the midpoints of the tetrahedral edges, in the same way as in the 2D case in Fig. 13A. The P2-E1 element in both 2D and 3D would also satisfy the solvability condition since the arguments in §B2 apply here as well because pressures are again located midway between vertical velocities along all vertical edges.