Numerical issues in modeling ~~ice sheet instabilities such as binge-purge type cyclic~~ thermally and hydraulically driven ice stream ~~surging~~surge cycling

Kevin Hank[1], Lev Tarasov[1], and Elisa Mantelli[2,3,4]

[1]Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, NL, A1B 3X7,Canada
[2]Department of Earth and Environmental Sciences , Ludwig-Maximillians-Universitaet Munich, Theresienstr. 41, 80333 Munich, Germany
[3]Alfred Wegener Institute for Polar and Marine Research, Am Alten Hafen 26, 27568 Bremerhaven, Germany
[4]Institute for Marine and Antarctic Studies, University of Tasmania, 20 Castray Esplanade, Battery Point TAS 7004, Australia
[*]khank@mun.ca

Correspondence: Kevin Hank (khank@mun.ca)

Abstract. ~~As in any environmental system, modeling instabilities within the glacial system~~ Modeling ice sheet instabilities is a numerical challenge of potentially high real-world relevance. ~~Differentiating~~ Yet, differentiating between the impacts of ~~physical system processes and numerical noise~~ model physics, numerical implementation choices, and numerical errors is not straightforward. Here we use an idealized North American geometry and climate representation (similar to the HEINO experiments, Calov et al., 2010) to examine the numerical sensitivity of ice stream surge cycling in ~~glaciological~~ ice flow models. Through sensitivity tests, we identify some numerical requirements for a more robust model configuration for such contexts. To partly address model-specific dependencies, we use both the Glacial Systems Model (GSM) and Parallel Ice Sheet Model (PISM).

We show that modeled surge characteristics are resolution-dependent though converging (decreasing differences between resolutions) at higher horizontal grid resolutions. Discrepancies between high and coarse horizontal grid resolutions can be reduced by incorporating ~~a resolution-dependent basal temperature ramp for basal sliding thermal activation. Inclusion of a diffusive~~ sliding at sub-freezing temperatures. The inclusion of a bed thermal model ~~reduces the surge cycling ice volume change by ~ 33 % as the additional heat storage~~ markedly reduces the ice volume lost during surges in both the GSM and PISM, as the substrate heat storage capacity dampens the change in basal temperature during ~~surge events~~surges. The inclusion of basal hydrology, as well as a non-flat topography, leads to ~~increased ice volume change during surge events (~ 20 and 17 %, respectively)~~an increase in the ice volume lost during surges in both models. Therefore, we conclude that these latter three components are essential ~~if one is endeavoring to maximize~~ to maximizing physical fidelity in ice stream surge cycle modeling. ~~An abrupt transition between hard bedrock and soft sediment, as in the HEINO experiments, leads to ice stream propagation along this boundary but is not the cause of the main surge events.~~

Copyright statement. TEXT

1 Introduction

1.1 Motivation and background

The use of Ice Sheet Models ~~(ISMs)~~ has grown at least an order of magnitude over the last two decades. The relevance of such modeling studies to the actual physical system can be unclear without careful consideration and testing of numerical ~~components~~ aspects and implementations. ~~Model validation is particularly important when modeling~~ This is especially true when modeling the highly non-linear ice sheet ~~instabilities, for which it is hard to distinguish between numerical noise and physical phenomena. In addition, there are a number of~~ surge instability, which has significant implications not only for the ice sheet itself but also for the climate. In fact, it is often difficult to assess whether model results are physically significant (effects of physical system processes), a consequence of model-specific numerical choices, ~~such as for thermal activation of basal sliding, for which no model to date has documented sensitivities.~~

or a combination of both. This is especially important in the case of abrupt changes. Whether ice sheet instabilities observed in numerical simulations are the result of physical instabilities of the underlying continuum models or spurious effects of the discretization and numerical implementation of said models has long been debated (e.g., Payne et al., 2000; Hindmarsh, 2009) and is a consequential matter. The present study is concerned with characterizing the impact of model physics, numerical choices, and numerical errors on ice stream surge cycling.

~~One archetypal ice sheet instability is the binge-purge~~

Binge-purge ice stream cycling ~~explanation of Heinrich Events (HEs) originally proposed by MacAyeal (1993). Within this hypothesis, the~~ was first introduced in the glaciological literature by MacAyeal (1993) as an explanation for Heinrich Events arising from the former Laurentide Ice Sheet (LIS) in the Hudson Bay/Hudson Strait region. The key idea is that the ice stream gradually grows to a threshold thickness (binge phase) driven by surface accumulation. Once the ice ~~sheet~~ stream is thick enough to sufficiently isolate the ice ~~sheet~~ stream base from the cold surface, heat from geothermal and deformation work sources can slowly bring the basal temperature to the pressure melting point. The bottom layer of the ice ~~sheet~~ stream is no longer frozen to the bed and thus enables basal sliding. Localized warm-based ice streaming increases the ice ~~sheet~~ stream surface gradient (steeper slope) at the warm/cold-base transition point, leading to an increase in driving stress. The resultant increase of deformation work can warm the surrounding ice close to or to the pressure melting point~~. The~~, thus enabling (sub-temperate) basal sliding (Fowler, 1986). When the melting point is reached, the presence of water at the ice sheet/bed interface (decreased basal friction, Fowler and Schiavi, 1998) as well as in a deformable sediment layer (loosened sediment, Bueler and Brown, 2009) can further increase ~~ice velocities. Fowler and Schiavi (1998) call this sequence of events the propagation of an *"activation wave"*.~~ sliding velocities. Instead of the slow deformation flow (ice creep), the ice ~~sheet now destabilizes~~ stream now flows rapidly (purge phase). As a consequence of the high ice velocities, the ice ~~sheet~~ stream thins and cold ice is advected from either upstream or the lateral boundaries of the ice stream. Cold ice advection in combination with changing heat source contributions (from both deformation work and basal sliding) and lowering of the pressure

melting point as ice thins eventually leads to refreezing of the ice/bed interface. The first localized frozen patch of ice acts as a *pinning point*, supporting some of the driving stress and decreasing the velocities and heat production in the adjacent ice. This ~~*"deactivation wave"* (Fowler and Schiavi, 1998) ends the surgeand the ice sheet enters~~ marks the end of the surge, thus enabling the ice stream to enter the next binge phase. Whether hydraulically or thermally driven, these activation (purge) and stagnation (binge) phases can alternate in a quasi-periodic fashion (e.g., Robel et al., 2013) - this is what we refer to as 'ice stream surge cycling' in the remainder of this paper.

~~Meaningful modeling of binge-purge type HEs and surges in general~~ As a result of the involved physics and expected behaviors, modeling of ice stream surge cycling is challenging. ~~Sensitivity in model response to different numerical choices are evident (Calov et al., 2010; Roberts et al., 2016; Ziemen et al., 2019) and small perturbations of the system can significantly vary the form, amplitude, and period of binge-purge oscillations (Souček and Martinec, 2011; Mantelli et al., 2016).The exact cause of the numerical sensitivities is often unclear.To date, uncertainties associated with the numerical aspects of a model~~ The challenges entail, among others, rapid surge onset, high ice velocities, and non-linear (thermo-viscous, hydraulic, and thermo-frictional) feedbacks. In addition to the physical complexity, further challenges arise in the numerical modeling of ice stream surge cycling, whether in terms of model choices (e.g., choice of mechanical model, thermal modeling of the substrate, accounting for sub-glacial hydrology) and/or in terms of their numerical implementation (e.g., grid and time step size, convergence under grid refinement, etc.).

Our focus here is on the challenges arising from numerical modeling, both those related to the modeling choices and those related to the implementation. Numerical challenges have received limited attention in studies examining ice sheet surging~~(e.g., Payne, 1995; Marshall and Clarke, 1997; Calov et al., 2002; Papa et al., 2006; Steen-Larsen and Dahl-Jensen, 2008; Calov et al., . Souček and Martinec (2011) thus rightfully conclude that *"... the implementation of surge-type physics in large- scale ice-sheet models is rather problematic since the information about the physical instability may be lost in the numerics"*. Furthermore, the theory underpinning the understanding of the instability mechanisms is not fully developed, especially in the context of a spatially extended 3D system, thus precluding systematic benchmarking of numerical models.~~. The few studies to date that do examine numerical aspects of surge cycling suggest strong sensitivities in model response to implementation choices such as grid size (e.g., Calov et al., 2010; Roberts et al., 2016; Ziemen et al., 2019). However, the effects of different approximations of the Stokes equations have been previously addressed (e.g., Brinkerhoff and Johnson, 2015), and are therefore not discussed here.

The discretization and related numerical implementation choices (e.g., grid resolution, grid orientation, and time step size) have been shown to affect numerical results. As far as the choice of grid is concerned, Ziemen et al. (2019), for example, find a constantly active ice stream at 40 km grid resolution and oscillatory behavior at 20 km grid resolution. They argue that this higher grid resolution is necessary to resolve the Hudson Strait properly. However, ~~studies examining~~ only a few studies examine the effect of different grid resolutions on surge behavior ~~are sparse (Greve et al., 2006; Van Pelt and Oerlemans, 2012; Brinkerhoff~~ (e.g., Payne and Dongelmans, 1997; Greve et al., 2006; Van Pelt and Oerlemans, 2012; Brinkerhoff and Johnson, 2015; Roberts et al., 201 and an in-depth numerical analysis of Hudson Strait ~~binge-purge type surges~~ ice stream surge cycling (to whatever idealized

form) is ~~missing entirely.~~ absent. In terms of grid rotation, Greve et al. (2006) and Takahama (2006) show only a minor effect of grid rotation on the general features of the oscillations.

The effect of time stepping on modeled surge cycling appears to be weak or entirely absent. Greve and MacAyeal (1996) examined the impact of different time steps on ~~binge-purge~~ ice stream surge oscillations in a coupled dynamic/thermodynamic flowline model. They report ~~a~~ similar dynamic behavior across different time steps, but both test runs crashed. Later studies using a three-dimensional version of the same but further developed model find shorter periods and a slight decrease in surge amplitude but otherwise reasonable convergence as the time step decreases ~~. (Greve et al., 2006; Takahama, 2006) . The same two studies also show only a minor effect of grid rotation on the general features of the oscillations. None~~ (Greve et al., 2006; Takahama, 2006). Yet, none of the more recent studies on ~~binge-purge type surge events~~ ice stream surge cycling includes experiments with different time steps.

~~Accounting for discrepancies associated with modeling choices becomes even more important when investigating the effects of artificially induced noise~~An additional level of complexity in the modeling of ice sheet surge cycling arises from the fact that small perturbations of the initial or boundary conditions can significantly vary the surge characteristics (Souček and Martinec, 2011; Mantel . For example, Souček and Martinec (2011) show that low levels of surface temperature noise can lead to chaotic behavior in the ~~time lag between subsequent~~ periodicity of ice stream oscillations, with mean periods varying by ± 2 kyr (~ 20 % of characteristic period of the oscillations, Fig. 8 in Souček and Martinec (2011)). ~~Similarly, Mantelli et al. (2016) find oscillations for an otherwise steadily streaming ice stream when surface temperature (within -22 to -32 °C) and accumulation rate (0.2 to 0.3 m yr^{-1}) are stochastically forced with time-correlated (red), low amplitude noise. Under such imposed climate variability, they find ice stream oscillation amplitudes larger than obtained for a constant climate forcing and exceeding the noise intensity. Additionally~~Moreover, Souček and Martinec (2011) find differences in form, period, and amplitude of oscillations when using two different numerical implementations for calculating the basal temperature for thermal activation of basal sliding~~(weighted average of basal temperature at surrounding grid points vs. no temperature averaging). Note that their model is based on an Arakawa A grid (velocities and temperatures are calculated on the same node (Arakawa and Lamb, 1977)), enabling a thermal activation scheme without temperature averaging.~~ . However, whether this observed sensitivity arises from physical grounds (e.g., as in Mantelli et al., 2016) or is a spurious numerical effect, the numerical error remains unclear. Souček and Martinec (2011) thus rightfully conclude that *"... the implementation of surge-type physics in large-scale ice-sheet models is rather problematic since the information about the physical instability may be lost in the numerics"*.

1.2 ~~Research questions~~Study overview

~~Given the above context~~With this as a starting point, in this ~~paper we examine~~ study we seek to disentangle the effects of numerical choices (both in terms of model components and in terms of their implementation) on ice sheet surges~~as framed in the research questions detailed below. All research questions are investigated in Sec. 3 and a concise summary is provided in Sec. 4. We primarly~~ . We will do so through a numerical modeling study.

In terms of ice flow models, we primarily use the 3D glacial systems model ~~(GSM). To partly address potential non-linear dependencies of surge cycling on model parameters, we use a high variance subset of 5 base GSM parameter vectors (each comprising 8 model input parameters) for our numerical experiments.~~

~~To~~ (GSM, Tarasov et al., 2023). However, to mitigate the possibility that ~~these results are just due to~~ our conclusions are biased by specific numerical/modeling choices within the GSM, we repeat experiments that do not require ~~source code changes~~ implementation of novel physics with the widely used Parallel Ice Sheet Model ~~(PISM) (Bueler and Brown, 2009; Winkelmann et al., 2011)~~ (PISM, Bueler and Brown, 2009; Winkelmann et al., 2011). As the two model setups and physics are somewhat different (see Table 2 for details), we do not intend to compare model results directly. Instead, ~~we aim~~ our aim is to increase confidence in model results by showing that the same conclusions can be drawn from two different models.

In order to partly address potential non-linear dependencies of surge cycling on model parameters, we run each of our numerical experiments with a high variance ensemble of 5 GSM parameter vectors (each comprising 8 model input parameters) and 9 PISM parameter vectors (each comprising 6 model input parameters).

In terms of different numerical choices, the impact on model results is usually determined by calculating the model error to the exact analytical solution. However, the theory behind the surge instability is not fully developed (no analytical solution exists), especially in the context of a spatially extended 3D system, thus precluding systematic benchmarking of numerical models.

To ~~provide~~ overcome this issue and provide at least a minimum estimate of the numerical ~~noise in the models, we examine the differences~~ model error, we first determine '**M**inimum **N**umerical **E**rror **E**stimates' (MNEEs). This is a new metric that aims to minimally resolve whether a change in surge characteristics ~~when applying a stricter numerical convergence in the GSM and adjusting the matrix solver used in PISM.These noise estimates set a minimum threshold for discerning physical significance of changes in surge characteristics due to physical model components (e.g.,~~ due to changes in the model configuration is significant (see Sec. 2.3 for details).

Equipped with these tools, we set out to tackle the research questions detailed in Sec. 1.3, which we denote with labels $Q_1 - Q_{11}$. The remainder of the paper is then structured as follows: we start by describing our models and experimental setups in Sec. 2. We then present detailed results that allow us to answer our research questions in Sec. 3, with a concise summary provided in Sec. 4. The results are organized into the following main themes: 1) key surge characteristics of the reference setup (Sec. 3.1), 2) MNEEs (Sec. 3.2), 3) sensitivity experiments with and without ~~subglacial hydrology) : *Q1 - What is the threshold of numerical noise in the two models (Sec. 3.2)?*~~ a significant (with respect to the MNEEs) effect on the results (Sec. 3.3.1), and 4) convergence study (Sec. 3.4).

1.3 Key research questions

~~An abrupt transition from hard bedrock to soft sediment can lead to additional localized shear heating caused by the difference in basal resistance and therefore sliding velocities at that transition. We explore the impact of the bed-type transition on surge characteristics by incorporating a smooth transition from 0 % sediment cover (hard bedrock) to 100 % sediment cover effectively changing~~ In this subsection, we detail the key research questions that we address through numerical experiments.

Following the above-described structure in the description of the results, the research questions are divided into three sub-categories: minimum numerical error estimates (MNEEs), sensitivity experiments, and convergence study.

Minimum numerical error estimates

Q_1 *What is the threshold of MNEEs in the two models (Sec. 3.2)?*

Surge cycling is sensitive to numerical aspects (e.g., numerical solver error). Since we can not determine the model error, we provide a minimum estimate of the numerical error in the models (see Sec. 2.3 for details). These MNEEs set a minimum threshold for discerning whether the model response to a change in model configuration is significant.

Sensitivity experiments

Here we aim to determine the significance of different model configurations on the surge characteristics. We are particularly interested in model configurations affecting the basal temperature and thus the surge behavior. Therefore, we first discuss the change in surge characteristics due to a bed thermal model (Q_2) and modeling choices affecting the basal temperature at the grid cell interface where the ice velocities are calculated (Q_3 and Q_4), including the basal sliding ~~coefficient C in Eq. (6b): *Q2 - Does the abrupt transition between a soft and hard bed significantly affect surge characteristics (Sec. 3.3.4)?* Given the topographic lateral bounds of Hudson Strait~~thermal activation criterion (Q_5). Previous studies examining the effects of ice stream behavior are often based on an idealized basal topography and sediment distribution and do not consider sub-glacial hydrology (e.g., Calov et al., 2010; Brinkerhoff and Johnson, 2015). Therefore, we determine the change in surge characteristics due to these aspects in Q_6, Q_7, Q_8, and Q_9, respectively. Since thermally and hydraulically driven ice stream surges are not exclusive, we also ~~examine: *Q3 - How does a non-flat topography affect the surge behavior (Sec. 3.3.4)?*~~ investigate the differences between the two mechanisms when used as the primary smoothing mechanism at the warm/cold-based transition zone (Q_{10}).

Q_2 *Is the inclusion of a bed thermal model a controlling factor for surge activity (Sec. 3.3.1)?*

Except for PISM, all models in the HEINO experiments did not include a bed thermal model (Calov et al., 2010). PISM is one of the few models that did not show oscillatory behavior in the HEINO experiments (except for experiment T1 (10 K colder minimum surface temperature, Calov et al. (2010))). We explore the role of the additional heat storage on surge activity by ~~incorporating~~ deactivating a 1 km deep bed thermal model in the GSM and PISM~~: *Q4 - Is the inclusion of a bed thermal model a controlling factor for surge activity (Sec. 3.3.1)?*~~ .

Q_3 *Do different approaches to determining the grid cell interface basal temperature significantly affect surge behavior, and if yes, which one should be implemented (Sec. 3.3.2)?*

Ambiguity arises when determining the basal temperature at the grid cell interface. On a staggered grid ~~(commonly Arakawa C grid (Arakawa and Lamb, 1977))~~(commonly Arakawa C grid, Arakawa and Lamb, 1977), the velocities are calculated at the grid cell interfaces, whereas basal temperatures are situated in the grid cell center. Therefore, the basal

temperature at the grid cell interface needed for the thermal activation of basal sliding needs to be determined as a function of the basal temperatures at the adjacent grid cell centers. ~~We~~ Here we examine surge sensitivity to different interpolation schemes (see Sec. **??** ~~for details on the three approaches): *Q5 - Do different approaches to determining the grid cell interface basal temperature significantly affect surge behavior, and if yes, which one should be implemented (Sec. 3.3.2)?*~~ 3.3.2).

Q_4 *How much of the ice flow should be blocked by upstream or downstream cold-based ice, or equivalently, what weight should be given to the adjacent minimum basal temperature (Sec. S8.1)?*

At relatively coarse horizontal grid resolutions (e.g., 25 km), the basal temperatures at the adjacent grid cell centers are of physical relevance~~as well~~. For example, a cold-based grid cell in the downstream direction should block at least part of the ice flow across a 25 km long warm-based interface (Eq. (S1)). ~~Therefore: *Q6 - How much of the ice flow should be blocked by upstream or downstream cold-based ice, or equivalently, what weight should be given to the adjacent minimum basal temperature (Sec. **??**)?*~~ Here we examine surge sensitivity to a change in the weight of the adjacent (grid cell center) minimum basal temperature when calculating the grid cell interface temperature.

Q_5 *How different are the model results for different basal temperature ramps and what ramp should be used (Sec. 3.3.3)?*

Another issue that is often ignored is the basal sliding thermal activation criterion. Based on the results of Souček and Martinec (2011), the basal temperature is a critical factor in the onset and termination of (surging) ice streams. Mantelli et al. (2019) show that an abrupt onset of sliding at the transition from a cold-based ice sheet to an ice sheet bed at the pressure melting point causes refreezing on the warm-based side and, therefore, cannot ~~exists~~exist. Observational and experimental evidence for ~~subtemperate~~ sub-temperate sliding further supports a smooth transition from cold-based no-sliding conditions to fully warm-based sliding, with sliding velocities increasing as the basal temperature approaches the pressure melting point (Barnes et al., 1971; Shreve, 1984; Echelmeyer and Zhongxiang, 1987; Cuffey et al., 1999; McCarthy et al., 2017).

An additional argument for sub-temperate sliding can be made on numerical grounds for coarse horizontal grid resolutions. It is unlikely that an entire grid cell reaches the pressure melting point within one time step (e.g., 25x25 km in 1 yr). As such, the activation of basal sliding should start at grid-cell basal temperatures below the pressure-melting point and ramp up as the pressure-melting point is approached. ~~For higher horizontal grid resolutions, the average grid cell temperature is a better representation of the subgrid temperatures~~ As the horizontal grid resolution becomes finer, the range of sub-grid temperatures in a grid cell decreases (e.g., Figs. 10, S26, and S27). Consequently, the thermal activation ramp should be sharper (smaller transition zone) for higher horizontal grid resolutions, but the exact width of the ramp at the highest horizontal grid resolutions is unknown.

Experimental work (e.g., Barnes et al., 1971; McCarthy et al., 2017) supports the notion of sub-temperate sliding within a narrow range of temperatures below the pressure melting point ($< 5°$C). A wide temperature ramp (e.g., $T_{ramp} = 1°$C,

see Eq. (9)) enables an earlier sliding onset (for increasing basal temperature), spatially extended sliding, and a prolonged sliding duration (for decreasing basal temperature).

While on theoretical ground, the sensitivity of sliding speeds on temperature is expected to ~~be strong near~~ grow with proximity to the pressure melting point, the appropriate functional form of the temperature ramp is not well constrained nor is the sensitivity to the functional form well documented. Herein, we use basal temperature gradients in high-resolution runs and approximations of the sub-grid warm-based connectivity between the faces of, e.g., a 25 km grid cell (there should be no ice streaming across the grid cell if a frozen sub-grid area disconnects warm-based patches) to constrain an apriori functional form of the basal temperature ramp. We then use upscaling and resolution-scaling experiments to constrain the dependency of the ramp on horizontal grid resolution~~: *Q7 - How different are the model results for different basal temperature ramps and what ramp should be used (Sec. 3.3.3)?*~~ .

Q_6 *Does the abrupt transition between a soft and hard bed significantly affect surge characteristics (Sec. 3.3.4)?*

An abrupt transition from hard bedrock to soft sediment can lead to additional localized shear heating caused by the difference in basal resistance and therefore sliding velocities at that transition. We explore the impact of the bed-type transition on surge characteristics by incorporating a smooth transition from 0 % sediment cover (hard bedrock) to 100 % (soft) sediment cover effectively changing the basal sliding coefficient C in Eq. (6b).

Q_7 *How does a non-flat topography affect the surge behavior (Sec. 3.3.4)?*

Given the topographic lateral bounds of Hudson Strait, we examine the effects of a non-flat topography on the surge characteristics.

Q_8 *What is the effect of a simplified basal hydrology on surge characteristics in the GSM (Sec. 3.3.5)?*

The implementation of a fully-coupled basal hydrology model changes the basal drag and, therefore, has the potential to affect the surge characteristics. A basal hydrology model coupled to an effective-pressure dependent sliding law, or a Coulomb-plastic bed (as in PISM), introduces a positive feedback such that larger sliding speeds increase frictional heating, and thus meltwater availability which further weakens the bed and leads to even faster sliding. Different basal hydrology process representations have been proposed in the literature (e.g., a 0D (Gandy et al., 2019), poroelastic (Flowers et al., 2003), or linked cavity hydrology model (Werder et al., 2013)), and in-depth comparison is currently under ~~open~~ review (Drew and Tarasov, 2022). Here we compare GSM surge statistics with and without a fully coupled 0D hydrology model~~:~~.

Q_9 ~~*Q8 - What is*~~ *How significant are the* ~~*effect*~~ *details of* ~~*a simplified*~~ *the basal hydrology model on surge characteristics in PISM (Sec.* ~~**??**~~*S8.2)?* ~~*PISM event*~~

PISM surge characteristics are compared for local and mass-conserving horizontal transport hydrology models~~: *Q9 - How significant are the details of the basal hydrology model on surge characteristics in PISM (Sec.* **??***)?*~~ .

~~While both subglacial~~

Q_{10} *What are the differences (if any) in surge characteristics between local basal hydrology and a basal temperature ramp as the primary smoothing mechanism at the warm/cold-based transition zone (Sec. S8.3)?*

While both sub-glacial hydrology and a basal temperature ramp provide a means for a smooth increase in sliding velocities, these processes operate in slightly different temperature regimes. The basal temperature ramp enables sub-temperate sliding and the maximum velocities occur once the pressure melting point is reached. In contrast, a local basal hydrology model increases sliding velocities once the basal temperature reaches the pressure melting point (basal melting), and basal ice velocities further ramp up with decreasing effective pressure (ice overburden pressure minus basal water pressure). Note that ~~subglacial~~ sub-glacial hydrology is not an alternative for a basal temperature ramp. The ramp is still needed to prevent refreezing even when a description of ~~subglacial~~ sub-glacial hydrology is included (Mantelli et al., 2019). ~~As such: *Q10 - What are the differences (if any) in surge characteristics between local basal hydrology and a basal temperature ramp as the primary smoothing mechanism at the warm/cold based transition zone (Sec. **??**)?*~~

Convergence study

Q_{11} *Do model results converge (decreasing differences when increasing horizontal grid resolution and decreasing maximum time step, Sec. 3.4)?*

Significant surge pattern differences can occur when computationally more feasible (coarser) horizontal grid resolutions are used (e.g., $\sim$ 200-fold increase in the GSM run time when increasing the horizontal grid resolution from 25 km to 3.125 km). Incorporating the findings of the above experiments, we study numerical convergence with respect to horizontal grid resolution and time step for surge cycling. By convergence, we mean decreasing differences between simulations when increasing the resolution or decreasing the time step. True model ~~validation~~ verification can only come from comparison with continuum model results, which are, however, only available to a very limited degree and do not encompass the process complexity considered here, so we limit ourselves to a formal study of numerical convergence~~:~~ ~~*Q11 - Do model results converge (decreasing differences when increasing horizontal grid resolution and decreasing maximum time step, Sec. 3.4)?*~~ .

2 Methods

2.1 GSM

2.1.1 GSM model description

The 3D thermo-mechanically coupled glacial systems model (GSM) has developed over many years (e.g., Tarasov and Peltier, 1997; Tarasov et al., 2012; Bahadory and Tarasov, 2018). It includes an energy-conserving finite volume ice and bed thermodynamics solver. The current hybrid shallow shelf shallow ice physics ~~(Tarasov et al., manuscript in preparation)~~ is based on a slight variant of the ice dynamical core of Pollard and DeConto (2012). ~~The~~ As is standard for thermo-mechanically coupled

glaciological ice sheet models, the GSM has a default explicit time step coupling between the thermodynamics and ice dynamics but also includes an optional implicit coupling scheme (c.f. Sec. 3.2.2). Ice dynamical ~~time-stepping~~ time stepping is subject to CFL (Courant–Friedrichs–Lewy) constraint (Courant et al., 1928) with further reductions upon ice dynamical solver convergence failure. ~~A complete description of the GSM will soon be submitted for publication (Tarasov et al., manuscript in preparation).~~ The source code of the model version used in this manuscript can be found in the supplementary material (Tarasov et al., 2023).

The GSM is run with an idealized down-scaled North American geometry (Fig. 1, modified after the ISMIP-HEINO setup (Calov and Greve, 2006)) and simplified climate representation. The ~~temperature forcing is defined by a domain wide surface temperature (rTnorth, Tab. 1) and a specified vertical temperature gradient (~~surface temperature forcing in the GSM is given by

$$T_{\mathrm{surf}} = \mathrm{rTsurf} + \mathrm{lapsr} \cdot H + T_{\mathrm{asym}}, \tag{1}$$

where rTsurf and lapsr are input parameters for the domain-wide surface temperature constant and atmospheric lapse rate~~(lapsr in Tab.~~, respectively (Table 1)~~)~~. ~~The surface temperature forcing is asymmetric in time (~~, H the ice sheet thickness, and T_{asym} the asymmetric (in time) temperature forcing (maximum difference of 10°C, orange line in Fig. S1) ~~, enabling~~ calculated according to

$$T_{\mathrm{asym}} = \left| \left(\frac{t}{200\ \mathrm{kyr}} \cdot 3 + 2 \right) - 1 \right| \cdot 5^{\circ}\mathrm{C}, \tag{2}$$

where t is the model time ranging from -200 kyr to 0 kyr (instead of 0 kyr to 200 kyr). The asymmetric temperature forcing enables the analysis of the timing of cycling onset and termination under different physical and numerical conditions (a comparison of ice stream ice volume evolution under constant and ~~assymetric~~ asymmetric temperature forcing is shown in Fig. S2 for one parameter vector).

The surface mass balance forcing is then determined by

$$M_{\mathrm{tot}} = M_{\mathrm{acc}} - M_{\mathrm{melt}}, \tag{3}$$

where M_{acc} and M_{melt} are the surface accumulation and melt, respectively. The surface accumulation is defined by

$$M_{\mathrm{acc}} = \mathrm{precRef} \cdot \exp\left(\mathrm{hpre} \cdot T_{\mathrm{surf}}\right), \tag{4}$$

where precRef and hpre are the precipitation coefficient input parameters. Surface melt is calculated according to a Positive Degree Day (PDD) approach:

$$M_{\mathrm{melt}} = \mathrm{rPDDmelt} \cdot \max\left(0.0, \mathrm{POSdays} \cdot (T_{\mathrm{surf}} + 10.0^{\circ}\mathrm{C})\right), \tag{5}$$

where rPDDmelt is the input parameter for melt per PDD and the PDD constant POSdays is set to 100 days yr^{-1}. Note that we set $T_{\mathrm{surf}} = 0.1^{\circ}$C and $M_{\mathrm{tot}} = -100$ m yr^{-1} for ocean grid cells, and $T_{\mathrm{surf}} = 0.1^{\circ}$C and $M_{\mathrm{tot}} = -200$ m yr^{-1} at the boundaries of the model domain.

The GSM is initialized from ice-free conditions and the ~~model runs last for 200 kyr, of which the~~ first 20 kyr of each model run are considered as spin-up interval, which is not included in the analysis of the surge characteristics. Note that this is a very conservative spin-up interval~~and that most~~. Most runs reach their mean pseudo-Hudson Strait ice volume after ~ 5 kyr (e.g., Fig. 11). The coarsest horizontal grid resolution is 25x25 km and is progressively refined (halved) to 3.125x3.125 km. This gives a total of 4 different horizontal grid resolutions. The maximum time step is 1 yr (automatically decreased as needed to meet CFL constraint or when convergence fails).

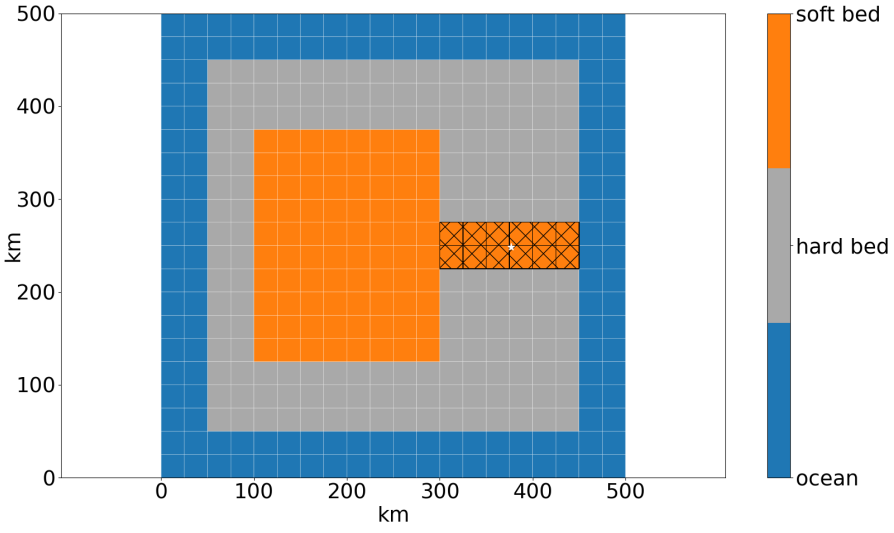


Figure 1. Modified ISMIP-HEINO geometry (Calov and Greve, 2006). The model domain is reduced to 500x500 km to enable horizontal grid resolutions up to 3.125 km. The shown grid resolution is 25x25 km. The basal topography is flat and the hatched area marks the soft-bedded pseudo-Hudson Strait. The white star indicates the location of the grid cell shown in Fig. 8+S21.

While Mantelli et al. (2019) conclude that Stokes mechanics are needed to arrive at a mathematically well-posed model, running numerical experiments with a thermo-mechanically coupled Stokes model is to date unfeasible for a glacial cycle ~~contexts~~context. Previous ~~HE~~ ice stream surge modeling studies are often based on zeroth-order, thin-film approximations of the Stokes problem, like the shallow-ice Approximation (SIA, e.g., 8 out of 9 models in the ISMIP HEINO experiments (Calov et al., 2010)). While resolving vertical shear, which is the dominant mode of motion in slow flowing regions, SIA-based models neglect longitudinal stress gradients and horizontal shear, which are known to be important for fast ice streams (Hindmarsh, 2009) and are instead captured by the zeroth-order shallow-shelf approximation (SSA).

To ~~improve on the drawbacks~~ partially offset the limitations of the zeroth-order approximations, the GSM uses hybrid SIA/SSA ice dynamics. This heuristic combination links the two sets of equations by including a shear softening term in the calculation of the effective viscosity in the respective other set (SIA internal shear in the SSA viscosity calculation and SSA vertical-mean longitudinal stretching in the SIA viscosity calculation) (Pollard and DeConto, 2007, 2012). Additionally,

horizontal shear and longitudinal stress gradient terms from the SSA equations reduce the driving stress in the SIA equations (Pollard and DeConto, 2007, 2012). A third coupling option adds the distinction between the depth-averaged internal-shear and basal velocity to the SSA basal stress term (Pollard and DeConto, 2007, 2012). This coupling term, however, tends to weaken numerical convergence without having much impact on ice sheet history and was therefore not used for the experiments in this paper. The hybrid SIA/SSA ice dynamics are activated for grid cells with a SIA velocity exceeding $30\ \mathrm{m\ yr^{-1}}$ ~~for soft beds and $200\ \mathrm{m\ yr^{-1}}$ for hard beds. The full set of equations is described in Tarasov et al. (manuscript in preparation).~~ . Changing these activation velocities has no significant effect on the surge characteristics. Activating the SSA everywhere leads to more, shorter, and weaker surges because no threshold velocity needs to be overcome to initiate basal sliding (Sec. S1.2). Note that we set an upper limit of $40\ \mathrm{km\ yr^{-1}}$ for the SSA velocity to ensure that sliding velocities stay within a physically reasonable range.

We set the GSM with a $1\ \mathrm{km}$ deep (17 non-linearly-spaced levels) bed thermal model. A basal temperature ramp is used to ensure a smooth transition between cold-based regions of no sliding and temperate sliding, account for observational evidence of sub-temperate sliding, and more accurately represent the sub-grid warm-based ~~area under an ice sheet and therefore~~ ice fraction in a grid cell and therefore more accurately represent sliding onset for coarse grid resolutions (Q_5 in Sec. 1.3). However, the shape of such a basal temperature ramp is not well constrained. In the GSM, the basal temperature ramp is incorporated into a Weertman-type power law

$$\boldsymbol{u_b} = \boldsymbol{C_b}\,|\boldsymbol{\tau_b}|^{n_b-1}\,\boldsymbol{\tau_b} \tag{6a}$$

as a dependence of the basal sliding coefficient $\boldsymbol{C_b}$ on the estimated warm-based fraction of a grid cell (indirectly accounting for sub-temperate sliding) F_{warm} (Eq. (8))

$$\boldsymbol{C_b} = (1 - F_{warm})\,C_{froz} + F_{warm}\boldsymbol{C}, \tag{6b}$$

where $\boldsymbol{u_b}$ is the basal sliding velocity, $\boldsymbol{\tau_b}$ the basal stress, n_b the bed power strength (~~Tab.~~Table 1), and $\boldsymbol{C}$ the fully warm-based sliding coefficient (depends on the bed properties, see also Fig. S4). C_{froz} is the fully cold-based sliding coefficient for numerical regularization:

$$C_{froz} = 2 \cdot 10^{-3}\ \mathrm{m\ yr^{-1}} \left(5 \cdot 10^{-6}\ \mathrm{Pa^{-1}}\right)^{n_b}. \tag{7}$$

F_{warm} is calculated according to

$$F_{warm} = \max\left[0, \min\left(1, \frac{T_{bp,I} + T_{ramp}}{T_{ramp}}\right)\right]^{T_{exp}}, \tag{8}$$

where $T_{bp,I}$ is the grid cell interface basal temperature relative to the pressure melting point, negative T_{ramp} the temperature below which the entire grid cell is cold-based, and T_{exp} the exponent used for the ramp. The ~~legacy values~~ values used in previous GSM modeling studies ($T_{ramp} = 1.0\ ^{\circ}\mathrm{C}$ and $T_{exp} = 28$ (e.g., Bahadory and Tarasov, 2018)) were based on horizontal basal temperature gradients around the basal sliding activation zone ~~and arguments of~~ with consideration of the sub-grid

warm-based connectivity between grid cell interfaces (as basal sliding requires a connected ~~subgrid~~ sub-grid warm-based path). Different values for T_{ramp} and T_{exp} are explored within this paper. T_{ramp} can be chosen as either constant or depending on the horizontal grid resolution (res, equal extent in x- and y-direction):

$$T_{ramp} = FP_{T_{ramp}} \cdot \frac{\text{res}}{50\ \text{km}}\ ^{\circ}\text{C} \tag{9}$$

This choice of resolution dependence leads to a sharper temperature ramp for higher horizontal grid resolutions. The ~~factor $F_{T_{ramp}}$~~ parameter $P_{T_{ramp}}$ is used to conduct experiments with different temperature ramps at the same horizontal grid resolution (Sec. 3.3.3). The temperature ramps for all 4 horizontal grid resolutions and ~~$F_{T_{ramp}} = 1$~~ $P_{T_{ramp}} = 1$ (default value) are shown in Fig. 2. ~~A~~ For comparison, a temperature ramp similar to the one suggested by Fowler (1986) and later Mantelli et al. (2019)

$$F_{warm} = \exp\left(\frac{T_{bp,I}}{\delta}\right) \quad \text{for} \quad T_{bp,I} \leq 0 \tag{10}$$

is shown for $\delta = 0.01$, where δ is a parameter controlling the width of the transition zone. Based on experiments conducted by Barnes et al. (1971), Mantelli et al. (2019) expect δ to be small.

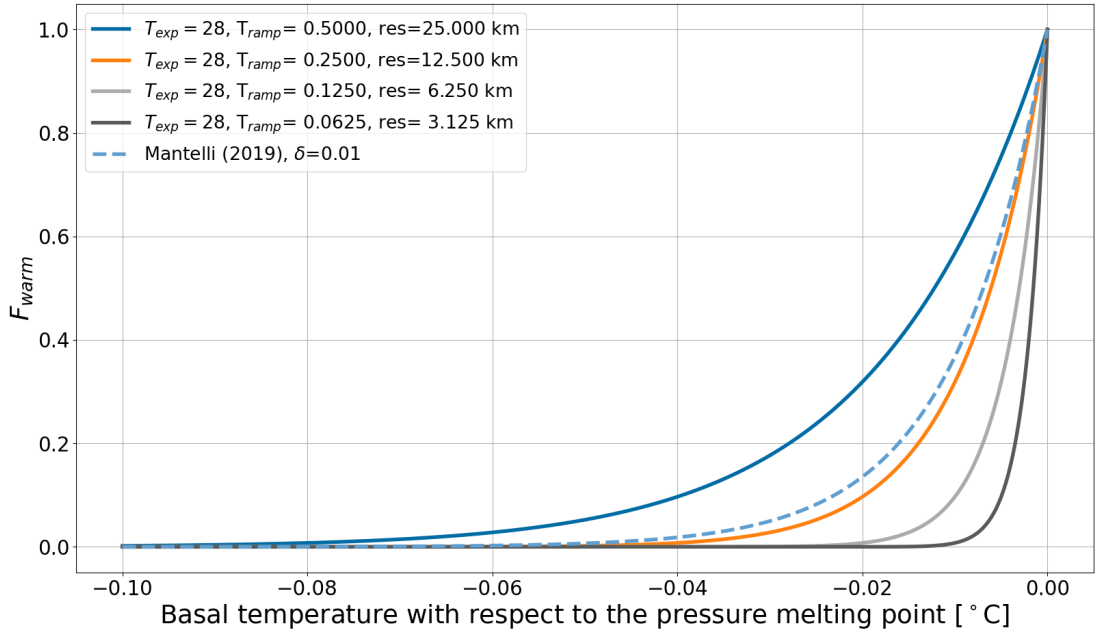


Figure 2. Temperature ramps for different values of T_{ramp} which depend on the horizontal grid resolution. A temperature ramp similar to the one suggested by Mantelli et al. (2019) (Eq. (10)) is shown for $\delta = 0.01$.

2.1.2 GSM ensemble input parameter vectors

Each GSM experiment is run with an ensemble based on 5 input parameter vectors. The current idealized setup encompasses a maximum of 8 input parameters (~~Tab.~~Table 1) per parameter vector. The 5 ~~base~~ parameter vectors used in this study are hand-picked from an exploratory ensemble (Fig. S3). The criteria for these 5 parameter vectors was the highest subset variance in ~~HE~~ surge characteristics and soft bed sliding law exponent. Note that the soft and hard bed sliding law ~~exponent~~ exponents in this study are equal (n_b in ~~Tab.~~Table 1). Due to the significantly increased model run time, exponents larger than 3 are not considered here. ~~The GSM base~~ To isolate interactions, the GSM reference setup used in this paper does not incorporate basal hydrology and glacial isostatic adjustment (GIA). ~~Determining the effects of GIA is beyond the scope of this paper and will be addressed in future work.~~ Processes associated with basal hydrology, such as lubrication of the bed and decoupling of the ice sheet from the bed, are likely to have a major effect on surge patterns. To determine the impact of these effects, we run ~~a GSM setup~~ the GSM with local basal hydrology enabled (Eq. (??20) to (??22), Sec. ??3.3.5) and examine resolution scaling (Sec. ??S9.2). However, experiments done with and without basal hydrology lead to qualitatively similar results (e.g., same conclusions from upscaling experiments in Sec. 3.3.3). We therefore omit ~~subglacial~~ sub-glacial hydrology coupling for the main analysis.

2.1.3 GSM model setups

The ~~base setup (Tab.~~reference setup (Table 2) has a 3.125 km horizontal grid resolution and 1 year maximum time step. The bed topography is flat (at sea level) and an asymmetric temperature forcing is used (Fig. S1). For the sake of generality, we chose a flat topography for the ~~base~~ reference setup, while the effect of a basal trough is investigated at a later stage (Sec. 3.3.4). Branching off this ~~base~~ reference set-up, we carry out ~~one at a time~~ one-at-a-time sensitivity experiments to isolate numerical and process impacts. These experiments~~in turn examine: sediment cover and~~ , in turn, examine the response to: 3 numerical aspects related to the MNEEs, 4 model aspects affecting the thermal onset of basal sliding, a change in sediment cover, a non-flat topography~~(Sec. S1.4), a~~ , the addition of local basal hydrology, different horizontal grid resolutions [25 km, 12.5 km, 6.25 km], and different maximum time steps [0.5 year, 0.25 year]. The 3 numerical aspects are stricter numerical convergence criteria, the addition of surface temperature noise (± 0.1°C and ± 0.5°C), and an approximate implicit time-step coupling between the thermodynamics and ice dynamics. The 4 thermal model aspects are switching to thin (20 m) bed thermal model, ~~approach~~ different approaches to determining the basal temperature at the grid cell interface~~(Sec. ??), weight~~ , different weights of the adjacent minimum basal temperature for the basal sliding temperature ramp ($W_{Tb,\mathrm{min}}$~~, Sec. S1.5), basal temperature ramp~~), and different basal temperature ramps (T_{ramp} and T_{exp}) ~~, horizontal grid resolution 25 km, 12.5 km, 6.25 km, maximum time step 0.5 year, 0.25 year, and local basal hydrology (Sec. ??). See Tab.~~ for thermal activation of basal sliding. See Table 1 for details ~~in~~ on parameter ranges.

Category	Parameter	Description	Range	Unit
Ensemble parameter - ISM	C_{rmu}	soft bed sliding coefficient	0.3 - 1	
	C_{fslid}	hard bed sliding coefficient	0 - 3	
	lapsr	atmospheric lapse rate	-5 - -10	~~K~~ °C km^{-1}
	PDDmelt	melt per Positive Degree Day (PDD)	0.005 - 0.012	m PDD^{-1} (°C)$^{-1}$
	hpre	precipitation coefficient	0.02 - 0.2	(°C)$^{-1}$
	PrecRef	precipitation coefficient	1 - 3	m yr^{-1}
	~~rTnorth~~ rTsurf	~~northern~~ domain wide surface temperature constant	-9 - -15	°C
	n_b	soft and hard bed sliding law exponent, bed power strength	1 - 3	
Hydrology parameters	$\text{h}_{\text{wb,Crit}}$	effective bed roughness scale (Eq. (??20))	0.01 - 1	m
	rBedDrainRate	constant bed drainage rate	0.001 - 0.01	m yr^{-1}
	$N_{\text{eff,Fact}}$	effective pressure factor (Eq. (??22))	$2 \cdot 10^4$ - $2 \cdot 10^5$	Pa
Additional parameters	~~$F_{T_{ramp}}$~~ $P_{T_{ramp}}$	basal temperature ramp scaling factor (Eq. (9))	0.125 - 16 (**1**)	
	T_{ramp}	basal temperature (with respect to the pressure melting point) at which sub-temperate sliding becomes important (Eq. (8), (9))	0.03125 - 1 (**0.0625**)	°C
	T_{exp}	basal temperature ramp exponent (Eq. (8))	5 - 56 (**28**)	
	$W_{Tb,\text{min}}$	weight of adjacent minimum basal temperature for basal sliding temperature ramp (Eq. (S1))	0.0 - 1.0 (**0.5**)	

Table 1. Model parameters are listed with respect to their purpose/category. Ice Sheet Model - ISM. Hydrology parameters used when running the GSM with local basal hydrology. Additional (non-regular) input parameters that are usually set to a fixed value. The default values of the 3.125 km horizontal grid resolution ~~base~~ reference setup are shown in the brackets for the additional parameters.

2.2 PISM

2.2.1 PISM model description

~~The GSMis an ice sheet model developed specifically~~ In contrast to the GSM, the Parallel Ice Sheet Model (PISM) is not specifically developed for glacial cycle ensemble ~~modelling. The GSM is therefore numerically optimized~~ modeling. Therefore, the two models use distinct sets of numerical optimizations for computational speed. To ~~increase confidence in the GSM model results and reduce~~ minimize the model dependency of our analysis, experiments are also carried out with v2.0.2 of the ~~Parallel Ice Sheet Model (PISM)~~PISM. Note that these experiments are not intended to be a direct comparison of the two models, but rather to show that the same conclusions can be achieved with different models (despite their differences in model setups, physics and numerics).

Similar to the GSM, PISM is a 3D thermodynamically-coupled ice sheet model and the SSA is used as a 'sliding law' once the sliding velocity exceeds 100 m yr^{-1}. For further details on the model itself, refer to Bueler and Brown (2009); Winkelmann et al. (2011). The details on the default PISM setup, together with the default GSM values, are listed in ~~Tab.~~Table 2. Given the higher computational cost of the PISM experiments, the relatively high sensitivity of PISM to the number of parallelized cores for these experiments (Table 6), and run time limitations of the computational cluster, the ~~base~~ reference setup is run at 25 km horizontal grid resolution.

For stability reasons, the PISM adaptive time stepping ratio (used in the explicit scheme for the mass balance equation) was reduced to 0.01 when using small till friction angles (Constantine Khrulev, personal communication).

The default sliding law in PISM is a purely-plastic (Coulomb) model where

$$|\boldsymbol{\tau_b}| \leq \tau_c \quad \text{and} \quad \boldsymbol{\tau_b} = -\tau_c \frac{\boldsymbol{u}}{|\boldsymbol{u}|} \quad \text{if} \quad |\boldsymbol{u}| > 0. \tag{11}$$

Therefore, the basal shear stress $\boldsymbol{\tau_b}$ can never exceed the yield stress τ_c, and basal sliding only occurs when $\boldsymbol{\tau_b}$ reaches τ_c.

Setup component	GSM	PISM
horizontal grid resolution	3.125 km x 3.125 km	25 km x 25 km
number of grid cells	160 x 160	120 x 120
model domain	500 km x 500 km	3000 km x 3000 km
vertical layers	65	60
run time	200 kyr	200 kyr
maximum time step	1 yr	1 yr
number of cores/processes	1	8
ice dynamics	hybrid SIA/SSA	hybrid SIA/SSA (maximum SIA diffusivity of 1000 m^2 s^{-1})
sliding law	Weertman-type power law (Eq. (6a))	Coulomb friction law (Eq. (11))
bed topography	flat (at sea level)	flat (at sea level)
bed thermal model	1 km deep (17 non-linearly-spaced levels)	1 km deep (20 equally-spaced levels)
basal hydrology	not included	local basal hydrology model based on an undrained plastic bed model (Tulaczyk et al., 2000a)

Table 2. Comparison between the GSM and PISM ~~base~~ reference setup.

2.2.2 PISM ensemble input parameter vectors

The PISM configuration encompasses 6 model input parameters (~~Tab.~~Table 3). These parameters define the input fields for surface temperature, surface accumulation, and till friction angle. Similar to Calov and Greve (2006), the surface temperature

at every grid cell is calculated as

$$T_{surf} = T_{min} + S_t \cdot d^3, \tag{12}$$

where ~~S_b~~ S_t represents the horizontal surface temperature gradient, d the distance from the domain center (x_{center},y_{center}) in km, defined as:

$$d = \sqrt{(x - x_{center})^2 + (y - y_{center})^2} < R, \tag{13}$$

and R denotes the radius and sets an upper limit for d. A comparable equation is used to calculate the surface mass balance (accumulation/ablation) rate input field.

$$B_{surf} = B_{max} - S_b \cdot d^5, \tag{14}$$

where S_b is the horizontal surface mass balance gradient. The input field for the till friction angle is defined by simple grid assignment and a somewhat smoothed transition between the soft and hard bed region. Input fields for one parameter vector are shown for surface temperature, surface accumulation, and till friction angle in Fig. S6, S7, and S8, respectively.

Category	Parameter	Description	Range	Unit
Ensemble parameters	$soft$	soft bed till friction angle	0.5 - 12.0	°
	$hard$	hard bed till friction angle	15.0 - 30.0	°
	B_{max}	maximum surface mass balance (accumulation/ablation) rate	50 - 450	kg m^{-2} yr^{-1}
	S_b	horizontal surface mass balance gradient	(0.15 - 1.00) $\cdot 10^{-11}$	kg m^{-2} yr^{-1} km^{-5}
	T_{min}	minimum surface temperature	220 - 245	K
	S_t	horizontal surface temperature gradient	(0.10 - 1.0) $\cdot 10^{-8}$	K km^{-3}
Constant parameters	x_{center}	location of the domain center in x-direction	1500	km
	y_{center}	location of the domain center in y-direction	1500	km
	R	maximum radius of the domain	1500	km

Table 3. Parameters used to generate the PISM input fields.

The 6 model ensemble parameters (~~Tab.~~Table 3) were ~~sampled via a Latin-Hypercube for specifying a 100 run ensemble. After filtering for runs~~ selected via Latin Hypercube sampling. After sieving an ensemble of 100 runs for those that show oscillatory behavior, a ~~10-member~~ 9-member high-variance (with respect to the surge characteristics) subset was extracted by visual identification (Fig. S10). Each PISM experiment is run with an ensemble based on these 9 input parameter vectors.

2.2.3 PISM bed properties

~~Note that~~ A PISM ensemble parameter restriction arose as experiments carried out with PISM only show oscillatory behavior for small yield stresses τ_c. This can be achieved by either a small till friction angle Φ or low effective pressure on the till (N_{till}) (Bueler and Van Pelt, 2015):

$$\tau_c = c_0 + \tan(\Phi)\, N_{\text{till}}, \tag{15}$$

where $c_0 = 0$ Pa is the till cohesion (Tulaczyk et al., 2000b). N_{till} is given by

$$N_{\text{till}} = N_0 \left(\frac{\delta_e P_0}{N_0} \right)^s 10^{\left(\frac{e_0}{C_c}\right)(1-s)}, \tag{16}$$

where $N_0 = 1$ kPa is the reference effective pressure, $e_0 = 0.69$ the void ratio at N_0, $C_c = 0.12$ the dimensionless coefficient of compressibility, δ_e the effective fraction of the overburden pressure, P_0 the ice overburden pressure, and s the ratio $\frac{W_{\text{till}}}{W_{\text{till}}^{\max}}$ (Tulaczyk et al., 2000b; Bueler and Van Pelt, 2015). W_{till} and $W_{\text{till}}^{\max} = 2$ m are the effective and maximum thickness of water in the till, respectively. The values listed here are the PISM defaults. C_c is on the lower end of measured values (Tulaczyk et al., 2000b) with significantly larger (up to 17) values reported (Sauer et al., 1993; Mitchell and Soga, 2005). e_0 can vary between 0.45 (Tulaczyk et al., 2000b) and approximately 4 (Fig. 10.2 in Mitchell and Soga, 2005). The default value of δ_e is based on Greenland and Antarctic model runs, but δ_e is generally considered as a tuning parameter to match observed surface velocities, which are not available in a paleo context (Andy Aschwanden, personal communication).

When only changing the till friction angle, oscillations do not occur unless $\Phi < 1°$ (Fig. S13). This is well below the measured values of about 10 to 40° (K.M. Cuffey and W.S.B. Paterson., 2010). However, similar oscillatory results are obtained for till friction angles between 5 and 10° when slightly adjusting the values of $C_c = 0.2$, $e_0 = 0.6$, and $\delta_e = 0.01$ to favor sliding (compare Fig. S11 and S12). These values are all well within the ranges set by laboratory measurements. For convenience, we decide to vary only the till friction angle between 0.5 and 1°, for which PISM shows oscillatory behavior, and otherwise use the PISM default values.

~~However, these experiments incorporate~~ These resulting very slippery beds ~~enabling~~ enabled occasional maximum sliding velocities of up to ~ 600 km yr^{-1} ~~for single grid cells and time steps~~ in the simulations (Fig. S11, Sec. S2.3). For comparison, observed ~~surge velocities range from tens of meters per year for several years to hundreds of meters per day for short periods (K.M. Cuffey and W.S.B. Paterson., 2010).~~ outlet glacier velocities at Jakobshavn Isbræ(Greenland) approach 20 km yr^{-1} (Joughin et al., 2012, 2014). Similar to the GSM, we, therefore, set an upper limit of 40 km yr^{-1} for the SSA velocity.

~~Excluding runs that show maximum sliding velocities > 50 km yr^{-1} from the analysis yields similar results to the full 10-member ensemble (Sec. S2.3 and Fig. **??**), indicating a stable solution of the numerical matrix solver even for runs with very high velocities. In addition, the 50 km yr^{-1} is exceeded no more than 7 times per 200 kyr run (100 yr output) and the maximum sliding velocities are generally within the observed range (Fig. S11).~~

2.2.4 PISM model setups

~~The default PISM setup (Tab.~~As for the GSM, we carry out one-at-a-time sensitivity experiments branching off the PISM reference set-up (Table 2) ~~is rerun for all 10 parameter vectorswith different number of cores/processes ($n = 2, 4, 16, 32$), without a bed thermal model~~ for all 9 parameter vectors. These experiments, in turn, examine the response to: 2 numerical aspects related to the MNEEs, removing the bed thermal model, an abrupt sediment transition zone, a non-flat topography (Fig. S9), a mass-conserving horizontal transport model for basal hydrology (Bueler and Van Pelt, 2015), different horizontal grid resolutions (50 km,12.5 km), and a different maximum time step (0.5 and 0.25 year)~~, and with a mass-conserving horizontal transport model for basal hydrology (Bueler and Van Pelt, 2015).~~ . The 2 numerical aspects are different number of cores/processes ($n = 2, 4, 16, 32$), and addition of surface temperature noise (± 0.1°C and ± 0.5°C).

2.3 Run analysis approach

For both models, we use the Python module *scipy* (version 1.5.2 on GSM cluster and 1.7.0 on PISM cluster, different versions due to the availability on computational clusters) and its built-in function *scipy.signal.find_peaks* on the ice volume output to determine the ~~event~~ surge characteristics. The standard output time steps in the GSM and PISM are 0.1 and 1 kyr, respectively. Note that these time steps might not exactly capture the minimum ice volume but are generally a good compromise between storage requirements and temporal resolution (e.g., Fig. S16 and S17). The Python analysis scripts are provided as supplementary material.

Surge characteristics

The quantities being analyzed are: the number of surges, the surge duration, the ice volume change during a surge~~.~~ , and the period between surges ~~. The event~~ (Fig. 3). The surge time is defined as the time of minimum (pseudo-Hudson Strait) ice volume, and the duration of ~~an event~~ a surge includes the surge itself as well as the time it takes the ice sheet to recover approximately half the ice volume lost during the surge (as per *scipy.signal.peak_widths*). The calculated ice volume change is the difference between the ~~pre-event~~ pre-surge and minimum (pseudo-Hudson Strait) ice volume in that particular ~~event~~ surge (as per *scipy.signal.peak_prominences*). The period between ~~events~~ surges is the time span between two subsequent occurrences of minimum (pseudo-Hudson Strait) ice volume (not defined for the very last event). The spin-up interval (first 20 kyr of every run) is not incorporated in the analysis, and only ~~events~~ surges with a (pseudo-Hudson Strait) ice volume change of more than 500 km^3 and 40^4 km^3 are considered in the GSM and PISM analyses, respecitvely (~ 5 % of mean ice volume across all runs). ~~A graphical illustration of the different event characteristics can be found in Fig. **??**.~~

~~Additionally~~

In addition to the surge characteristics, the Root Mean Square Error (RMSE) and mean bias are calculated as a percentage deviation from the ~~base~~ reference (pseudo-Hudson Strait) ice volume time series for all setups (each parameter individually) and then averaged over the 5 parameter vectors (Eq. (S2) and (S3)). The full run time is considered (no spin-up interval).

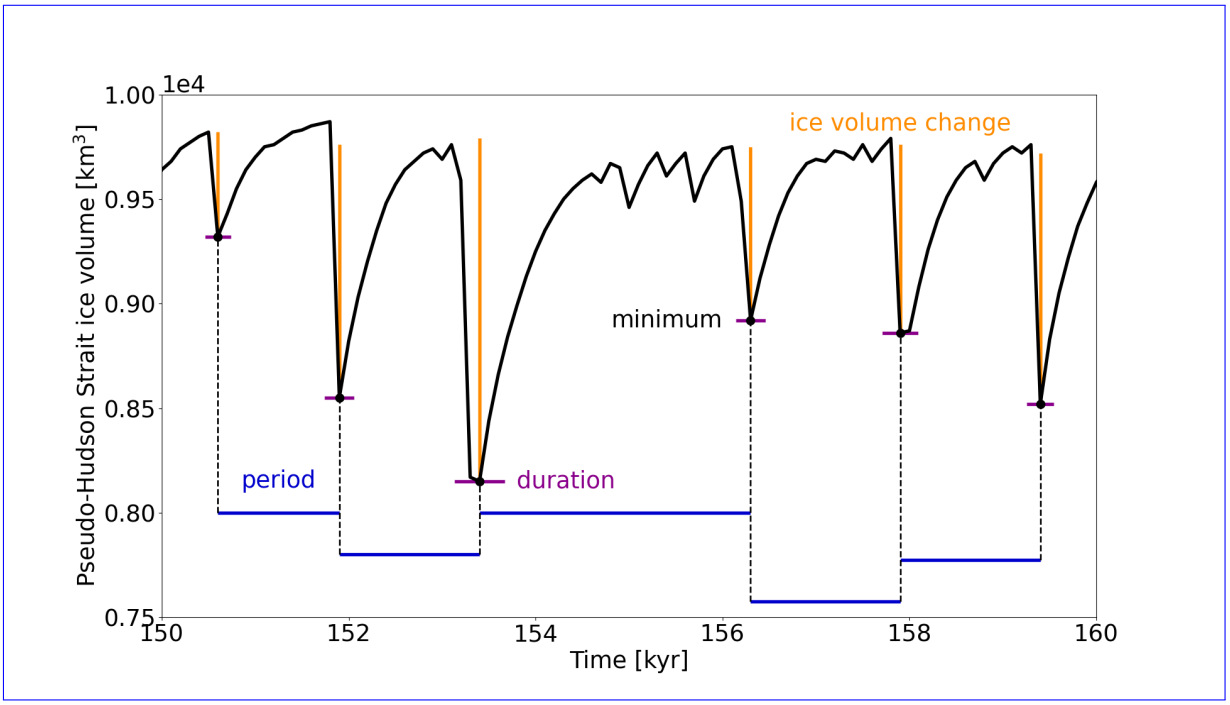


Figure 3. Pseudo-Hudson Strait ice volume of a GSM model run with visual illustration of the surge characteristics used to compare different model setups. The horizontal grid resolution is 3.125 km.

Percentage differences

We compare different model setups by calculating the percentage difference between the ~~base~~ reference setup and all other setups for every parameter vector individually and then average this difference over the 5 parameter vectors. Crashed runs are not considered and runs with less than 2 ~~events~~ surges require special treatment (see Sec. S5 for further details on the analysis).

Surge area

In the GSM, the whole pseudo-Hudson Strait (Fig. 1) is ice-covered and at maximum ice volume at the beginning of a surge. Surges in the GSM, therefore, consistently appear as ice volume minima, which allows us to directly use the pseudo-Hudson Strait ice volume for the GSM results.

~~Due to the setup of the PISMmodel domain (Sec. 2.2.1 and S2.1)~~For PISM, a large fraction of the pseudo-Hudson Strait area is ~~ice-free when no surge occurs~~ only ice-covered when a surge occurs (e.g., Fig. 5), leading to an inconsistency in the surge detection. This issue ~~can be avoided~~ is addressed by including the ice volume over the eastern half of the pseudo-Hudson Bay, the area most affected by the surge drained through the pseudo-Hudson Strait. See Sec. S2.4 for further details and a comparison between the two approaches.

Minimum numerical error estimates

We compute the new '**M**inimum **N**umerical **E**rror **E**stimates' (MNEEs) metric by examining the model response to changes in the model configuration that are not part of the physical system. The MNEEs are defined as the percentage differences in surge characteristics when applying a stricter (than default) numerical convergence in the GSM and adjusting the matrix solver used in PISM (changing the number of processor cores used). They are then used as a threshold to determine if model sensitivities to changes in the model configuration that affect the physical system (e.g., the inclusion of a bed thermal model or sliding dependence on effective pressure from basal hydrology) are above the level of background noise induced by iterative numerical solvers in the model. We refrain from drawing conclusions about the effects of a change in model configuration with physical relevance when the model sensitivities in question are smaller than the MNEEs. In these cases, the actual physical response of the model might be hidden within the numerics.

While the MNEEs are useful to our purpose, we wish to emphasize that they can not replace proper model verification and validation and are missing uncertainties due to, e.g., different approximations of the Stokes equations and other physical processes not included in the models. Nonetheless, they provide a minimum estimate of the numerical model error, which is still a significant improvement over ignoring this issue entirely.

3 Results

3.1 Key ~~features~~ surge characteristics of the ~~base~~ reference setup

3.1.1 Surge onset, propagation, and termination

Before analyzing ensemble characteristics, it is crucial to understand how surges initiate, propagate and terminate. Surges in the GSM originate at the pseudo-Hudson Strait mouth ($x = 450$ km, $y = 225$ to 275 km) and propagate towards the center of the pseudo-Hudson Bay ($x = 200$ km, $y = 250$ km, Fig. 1 and 4). The surging onset is a complex interplay between heating at the ice sheet bed, basal temperature, and ice sheet velocity. The beginning of a surge is shown in an online video (video 01 of Hank (2023)) and Fig. 4. Just before the start of the surge, the entire South-North extent of pseudo-Hudson Strait grid cells close to the ocean is warm-based. At $t = 6.69$ kyr, the SIA velocities exceed 30 m yr^{-1} and the SSA is activated (Sec. 2.1.1)~~, and the~~. The longitudinal stress gradient and horizontal shear terms provide additional heating (heating due to shelfy stream dynamics in video 01 of Hank (2023)). This leads to several small ice streams with relatively strong heating due to ~~the basal shear stress~~ basal sliding ($\sim 10^7$ J m^{-2} yr^{-1}) at $t = 6.70$ kyr in the video. This is an order of magnitude larger than heat production from deformation work. The additional heat fosters higher ice velocities, leading to even more heating, the extension of the warm-based area to the West, and therefore the upstream propagation of the small ice streams ($t = 6.71$ kyr). The narrow ice streams draw in warm-based ice from the surrounding grid cells, increasing the velocities and heat production in the area between the ice streams. This leads to a merger of the ice streams with now high velocities occurring over the full South-North extend of

the pseudo-Hudson Strait ($t = 6.72$ kyr). The warm-based area rapidly extends towards the West due to the strong heating and high ice velocities, causing a pseudo-Hudson Strait surge.

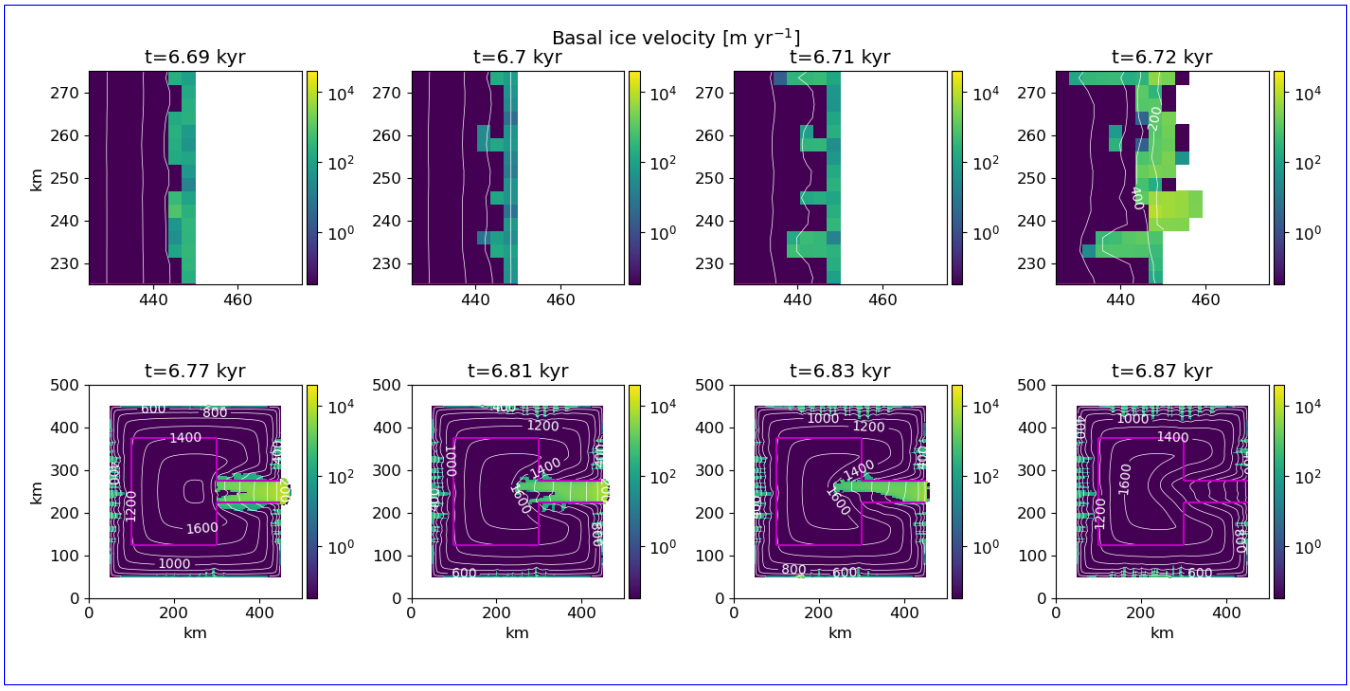


Figure 4. Basal ice velocity for parameter vector 1 at different time steps using the GSM. The horizontal grid resolution is 3.125 km and the maximum model time step is 1 yr. The contour lines show the ice sheet surface elevation in m. The magenta line outlines the soft-bedded pseudo-Hudson Bay and Hudson Strait. Note that the top and bottom rows show different areas of the domain, with the top ~~being zoomed~~ zooming in on the surge onset area.

The surge propagates nearly symmetrically until the pseudo-Hudson Bay area is reached ($t = 6.77$ kyr in Fig. 4 and video 02 of Hank (2023)). After this point, the northern branch of the ice stream propagates more rapidly and extends further to the West than the southern branch. While the smaller southern branch starts to shrink at $t = 6.81$ kyr, the northern part propagates until $t = 6.83$ kyr. At this time, the southern branch has vanished almost completely due to a thinner ice sheet (than at the start of the surge) and the advection of cold ice into the surge area. After $t = 6.83$ kyr, the available heating is no longer sufficient to keep the ice sheet bed at the pressure melting point, and the northern part collapses as well. The surge ends after 150 yr (at $t = 6.87$ kyr).

Since the GSM setup and climate forcing are symmetric about the horizontal axis in the middle of the pseudo-Hudson Strait ($y = 250$ km in Fig. 1), we interpret the ~~the~~ induced asymmetry as a numerical ~~noise~~ induced bifurcation. We define the asymmetry as positive when the surge is stronger Northward (Fig. 4 and video 02 of Hank (2023)) or shifted Northward. The asymmetry sign varies across the first ~~surge events~~ surges (i.e., the ~~event~~ surge least biased by previous asymmetries) ~~across~~ of the 5 ~~base~~ reference runs, ruling out any persistent numerical bias.

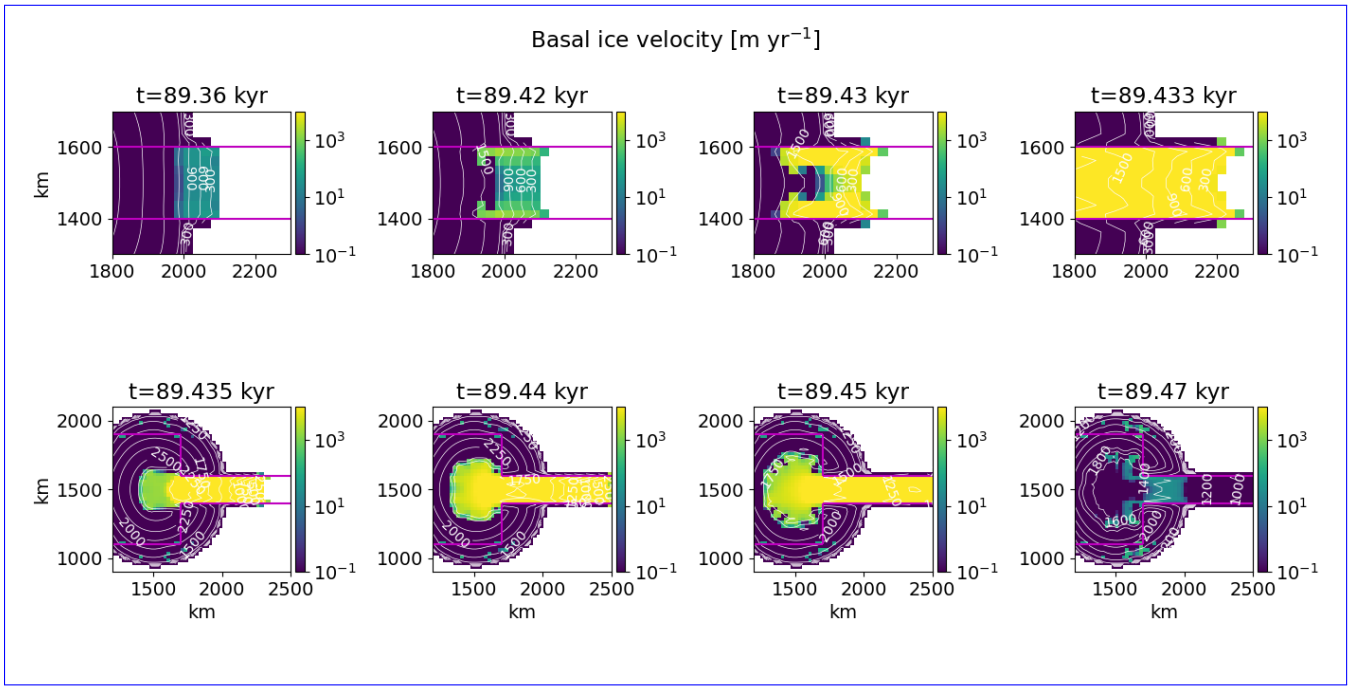


Figure 5. Basal ice velocity for parameter vector 8 at different time steps using the PISM. The horizontal grid resolution is 25 km and the maximum model time step is 1 yr. Otherwise as in Fig. 4.

Surges in the PISM originate at the ice sheet margin in the soft-bedded pseudo-Hudson Strait (exact position varies between runs) and propagate towards the center of the pseudo-Hudson Bay ($x = 1300$ km, $y = 1500$ km, Fig. S8 and 5). The ice near the margin is already flowing downstream before the start of the surge ($t = 89.36$ kyr). However, the basal temperature is below the pressure melting point, and the ice velocities are low (< 100 m yr^{-1}). As the ice sheet upstream of the margin thickens, the warm-based area extends further downstream, particularly along the 100 % soft-bedded contour line (magenta line in Fig. 5). Once the warm-based area connects with the margin ($t = 89.42$ kyr), the ice velocities increase beyond 100 m yr^{-1}, activating the SSA (Sec. 2.2.1). Similar to the surges in the GSM, the sliding velocities then increase rapidly, quickly extending the warm-based area ($t = 89.43$ kyr and $t = 89.433$ kyr). The surge propagates upstream into the pseudo-Hudson Bay and the ice is transported along the pseudo-Hudson Strait into regions with increasingly negative surface mass balance rates ($t = 89.435$ kyr to $t = 89.45$ kyr, Fig. S7). The ice sheet thins, the basal temperature at the margin falls below the pressure melting point, blocking parts of the upstream ice stream, and the surge ceases at $t = 89.47$ kyr (~ 100 yr surge duration). The ice volume in the surge-affected area continues to decrease for, on average, another 2.5 kyr due to the large amounts of ice in the negative surface mass balance regions. In contrast to the GSM, the PISM results remain symmetrical about $y = 1500$ km throughout the surge.

3.1.2 ~~Event~~ Surge characteristics of the GSM and PISM ~~base~~ reference setup

Due to the differences in model setup, physics, and numerics (~~Tab.~~Table 2), the GSM and PISM ~~base~~ reference setup yield different ~~event characteristics (Tab.~~surge characteristics (Table 4). While resembling the inferred ice-rafted debris (IRD) interval duration as closely as possible is not a goal of this study, the modeled values are in agreement with the literature (200 to 2280 yr (Hemming, 2004)). ~~Due to the downscaled GSM domain, the~~ The mean modeled GSM period is shorter than the observed period of, on average, 7 kyr (K.M. Cuffey and W.S.B. Paterson., 2010). However, exploratory GSM runs with a dimensionally accurate (not downscaled) model domain (but otherwise identical experimental setup) yielded periods within the range of geological inferences. The mean modeled PISM period is within limits set by the literature~~(Hemming, 2004)~~. The mean (pseudo-Hudson Strait) ice volume change in the GSM corresponds to 15 % of a 1.5 km thick ice sheet covering the downscaled pseudo-Hudson Strait area (150x50 km). In PISM, the mean ice volume change is ~~7.4 %~~ 7.1 % of the mean (across ~~base~~ reference setup runs) maximum ice volume in the eastern half of the pseudo-Hudson Bay and pseudo-Hudson Strait.

Metric	GSM ~~base~~ reference setup	PISM ~~base~~ reference setup
number of ~~events~~ surges	180 ± 100	~~28 ± 17~~ 35 ± 25
mean period	1.1 ± 0.5 kyr	~~10 ± 12 kyr~~ 10 ± 10 kyr
mean duration	0.3 ± 0.1 kyr	3 ± 2 kyr
mean pseudo-Hudson Strait ice volume change	$1.7 \pm 0.2 \cdot 10^3$ km^3	~~$1.2 \pm 0.3 \cdot 10^5$ km^3~~ $1.1 \pm 0.3 \cdot 10^5$ km^3

Table 4. ~~Event~~ Surge characteristics of the GSM ($T_{ramp} = 0.0625$°C, $T_{exp} = 28$ (black line in Fig. 2), $W_{Tb,\min} = 0.5$, TpmTrans for the interface calculation, sharp transition between hard and soft bed) and PISM ~~base~~ reference setup (~~Tab.~~Table ~~4~~2). No runs crashed and all runs had more than 1 surge event. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above.~~The basal temperature ramp in GSM uses $T_{ramp} = 0.0625$ and $T_{exp} = 28$ (black line in Fig. 2.)~~

3.1.3 ~~Numerical noise estimation~~

3.2 Minimum numerical error estimates

Differences in ~~event characteristics~~ surge characteristics (compared to the reference setup) are considered significant when they exceed the ~~numerical noise estimates given in Tab.~~MNEEs given in Table 5 and 6 for the GSM and PISM, respectively. However, this does not necessarily mean that smaller changes have no physical relevance but rather that their interpretation is difficult (if not impossible) because the physical response is hidden within the numerical ~~noise~~sensitivities. Likely sources of the ~~numerical noise~~ MNEEs are the iterative SSA solutions and floating point accuracy.

~~To estimate the numerical noise~~To determine a minimum significant threshold in the GSM, we re-run a set of GSM runs with 3.125 km horizontal grid resolution, imposing a stricter numerical convergence (decreasing final iteration thresholds~~, see Tarasov et al. (manuscript in preparation) for further details).~~). In a second experiment, we additionally increase the maximum iterations from 2 to 3 for the outer Picard loop (ice dynamics) and from 2 to 4 when solving the non-linear elliptic SSA equation.

Note that the goal of these experiments is not necessarily to decrease the model error, especially since we do not know the exact solution and, therefore, can not determine the model error. Instead, we aim to show (by changing purely numerical aspects) what the minimum numerical errors are for each surge characteristic.

The largest differences between simulations ~~with base and stricter convergence thresholds~~ occur for the mean period (~~$\sim 7 \pm 11$ %, Tab.~~7 %, Table 5) when using stricter convergence thresholds (no change in the maximum number of iterations). The standard deviations are on the same order of magnitude as the values themselves, indicating different responses across the 5 parameter vectors. Note that ~~estimating the numerical noise~~ determining the MNEEs at 12.5 km instead of 3.125 km horizontal grid resolution yields similar results, except for the mean pseudo-Hudson Strait ice volume change (~~$\sim 21 \pm 53$ %, Tab.~~21 %, Table S2).

Metric	~~base~~ reference setup	stricter numerical convergence [% difference]	stricter numerical convergence with increased maximum iterations [% difference]
number of ~~events~~ surges	180 ± 100	~~-4.1 ± 4.9~~ **-4.1 ± 4.9**	-0.9 ± 3.6
mean period	1.1 ± 0.5 kyr	~~7.0 ± 10.6~~ **7.0 ± 10.6**	4.7 ± 10.6
mean duration	0.3 ± 0.1 kyr	2.5 ± 3.2	**3.9 ± 4.8**
mean pseudo-Hudson Strait ice volume change	$1.7 \pm 0.2 \cdot 10^3$ km^3	-1.1 ± 3.1	**4.6 ± 4.6**

Table 5. Percentage differences of ~~event~~ surge characteristics between GSM runs with regular and stricter numerical convergence and increased maximum iterations for the ice dynamics loops at 3.125 km horizontal grid resolution. The values represent the average of 5 parameter vectors. No runs crashed and all runs had more than 1 surge event. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above. The bold numbers mark the largest MNEE for each surge characteristic.

~~Numerical noise in PISM is estimated~~ MNEEs in PISM are determined by comparing runs with different numbers of cores. Although most parameter vectors show similar results at the beginning of the runs, minor differences can slowly accumulate and lead to significant discrepancies in surge activity by the end of the run (Fig. S18). The largest differences occur for the ~~mean period (~ 14 % for nCores$=16$) and duration (~ 9 %~~ number of surges (16 %) and mean ice volume change (16 %) for nCores$=32$~~)~~, but the standard deviations are large due to a more than ~ 200 % increase in both surge characteristics for parameter vector 6.

The differences in surge characteristics between different numbers of cores can be minimized by decreasing the relative Picard tolerance in the calculation of the vertically-averaged effective viscosity (10^{-4} to 10^{-7}) and the relative tolerance for the Krylov linear solver used at each Picard iteration (10^{-7} to 10^{-12}, Table S5 and Fig. S19). However, this leads to an

Setup	number of ~~events~~ surges	mean period	mean duration	mean ice volume change	~~nE1~~ nS1
25 km ~~base~~ reference setup	~~28 ± 17~~ 35 ± 25	~~10 ± 12 kyr~~ 10 ± 10 kyr	3 ± 2 kyr	~~$1.2 \pm 0.3 \cdot 10^5$ km^3~~ $1.1 \pm 0.3 \cdot 10^5$ km^3	0
nCores= 2	~~0.8 ± 25.0~~ −7.1 ± 19.5	~~8.3 ± 45.0~~ 6.8 ± 36.2	~~−2.3 ± 19.8~~ −0.4 ± 9.5	~~1.5 ± 6.2~~ 1.5 ± 10.3	0
nCores= 4	~~1.9 ± 34.1~~ −8.2 ± 22.9	~~−7.0 ± 15.0~~ −3.8 ± 6.6	~~−8.8 ± 15.0~~ 2.8 ± 18.3	~~0.8 ± 4.4~~ 0.6 ± 4.8	1
nCores= 16	~~−1.2 ± 16.5~~ −10.9 ± 26.0	**−8.2 ± 14.7**	~~6.2 ± 42.1~~ 7.6 ± 21.2	−0.7 ± 13.3	~~0~~ 1
nCores= 32	**16.0 ± 56.2**	~~−6.3 ± 15.3~~ 6.9 ± 48.5	**−8.0 ± 17.4**	~~0.3 ± 12.7~~ **16.3 ± 35.1**	0

Table 6. ~~Change in event~~ Percentage differences of surge characteristics ~~compared to the 25 km PISM base setup in percent~~ (except first row) between the PISM reference setup and setups with different numbers of cores at 25 km horizontal grid resolution. The values represent the average of ~~10~~ 9 parameter vectors. No runs crashed and all runs showed at least 1 event. Runs with just one ~~event~~ surge (~~nE1~~nS1) are ignored when ~~caluclating~~ calculating the change in mean period. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above. The bold numbers mark the largest ~~noise estimate~~ MNEE for each ~~event~~ surge characteristic.

unreasonable increase in model run time (∼ 300 %) that is not feasible for an ensemble-based approach (more than 50 % of all runs did not finish within the time limit of the computational cluster). Intermediate decreases in the relative tolerances still lead to significant differences in surge characteristics while increasing the model run time and are, therefore, not used in the PISM reference setup.

3.2.1 ~~Surface~~ Adding surface temperature noise

Low levels of surface temperature noise have previously been shown to cause chaotic behavior in the mean periods of oscillations (Souček and Martinec, 2011). Adding low levels of uniformly distributed surface temperature noise (±0.1°C and ±0.5°C) to the climate forcing does not significantly affect the ~~event~~ surge characteristics for the GSM (~~Tab.~~Table S3). For example, the effect of adding ±0.5°C surface temperature noise on the mean period is only ~~∼ 4 %~~ 4 % (compared to the ∼ 20 % for ±0.01°C reported by Souček and Martinec (2011)). Adding the same levels of uniformly distributed surface temperature noise to PISM increases the mean duration by 12 % (for ±0.1°C), but has no significant effect on the other surge characteristics (Table S6).

3.2.2 Implicit thermodynamics/ice dynamics coupling

Here In contrast to the commonly used explicit time step coupling between the thermodynamics and ice dynamics in glaciological ice sheet models, we test the impact of implicit coupling (via an iterative implementation) between the thermodynamics and ice dynamics in the GSM (see Tarasov et al. (manuscript in preparation) for details)approximate implicit time step coupling via an iteration between the two calculations for each time step. The implicit coupling decreases the mean duration and pseudo-Hudson Strait ice volume change (~ -13 % and ~ -25 %-13 % and -25 %, respectively). The number of events surges and mean period show no significant change (Tab. S3Table S4). While the changes in mean duration and pseudo-Hudson Strait ice volume change are larger than the numerical noise estimatesMNEEs, they do not justify an increase in run time of ~ 265 % and the implicit coupling is therefore omitted for the GSM base reference setup.

3.3 Sensitivity experiments

3.3.1 Smooth sediment transition zone and non-flat topography

The effects of a smooth sediment transition zone (instead of an abrupt transition from hard bedrock to soft sediment) and a non-flat topography on surge characteristics are examined here. The smooth transition zone does not significantly affect the surge characteristics . A non-flat topography leads to more, longer, and stronger (larger mean pseudo-Hudson Strait ice volume change) surges.

The abrupt transition from hard bedrock to soft sediment (pseudo-Hudson Bay and Hudson Strait) in the GSM base setupand the corresponding difference in basal sliding coefficient provide an additional heating source due to shearing between slow and fast-moving ice. This additional heat appears to foster the propagation of small surge events along the transition zone (e.g., 6 to 6.3 kyr in the upper row of video 03 of Hank (2023)).Incorporating a smooth transition zone (3.125 km or 25 km wide) affects the location of the small-scale surges (not considered in event characteristics) but shows only minor differences for the major surge events (< 7.5 % for all event characteristics, Tab. **??**). The mean bias for both widths is < 1 %, indicating only minor differences in ice volume between an abrupt and smooth transition. However, the timing of events varies for different transition zones (RMSE ≤ 8 %, Fig. 11). A wider transition zone (more sediment surrounding the pseudo-Hudson Strait and Hudson Bay)generally favors an earlier sliding onset, but the details depend on the parameter vector in question (e.g., Fig. 11).

Pseudo-Hudson Strait ice volume for parameter vector 1 and three different bed configurations. The horizontal grid resolution is 3.1125 km. Note that the width of the topographical transition zone matches the width of the soft to hard bed transition zone. In experiments with a pseudo-Hudson Bay/Hudson Strait (HB/HS) topography, the pseudo-Hudson Strait topography is below sea level, increasing the time required for glaciation. A wider transition zone (larger area below sea level) leads to a later glaciation.

Imposing non-flat basal topography has a more significant effect than the Here we discuss differences in surge characteristics due to changes in the model setup. An overview of the results can be found in Fig. 6 and 7 for the GSM and PISM, respectively. The exact values of the percentage differences and information on crashed runs or runs without oscillations are provided in the

supplement. We first examine the 4 model aspects affecting the thermal activation of basal sliding (Sec. 3.3.1 to 3.3.3), followed by the analysis of a smooth sediment transition zone. In general, the number of events, mean duration, and mean pseudo-Hudson Strait ice volume change all increase compared to a flat topography(Tab. **??**). Note that Tab. **??** also shows an increase in the mean period, but this is somewhat misleading due to the now early surges for parameter vector 0 and the subsequent large increase in the mean period (∼ 200 %, no surges in the middle part of the run due to cold surface temperatures (Fig. S28)). All other parameter vectors show a decrease in the mean period for both widths of the transition zone. The mean bias indicates a decrease in ice volume of ∼ 6.5 % for runs with a non-flat topography caused by the larger surge events. A wider transition zone (smaller slope) leads to fewer (difference of ∼ 16 %) but stronger events (difference in mean pseudo-Hudson Strait ice volume change and mean duration of ∼ 9 % , non-flat topography, and ∼ 14 %, respectively, Tab. **??**local basal hydrology (Sec 3.3.4 and 3.3.5). Experiments without significant differences in the surge characteristics are only briefly mentioned here (Sec. 3.3.6). A detailed comparison of an individual run is presented in Sec. S7.4more in-depth discussion of these latter experiments is available in the supplement.

Since the topography will vary from ice stream to ice stream, we use a flat topography for the remaining experiments.

3.3.1 Bed thermal model

We First, we examine the effects of a 1 km deep bed thermal model on the basal temperature and the surge characteristics in the GSM as well as PISM. Both models show significant differences when limiting the bed thermal model to one layer (GSM) or removing it entirely (PISM).

Advection of cold ice near the end of a surge rapidly decreases the basal ice temperature and, therefore, increases the temperature gradient between the basal ice and the bed. In GSM runs with the 1 km deep (17 non-linearly-spaced levels) bed thermal model (base reference setup), this stronger gradient increases the heat flux from the bed into the ice and dampens the actual change in basal ice temperature. Similarly, a rapid increase in basal ice temperature due to higher basal ice velocities at the beginning of a surge reverses the existing temperature gradient at the base of the ice sheet, leading to a heat flux from the ice into the bed. Consequently, less heat is available to warm the surrounding cold-based ice, counteracting the surge propagation (Fig. 8).

With only one bed thermal layer (20 m deep, removing most of the heat storage), the variance of the average basal temperature with respect to the pressure melting point in the pseudo-Hudson Strait increases (Fig. S20) and more heat is available to warm the surrounding ice (no or smaller heat flux into the bed, Fig. S21). The additional heat increases the mean pseudo-Hudson Strait ice volume change and duration (∼ 50 % and ∼ 65 %50 % and 65 %, respectively, Tab. **??**Fig. 6). Due to the larger changes in pseudo-Hudson Strait ice volume and average basal temperature with respect to the pressure melting point, the ice sheet requires more time to reach the pre-surge state when only one bed thermal layer is used. Therefore, the period increases (∼ 60 %60 %) while the number of events drops(∼ 32 %)surges drops. These differences in event surge characteristics exceed the numerical noise estimates (Tab.MNEEs (Table 5). The stronger events surges (larger pseudo-Hudson Strait ice volume change) also lead to a mean ice volume bias of −2.3 % (Tab. **??**lead to overall less ice volume in the pseudo-Hudson Strait (Table S7).

~~Metric base setup 20 m deep (1 layer) bed thermal model number of events 180 ± 100 — 31.6 ± 5.6 mean period 1.1 ± 0.5 kyr 60.2 ± 22.4 mean duration 0.3 ± 0.1 kyr 65.2 ± 24.5 mean pseudo-Hudson Strait ice volume change $1.7 \pm 0.2 \cdot 10^3$ km^3 49.6 ± 14.6 RMSE - 10.4 ± 2.2 Mean Bias - — 2.3 ± 1.7 Percentage differences of event characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM base setup for runs with only one bed thermal layer (20 m deep). No runs crashed and all runs had more than 1 surge event. The first 20 kyr of each run are treated as a spin-up interval for the event characteristics (not the RMSE and mean bias).~~

Running PISM without the 1 km deep (20 equally-spaced levels) bed thermal model yields similar behavior as the GSM, further underlining the impact of a bed thermal model. The mean period, mean duration, and mean ice volume change all increase (~~~ 75 %, ~ 65 %, and ~ 375 %~~80 %, 70 %, and 396 %, respectively; ~~Tab. ??~~Fig. 7). In contrast to the GSM characteristics, the number of ~~events increases (~ 28 %)~~ surges increases for runs without a bed thermal model. However, the standard deviation is large ~~(~ 113 %)~~ and the change in the number of ~~events~~ surges is somewhat misleading. The number of ~~events~~ surges decreases for 6 out of ~~10~~ 9 runs. Parameter vectors showing an increase in the number of ~~events~~ surges without a bed thermal model show very few surges (e.g., Fig. S22) or transition to a constantly active ice stream when the bed thermal model is included. As for the GSM, the stronger ~~events~~ surges lead to an overall smaller ice sheet in the surge affected-area (~~mean ice volume bias of ~ -27 %, Tab. ??~~). Table S8).

~~Metric base setup no bed thermal model number of events 28 ± 17 27.9 ± 113.3 mean period 10 ± 12 kyr 74.8 ± 83.6 mean duration 3 ± 2 kyr 63.2 ± 69.9 mean ice volume change $1.2 \pm 0.3 \cdot 10^3$ km^3 374.5 ± 176.2 RMSE - 36.5 ± 4.9 Mean Bias - -26.6 ± 4.9 Percentage differences of event characteristics, ice volume RMSE and mean bias compared to the PISM base setup for runs without a bed thermal model. No runs crashed and all runs had more than 1 surge event. The first 20 kyr of each run are treated as a spin-up interval for the event characteristics (not the RMSE and mean bias).~~

3.3.2 Basal temperature at the grid cell interface

~~Here we compare the effects of different approaches~~ Another modeling choice that affects the thermal activation of basal sliding is the approach to determining the basal temperature at the grid cell interface. ~~The GSM base setup (no hydrology~~The most straightforward approach to determining the basal temperature with respect to the pressure melting point at the grid cell interface ($T_{bp,I}$) is to use the mean of the two adjacent basal **T**emperatures with respect to the **p**ressure **m**elting point at the grid cell **Cen**ters (TpmCen).

$$T_{bp,I} = 0.5 \cdot (T_{bp,L} + T_{bp,R}), \tag{17}$$

where $T_{bp,L}$ and $T_{bp,R}$ are the grid cell center basal temperatures with respect to the pressure melting point to the left and right of the interface, respectively. Similarly for upper and lower grid cells adjacent to a horizontally aligned interface. However, this approach does not explicitly account for ice thickness changes at the grid cell interface.

TpmInt, on the other hand, calculates the basal temperature at the **Int**erface (T_I) by averaging the adjacent grid cell center basal temperatures (T_L and T_R, Eq. (18a)). $T_{bp,I}$ is then determined by using the interface ice sheet thickness (average of

adjacent grid cell center ice thicknesses H_L and H_R, Eq. (18b)).

$$T_I = 0.5 \cdot (T_L + T_R) \tag{18a}$$

$$T_{bp,I} = T_I + \beta_P \frac{H_L + H_R}{2}, \tag{18b}$$

where $\beta_P = 8.7 \cdot 10^{-4}$ °C m^{-1} is the standard basal melting point depression coefficient. When TpmInt is used with the upwind scheme and the basal ice velocity exceeds 20 m yr^{-1}, Eq. 18a is replaced by $T_I = T_{up}$, where T_{up} is the upstream adjacent grid cell center basal temperature.

The last approach (TpmTrans) attempts to represent heat transfer from ~~subglacial~~ sub-glacial hydrology and ice advection by accounting for extra warming above the pressure melting point ~~(TpmTrans, Sec. ??) . This additional heat warms up adjacent colder ice~~given by

$$T_{\text{add}} = M_b \cdot \frac{L_H}{c_H} \cdot \frac{1}{H_b} \cdot \Delta t, \tag{19a}$$

where M_b is the basal mass balance in m yr^{-1} (positive for melt), $L_H = 3.35 \cdot 10^5$ J kg^{-1} the specific latent heat of fusion of water/ice, $c_H = 2097$ J kg^{-1} K^{-1} the heat capacity of ice at 273.03 K, H_b the basal ice layer thickness in m, and Δt the current model time step in yr. In an intermediate calculation step, the temporary basal temperature at the grid cell center $T_{Im,C}$ is calculated by accounting for the additional heating T_{add}

$$T_{Im,C} = T_C + T_{\text{add}}, \tag{19b}$$

where T_C is the basal temperature at the grid cell center. The basal temperature with respect to the pressure melting point at each adjacent grid cell center $T_{bp,Im,C}$ is then calculated using the interface ice thickness.

$$T_{bp,Im,C} = T_{Im,C} + \beta_P \frac{H_L + H_R}{2} \tag{19c}$$

In the intermediate steps to calculate the interface temperature (Eq. (19b) and (19c)), $T_{Im,C}$ and $T_{bp,Im,C}$ are allowed to exceed the pressure melting point. This temporary higher basal temperature is an attempt to account for heat transported to the interface by ice advection and basal water.

$$\text{IF } T_{bp,Im,C} > 0°\text{C}: \quad T_{bp,Im,C} = \min(0.5°\text{C}, 0.5 \cdot T_{bp,Im,C}) \tag{19d}$$

Averaging the adjacent basal temperatures with respect to the pressure melting point at the grid cell center ($T_{bp,Im,L}$ and $T_{bp,Im,R}$) yields the final basal temperature with respect to the pressure melting point at the interface ($T_{bp,I}$).

$$T_{bp,I} = 0.5 \cdot (T_{bp,Im,L} + T_{bp,Im,R}) \tag{19e}$$

Note that neither the grid cell center nor the interface basal temperature may exceed the pressure melting point (only the basal temperature in the intermediate calculation steps).

The GSM reference setup (no hydrology) uses TpmTrans. The additional heat embodied in T_{add} warms up the grid cell interface. Without the extra warming (TpmInt), 4 out 5 parameter vectors do not show any surges~~when using TpmInt (Sec. ??)~~.

For the only run that still has cyclic behavior (parameter vector 1), the number of ~~events decreases by ~ 84 %~~ surges decreases by 84 % (note that runs without surges are considered for the number of ~~events in Tab. ??~~surges in Fig. 6). Using TpmInt with an upwind ~~instead of a downwind~~ scheme leads to slightly more ~~events~~ surges (difference of ~~~ 7 %~~ 7 % and, therefore, on the same order of magnitude as the ~~numerical noise estimate ($\sim 4 \pm 5$ %, Tab.~~MNEE (4 %, Table 5)). Sporadic surges now occur in all but one run, leading to a large increase in the mean period (~~~ 1650 %, Tab. ??).~~ 1645 %, Fig. 6).

The most straightforward approach~~(TpmCen, Sec. ??) leads to ~ 75 % fewer events~~, TpmCen, leads to 75 % fewer surges, and an increase in mean period and mean duration (~~~ 610 % and ~ 43 %~~609 % and 43 %, respectively). The mean pseudo-Hudson Strait ice volume change decreases (~~~ -61 %~~-61 %). Note that the TpmInt, TpmInt uwpind, and TpmCen ~~event~~ surge characteristics are difficult to compare due to the different number of runs considered (except for the number of ~~events~~surges, decrease of ~~~ 97 % vs. ~ 90 % vs. ~ 75 %~~97 % vs. 90 % vs. 75 %, respectively).

~~Metric base setup (TpmTrans) TpmInt TpmInt, upwind TpmCen nC 0 0 0 1 nE0 0 4 1 0 nE1 0 0 1 0 number of events 180 ± 100 -96.9 ± 6.3 -90.2 ± 15.4 -74.6 ± 13.9 mean period 1.1 ± 0.5 kyr 106.2 ± 0.0 1645.4 ± 2136.8 609.4 ± 832.22 mean duration 0.3 ± 0.1 kyr -15.9 ± 0.0 11.1 ± 17.4 43.3 ± 71.1 mean pseudo-Hudson Strait ice volume change $1.7 \pm 0.2 \cdot 10^3$ km^3 -66.2 ± 0.0 -60.4 ± 6.5 -61.3 ± 5.6 RMSE - 7.4 ± 2.4 9.4 ± 2.6 6.9 ± 2.5 Mean Bias - 4.0 ± 1.6 6.7 ± 2.4 2.1 ± 2.1 Percentage differences of event characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM base setup for different approaches to calculate the basal temperature at the grid cell interface in percent (except first column). Crashed runs (nC) are not considered and runs without events (nE0) only contribute to the change in event number. Runs with only 1 event (nE1) are excluded from the calculation of the mean period. The first 20 kyr of each run are treated as a spin-up interval for the event characteristics (not the RMSE and mean bias).~~

Due to significantly fewer ~~events~~surges, the mean ~~bias in~~ pseudo-Hudson Strait ice volume increases for runs with TpmInt, TpmInt uwpind, and TpmCen (~~4 %, ~ 7 % and ~ 2 %, respectively). RMSE values are ~ 7 % for TpmInt and TpmCen, and ~ 9 % for TpmInt upwind (Tab. ??~~Table S9).

3.3.3 ~~Weight of adjacent minimum basal temperature~~

~~Here we compare the event characteristics for three different weights when calculating the basal interface temperature in the GSM (Eq. (S1)): no consideration of adjacent minimum basal temperature ($W_{Tb,\mathrm{min}} = 0.0$), basal temperature at the interface depends to 50 % on the adjacent minimum basal temperature at the grid cell center (base setup, $W_{Tb,\mathrm{min}} = 0.5$), and basal temperature at the interface is equal to the adjacent minimum basal temperature at the grid cell center ($W_{Tb,\mathrm{min}} = 1.0$).~~

~~Depending on the location of the adjacent minimum grid cell center basal temperature, either the ice flow (when the adjacent minimum basal temperature is downstream) or upstream propagation of the surge should be affected. For the large-scale surges, the adjacent minimum basal temperature is almost exclusively located upstream (e.g., video 02 of Hank (2023)). Changing the weight, therefore, affects the surge propagation rather than blocking parts of the ice flow.~~

~~The surge cycling response to changes in $W_{Tb,\min}$ is not coherent (Tab. S15). For instance, the mean surge period increases for both $W_{Tb,\min} = 0.$ and $W_{Tb,\min} = 1.0$ compared to the base $W_{Tb,\min} = 0.5$. However, standard deviations are large, indicating a different model response for different parameter vectors.~~

3.3.3 Basal temperature ramps at different resolutions

Here we examine the effect of different basal temperature ramps (thermal activation criteria for basal sliding) at 3.125 km horizontal grid resolution and determine ramps for the coarse resolution runs that best match the 3.125 km model results (later used in Sec. 3.4.1). For coarse resolutions, changing the basal temperature ramp can lead to a shift from oscillatory to non-oscillatory behavior (compare 25 km runs in Fig. S23 and 12).

When running the GSM at 3.125 km horizontal grid resolution, ~~surge events~~ surges are apparent for all tested basal temperature ramps. Due to an earlier sliding onset and easier surge propagation, increasing the width of the temperature ramp generally increases the mean pseudo-Hudson Strait ice volume change and duration (Fig. 9). The ice sheet takes longer to recover from the surge (~~binge-phase~~longer regrowth phase), increasing the mean period and decreasing the average number of ~~events~~surges. The largest differences ~~(compared to the base setup) in event~~ in surge characteristics occur for the widest ramp. Running the GSM without a basal temperature ramp leads to small but significant (according to Sec. 3.2) differences ~~(compared to the base setup)~~ in the mean duration (~~-7 ± 3 %) and mean pseudo-Hudson Strait ice volume change (-4 ± 6 %).~~ -7 %).

All ramps (wider and sharper than ~~base~~ reference setup) show fewer ~~events~~ surges and a longer mean period than the ~~base~~ reference temperature ramp setup ($T_{ramp} = 0.0625$, $T_{exp} = 28$). However, except for the 5 widest temperature ramps, changes in both ~~event characteristics are relatively small (~~surge characteristics are smaller than the MNEEs with standard deviations on the same order of magnitude ~~) for ramps with a similar width to the base setup~~ (Fig. ~~S24~~9). The mean duration and mean pseudo-Hudson Strait ice volume change show a consistent response (increase/decrease for both ~~event~~ surge characteristics for a wider/sharper ramp) except for the four basal temperature ramps with the smallest difference to the ~~base setup~~ reference setup (Fig. S24).

Except for the three widest ramps, the mean bias is less than one percent. The RMSE, on the other hand, is roughly 8 %, indicating that the average pseudo-Hudson Strait ice volume is similar, but the timing of surges varies even for small differences in the width of the ramp (~~Tab.~~Table S10).

We compare the different temperature ramps at 25 km, 12.5 km and 6.25 km horizontal grid resolution by calculating a single score for the mean and standard deviation of all ~~event~~ surge characteristics (Sec. S7.3). The ramps yielding the smallest differences compared to the 3.125 km ~~base~~ reference setup are listed in ~~Tab.~~Table S11 and shown in Fig. S25. These results may be different for a different ~~base~~ reference setup (see ~~Tab. ??~~ Table S22 for a comparison of different ~~base~~ reference setups with local basal hydrology).

At 25 km horizontal grid resolution, only 3 out of 12 basal temperature ramps remain after removing the ramps for which the sum of scores (score-mean + score-std, last column in Table S11) differs by more than 50 % from the minimum sum of scores (bold numbers in last column in Table S11). The minimum scores for the mean and standard deviation occur for the same ramp ($T_{exp} = 5$, $T_{ramp} = 0.5$), clearly identifying it as the ramp that best resembles the 3.125 km horizontal grid resolution

~~base~~ reference runs. For the two higher horizontal grid resolutions, the minimum mean and standard deviation scores arise for different temperature ramps, preventing the determination of a single best ramp.

We complement the above analysis by upscaling the 3.125 km ~~base~~ reference runs. For example, a 25x25 km grid cell contains a patch of 64 3.125x3.125 km grid cells. The scatter plot (e.g., Fig. 10) of the warm-based fraction (basal temperature with respect to the pressure melting point at 0 °C) and the mean basal temperature with respect to the pressure melting point of the patch can be used to estimate the parameters T_{ramp} and T_{exp} of the basal temperature ramp (Eq. (8)). However, this does not account for the connectivity between the faces of, e.g., a 25 km grid cell. Without a continuous warm-based channel from one grid cell interface to another, there should be effectively no basal sliding across the grid cell, even when the average basal temperature is close to the pressure melting point. Consequently, this estimate for the basal temperature ramp should be a lower bound to the points in the scatter plot. Furthermore, the upscaling results depend on the bed properties (soft sediment vs. hard bedrock) and the specific scenario (surge vs. quiescent phase). As such, the upscaling statistics only consider grid cells within the pseudo-Hudson Strait area during surges. Due to the limited storage capacity for the 10 yr output fields, only the first 10 kyr after the first surge are used for the upscaling experiments.

The upscaling results agree well with the score analysis at 25 km horizontal grid resolution. Both indicate that at this resolution, the ramp $T_{exp} = 5$, $T_{ramp} = 0.5$ (first row in Table S11, Fig. 10) gives results that best match those of the 3.125 km ~~base~~ reference run. The two approaches yield a similar range of temperature ramps at 12.5 and 6.25 km horizontal grid resolution, but the upscaling experiments generally favor wider temperature ramps (Table S11 and Fig. S26 and S27). This is likely a consequence of the above-mentioned role of sub-grid warm-based connectivity not accounted for in the upscaling analysis. When using the resolution-dependent ramp of Eq. (9), the upscaling experiments, therefore, provide a lower bound of $T_{exp} = 5$. ~~The upscaling~~ Upscaling experiments with local basal hydrology lead to similar results.

3.3.4 ~~Basal hydrology~~Smooth sediment transition zone and non-flat topography

The effects of a ~~simple local basal hydrology model in the GSM (Sec. ??)~~smooth sediment transition zone (instead of an abrupt transition from hard bedrock (0 % sediment cover) to 100 % (soft) sediment cover) and a non-flat topography on surge characteristics are examined here. In the GSM, the smooth transition zone alone does not significantly affect the surge characteristics. Additionally imposing a non-flat topography (Fig. S5) leads to more, longer, and stronger (larger mean pseudo-Hudson Strait ice volume change) surges (Fig. 6). The PISM experiments show fewer, longer and stronger surges for a non-flat topography (Fig. 7 and S9), but no significant effect for an abrupt sediment transition (instead of a more gradual transition, e.g., Fig. S8).

The abrupt transition from hard bedrock to soft sediment (pseudo-Hudson Bay and Hudson Strait) in the GSM reference setup and the corresponding difference in basal sliding coefficient provide an additional heating source due to shearing between slow and fast-moving ice. This additional heat appears to foster the propagation of small surges along the transition zone (e.g., 6 to 6.3 kyr in the upper row of video 03 of Hank (2023)). Incorporating a smooth transition zone (3.125 km or 25 km wide) affects the location of the small-scale surges (not considered in surge characteristics) but shows only minor differences for the major surges (< 7.5 % for all surge characteristics, Fig. 6). The mean bias for both widths is < 1 %, indicating only minor

differences in ice volume between an abrupt and smooth transition. However, the timing of surges varies for different transition zones (RMSE $< 8\,\%$, Fig. 11). A wider transition zone (more sediment surrounding the pseudo-Hudson Strait and Hudson Bay) generally favors an earlier sliding onset (e.g., Fig. 11), but the details depend on the parameter vector in question.

In the GSM, imposing a non-flat basal topography has a more significant effect than the sediment transition zone. In general, the number of surges, mean duration, and mean pseudo-Hudson Strait ice volume change all increase compared to a flat topography (Fig. 6). Note that Fig. 6 also shows an increase in the mean period, but this is somewhat misleading due to the now early surges for parameter vector 0 and the subsequent large increase in the mean period ($\sim 100\,\%$, no surges in the middle part of the run due to cold surface temperatures (Fig. S28)). All other parameter vectors show a decrease in the mean period for both widths of the transition zone. The mean bias indicates a decrease in ice volume of $\sim 6.5\,\%$ for runs with a non-flat topography caused by the larger surges.

A wider transition zone (smaller slope) leads to fewer (difference of $16\,\%$) but stronger surges (difference in mean pseudo-Hudson Strait ice volume change and mean duration of $9\,\%$ and ~~the comparison of a local and horizontal transport model in PISM~~ $14\,\%$, respectively, Fig. 6). A detailed comparison of an individual run is presented in Sec. S7.4.

The width of the transition zone ($-200\,\mathrm{m}$ to sea level) affects the position and width of the major surges (tendency towards wider surges for a wider transition zone (video 04 of Hank (2023))). The pseudo-Hudson Strait topography also suppresses the small surges otherwise observed in the vicinity of the pseudo-Hudson Strait.

Similar to the GSM results, the PISM percentage differences between a smooth (reference setup) and abrupt sediment transition show no significant effect, except a $22\,\%$ increase in surge duration (Fig. 7). While imposing a non-flat topography fosters surges in both models, the increase in mean ice volume change is much larger in PISM ($390\,\%$) than in the GSM (maximum $\sim 17\,\%$), leading to a longer regrowth-phase ($79\,\%$ increase in mean period) and overall less ice volume (mean bias $-30\,\%$, Table S13). The longer recovery times in the PISM outweigh the effect of earlier sliding onsets leading to more surges described above for the GSM. Therefore, the number of surges decreases in the PISM (while increasing in the GSM) when using a non-flat topography (Fig. 7).

Since the topography will vary from ice stream to ice stream, we stick to a flat topography for the remaining experiments.

3.3.5 Basal hydrology

The effects of adding a simple local basal hydrology model to the GSM are examined here. The local basal hydrology sets the basal water thickness by calculating the difference between the basal melt rate and a constant basal drainage rate (rBedDrainRate in Table 1). This sub-glacial hydrology provides a simple and computationally efficient way to capture changes in basal sliding velocities due to effective pressure variations (Drew and Tarasov, 2022, under review). However, it does not account for basal ice accumulation, englacial or supraglacial water input, or horizontal water transport.

The basal water thickness (h_{wb}) and an estimated effective bed roughness scale ($h_{wb,Crit}$ in Table 1) determine the effective pressure coefficient

$$N_{C,eff} = 1 - \min\left(\frac{h_{wb}}{h_{wb,Crit}}, 1.0\right)^{3.5} \tag{20}$$

The basal water thickness is limited to $h_{wb,Crit} = 10$ m and is set to $h_{wb} = 0$ m where the ice thickness is less than 10 m and where the temperature with respect to the pressure melting point is below -0.1°C. Experiments with $h_{wb,Crit} = 5$ m yield the same results, and removing all the water for $H < 1$ m, $H < 50$ m, and $T_{bp} < -0.5$°C does not significantly (according to Sec. 3.2) affect the model results. The effective pressure at the grid cell interface is then

$$N_{eff} = g\rho_{ice} \cdot 0.5\,(H_L N_{C,eff,L} + H_R N_{C,eff,R}), \tag{21}$$

where $g = 9.81$ m s^{-2} is the acceleration due to gravity, $\rho_{ice} = 910$ kg m^{-3} the ice density, H the ice thickness and the subscripts L and R denote the adjacent grid cells to the left and right of the interface, respectively (similarly for upper and lower grid cells adjacent to a horizontally aligned interface). We enforce that N_{eff} never falls below 10 kPa (denominator in Eq. (22), similar results for $N_{eff,min} = 5$ kPa). Finally, the effective pressure of each grid cell alters the basal sliding coefficient in the sliding law (Eq. (6a)) according to

$$C_b = C_b \cdot \min\left(10, \max\left(0.5, \frac{N_{eff,Fact}}{N_{eff} + 10^4\ \text{Pa}}\right)\right), \tag{22}$$

where $N_{eff,Fact}$ is the effective pressure factor (Table 1). The change of the basal sliding coefficient C_b is, therefore, limited to $C_b \cdot 0.2$ to $C_b \cdot 10$. Allowing a larger change of $C_b \cdot 0.1$ to $C_b \cdot 20$ does not significantly (according to Sec. 3.2) change the model results.

When running the GSM with the local ~~subglacial~~ sub-glacial hydrology model, intermediate values are used for all 3 parameters (the effective bed roughness scale $h_{wb,Crit} = 0.1$ m (Eq. (??20)), the constant bed drainage rate rBedDrainRate $\simeq 0.003$ m yr^{-1}, and the effective pressure factor $N_{eff,Fact} \simeq 63246$ Pa (Eq. (??22))) for all 5 parameter vectors. However, different values were tested for all 3 parameters (not shown). In general, a larger $N_{eff,Fact}$ increases the basal sliding coefficient (Eq. (??22)) and, therefore, leads to fewer but stronger ~~events~~surges. The results for $h_{wb,Crit}$ and rBedDrainRate are not as straightforward to interpret. The model response varies for the 2 tested parameter vectors, and the changes are generally smaller than the ~~numerical noise estimates~~ MNEEs of Table 5.

Adding ~~a~~ the local basal hydrology model to the GSM increases the mean ice volume change and duration by ~~~ 20 % and ~ 12 %~~20 % and 12 %, respectively (~~Table ??~~Fig. 6, exceeding the ~~numerical noise estimates~~ MNEEs (Table 5)). The stronger ~~surge events~~ surges are due to the reduction of effective pressure and, thus, increased sliding (Eq. (??22) and (6a)). The mean period ~~increase (~ 17 %~~increases (17 %) while the number of ~~events decreases (~ 4 %~~surges decreases (-4 %), but the standard deviations are large.

Since the local hydrology model effectively increases the basal sliding coefficient, we ~~also compare setups with higher soft and hard bed~~ test if this impact can be replicated simply by increasing the sliding coefficients (~~C_{rmu} and C_{fslid} in~~ Table 1~~,~~

respectively) and no) in a GSM configuration without basal hydrology. Doubling the soft bed sliding coefficient leads to a similar model response but with a smaller increase in the mean period (12 % vs. 17 %) and mean pseuod-Hudson pseudo-Hudson Strait ice volume change (~ 11 % vs. ~ 20 %) and a stronger effect on the number of events and mean period (~ −10 % vs. ~ 4 %11 % vs. 20 %) than the local hydrology model. Increasing the hard bed sliding coefficient has no significant effect on the event surge characteristics (pseudo-Hudson Bay and Hudson Strait are soft-bedded, Table ??Fig. 6). Simultaneously increasing the soft and hard bed sliding coefficient yields similar results to increasing the soft bed sliding coefficient alone (not shown).

3.3.6 Sensitivity experiments without a significant effect

Metric no hydrology local hydrology no hydrology, double C_{rmu} no hydrology, double C_{fslid} number of events 180 ± 100 -3.8 ± 23.8 -9.5 ± 3.9 -3.0 ± 8.8 mean period 1.1 ± 0.5 kyr 17.4 ± 44.9 12.4 ± 4.1 4.5 ± 10.3 mean duration 0.3 ± 0.1 kyr 11.6 ± 19.1 3.1 ± 5.6 2.3 ± 3.5 mean ice volume change $1.7 \pm 0.2 \cdot 10^3$ km^3 20.2 ± 44.7 10.5 ± 5.9 -0.9 ± 5.8 RMSE - 8.7 ± 2.6 8.5 ± 2.7 7.8 ± 2.2 Mean Bias - -0.9 ± 0.8 -0.4 ± 0.4 -0.1 ± 0.1 Percentage differences of event characteristics, ice volume RMSE and mean bias of GSM runs with a local basal hydrology model compared to runs without subglacial hydrology in percent (except first column). Additionally shown are the changes in event characteristics when doubling the values of the soft and hard bed sliding coefficient (C_{rmu} and C_{fslid} in Table 1, respectively). No runs crashed and all runs had more than 1 surge event. The first 20 kyr of each run are treated as a spin-up interval for the event characteristics (not the RMSE and mean bias).

PISM experiments with a mass-conserving horizontal transport hydrology model yield similar results to the local hydrology model (Table **??**). The mean duration, period, and ice volume change increase (0.5 %, 8.9 %, and 1.7 %, respectively), while the number of events decreases (3.5 %).These differences are generally within the numerical noise estimates (Table 6)and show large standard deviations, indicating a different model response for different parameter vectors. The ice volume RMSE and mean bias are also small (+3.2 % and −0.3 %, respectively).

Metric local hydrology horizontal transport number of events 28 ± 17 -3.5 ± 18.5 mean period 10 ± 12 kyr 8.9 ± 40.0 mean duration 3 ± 2 kyr 0.5 ± 23.4 mean ice volume change $1.2 \pm 0.3 \cdot 10^3$ km^3 1.7 ± 13.6 RMSE - 3.2 ± 3.1 Mean Bias - -0.3 ± 1.9 Percentage differences of event characteristics, ice volume (eastern half of pseudo-Hudson Bay and the pseudo-Hudson Strait) RMSE and mean bias of PISM runs with a mass-conserving horizontal transport hydrology model compared to the local hydrology model in percent (except first column). No runs crashed and all runs had at least 1 surge event. One run with the horizontal transport model showed just one event and was ignored when caluclating the change in mean period. The first 20 kyr of each run are treated as a spin-up interval for the event characteristics (not the RMSE and mean bias).

3.3.7 Basal hydrology instead of basal temperature ramp as the primary smoothing mechanism

We examine the effects of a local basal hydrology as main smoothing mechanism for basal sliding (compared to a basal temperature ramp) by using a very sharp ramp ($T_{ramp} = 0.001$, $T_{exp} = 28$), minimizing the smoothing effect The effect of an experiment is considered insignificant when the change in surge characteristics is smaller than the MNEEs (Sec. 3.2). This is the

case for different weights of the adjacent minimum basal temperature when calculating the basal interface temperature (Q_6), for different implementations of the basal ~~temperature ramp. The change in event characteristics between runs with local basal hydrology and the sharp temperature ramp and the GSM base setup is similar (maximum difference of 3 %; compare Table ?? and S17)to the runs with local basal hydrology and the base~~ hydrology (Q_9), and when using basal hydrology instead of the basal temperature ramp ~~($T_{ramp} = 0.0625$, $T_{exp} = 28$), indicating that the local basal hydrology is~~ as the primary smoothing mechanism ~~in both cases. The differences in the change of event characteristics between the base and the steeper ramp are smaller than the numerical noise estimates, preventing further analysis~~(Q_{10}). The details of these experiments are presented in Sec. S8.1, S8.2, and S8.3, respectively. We want to emphasize that experiments without a significant effect can still have physical relevance, but it is currently hidden within the numerical sensitivities.

3.4 ~~Resolution and time step dependence~~Convergence study

In this section, we examine the horizontal grid ~~resoltion~~ resolution and time step dependence of the GSM and PISM model results. Model results are considered as converging when the differences in ~~event~~ surge characteristics decrease with increasing horizontal grid resolutions and decreasing time steps. In general, both models show convergence, but the discrepancies between different horizontal grid resolutions are significant.

3.4.1 GSM convergence study~~without basal hydrology~~

~~In the GSM, significant surge pattern differences occur when computationally more feasible (coarser) horizontal grid resolutions are used (~ 200-fold increase in run time when increasing~~ Significant differences in surge characteristics occur when changing the horizontal grid resolution~~from 25 km to 3.125 km)~~. These differences can be as large as a highly oscillatory behavior at 3.125 km and no oscillations at 25 km horizontal grid resolution (Fig. S23). Changing the basal temperature ramp can somewhat counteract this discrepancy by enabling basal sliding at lower basal temperatures for coarser grid resolutions (Fig. 12 and video 05 of Hank (2023)). Further details on discrepancies between horizontal grid resolutions for individual parameter vectors are discussed in Sec. S9.1.

We compare the differences in ~~event~~ surge characteristics for different basal temperature ramps at each resolution. We examine: a constant ramp ($T_{ramp} = 0.0625$, $T_{exp} = 28$), a resolution-dependent temperature ramp ($T_{exp} = 28$, Fig. 2), and the ramp with the smallest ~~mean score (bold number~~ differences in surge characteristics (bold mean score in Table S11). Note that the large differences in mean period at 25 km resolution are caused by long time intervals without any oscillations in the coarse resolution runs (Table S18). 25 km, 12.5 km, and 6.25 km runs show progressively smaller differences for the constant and resolution-dependent ramp, indicating model convergence~~(Table ??)~~. Convergence of the GSM results with increasing grid resolutions is further supported by successively smaller pseudo-Hudson Strait ice volume RMSE and mean bias values (Table S19). RMSE and mean bias are smaller across all resolutions when using a resolution-dependent instead of a constant temperature ramp (except for the RMSE at 12.5 km horizontal grid resolution).

All three basal temperature ramps lead to similar differences in ~~event~~ surge characteristics at 6.25 km and 12.5 km horizontal grid resolution (Table ~~??~~S18). At 25 km resolution, the ramp with the minimum ~~mean score~~ differences in surge characteristics

significantly improves the agreement with the 3.125 km runs, with differences smaller than for any other ramp or resolution. This could either be a coincidence or indicate that despite thorough testing, the best ramp has not been found at 6.25 km and 12.5 km horizontal grid resolution. Since other ramps at 25 km horizontal grid resolution show only slightly larger ~~scores~~ differences in surge characteristics (e.g., difference of 0.23 in the mean score, Table S11), it is unlikely that it is just a coincidence. However, the sensitivity of the ~~event~~ surge characteristics to grid refinement remains, no matter the choice of the temperature ramp, with differences significantly exceeding the ~~numerical noise estimates~~ MNEEs from Sec. 3.2.

~~Setup number of events mean period mean duration mean pseudo-Hudson Strait ice volume change nE0 3.125 km base setup 180 ± 100 1.1 ± 0.5 kyr 0.3 ± 0.1 kyr 1.7 ± 0.2 · 10^3 km^3 0 25 km, constant ramp −95.1 ± 7.4 942.3 ± 517.70 300.0 ± 172.22 95.9 ± 52.6 3 25 km, resolution-dependent ramp −78.1 ± 18.2 414.5 ± 309.0 119.5 ± 17.6 91.9 ± 23.6 1 25 km, $T_{ramp} = 0.5$, $T_{exp} = 5$ −15.9 ± 20.4 29.7 ± 24.6 43.8 ± 36.6 3.5 ± 18.7 0 12.5 km, constant ramp −59.2 ± 16.5 129.0 ± 41.8 90.3 ± 17.9 50.3 ± 76.5 0 12.5 km, resolution-dependent ramp, also minimum mean score −56.5 ± 15.1 115.7 ± 46.8 101.1 ± 20.5 33.0 ± 66.3 0 6.25 km, constant ramp −24.2 ± 13.1 36.4 ± 20.9 24.8 ± 8.5 14.9 ± 14.2 0 6.25 km,~~ Since including a sub-glacial hydrology model significantly affects the surge characteristics, we also examine the horizontal grid resolution scaling with a local basal hydrology model (Sec. S9.2). The results show overall smaller differences (relative to the 3.125 km reference simulations) in surge characteristics than without (Table S22 vs. S18). The analysis of the convergence study (with and without basal hydrology) and the upscaling experiments in Sec. 3.3.3, therefore, suggest a resolution-dependent ~~ramp −27.9 ± 9.9 42.2 ± 18.9 32.1 ± 6.3 15.9 ± 12.3 0 6.25 km, $T_{ramp} = 0.125$, $T_{exp} = 45$ −25.3 ± 13.6 37.9 ± 26.7 28.2 ± 7.0 9.8 ± 11.6 0 0.5 year maximum time step −4.4 ± 4.5 5.4 ± 4.8 2.0 ± 2.4 −0.5 ± 5.6 0 0.25 year maximum time step −1.8 ± 3.2 2.6 ± 4.4 −0.1 ± 3.4 −0.1 ± 3.4 0 Change in event characteristics compared to the 3.125 km GSM base setup in percent (except first row). The values represent the average of 5 parameter vectors. No runs crashed and runs without events (nE0) only contribute to the change in event numbers. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above. The resolution-dependent ramps ($T_{exp} = 28$) and constant ramp (black line, $T_{ramp} = 0.0625$, $T_{exp} = 28$) are shown in Fig. 2. The third ramp listed for each resolution is the ramp with the smallest mean score (Table S11).~~ temperature ramp with T_{exp} between 5 and 10.

Experiments with different maximum time steps show only minor (~~less than 6 %~~ < 8 %) differences compared to the ~~base~~ reference runs for all ~~event~~ surge characteristics (Table ~~??~~). ~~The pseudo-Hudson Strait ice volume RMSE values for smaller maximum time steps are on the same order of magnitude as the ones for 6.25 km grid resolution (∼ 8 %), while the mean bias shows no difference for the given precision~~ S18), RMSE and mean bias (Table S19). Overall, changes due to different maximum time steps are considerably smaller than for different horizontal grid resolutions and within the range of ~~numerical noise~~ MNEEs (Table 5). Therefore, the implemented CFL condition is adequate for determining the ice dynamical time step, even though the condition is only sufficient for the solution of linear partial differential equations.

3.4.2 ~~GSM convergence study with basal hydrology~~

~~Since including a subglacial hydrology model significantly affects the event characteristics, we further examine the horizontal grid resolution scaling with a local basal hydrology model (Sec. ??). Based on the results without basal hydrology (Sec. 3.4.1),~~

5 basal temperature ramps ($T_{exp} = [5, 10, 15, 20, 28]$) with a resolution-dependent T_{ramp} (Eq. (9)) are tested for all resolutions. As it is unclear which basal temperature ramp should be used at the highest horizontal grid resolution (3.125 km), we test two different ramps ($T_{exp} = [5, 28]$). The experiments that yield the smallest differences in event characteristics (smallest mean score in Table S20 and S21) compared to the corresponding 3.125 km base runs (bold rows) are presented in Table **??**.

Setup number of events mean period mean duration mean pseudo-Hudson Strait ice volume change 25 km, $T_{ramp} = 0.5$, $T_{exp} = 5$ 9.7 ± 59.9 15.5 ± 42.3 24.3 ± 36.1 13.6 ± 46.7 12.5 km, $T_{ramp} = 0.25$, $T_{exp} = 5$ -36.1 ± 17.6 68.0 ± 49.8 97.1 ± 60.3 3.0 ± 26.4 6.25 km, $T_{ramp} = 0.125$, $T_{exp} = 28$ -13.2 ± 31.1 27.0 ± 40.6 25.7 ± 25.2 5.6 ± 27.5 25 km, $T_{ramp} = 0.5$, $T_{exp} = 5$ -2.4 ± 35.8 16.1 ± 31.4 20.7 ± 30.3 14.3 ± 35.8 12.5 km, $T_{ramp} = 0.25$, $T_{exp} = 10$ -37.7 ± 12.1 61.7 ± 44.1 63.4 ± 34.8 20.5 ± 39.0 6.25 km, $T_{ramp} = 0.125$, $T_{exp} = 5$ -25.6 ± 13.9 37.8 ± 23.8 41.1 ± 21.3 0.3 ± 19.8 Change in event characteristics compared to the 3.125 km GSM setups with local basal hydrology (bold rows, $T_{exp} = [5, 28]$) for the ramps with the smallest mean score (analysis steps described in Sec. S7.3) in percent. The values represent the average of 5 parameter vectors. No runs crashed and all runs had more than 1 surge event. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above.

Similar to the results without a basal hydrology model, the smallest differences in event characteristics (except the mean pseudo-Hudson Strait ice volume change) occur for the coarsest horizontal grid resolution (25 km, Table **??**). This likely indicates that the optimal ramps at 12.5 and 6.25 km horizontal grid resolution have not been found.

In general, the resolution-dependent ramp with $T_{exp} = 5$ leads to the smallest differences between coarse and high-resolution runs. The differences in event characteristics are significantly smaller than for a resolution-dependent temperature ramp without local basal hydrology (except for the mean pseudo-Hudson Strait ice volume change, Table **??** vs. **??**), further underlining the importance of the basal hydrology.

Except for 12.5 km horizontal grid resolution, the resolution-dependent ramp with $T_{exp} = 5$ yields a self-consistent response across all resolutions. At 12.5 km, the next closest exponent ($T_{exp} = 10$) has the minimum mean score. However, given that there is no single best ramp across all resolutions, we assess different ramps as to whether differences are within inferred numerical noise (DWINN). To this end, we calculate the differences between the ramp with the minimum mean score and all other ramps at each resolution and for all event characteristics (Table S20 and S21). We rule out ramps for which the differences exceed the maximum numerical noise estimates (maximum of Table 5 and S2) for more than one event characteristic (DWINN failures).

Under these criteria and when using $T_{exp} = 5$ at 3.125 km horizontal grid resolution, the resolution-dependent ramp with $T_{exp} = 10$ remains within the DWINN ensemble for all resolutions (Table S21). The results for $T_{exp} = 28$ at 3.125 km horizontal grid resolution do not yield a single ramp that remains within the DWINN ensemble at all resolutions (Table S20). However, except for 6.25 km, for which the differences between the tested basal temperature ramps are the smallest, $T_{exp} = 5$ yields the minimum mean-score. The above analysis and the upscaling experiments in Sec. 3.3.3, therefore, suggest a resolution-dependent temperature ramp with T_{exp} between 5 and 10.

The pseudo-Hudson Strait ice volume RMSE and mean bias show convergence (smaller differences) for both 3.125 km horizontal grid resolution setups (Table S23).

3.4.2 PISM convergence study

Similar to the results presented for the GSM~~(Table ?? and ??)~~, running PISM with different resolutions can lead to significant differences in surge behavior. ~~In contrast to the GSM resolution scaling results~~However, the PISM ~~event~~ surge characteristics do not show convergence for the three resolutions examined here (Table ~~??~~S25). Note that 4 out of ~~10 50 km runs crashed, indicating that this resolutions is too coarse to model binge-purge-type surges and skewing the statistics. The differences in event characteristics between the~~ 9 12.5 km ~~and 25 km PISM runs show a similar response than, e.g., the differences between the 3.125 km and 6.25 km runs in the GSM (Table ?? vs. Table ??).~~

~~The ice volume RMSE and mean bias converge (but not the event characteristics, see Table S26 and Fig. S30)~~runs did not finish within the time limit of the computational cluster and are considered as crashed runs (potentially skewing the statistics). Additionally, 1 12.5 km did not show any surges and was also excluded from the analysis. The differences in ~~event~~ surge characteristics for different grid resolutions are, in general, larger than the ~~numerical noise estimates (different numbers of cores)~~MNEEs, but can be smaller (mean ~~period of the 50 km~~ ice volume change of the 25 km runs).

~~Setup number of events mean period mean duration mean ice volume change nC nE0 50 km 10.0 ± 47.3 9.6 ± 48.5 10.6 ± 40.2 29.6 ± 45.7 4 0 25 km −27.5 ± 16.8 38.5 ± 40.7 21.7 ± 30.0 9.2 ± 14.4 0 0 0.5 year maximum time step −2.0 ± 37.0 −6.8 ± 13.0 −8.6 ± 16.1 −0.3 ± 3.2 0 1 0.25 year maximum time step −3.4 ± 16.5 −0.8 ± 25.8 1.5 ± 31.9 0.9 ± 6.6 0 0 Change in PISM event characteristics due to different horizontal grid resolutions and maximum time steps. Note that the 12.5 km (highest resolution tested) is used as base for the grid resolution convergence study, whereas the 25 km setup is used for the maximum time step experiments. Except for the two rows containing the base setups (bold), all values are in percent. The values represent the average of 10 parameter vectors (8 for the resolution convergence study because one of the 12.5 km runs crashed and one did not show any oscillations (infinite difference)). Crashed runs (nC) are not considered and runs without events (nE0) only contribute to the change in event numbers. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above.~~ The ice volume RMSE and mean bias converge (but not the surge characteristics, see Table S26).

Similar to the results for different numbers of cores, small differences can slowly accumulate for different maximum time steps, leading to a different pattern at the end of the run (e.g., Fig. S31). As for the GSM, different horizontal grid resolutions have a larger effect on surge behavior than the maximum time step for all ~~event~~ surge characteristics (Table ~~??~~S25).

4 Results Summary and Discussion

This section summarizes our ~~modelling~~ modeling results in the context of the research questions outlined in Sec. 1.3 and previous modeling studies.

For the sensitivity experiments, changing the approach for determining the basal temperature at the grid cell interface shows the largest differences in surge characteristics compared to the reference setup. Including a bed thermal model has overall the second strongest effect. A non-flat basal topography and sub-glacial hydrology show moderate differences (Fig. 6 and 7). The effects of different basal temperature ramps depend on the width of the ramp but are generally moderate (Fig. 9). The weight of the adjacent minimum basal temperature, the details of the basal hydrology model, and the primary smoothing mechanism

at the warm/cold-based transition zone (sub-glacial hydrology vs. basal temperature ramp) do not have an effect above our reference MNEE threshold.

In general, changing the horizontal grid resolution shows moderate differences. However, depending on the resolution and the temperature ramp, the differences can be as large as for the most impactful sensitivity experiments (Table S19). Experiments with different maximum time steps show differences smaller than the reference MNEE threshold.

Minimum numerical error estimates

Q_1 ~~*Q1 -*~~ *What is the threshold of* ~~*numerical noise*~~ *MNEEs in the two models?* ~~*When modeling binge-purge type surge events, numerical noise is*~~

Since the model error to the exact analytical solution can not be determined in the context of this study, we use MNEEs to determine a minimum threshold for the significance of a change in the model configuration. When modeling ice stream surge cycling, numerical sensitivities are apparent in both the GSM and PISM. The differences in ~~event~~ surge characteristics when applying ~~a~~ stricter numerical convergence criteria in the GSM can be as large as ~~$\sim 7 \pm 11\,\%$~~ $7\,\%$ (Table 5). Adjusting the matrix solver used in PISM (different number of cores) leads to differences in ~~event~~ surge characteristics of up to ~~$\sim 14 \pm 46\,\%$~~ $16\,\%$ (Table 6). Consequently, the ~~effects of physical model components~~ model sensitivity to physical model aspects cannot be determined if the differences in ~~event~~ surge characteristics are smaller than ~~these numerical noise estimates~~ the MNEEs and are considered insignificant.

In contrast to the findings of Souček and Martinec (2011), adding low levels of surface temperature noise does not significantly affect the GSM and PISM results (Table S3 and S6). Potential reasons for the different model responses are the use of an Arakawa A grid ~~in Souček and Martinec (2011), the different implementations of the~~ (velocities and temperatures are calculated and the JOSH (JOint Shallow-ice/Higher-order model) ice sheet dynamics ~~, and the asymmetric in time temperature forcing in this paper (compared to a constant forcing in Souček and Martinec (2011)).~~ in Souček and Martinec (2011).

~~*Q2 - Does the abrupt transition between a soft and hard bed significantly affect surge characteristics?* The abrupt transition between the soft and hard bed sliding law is not the cause of the major surge events (Table **??**). However, incorporating a smooth transition zone with two different widths (3.125 km and 25 km) does affect the location of proximal small-scale ice streams (video 03 of Hank (2023)).~~

~~*Q3 - How does a non-flat topography affect the surge behavior?* Adding a 200 m deep pseudo-Hudson Strait and Hudson Bay with a smooth transition zone and 500 m deep ocean displaces the origin of surges slightly further inland. Due to both the resultant warmer basal temperature and depressed presure melting point, the surges propagate faster, last longer, and evacuate more ice volume (Table **??**). The topography slopes down towards the pseudo-Hudson Strait, increasing the ice inflow from the surroundings. The ice sheet recovers faster from the previous surge, decreasing the mean period.~~

~~Comparing the results for two different widths of the transition zone indicates fewer but larger events for a wider transition zone. Due to the gentle slope, the topography affects a larger area, increasing the width of the ice stream. More ice is available for evacuation, prolonging the surge and decreasing the pseudo-Hudson Strait ice volume at the end of~~

the surge. The stronger surges for a wider transition zone increase the recovery time, leading to a smaller increase in the number of events than for the narrow transition zone (Table **??**). We expect other ice sheet models with a comparable experimental design and ice dynamics to show similar levels of MNEEs. To minimize the possibility of interpreting numerical errors as a physical response to a change in model setup, it is crucial to determine MNEEs (or a comparable metric).

Sensitivity experiments with a significant effect

Q_2 *Q4 - Is the inclusion of a bed thermal model a controlling factor for surge activity?*

Including a 1 km deep bed thermal model significantly (according to the MNEEs in Sec. 3.2) affects the surge characteristics in the GSM and PISM. The additional heat stored in the bed changes the thermal conditions at the ice-bed boundary, dampening the ice volume change during a surge event (Table **??** and **??**(Fig. 6 and 7). Models with similar setups but without a bed thermal model likely overestimate the ice volume change during a surge (e.g., Calov et al., 2010; Brinkerhoff and Johnson, 2015). Therefore, the inclusion of a bed thermal model is a key aspect of modeling ice stream surge cycling.

Q_3 *Q5 - Do different approaches for determining the grid cell interface basal temperature significantly affect surge behavior, and if yes, which one should be implemented?*

The choice of approach for determining the basal temperature at the grid cell interface significantly changes the event surge characteristics. Without considering additional heat transfer to the grid cell interface (as an attempt to represent heat contributions from subglacial sub-glacial hydrology and ice advection), 4 out of 5 runs do not show any eventssurges, and the number of events surges for the remaining run decreases by ~ 97 % when using TpmInt (Table **??**84 % (TpmInt). The additional heat is, therefore, an essential component for modeling surge events surges in the GSM. Using an upwind instead of downwind scheme for TpmInt has no significant effect (Table **??**Fig. 6).

Q6 - How much of the ice flow should be blocked by upstream or downstream cold-based ice, or equivalently, what weight should be given to the adjacent minimum basal temperature? Changing the weight of the adjacent minimum basal temperature for the basal sliding temperature ramp in the GSM yields a maximum difference of ~ 15 % (Table S15). These somewhat small effects on event characteristics are likely due to the fact that most surges propagate upstream (from the ocean to the pseudo-Hudson Bay) and the adjacent minimum basal temperatures have little potential to affect (e.g., partly block)the ice flow. The additional heat transfer to the grid cell interface is comparable to spreading 50 % of the basal heating effect from sliding in a grid cell to the surrounding grid cells used in mPISM (latest version based on PISM v0.7.3) (e.g., Ziemen et al., 2014, 2019; Schannwell et al., 2023). This spreading of basal heating warms the grid cells adjacent to an ice stream and was necessary to model Heinrich Event-like surges (Florian Ziemen, personal communication). While no additional heat transfer was added to PISM v2.0.2 used within this study, the till friction angles had to be reduced to model surges.

Q_5 *How different are the model results for different basal temperature ramps and what ramp should be used?*

~~*Q7 - How different are the model results for different basal temperature ramps and what ramp should be used?*~~ Similar to Souček and Martinec (2011), we find significant differences in the period and amplitude of surges at all tested resolutions when using different implementations for thermal activation of basal sliding (the basal temperature ramp). ~~A~~ In the GSM, a wider temperature ramp enables sliding onset at lower temperatures, fostering surge propagation and leading to stronger ~~surge events~~surges. However, the choice of the most appropriate ~~3.125 km temperature ramp (Table S10~~temperature ramp at the highest resolution tested (3.125 km, Fig. 9) is unclear and identifying a single best ramp (fit of coarse ~~resultions~~ resolutions runs to 3.125 km runs) is challenging (Table S11). In general, a resolution-dependent ramp with T_{exp} between 5 and 10 (Eq. (8) and (9)) yields the smallest differences between high and low resolution simulations. Even at the highest tested horizontal grid resolution (sharpest temperature ramp), running the GSM without a basal temperature ramp leads to significant (according to Sec. 3.2) differences in the mean duration~~and mean pseudo-Hudson Strait ice volume change~~, underlining the importance of the basal temperature ramp across all resolutions.

To account for observational and experimental evidence of sub-temperate sliding (Barnes et al., 1971; Shreve, 1984; Echelmeyer and , avoid an abrupt onset of sliding at the warm/cold-based transition that causes refreezing on the warm-based side (Mantelli et al., 2019), and minimize resolution dependencies, a basal temperature ramp (or similar mechanism) should be implemented in all ice sheet models for contexts where surge onset/termination are important.

Q_6 *Does the abrupt transition between a soft and hard bed significantly affect surge characteristics?*

An abrupt transition between the soft (100 % sediment cover) and hard bed (0 % sediment cover) sliding law (as, e.g., used in the HEI can lead to additional localized shear heating due to differences in basal resistance. Incorporating a smooth transition zone with two different widths (3.125 km and 25 km) in the GSM does affect the location of proximal small-scale ice streams (video 03 of Hank (2023)). However, the abrupt transition is not the cause of the major surges (Fig. 6). PISM experiments with an abrupt transition (instead of the smooth transition used in the reference setup) show a slight increase in surge duration (22 %, Fig. 7) but otherwise no significant differences in the surge characteristics. Since the sediment cover can change within a few kilometers (e.g., Andrews and MacLean, 2003), we conclude that, despite the minor differences, an abrupt transition between soft and hard beds is a reasonable simplification, especially considering horizontal grid cell dimensions of 25 km or larger.

Q_7 *How does a non-flat topography affect the surge behavior?*

Adding a 200 m deep pseudo-Hudson Strait and Hudson Bay with a smooth transition zone and 500 m deep ocean to the GSM setup displaces the origin of surges slightly further inland. Due to both the resultant warmer basal temperature and depressed pressure melting point, the surges propagate faster, last longer, and evacuate more ice volume (Fig. 6). The topography slopes down towards the pseudo-Hudson Strait, increasing the ice inflow from the surroundings. The ice sheet recovers faster from the previous surge, decreasing the mean period.

Comparing the results for two different widths of the topographic transition zone indicates fewer but larger surges for a wider transition zone. Due to the gentle slope, the topography affects a larger area, increasing the width of the ice stream.

More ice is available for evacuation, prolonging the surge and decreasing the pseudo-Hudson Strait ice volume at the end of the surge. The stronger surges for a wider transition zone increase the recovery time, leading to a smaller increase in the number of surges than for the narrow transition zone (difference of 16 %, Fig. 6).

Imposing a non-flat topography in the PISM also leads to longer and stronger surges (Fig. 7). However, the increase in mean ice volume change is much higher than in the GSM (390 % vs. $\sim$ 17 %), prolonging the regrowth phase by $\sim$ 80 % and reducing the number of surges.

In agreement with previous modeling studies (e.g., Winsborrow et al., 2010, and references within), the topography is a key aspect of ice stream modeling. When interested in a comparison with observational data or proxy reconstructions, a more realistic topography (in contrast to the idealized flat topography) should be used.

Q_8 *What is the effect of a simplified basal hydrology on surge characteristics?*

~~*Q8 - What is the effect of a simplified basal hydrology on surge characteristics?*~~ The local basal hydrology model (including the addition of effective pressure dependence into the sliding law) in the GSM increases the mean ice volume change, mean period, and mean duration while the number of ~~events~~ surges slightly decreases (~~Table **??**~~Fig. 6). Somewhat stronger ~~events~~ surges are expected due to the reduction in effective pressure introduced by the ~~subglacial~~ sub-glacial water. Model runs without ~~subglacial~~ sub-glacial hydrology will therefore tend to underestimate the amplitude of surges (mean ice volume change and duration).

Increasing the soft bed sliding coefficient in model runs without basal hydrology has a smaller increase in the mean duration, mean period, and pseudo-Hudson Strait ice volume change than including the local basal hydrology model, but a stronger effect on the number of ~~events (Table **??**~~surges (Fig. 6). Therefore, simply changing the basal sliding coefficient cannot replace the basal hydrology model. The importance of ~~subglacial~~ sub-glacial hydrology has also been shown in several other studies ~~examing~~ examining the effects of ice sheet surges and ice streaming within a continuum model approach ~~(Fowler and Johnson, 1995; Fowler and Schiavi, 1998; Benn et al., 2019, e.g.,).~~ (e.g., Fowler and Johnson, 1995; Fowler and

.

Sensitivity experiments without a significant effect

Q_4 *How much of the ice flow should be blocked by upstream or downstream cold-based ice, or equivalently, what weight should be given to the adjacent minimum basal temperature?*

Changing the weight of the adjacent minimum basal temperature for the basal sliding temperature ramp in the GSM yields a maximum difference of 15 % (Table S15). These somewhat small effects on surge characteristics are likely due to the fact that most surges propagate upstream (from the ocean to the pseudo-Hudson Bay) and the adjacent minimum basal temperatures (almost exclusively located upstream) have little potential to affect (e.g., partly block) the ice flow.

Q_9 ~~*Q9 -*~~ *How significant are the details of the basal hydrology model on surge characteristics in PISM?*

Incorporating a mass-conserving horizontal transport hydrology model does not significantly change the surge characteristics in PISM (~~Table ??~~Fig. 7), indicating that the computationally ~~more efficient~~ much cheaper local hydrology model is a reasonable simplification for this context. More nuanced results, depending on the surge characteristics examined, are observed for the GSM ~~(Drew and Tarasov, 2022, under open review)~~(Drew and Tarasov, 2022, under review).

Q_{10} ~~*Q10 -*~~ *What are the differences (if any) in surge characteristics between local basal hydrology and a basal temperature ramp as the primary smoothing mechanism at the warm/~~cold based~~ cold-based transition zone?* ~~Event~~ Surge characteristics in runs with an active local basal hydrology and a sharp temperature ramp ($T_{ramp} = 0.001$, $T_{exp} = 28$, minimizing the smoothing effect of the basal temperature ramp, Table S17) show only minor differences compared to the GSM setup with a local hydrology and the ~~base~~ reference temperature ramp ($T_{ramp} = 0.0625$, $T_{exp} = 28$, ~~Table ??~~Fig. 6). Once included, the local basal hydrology is the primary smoothing mechanism. However, since the two smoothing mechanisms operate in different temperature regimes, a basal temperature ramp (representing sub-temperate sliding) cannot be replaced by a basal hydrology scheme (as in, e.g., Robel et al., 2013; Kyrke-Smith et al., 2014; Brinkerhoff and Johnson, 2015). The numerical ~~noise prevents~~ sensitivities prevent further analysis.

Convergence study

Q_{11} *Do model results converge (decreasing differences when increasing horizontal grid resolution and decreasing maximum time step)?*

~~*Q11 - Do model results converge (decreasing differences when increasing horizontal grid resolution and decreasing maximum time step)?*~~ Systematic grid refinement shows a converging overall ice volume (mean bias) in both models (Table S19+S23 and S26). ~~Event characteristics, on the other hand, converge for a~~ However, surge characteristics converge for constant and resolution-dependent ~~ramp~~ basal sliding activation ramps in the GSM (Table ~~??~~S18), but not in PISM (Table ~~??~~S25). This clearly illustrates that mean ice volume and, consequently, mean ice thickness, as presented, e.g., in Van Pelt and Oerlemans (2012), are insufficient metrics to determine whether cyclic model results exhibit a resolution dependency. The highest horizontal grid resolution used for PISM is 4 times coarser than the highest resolution in the GSM (12.5 km vs. 3.125 km), which might explain why PISM results do not converge. In the GSM, the agreement between coarse and high-resolution runs can be significantly improved when applying a resolution-dependent temperature ramp (Table ~~??~~ S18 and Sec. S7.3). ~~GSM resolution scaling experiments with activated local basal hydrology lead to overall smaller differences (relative to the 3.125 km reference simulations) in event characteristics than without (Table ?? vs. ??).~~

~~Event~~ Surge characteristics in both the GSM and PISM show a strong resolution dependence for all sensitivity tests (Table ~~??~~S18+~~?? and ??~~). ~~This is in contrast to the findings of~~ S22 and S25). While other studies examining thermally induced ice streaming ~~Hindmarsh (2009); Brinkerhoff and Johnson (2015). However, both of these studies analyze just one parameter vector, and it is relatively easy to find a parameter vector for which, e. g. , the GSM exhibits only a minor resolution dependence~~do not find a strong resolution dependence (Hindmarsh, 2009; Brinkerhoff and Johnson, 2015),

these studies are not directly comparable. The different results are likely due to differences in the experimental design. For example, neither Hindmarsh (2009) nor Brinkerhoff and Johnson (2015) consider a bed thermal model. While Hindmarsh (2009) considers sub-temperate sliding, his model allows sliding far below the pressure melting point (order of $\delta = 1$ compared to $\delta = 0.01$ within this study, Eq. (10)) and focuses on steady ice streams, not ~~binge-purge-type surges~~ice stream surge cycling. Over 200 kyr, even minor differences at the beginning of a run can slowly accumulate and yield overall different surge characteristics (e.g., Fig. ~~S31~~S30). Furthermore, Brinkerhoff and Johnson (2015) examine ice stream statistics over the whole domain and not a specific soft-bedded region. ~~Neither Hindmarsh (2009) nor Brinkerhoff and Johnson (2015) consider a bed thermal model~~Additionally, both of these studies analyze just one parameter vector, and there are some parameter vectors for which, e.g., the GSM exhibits only a minor resolution dependence.

Even though the studies are not directly comparable, the results of Brinkerhoff and Johnson (2015) offer some insight relevant to this study. For example, they suggest membrane stresses are necessary for convergence under horizontal grid refinement. The hybrid SIA/SSA ice dynamics used in the GSM and PISM might be insufficiently 'higher-order' and lead to a stronger resolution dependence than the schemes used in Hindmarsh (2009); Brinkerhoff and Johnson (2015). However, GSM experiments with the SSA active everywhere show a resolution dependence comparable to the velocity-dependent SSA activation criteria (Table S24 and S18, respectively), indicating that the hybrid SIA/SSA ice dynamics are not the sole reason for the strong resolution dependence.

Although claiming that their model does not show a strong resolution dependency, Roberts et al. (2016) show differences in the surge timing, ice volume change, and period in their supplement. Greve et al. (2006) also conclude that different horizontal grid resolutions have no significant effect on the surge characteristics, but the two time series are difficult to compare (not shown in the same plot).

Decreasing the maximum time step leads to only minor (~~$< 9\,\%$~~$< 6\,\%$ except the $-15\,\%$ difference in the number of surges for the 0.5 year maximum time step setup in PISM, Table ~~?? and ??~~S18 and S25) changes in ~~event~~ surge characteristics for both PISM and the GSM. This is in agreement with the findings of earlier studies (Greve and MacAyeal, 1996; Greve et al., 2006; Takahama, 2006). However, individual parameter vectors might still show a different surge pattern (e.g., Fig. S31).

5 Conclusions

~~We~~ Within the limitations of hybrid SIA/SSA ice dynamics, we investigate the effect of ice sheet model numerics on surge characteristics ~~and determine key model components for simulating binge-purge type surge events. Numerical noise estimates (~~often neglected in ice sheet modeling studies. Minimum numerical error estimates (MNEEs, differences in surge characteristics of up to ~~$\sim 15\,\%$~~$16\,\%$ when changing the settings of the numerical solver) are used to discern the ~~physical~~ significance of the process in question. For some experiments (e.g., the weight of the adjacent minimum basal temperature), the ~~numerical~~

~~noise estimates~~ MNEEs are on the same order of magnitude or larger than the modeled differences in surge characteristics, hindering the analysis of the underlying physical process.

Experiments showing only minor changes in surge characteristics (generally smaller than the MNEEs) include: a mass-conserving horizontal transport hydrology model (instead of a local hydrology model), a smoothed transition between regions of soft sediment and hard bedrock (instead of an abrupt transition), and smaller (than $1\ \mathrm{yr}$) maximum time steps in the CFL condition.

On the other hand, surge characteristics are sensitive to the basal sliding activation function and show a strong resolution dependency. Since both the GSM and PISM show a resolution dependency, it is likely that it also exists in other ice sheet models with similar approximations. Incorporating a resolution-dependent basal temperature ramp for basal sliding thermal activation reduces the resolution dependency in the GSM. Based on our results, we suggest that those interested in modeling ice stream cycling at horizontal grid resolutions $> 3\ \mathrm{km}$ should use a resolution-dependent ramp with $T_{exp} = 10$ as a ~~base~~ reference test configuration. However, we strongly recommend resolution testing to determine the configuration with the smallest resolution dependency. Additionally, our results indicate that ~~modelling of binge-purge~~ modeling of ice stream surge instabilities that aims to reflect the physical ~~behaviour~~ behavior of actual ice streams should include a non-flat topography, a bed thermal model, and a basal hydrology model.

Basal temperature spokes, such as the ones modeled in the EISMINT-F and H experiments (Payne et al., 2000), are not apparent in the PISM experiments. The GSM runs show some warm-based areas at the margins interspersed by colder regions, but this is likely due to a steep surface slope leading to a large driving stress, high velocity, and then consequently, a basal temperature increase. Therefore, neither the PISM nor GSM instabilities discussed here are comparable to the EISMINT temperature spokes. The absence of basal temperature spokes is likely due to the inclusion of membrane stresses in the ice dynamics of both models (Bueler et al., 2007; Bueler and Brown, 2009).

The key takeaway of this study is the numerical ~~sensitivities~~ sensitivity that must be considered when numerically modeling ~~binge-purge-type oscillations of ice streams~~ice stream surge oscillations. Our analyses offer guidance in minimizing these sensitivities for research contexts that limit horizontal grid cell resolution to larger than about $3\ \mathrm{km}$. Significant (albeit smaller) ~~numerical sensitivities~~ MNEEs to the choice of thermal activation ramp remain at our highest tested horizontal grid resolution ($3.125\ \mathrm{km}$). Analytical examination (where possible) and/or higher-resolution numerical modeling with higher-order glaciological models is needed to further verify that modeling approaches represent the actual physical system for this context.

Code availability. TEXT

Data availability. TEXT

Code and data availability. The GSM source code (v01.31.2023) and run instructions are available at https://doi.org/10.5281/zenodo.7668472 (Tarasov et al., 2023). Instructions on how to install and run PISM and the PISM source code (v2.0.2) can be acquired from the repository at https://zenodo.org/record/6001196. Further information on how to recreate this work's results, input files, parameter vectors, and the analysis scripts used to determine the surge characteristcs can be found at https://doi.org/10.5281/zenodo.7905404 (Hank, 2023).

Sample availability. TEXT

Author contributions. TEXT

K.H. and L.T. conceptualized the ideas behind this study. All authors were involved in designing the experimental setup of the GSM. K.H. designed the experimental setup for PISM and performed the modeling analysis for both models under the supervision of L.T. All authors contributed to the results, interpretation, and writing of the manuscript.

Competing interests. The authors have no competing interests.

Disclaimer. TEXT

Acknowledgements. The authors thank *Andy Aschwanden*, *Ed Bueler* and *Constantine Khrulev* for support with the Parallel Ice Sheet Model (PISM). We thank *Ed Bueler* ~~for a fruitful discussion~~ and *Daniel F Martin* for fruitful discussions about the bed thermal model and the numerical tolerances, respectively. We also thank *Florian Ziemen* and *Clemens Schannwell* for insightful discussions on modeling Heinrich Event-like surges. This research has been supported by ~~the Bundesministerium für Bildung und Forschung (~~an NSERC Discovery Grant (number RGPIN-2018-06658), the Canadian Foundation for Innovation, and the German Federal Ministry of Education and Research (BMBF) as a Research for Sustainability initiative ~~, FONA, (~~FONA) through the PalMod project~~) and an NSERC Discovery Grant held by LT~~. EM was supported by the European Union (ERC-2022-STG, grant number 101076793). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Andrews, J. T. and MacLean, B.: Hudson Strait ice streams: A review of stratigraphy, chronology and links with North Atlantic Heinrich events, Boreas, 32, 4–17, https://doi.org/10.1080/03009480310001010, 2003.

Arakawa, A. and Lamb, V. R.: Computational Design of the Basic Dynamical Processes of the UCLA General Circulation Model, in: General Circulation Models of the Atmosphere, edited by CHANG, J., vol. 17 of *Methods in Computational Physics: Advances in Research and Applications*, pp. 173–265, Elsevier, https://doi.org/https://doi.org/10.1016/B978-0-12-460817-7.50009-4, 1977.

Bahadory, T. and Tarasov, L.: LCice 1.0-a generalized Ice Sheet System Model coupler for LOVECLIM version 1.3: Description, sensitivities, and validation with the Glacial Systems Model (GSM version D2017.aug17), Geoscientific Model Development, 11, 3883–3902, https://doi.org/10.5194/gmd-11-3883-2018, 2018.

Barnes, P., Tabor, D., and Walker, J. C. F.: The Friction and Creep of Polycrystalline Ice, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 324, 127–155, http://www.jstor.org/stable/77933, 1971.

Benn, D. I., Fowler, A. C., Hewitt, I., and Sevestre, H.: A general theory of glacier surges, Journal of Glaciology, 65, 701–716, https://doi.org/10.1017/jog.2019.62, 2019.

Brinkerhoff, D. J. and Johnson, J. V.: Dynamics of thermally induced ice streams simulated with a higher-order flow model, Journal of Geophysical Research F: Earth Surface, 120, 1743–1770, https://doi.org/10.1002/2015JF003499, 2015.

Bueler, E. and Brown, J.: Shallow shelf approximation as a "sliding law" in a thermodynamically coupled ice sheet model, J. Geophys. Res., 114, https://doi.org/10.1029/2008JF001179, 2009.

Bueler, E. and Van Pelt, W.: Mass-conserving subglacial hydrology in the Parallel Ice Sheet Model version 0.6, Geoscientific Model Development, 8, 1613–1635, https://doi.org/10.5194/gmd-8-1613-2015, 2015.

Bueler, E., Brown, J., and Lingle, C.: Exact solutions to the thermomechanically coupled shallow-ice approximation: Effective tools for verification, Journal of Glaciology, 53, 499–516, https://doi.org/10.3189/002214307783258396, 2007.

Calov, R. and Greve, R.: ISMIP HEINO. Ice Sheet Model Intercomparison Project - Heinrich Event INtercOmparison, pp. 1–15, http://www.pik-potsdam.de/~calov/heino/he_setup_2006_11_02.pdf, 2006.

Calov, R., Ganopolski, A., Petoukhov, V., Claussen, M., and Greve, R.: Large-scale instabilities of the Laurentide ice sheet simulated in a fully coupled climate-system model, Geophysical Research Letters, 29, 1–4, https://doi.org/10.1029/2002GL016078, 2002.

Calov, R., Greve, R., Abe-Ouchi, A., Bueler, E., Huybrechts, P., Johnson, J. V., Pattyn, F., Pollard, D., Ritz, C., Saito, F., and Tarasov, L.: Results from the Ice-Sheet Model Intercomparison Project-Heinrich Event INtercOmparison (ISMIP HEINO), Journal of Glaciology, 56, 371–383, https://doi.org/10.3189/002214310792447789, 2010.

Courant, R., Friedrichs, K., and Lewy, H.: Über die partiellen Differenzengleichungen der mathematischen Physik, Mathematische Annalen, 100, https://doi.org/10.1007/BF01448839, 1928.

Cuffey, K. M., Conway, H., Hallet, B., Gades, A. M., and Raymond, C. F.: Interfacial water in polar glaciers and glacier sliding at -17 °C, Geophysical Research Letters, 26, 751–754, https://doi.org/10.1029/1999GL900096, 1999.

Drew, M. and Tarasov, L.: Surging of a Hudson Strait Scale Ice Stream: Subglacial hydrology matters but the process details don't, The Cryosphere Discussions, 2022, 1–41, https://doi.org/10.5194/tc-2022-226, 2022.

Echelmeyer, K. and Zhongxiang, W.: Direct Observation of Basal Sliding and Deformation of Basal Drift at Sub-Freezing Temperatures, Journal of Glaciology, 33, 83–98, https://doi.org/10.3189/s0022143000005396, 1987.

Feldmann, J. and Levermann, A.: From cyclic ice streaming to Heinrich-like events : the grow-and-surge instability in the Parallel Ice Sheet Model, The Cryosphere, 11, 1913–1932, https://doi.org/10.5194/tc-11-1913-2017, 2017.

Flowers, G. E., Björnsson, H., and Pálsson, F.: New insights into the subglacial and periglacial hydrology of Vatnajökull, Iceland, from a distributed physical model, Journal of Glaciology, 49, 257–270, https://doi.org/10.3189/172756503781830827, 2003.

Fowler, A. C.: Sub-Temperate Basal Sliding, Journal of Glaciology, 32, 3–5, https://doi.org/10.3189/S0022143000006808, 1986.

Fowler, A. C. and Johnson, C.: Hydraulic run-away: a mechanism for thermally regulated surges of ice sheets, Journal of Glaciology, 41, 554–561, https://doi.org/10.3189/S002214300003478X, 1995.

Fowler, A. C. and Schiavi, E.: A theory of ice-sheet surges, Journal of Glaciology, 44, 104–118, https://doi.org/10.3189/s0022143000002409, 1998.

Gandy, N., Gregoire, L. J., Ely, J. C., Cornford, S. L., Clark, C. D., and Hodgson, D. M.: Exploring the ingredients required to successfully model the placement, generation, and evolution of ice streams in the British-Irish Ice Sheet, Quaternary Science Reviews, 223, 105 915, https://doi.org/10.1016/j.quascirev.2019.105915, 2019.

Greve, R. and MacAyeal, D. R.: Dynamic/thermodynamic simulations of Laurentide ice-sheet instability, Annals of Glaciology, 23, 328–335, https://doi.org/10.3189/S0260305500013604, 1996.

Greve, R., Takahama, R., and Calov, R.: Simulation of large-scale ice-sheet surges: The ISMIP HEINO experiments, Polar Meteorology and Glaciology, pp. 1–15, http://hdl.handle.net/2115/30205, 2006.

Hank, K.: Supplementary material for "Numerical issues in modeling ice stream surge cycling", https://doi.org/10.5281/zenodo.7905404, 2023.

Hemming, S.: Heinrich events: Massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint, Reviews of Geophysics - REV GEOPHYS, 42, https://doi.org/10.1029/2003RG000128, 2004.

Hindmarsh, R. C.: Consistent generation of ice-streams via thermo-viscous instabilities modulated by membrane stresses, Geophysical Research Letters, 36, 1–6, https://doi.org/10.1029/2008GL036877, 2009.

Joughin, I., Smith, B. E., Howat, I. M., Floricioiu, D., Alley, R. B., Truffer, M., and Fahnestock, M.: Seasonal to decadal scale variations in the surface velocity of Jakobshavn Isbrae, Greenland: Observation and model-based analysis, Journal of Geophysical Research: Earth Surface, 117, 1–20, https://doi.org/10.1029/2011JF002110, 2012.

Joughin, I., Smith, B. E., Shean, D. E., and Floricioiu, D.: Brief communication: Further summer speedup of Jakobshavn Isbræ, Cryosphere, 8, 209–214, https://doi.org/10.5194/tc-8-209-2014, 2014.

K.M. Cuffey and W.S.B. Paterson.: The Physics of Glaciers, Butterworth-Heinemann/Elsevier, Burlington, MA, 4th edn., 2010.

Kyrke-Smith, T. M., Katz, R. F., and Fowler, A. C.: Subglacial hydrology and the formation of ice streams, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 470, https://doi.org/10.1098/rspa.2013.0494, 2014.

MacAyeal, D. R.: Binge/purge oscillations of the Laurentide Ice Sheet as a cause of the North Atlantic's Heinrich events, Paleoceanography, 8, 775–784, https://doi.org/10.1029/93PA02200, 1993.

Mantelli, E., Bertagni, M. B., and Ridolfi, L.: Stochastic ice stream dynamics, Proceedings of the National Academy of Sciences, 113, E4594–E4600, https://doi.org/10.1073/pnas.1600362113, 2016.

Mantelli, E., Haseloff, M., and Schoof, C.: Ice sheet flow with thermally activated sliding. Part 1: the role of advection, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 475, https://doi.org/10.1098/rspa.2019.0410, 2019.

Marshall, S. J. and Clarke, G. K. C.: A continuum mixture model of ice stream thermomechanics in the Laurentide Ice Sheet 2. Application to the Hudson Strait Ice Stream, Journal of Geophysical Research: Solid Earth, 102, 20 615–20 637, https://doi.org/10.1029/97jb01189, 1997.

McCarthy, C., Savage, H., and Nettles, M.: Temperature dependence of ice-on-rock friction at realistic glacier conditions, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375, https://doi.org/10.1098/rsta.2015.0348, 2017.

Mitchell, J. and Soga, K.: Fundamentals of Soil Behavior, John Wiley & Sons, Inc., 3ed edn., 2005.

Papa, B. D., Mysak, L. A., and Wang, Z.: Intermittent ice sheet discharge events in northeastern North America during the last glacial period, Climate Dynamics, 26, 201–216, https://doi.org/10.1007/s00382-005-0078-4, 2006.

Payne, A. J.: Limit cycles in the basal thermal regime of ice sheets, 1995.

Payne, A. J. and Dongelmans, P. W.: Self-organization in the thermomechanical flow of ice sheets, Journal of Geophysical Research: Solid Earth, 102, 12 219–12 233, https://doi.org/10.1029/97jb00513, 1997.

Payne, A. J., Huybrechts, P., Abe-Ouchi, A., Calov, R., Fastook, J. L., Greve, R., Marshall, S. J., Marsiat, I., Ritz, C., Tarasov, L., and Thomassen, M. P.: Results from the EISMINT model intercomparison: The effects of thermomechanical coupling, Journal of Glaciology, 46, 227–238, https://doi.org/10.3189/172756500781832891, 2000.

Pollard, D. and DeConto, R. M.: A Coupled Ice-Sheet/Ice-Shelf/Sediment Model Applied to a Marine-Margin Flowline: Forced and Unforced Variations, Glacial Sedimentary Processes and Products, pp. 37–52, https://doi.org/10.1002/9781444304435.ch4, 2007.

Pollard, D. and DeConto, R. M.: Description of a hybrid ice sheet-shelf model, and application to Antarctica, Geoscientific Model Development, 5, 1273–1295, https://doi.org/10.5194/gmd-5-1273-2012, 2012.

Robel, A. A., Degiuli, E., Schoof, C., and Tziperman, E.: Dynamics of ice stream temporal variability: Modes, scales, and hysteresis, Journal of Geophysical Research: Earth Surface, 118, 925–936, https://doi.org/10.1002/jgrf.20072, 2013.

Roberts, W. H., Payne, A. J., and Valdes, P. J.: The role of basal hydrology in the surging of the Laurentide Ice Sheet, Climate of the Past, 12, 1601–1617, https://doi.org/10.5194/cp-12-1601-2016, 2016.

Sauer, E. K., Egeland, A. K., and Christiansen, E. A.: Preconsolidation of tills and intertill clays by glacial loading in southern Saskatchewan, Canada, Canadian Journal of Earth Sciences, 30, 420–433, https://doi.org/10.1139/e93-031, 1993.

Schannwell, C., Mikolajewicz, U., Ziemen, F., and Kapsch, M.-L.: Sensitivity of Heinrich-type ice-sheet surge characteristics to boundary forcing perturbations, Climate of the Past, 19, 179–198, https://doi.org/10.5194/cp-19-179-2023, 2023.

Shreve, R. L.: Glacier sliding at subfreezing temperatures., Journal of Glaciology, 30, 341–347, https://doi.org/10.1017/S0022143000006195, 1984.

Souček, O. and Martinec, Z.: ISMIP-HEINO experiment revisited: Effect of higher-order approximation and sensitivity study, Journal of Glaciology, 57, 1158–1170, https://doi.org/10.3189/002214311798843278, 2011.

Steen-Larsen, H. C. and Dahl-Jensen, D.: Modelling binge-purge oscillations of the Laurentide ice sheet using a plastic ice sheet, Annals of Glaciology, 48, 177–182, https://doi.org/10.3189/172756408784700635, 2008.

Takahama, R.: Heinrich Event Intercomparison with the ice-sheet model SICOPOLIS, Master's thesis, http://hdl.handle.net/2115/28749, 2006.

Tarasov, L. and Peltier, W. R.: A high-resolution model of the 100 ka ice-age cycle, Annals of Glaciology, 25, 0–7, https://doi.org/10.3189/s026030550001380x, 1997.

Tarasov, L., Dyke, A. S., Neal, R. M., and Peltier, W. R.: A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling, Earth and Planetary Science Letters, 315-316, 30–40, https://doi.org/10.1016/j.epsl.2011.09.010, 2012.

Tarasov, L., Hank, K., and Lecavalier, B. S.: GSMv01.31.2023 code archive for LISsq experiments, https://doi.org/10.5281/zenodo.7668472, 2023.

Tulaczyk, S., Kamb, W. B., and Engelhardt, H. F.: Basal mechanics of Ice Stream B, west Antarctica: 2. Undrained plastic bed model, Journal of Geophysical Research: Solid Earth, 105, 483–494, https://doi.org/https://doi.org/10.1029/1999JB900328, 2000a.

Tulaczyk, S., Kamb, W. B., and Engelhardt, H. F.: Basal mechanics of Ice Stream B, West Antarctica 1. Till mechanics, Journal of Geophysical Research: Solid Earth, 105, 463–481, https://doi.org/10.1029/1999jb900329, 2000b.

Van Pelt, W. J. and Oerlemans, J.: Numerical simulations of cyclic behaviour in the Parallel Ice Sheet Model (PISM, Journal of Glaciology, 58, 347–360, https://doi.org/10.3189/2012JoG11J217, 2012.

Werder, M. A., Hewitt, I. J., Schoof, C. G., and Flowers, G. E.: Modeling channelized and distributed subglacial drainage in two dimensions, Journal of Geophysical Research: Earth Surface, 118, 2140–2158, https://doi.org/10.1002/jgrf.20146, 2013.

Winkelmann, R., Martin, M. A., Haseloff, M., Albrecht, T., Bueler, E., Khroulev, C., and Levermann, A.: The Potsdam Parallel Ice Sheet Model (PISM-PIK) Part 1: Model description, The Cryosphere, 5, 715–726, http://www.the-cryosphere.net/5/715/2011/tc-5-715-2011.pdf, 2011.

Winsborrow, M. C., Clark, C. D., and Stokes, C. R.: What controls the location of ice streams?, Earth-Science Reviews, 103, 45–59, https://doi.org/10.1016/j.earscirev.2010.07.003, 2010.

Ziemen, F. A., Rodehacke, C. B., and Mikolajewicz, U.: Coupled ice sheet-climate modeling under glacial and pre-industrial boundary conditions, Climate of the Past, 10, 1817–1836, https://doi.org/10.5194/cp-10-1817-2014, 2014.

Ziemen, F. A., Kapsch, M. L., Klockmann, M., and Mikolajewicz, U.: Heinrich events show two-stage climate response in transient glacial simulations, Climate of the Past, 15, 153–168, https://doi.org/10.5194/cp-15-153-2019, 2019.

Metric base setup (abrupt transition) reference values 3.125 km wide transition 25 km wide transition 3.125 km wide transition with HB/HS topography 25 km wide transition with HB/HS topography number of events 180 ± 100 -4.2 ± 8.9 1.0 ± 11.4 36.3 ± 17.3 19.9 ± 22.6 mean period 1.1 ± 0.5 kyr 4.6 ± 9.2 -0.4 ± 10.4 2.2 ± 48.0 14.5 ± 45.1 mean duration 0.3 ± 0.1 kyr 2.7 ± 3.6 7.2 ± 4.4 10.2 ± 17.5 24.3 ± 9.0mean pseudo-Hudson Strait ice volume change $1.7 \pm 0.2 \cdot 10^3$ km^3 0.2 ± 4.8 -1.7 ± 4.1 8.4 ± 10.0 17.3 ± 15.6RMSE - 7.9 ± 2.3 8.0 ± 2.2 11.2 ± 1.8 12.2 ± 2.0Mean Bias - 0.0 ± 0.2 -0.6 ± 0.5 -6.2 ± 1.9 -6.6 ± 2.2

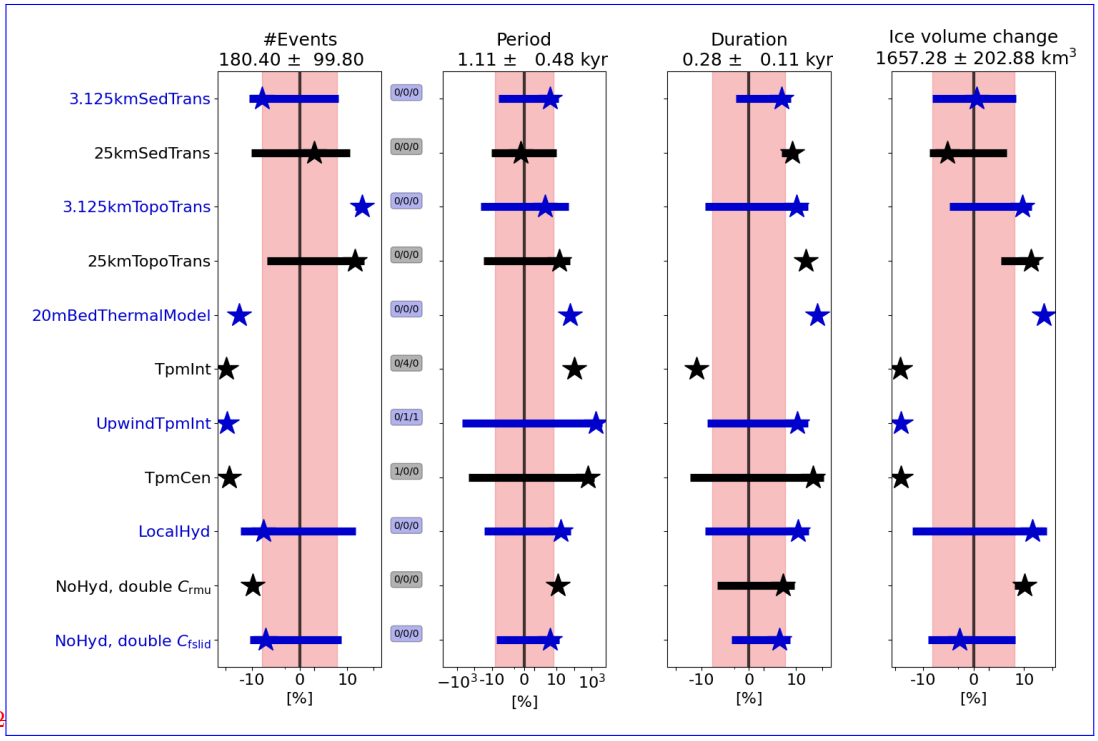


Figure 6. Percentage differences of event in surge characteristics , pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM base reference setup for runs with a smooth transition between hard bedrock and soft sedimentmodel setups discussed in Sec. 3.3.1 (average of the 5 parameter vectors). The horizontal grid resolution is 3.125 km. The different colors were added for visual alignment of the individual model setups, the stars are the ensemble mean percentage differences, and runs with a pseudo-Hudson Bay/Hudson Strait the horizontal bars represent the ensemble standard deviations. The shaded pink regions mark the MNEEs (HB/HSTable 5) topographyand the black numbers in the title of each subplot represent the mean values of the reference setup. No runs The 3 small numbers between the first two columns represent the number of crashed runs (nC), the number of runs without a surge (nS0), and all the number of runs had more than 1 with only one surge event(nS1), respectively. The first 20 kyr of each run are treated as a spin-up interval for and are not considered in the event characteristics above. The x-axis is logarithmic. Further details of each individual experiments are provided in the subsequent sections and the supplement. The model setups, from top to bottom, are: 3.125 km wide sediment transition zone (except for instead of an abrupt transition in the RMSE reference setup), 25 km wide sediment transition zone, 3.125 km wide sediment transition zone with pseudo-Hudson Bay/Hudson Strait topography (instead of a flat topography in the reference setup), 25 km sediment transition zone with pseudo-Hudson Bay/Hudson Strait topography, 20 m deep (1 layer) bed thermal model (instead of a 1 km deep bed thermal model (17 non-linearly-spaced layers) in the reference setup), 3 different approaches to calculate basal grid cell interface temperature (TpmInt, upwind TpmInt, TpmCen), local hydrology (instead of no hydrology), and mean biasdoubling the values of the soft and hard bed sliding coefficients (as an attempt to represent basal hydrology without actually adding it).

The width of the transition zone ($\sim$200 m to sea level) affects the position and width of the major surge events (tendency towards wider surges for a wider transition zone (video 04 of Hank (2023))). The pseudo-Hudson Strait topography also suppresses the small surge events otherwise observed in the vicinity of the pseudo-Hudson Strait itself.

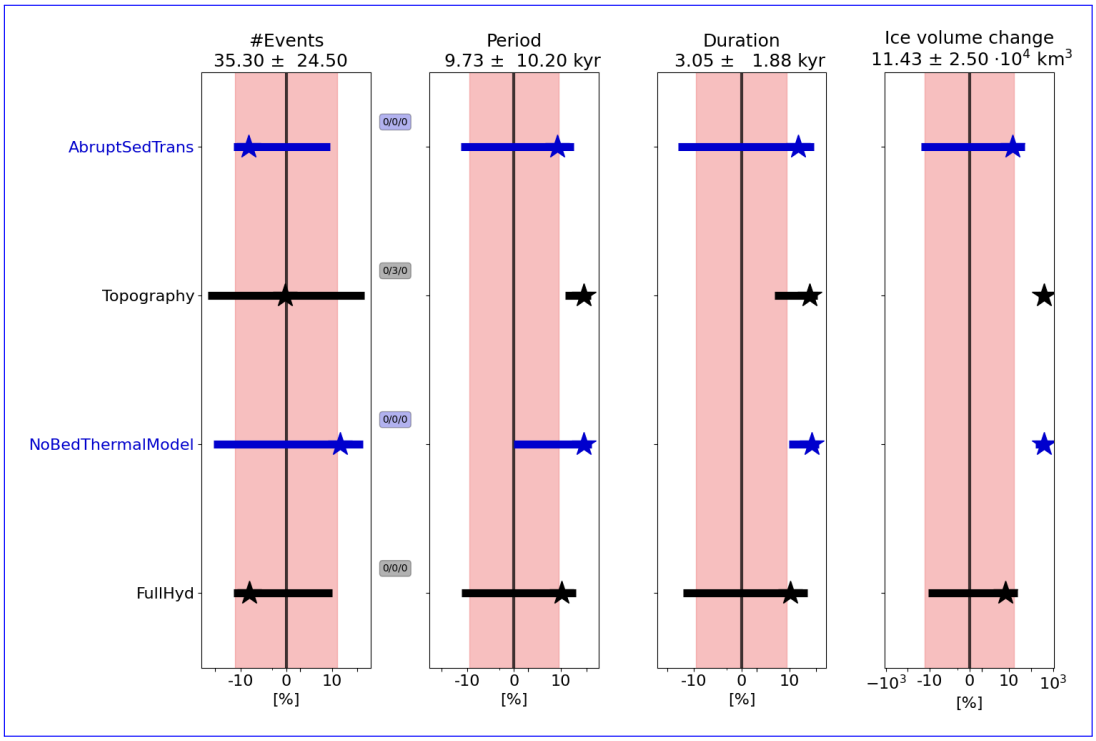


Figure 7. Percentage differences in surge characteristics compared to the PISM reference setup for model setups discussed in Sec. 3.3.1 (average of the 9 parameter vectors). Otherwise same as Fig. 6. The model setups, from top to bottom, are: abrupt sediment transition (instead of the transition shown in, e.g., Fig. S8), pseudo-Hudson Bay/Hudson Strait topography (instead of a flat topography in the reference setup, Fig. S9), no bed thermal model (instead of a 1 km deep bed thermal model (20 equally-spaced layers) in the reference setup), and a mass-conserving horizontal transport model for basal hydrology (instead of a local hydrology).

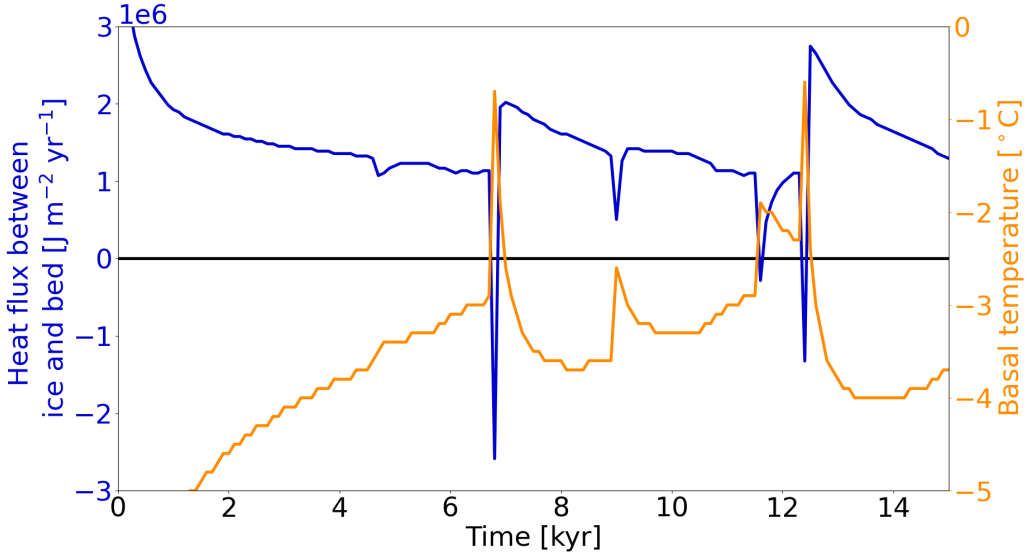


Figure 8. Heat flux at the base of the ice sheet (positive from bed into ice) and basal ice temperature for a grid cell in the center of the pseudo-Hudson Strait (grid cell center at $x = 376.5625$ km and $y = 248.4375$ km, white star in Fig. 1) and parameter vector 1 with the 1 km deep bed thermal model (17 non-linearly-spaced levels) using the GSM. The horizontal grid resolution is 3.125 km.

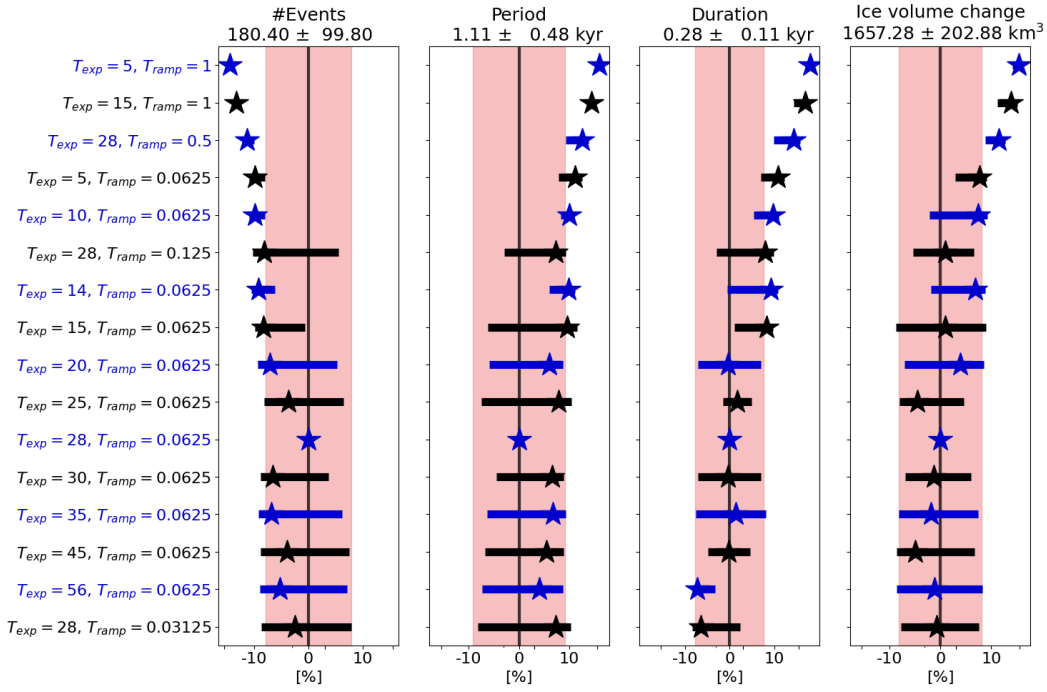


Figure 9. Percentage differences in ~~event~~ surge characteristics compared to the GSM ~~base~~ reference setup ($T_{ramp} = 0.0625$, $T_{exp} = 28$) for different basal temperature ramps at 3.125 km horizontal grid resolution (average of the 5 parameter vectors). The ramps are sorted from widest (first row) to sharpest (last row, see Fig. S24 for a visualization of all ramps). ~~The shaded pink regions mark the numerical noise estimates (Tab~~Otherwise same as Fig. ~~5) and the black numbers in the title of each subplot represent the mean values of the base setup~~6. No runs crashed and all runs had more than 1 surge event. The ~~first 20 kyr of each run are treated as a spin-up interval and are not considered in the above. The x-axis is logarithmic. The~~ exact values are given in ~~Tab.~~Table S10.

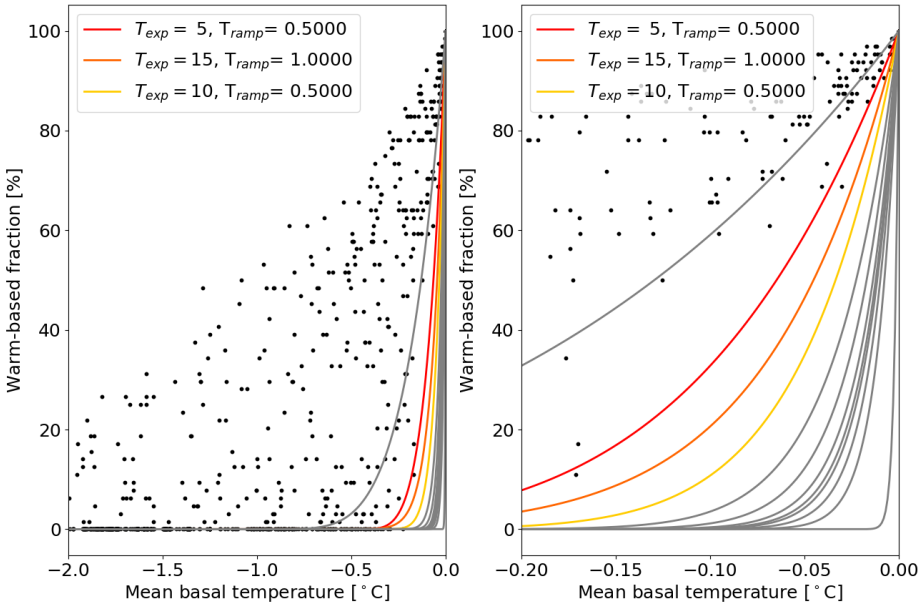


Figure 10. Warm-based fraction (basal temperature with respect to the pressure melting point at 0 °C) vs. mean basal temperature with respect to the pressure melting point when upscaling a 3.125 km run to 25 km horizontal grid resolution including all 5 parameter vectors using the GSM. For example, an upscaled 25 km patch (containing 64 3.125 km grid cells) with 32 3.125 km grid cells at the pressure melting point and 32 3.125 km grid cells at −1°C with respect to the pressure melting point has a warm-based fraction of 50 % and a mean basal temperature of −0.5°C. Only grid cells within the pseudo-Hudson Strait and time steps within the surges of the 10 kyr after the first surge are considered. The restriction to the 10 kyr after the first surge for these experiments is set by storage limitations due to the high temporal resolution of the model output fields (10 yr). The colored ramps correspond to the 25 km horizontal grid resolution basal temperature ramps in Table S11 and the gray lines show all other ramps that were tested at this resolution.

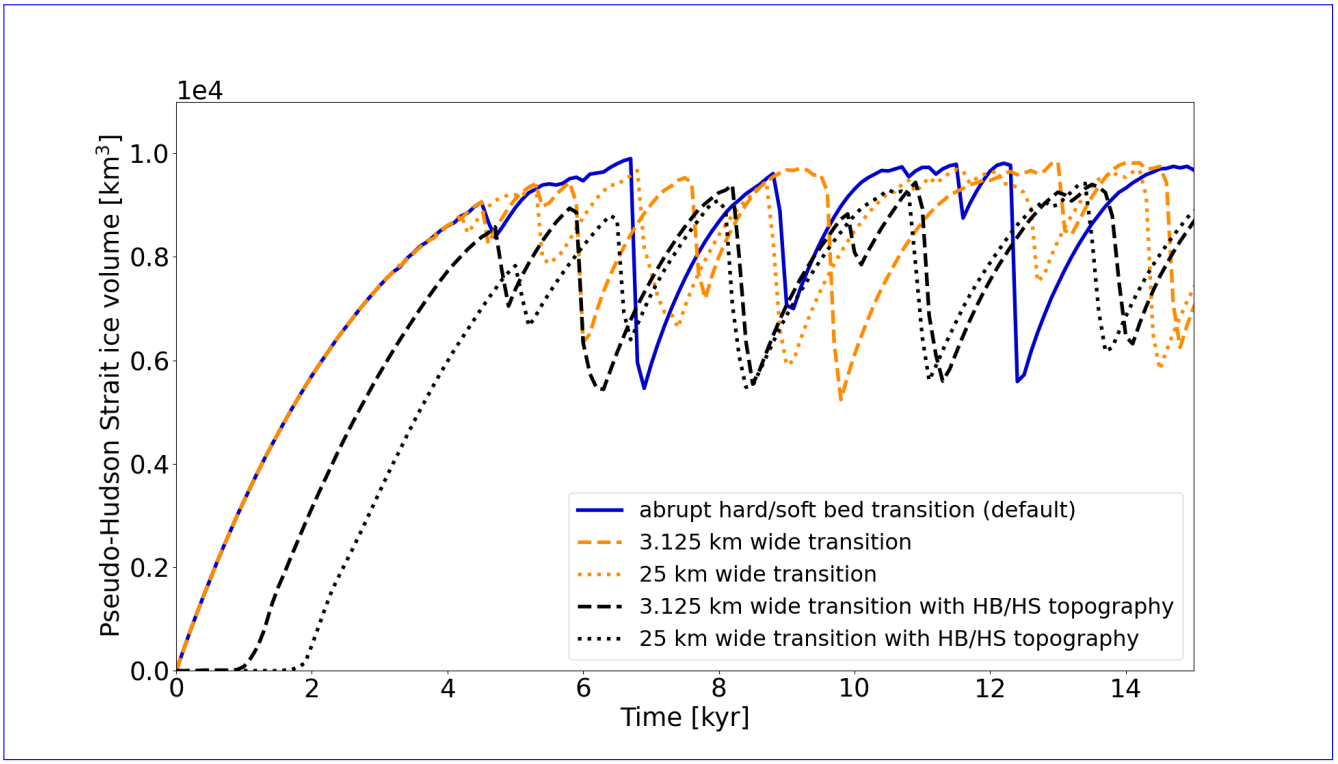


Figure 11. Pseudo-Hudson Strait ice volume for GSM parameter vector 1 and three different bed configurations. The horizontal grid resolution is 3.125 km. Note that the width of the topographical transition zone matches the width of the soft to hard bed transition zone. In experiments with a pseudo-Hudson Bay/Hudson Strait (HB/HS) topography, the pseudo-Hudson Strait topography is below sea level, increasing the time required for glaciation. A wider transition zone (larger area below sea level) leads to a later glaciation.

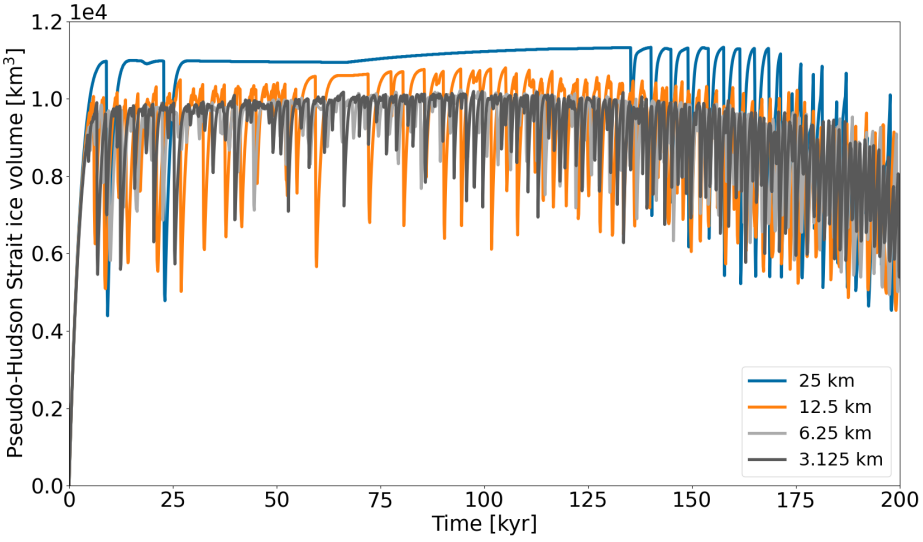


Figure 12. Pseudo-Hudson Strait ice volume for parameter vector 1 and different horizontal grid resolutions using the GSM. A resolution-dependent temperature (Eq. (9)) with ~~$F_{T_{ramp}} = 1$~~ $P_{T_{ramp}} = 1$ and $T_{exp} = 28$ is used for all horizontal grid resolutions (matching colors in Fig. 2).

Supplement to "Numerical issues in modeling ~~ice sheet instabilities such as binge-purge type cyclic~~ thermally and hydraulically driven ice stream ~~surging~~surge cycling"

Kevin Hank[1], Lev Tarasov[1], and Elisa Mantelli[2,3,4]

[1]Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, NL, A1B 3X7,Canada
[2]Department of Earth and Environmental Sciences , Ludwig-Maximillians-Universitaet Munich, Theresienstr. 41, 80333 Munich, Germany
[3]Alfred Wegener Institute for Polar and Marine Research, Am Alten Hafen 26, 27568 Bremerhaven, Germany
[4]Institute for Marine and Antarctic Studies, University of Tasmania, 20 Castray Esplanade, Battery Point TAS 7004, Australia
[*]khank@mun.ca

Correspondence: Kevin Hank (khank@mun.ca)

S1 GSM - ~~Climate forcing~~Details of different model aspects

~~The asymmetric temperature forcing in the GSM (orange line in Fig. S1) is calculated according to~~

$$T_{asym} = \left| \left(\frac{t}{200\ \mathrm{kyr}} \cdot 3 + 2 \right) - 1 \right| \cdot 5$$

~~, where t is the model time ranging from -200 kyr to 0 kyr (instead of 0 kyr to 200 kyr).~~

S1.1 Climate forcing

S1.2 SSA activation velocities

S2 ~~GSM - Parameter vectors~~

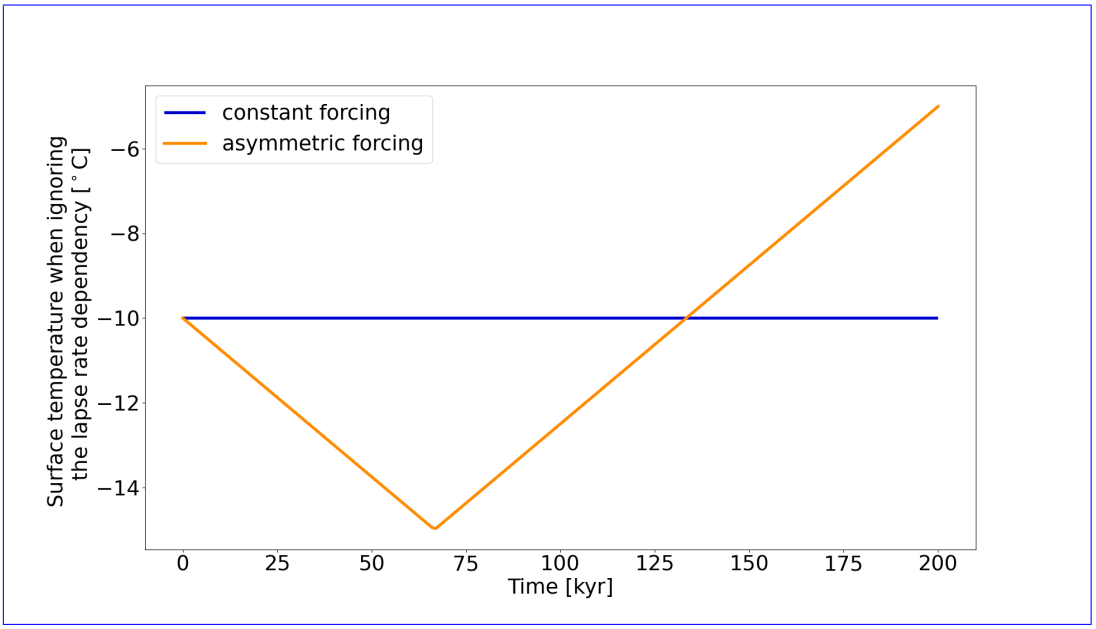


Figure S1. Constant and asymmetric temperature forcing in the GSM. The coldest temperature is reached at 66.7 kyr. For the case shown here, the surface temperature constant is set to ~~rTnorth~~rTsurf $= -10°$C (~~Tab.~~Table 1). All model runs within this paper use the asymmetric forcing.

Setup	number of surges	mean duration	mean period	mean pseudo-Hudson Strait ice volume change
reference setup	180 ± 100	1.1 ± 0.5 kyr	0.3 ± 0.1 kyr	$1.7 \pm 0.2 \cdot 10^3$ km^3
$v_{\mathrm{SIA,crit}} = 20$ m yr^{-1}	-3.7 ± 7.0	3.2 ± 6.4	1.5 ± 2.1	3.2 ± 2.4
$v_{\mathrm{SIA,crit}} = 40$ m yr^{-1}	-5.5 ± 5.4	6.1 ± 6.8	2.4 ± 5.7	3.5 ± 9.0
SSA everywhere	7.3 ± 24.8	1.7 ± 27.6	-9.3 ± 14.1	-17.7 ± 29.7

Table S1. Percentage differences of surge characteristics between the GSM reference setup (first row) and runs with different SSA activation velocities at 3.125 km. By default, the SSA is activated once the SIA velocity exceeds $v_{\mathrm{SIA,crit}} = 30$ m yr^{-1}. No runs crashed and all runs had more than 1 surge. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above.

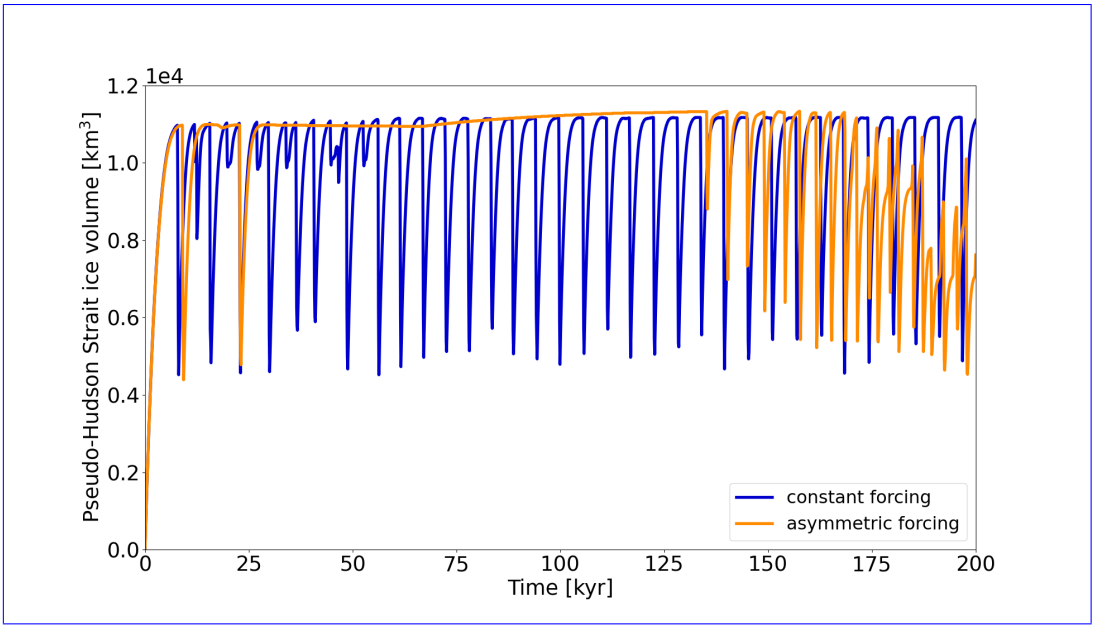


Figure S2. Pseudo-Hudson Strait ice volume for a constant and asymmetric temperature forcing in the GSM (Fig. S1). This plot shows parameter vector 1 with a horizontal grid resolution of 25 km.

S1.1 Parameter vectors

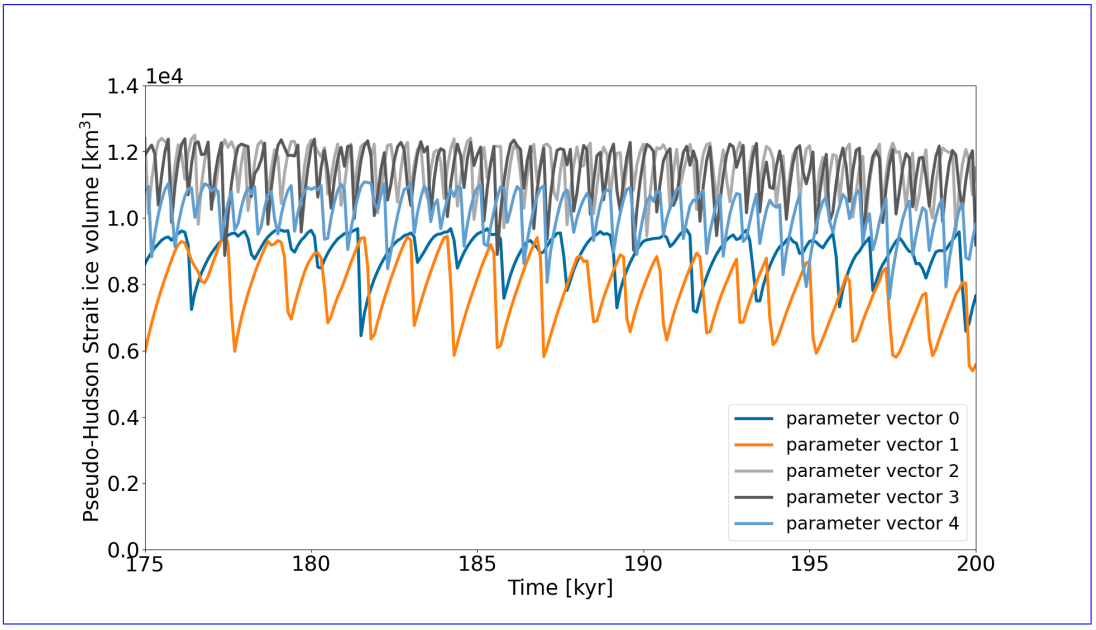


Figure S3. Pseudo-Hudson Strait ice volume for the last 25 kyr of all 5 GSM parameter vectors when using the ~~base~~ reference setup. Note that only the last 25 kyr are shown for better visibility of the individual oscillation pattern.

S2 ~~GSM - Details of different model setups~~

S1.1 Bed properties

The effects of an abrupt transition from hard bedrock (0 % sediment cover) to soft sediment (100 % sediment cover) are examined by adding a smooth transition zone (Fig. S5 a)). Two widths of this transition zone (25 km and 3.125 km) are investigated. The basal velocity (or more precisely the sliding coefficient C in Eq. (6b)) then depends on the sediment cover within a grid cell (Fig. S4). In the experiments with a non-flat topography, the bed of the pseudo-Hudson Bay and Hudson Strait is placed 200 m and the surrounding ocean 500 m below the sea level (Fig. S5 b). The topographic transition zones (25 km and 3.125 km wide) align with the sediment transition zones.

~~Basal velocity at 50 kPa basal drag for variable sediment cover and a power-law exponent of 3 (n_b in Tab. 1).~~

~~Sediment cover and topography map for a 25 km wide transition zone at 3.125 km horizontal grid resolution. The transition zones for topography and sediment cover are at the same locations. The magenta line outlines the 100 % soft-bedded pseudo-Hudson Bay and Hudson Strait.~~

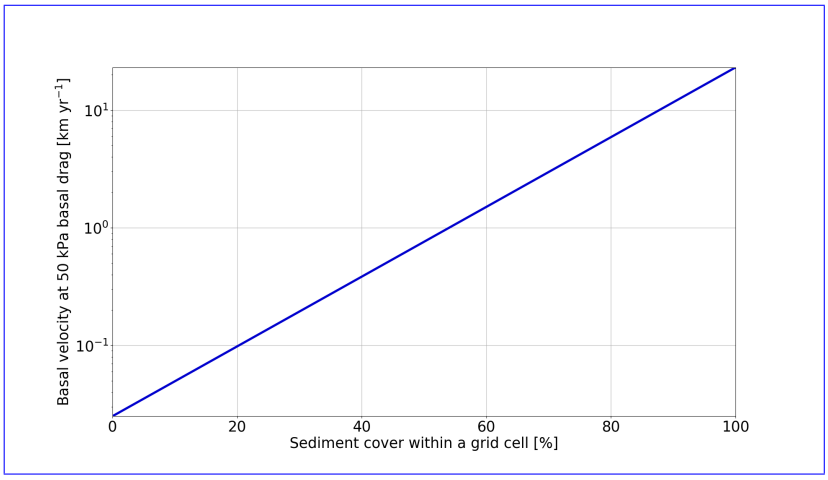
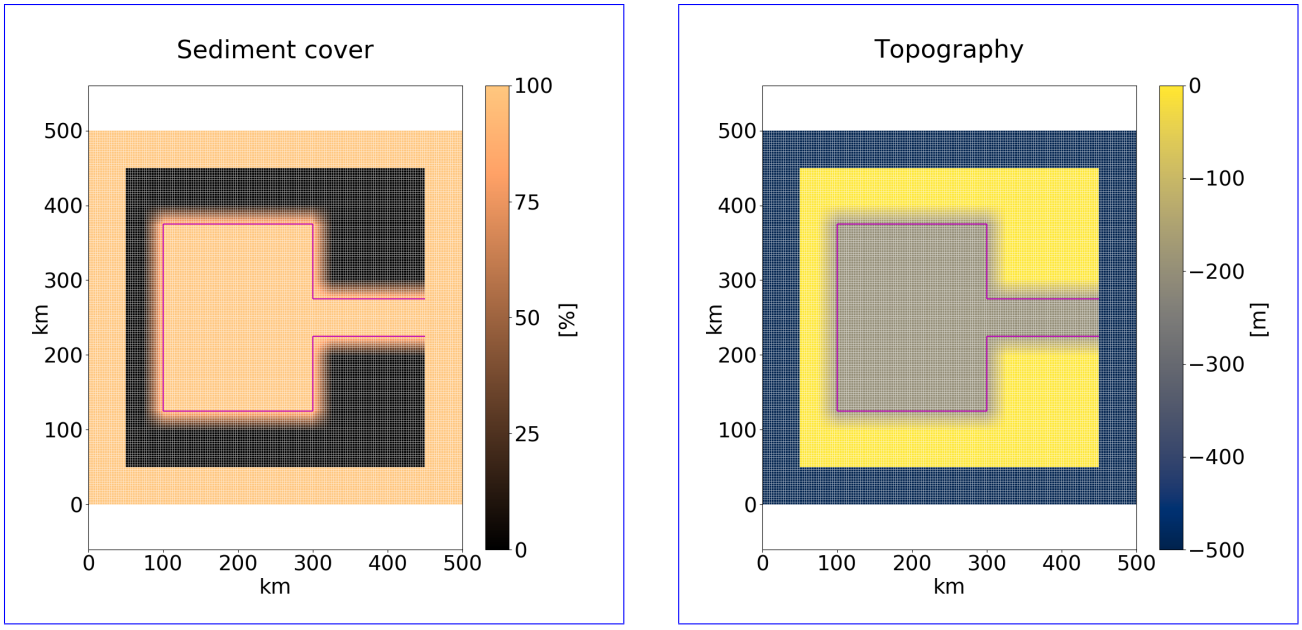


Figure S4. Basal velocity at 50 kPa basal drag for variable sediment cover and a power-law exponent of 3 (n_b in Table 1).



(a) (b)

Figure S5. Sediment cover and topography map for a 25 km wide transition zone at 3.125 km horizontal grid resolution. The transition zones for topography and sediment cover are at the same locations. The magenta line outlines the 100 % soft-bedded pseudo-Hudson Bay and Hudson Strait.

S1.2 ~~Approches to determine the basal temperature at the grid cell interface~~

The most straightforward approach to determining the basal temperature with respect to the pressure melting point at the grid cell interface ($T_{bp,I}$) is to use the mean of the two adjacent basal **T**emperatures with respect to the **p**ressure **m**elting point at the grid cell **Cen**ters (TpmCen).

$$T_{bp,I} = 0.5 \cdot (T_{bp,L} + T_{bp,R})$$

where $T_{bp,L}$ and $T_{bp,R}$ are the grid cell center basal temperatures with respect to the pressure melting point to the left and right of the interface, respectively. Similarly for upper and lower grid cells adjacent to a horizontally aligned interface. However, this approach does not explicitly account for ice thickness changes at the grid cell interface.

TpmInt, on the other hand, calculates the basal temperature at the **Int**erface (T_I) by averaging the adjacent grid cell center basal temperatures (T_L and T_R, Eq. (S1)). $T_{bp,I}$ is then determined by using the interface ice sheet thickness (average of adjacent grid cell center ice thicknesses H_L and H_R, Eq. (**??**)).

$$T_I = 0.5 \cdot (T_L + T_R)$$

$$T_{bp,I} = T_I + \beta_P \frac{H_L + H_R}{2}$$

where $\beta_P = 8.7 \cdot 10^{-4}\,^\circ\mathrm{C\ m^{-1}}$ is the standard basal melting point depression coefficient.

The last approach (TpmTrans) accounts for extra heat available beyond that needed to reach the local pressure melting point (T_{add}).

$$T_{Im,C} = T_C + T_{add} \tag{S1a}$$

, where T_C and $T_{Im,C}$ are the basal temperature at the grid cell center and the basal temperature in the intermediate calculation step, respectively. The basal temperature with respect to the pressure melting point at each adjacent grid cell center $T_{bp,Im,C}$ is then calculated using the interface ice thickness.

$$T_{bp,Im,C} = T_{Im,C} + \beta_P \frac{H_L + H_R}{2} \tag{S1b}$$

In the intermediate steps to calculate the interface temperature (Eq. (S1a) and (S1b)), $T_{Im,C}$ and $T_{bp,Im,C}$ are allowed to exceed the pressure melting point. This temporary higher basal temperature is an attempt to account for heat transported to the interface by ice advection and basal water.

$$\text{IF } T_{bp,Im,C} > 0^\circ\mathrm{C}: \quad T_{bp,Im,C} = \min\left(0.5^\circ\mathrm{C}, 0.5 \cdot T_{bp,Im,C}\right) \tag{S1c}$$

Averaging the adjacent basal temperatures with respect to the pressure melting point at the grid cell center ($T_{bp,Im,L}$ and $T_{bp,Im,R}$) yields the final basal temperature with respect to the pressure melting point at the interface ($T_{bp,I}$).

$$T_{bp,I} = 0.5 \cdot (T_{bp,Im,L} + T_{bp,Im,R}) \tag{S1d}$$

Note that neither the grid cell center nor the interface basal temperature may exceed the pressure melting point (only the basal temperature in the intermediate calculation steps). The GSM base setup uses TpmTrans.

S1.2 Weighting function of the adjacent minimum basal temperature

A weighting function takes into account the adjacent minimum basal temperature for the basal sliding temperature ramp.

$$T_{bp,I} = W_{Tb,\min} \cdot \min\left[T_{bp,L}, T_{bp,R}\right] + T_{bp,I} \cdot \left(1 - W_{Tb,\min}\right), \tag{S2}$$

, where $T_{bp,I}$ is the basal temperature with respect to the pressure melting point at the grid cell interface, and $T_{bp,L}$ and $T_{bp,R}$ are the basal temperatures with respect to the pressure melting point at the adjacent grid cell centers. Note that $T_{bp,Im,L}$ and $T_{bp,Im,R}$ instead of $T_{bp,L}$ and $T_{bp,R}$ are used when calculating $T_{bp,I}$ according to TpmTrans (Eq. (S119)). In this way, the additional heat T_{add} is still considered even when $W_{Tb,\min} = 1$.

S1.3 Local basal hydrology

The local basal hydrology sets the basal water thickness by calculating the difference between the basal melt rate and a constant basal drainage rate (rBedDrainRate in Tab. 1). This subglacial hydrology provides a simple and computationally efficient way to capture changes in basal sliding velocities due to effective pressure variations (Drew and Tarasov, 2022, under open review). However, it does not account for basal ice accumulation, englacial or supraglacial water input, or horizontal water transport.

The basal water thickness ($\mathrm{h_{wb}}$) and an estimated effective bed roughness scale ($\mathrm{h_{wb,Crit}}$ in Tab. 1) determine the effective pressure coefficient

$$N_{\mathrm{C,eff}} = 1 - \min\left(\frac{\mathrm{h_{wb}}}{\mathrm{h_{wb,Crit}}}, 1.0\right)^{3.5}$$

The basal water thickness is limited to $\mathrm{h_{wb,Crit}} = 10$ m and is set to $\mathrm{h_{wb}} = 0$ m where the ice thickness is less than 10 m and where the temperature with respect to the pressure melting point is below -0.1°C. Experiments with $\mathrm{h_{wb,Crit}} = 5$ m yield the same results, and removing all the water for $H < 1$ m, $H < 50$ m, and $T_{bp} < -0.5$°C does not significantly (according to Sec. 3.2) affect the model results. The effective pressure at the grid cell interface is then

$$N_{\mathrm{eff}} = g\rho_{\mathrm{ice}} \cdot 0.5\left(H_L N_{\mathrm{C,eff,L}} + H_R N_{\mathrm{C,eff,R}}\right)$$

, where $g = 9.81\ \mathrm{m\,s^{-2}}$ is the acceleration due to gravity, $\rho_{\mathrm{ice}} = 910\ \mathrm{kg\,m^{-3}}$ the ice density, H the ice thickness and the subscripts L and R denote the adjacent grid cells to the left and right of the interface, respectively (similarly for upper and lower grid cells adjacent to a horizontally aligned interface). We enforce that N_{eff} never falls below 10 kPa (denominator in Eq. (??), similar results for $N_{\mathrm{eff,min}} = 5$ kPa). Finally, the effective pressure of each grid cell alters the basal sliding coefficient in the sliding law (Eq. (6a)) according to

$$C_b = C_b \cdot \min\left(10, \max\left(0.5, \frac{N_{\mathrm{eff,Fact}}}{N_{\mathrm{eff}} + 10^4\ \mathrm{Pa}}\right)\right)$$

, where $N_{\mathrm{eff,Fact}}$ is the effective pressure factor (Tab. 1). The change of the basal sliding coefficient C_b is, therefore, limited to $C_b \cdot 0.2$ to $C_b \cdot 10$. Allowing a larger change of $C_b \cdot 0.1$ to $C_b \cdot 20$ does not significantly (according to Sec. 3.2) change the model results.

S2 PISM - ~~Input fields~~Details of different model aspects

S2.1 Input fields

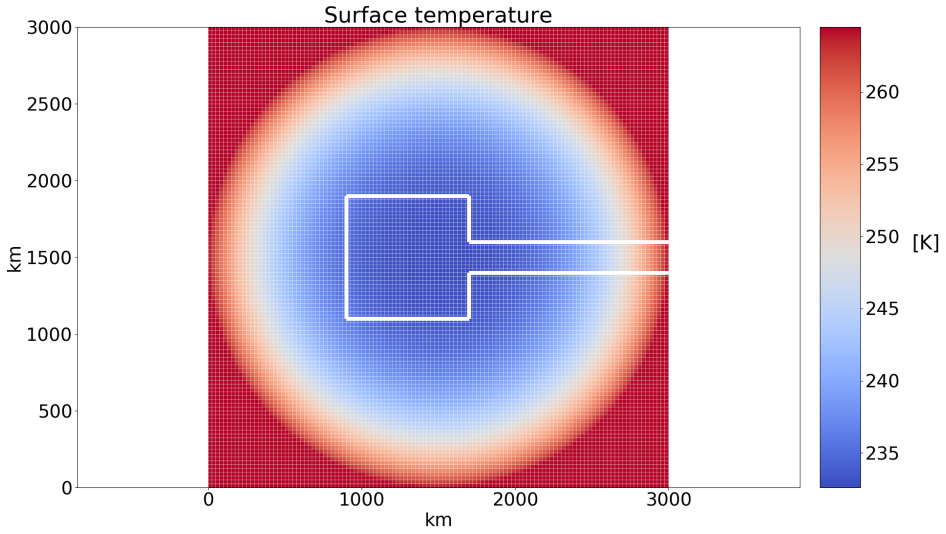


Figure S6. PISM surface temperature input field for parameter vector 1. The corresponding parameter values of T_{min} and S_t are 232.60 K and $9.45 \cdot 10^{-9}$ K km^{-3}, respectively. Thick white lines outline the simplified soft-bedded pseudo-Hudson Bay/Hudson Strait area. The horizontal grid resolution is 25x25 km.

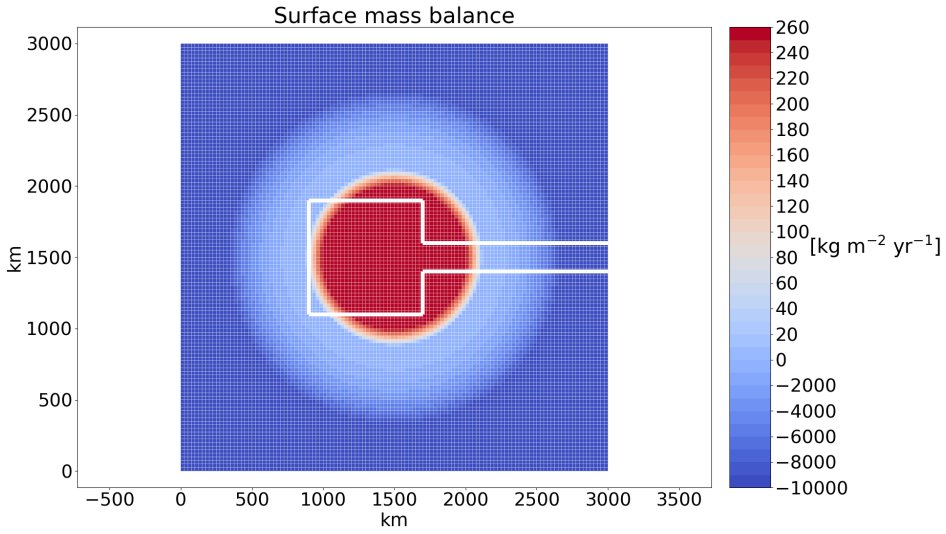


Figure S7. PISM surface mass balance input field for parameter vector 1. The corresponding parameter values of B_{max} and S_b are 408.81 kg m^{-2} yr^{-1} and $4.55 \cdot 10^{-12}$ kg m^{-2} yr^{-1} km^{-5}, respectively. Thick white lines outline the simplified soft-bedded pseudo-Hudson Bay/Hudson Strait area. The horizontal grid resolution is 25x25 km.

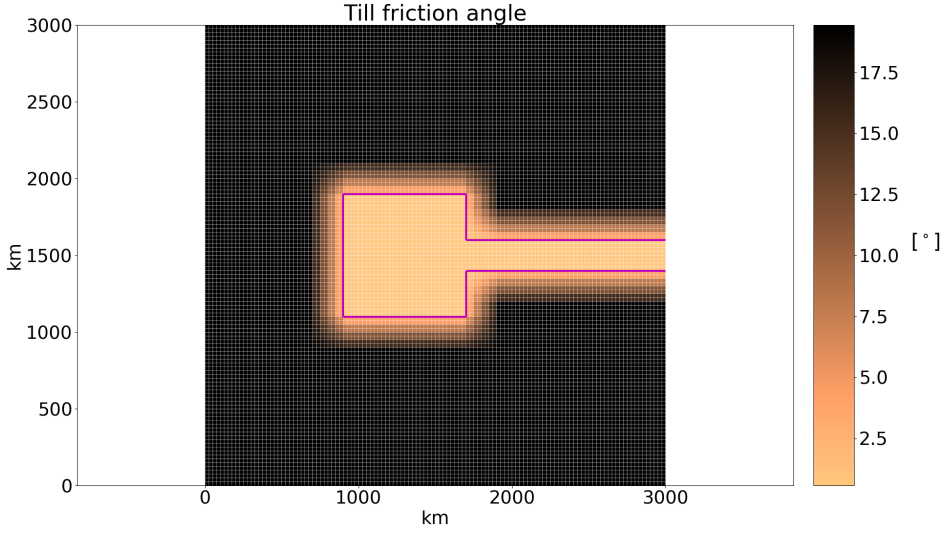


Figure S8. PISM till friction angle input field for parameter vector 1. The corresponding parameter values of $soft$ and $hard$ are 0.56°C and 19.44°C, respectively. Magenta lines outline the simplified soft-bedded pseudo-Hudson Bay/Hudson Strait area. The horizontal grid resolution is 25x25 km.

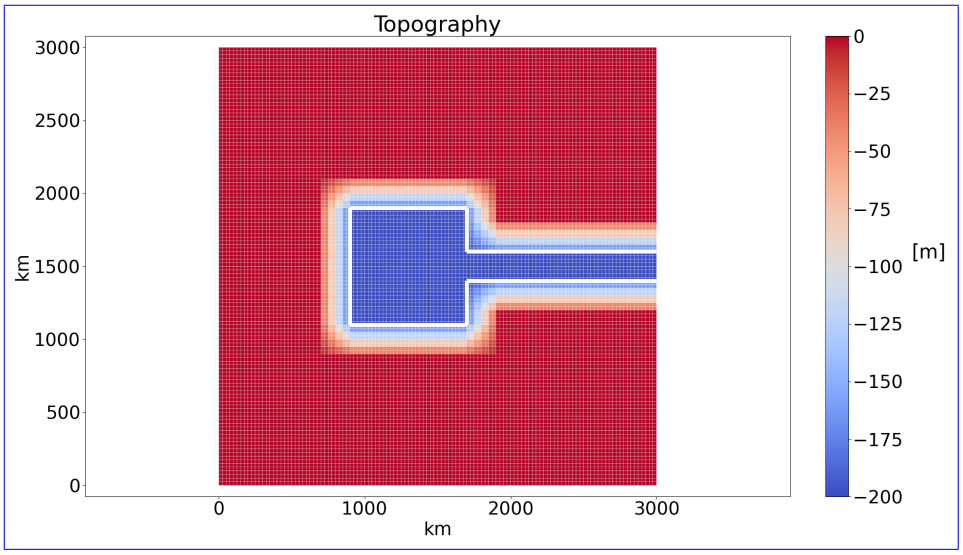


Figure S9. PISM topography input field (same for all parameter vectors). The white lines outline the simplified soft-bedded pseudo-Hudson Bay/Hudson Strait area. The horizontal grid resolution is 25x25 km.

S3 ~~PISM - Parameter vectors~~

S2.1 Parameter vectors

S3 ~~PISM - Maximum magnitude of basal ice velocity~~

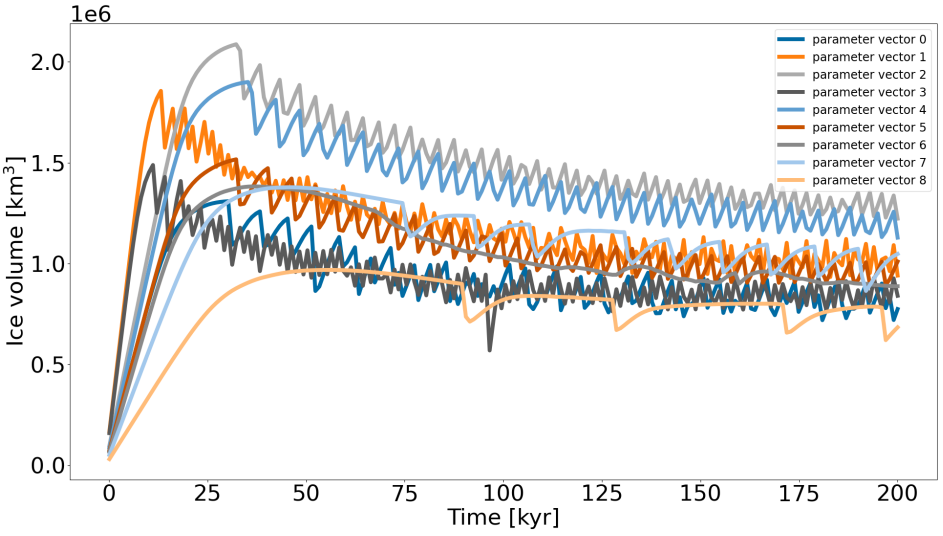


Figure S10. Ice volume in the eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait for all ~~10~~ 9 PISM parameter vectors when using the ~~base~~ reference setup.

S2.1 Maximum magnitude of basal ice velocity

Small till friction angles (0.5 to 1.0°) lead to slippery beds and high maximum basal sliding velocities (up to ~ 600 km yr^{-1}) for a small number of time steps in some runs. A maximum of 7 out of 2000 time steps exceeds 50 km yr^{-1} (parameter vector 1 in Fig. S11). While observed velocities can reach several hundreds of meters per day for short periods (K.M. Cuffey and W.S.B. Paterson. (2010), e.g., 300 m d^{-1} = 109.5 km yr^{-1}), high modeled velocities might lead to instabilities in the numerical matrix solver.

~~To test if the model response to different setups is affected by such numerical instabilities, we create a subensemble of the 10 base parameter vectors, including only runs with a maximum basal sliding velocity $v_{max} < 50$ km yr^{-1} (parameter vectors 6, 7, and 8 in Fig. S11). We repeat the analysis in Sec. 2.3 for the subensemble and compare the event characteristics to the full ensemble~~Therefore, we set an upper limit of 40 km yr^{-1} for the SSA velocity.

~~Setup #Events mean period mean duration mean ice volume change nE1 full ensemble base setup (nCores= 8) 28 ± 17 10 ± 12 kyr 3 ± 2 kyr $1.2 \pm 0.3 \cdot 10^5$ km^3 0 subensemble base setup (nCores= 8) 5 ± 2 26 ± 11 kyr 7 ± 2 kyr $1.2 \pm 0.5 \cdot 10^5$ km^3 0 nCores= 2, full ensemble 0.8 ± 25.0 -8.3 ± 45.0 -2.3 ± 19.8 1.5 ± 6.2 0 nCores= 2, subensemble -22.2 ± 15.7 11.5 ± 15.3 7.5 ± 7.2 0 nCores= 4, full ensemble 1.9 ± 34.1 -7.0 ± 15.0 -8.8 ± 15.0 0.8 ± 4.4 1 nCores= 4, subensemble -15.8 ± 17.6 1.1 ± 5.2 1 nCores= 16, full ensemble -1.2 ± 16.5 6.2 ± 42.1 0 nCores= 16, subensemble -11.1 ± 15.7 37.5 ± 68.6 -1.4 ± 18.5 -6.1 ± 4.5 0 nCores= 32, full ensemble -6.3 ± 15.3 0.3 ± 12.7 0 nCores= 32, subensemble -4.2 ± 5.9 -1.2 ± 17.7 -9.8 ± 5.3 0 N_{diff} 35.2 51.8 21.0 8.9 - Noise estimates in percent (except first two rows) for the full ensemble (all 10 parameter vectors) and the subensemble including only the 3 runs with a maximum basal sliding velocity $v_{max} < 50$ km yr^{-1} (Fig. S11). No runs crashed and all runs showed at least 1 event. Runs with just one event (nE1) are ignored when caluclating the change in mean period. The maximum changes across all 4 experiments are marked as bold numbers for the the full- and subensemble. The last row shows the maximum noise estimates for the difference in event characteristics between the full- and subensemble described in the text and used in Fig. **??**.~~

~~The maximum noise estimates for each event characteristic are on the same order of magnitude (mean duration and ice volume change) or larger for the low-velocity subensemble (number of events and mean period, bold numbers in Tab. **??**). Note that the larger noise estimates might be partly due to the lower number of runs considered in the subensemble. Since the 4 noise experiments show positive and negative changes for the same event characteristic (e.g., for the mean duration of the subensemble), we assume that the maximum noise estimates also represent changes in both directions (e.g., mean duration ± 12.2 %). When calculating the difference in model response between the full ensemble and subensemble, the total maximum noise estimate N_{diff} is, therefore, the sum of the absolute values of the noise estimates of the full- and subensemble (e.g., for the number of events $N_{\text{diff}} = \pm(2.8\ \% + 33.3\ \%) = \pm 35.2\ \%$, last row in Tab. **??** and red area in Fig. **??**).~~

~~The differences in event characteristics between the full- and subensemble are generally within the N_{diff} noise estimates indicating no significant change in model response when excluding runs with $v_{max} \geq 50$ km yr^{-1} (Fig. **??**). Exceptions occur for increasing the horizontal grid resolutions to 50 km and removing the bed thermal model. The large differences for the 50 km runs are likely due to the large number of crashed runs (3 out of 7 and 1 out of 3 runs for the full- and subensemble,~~

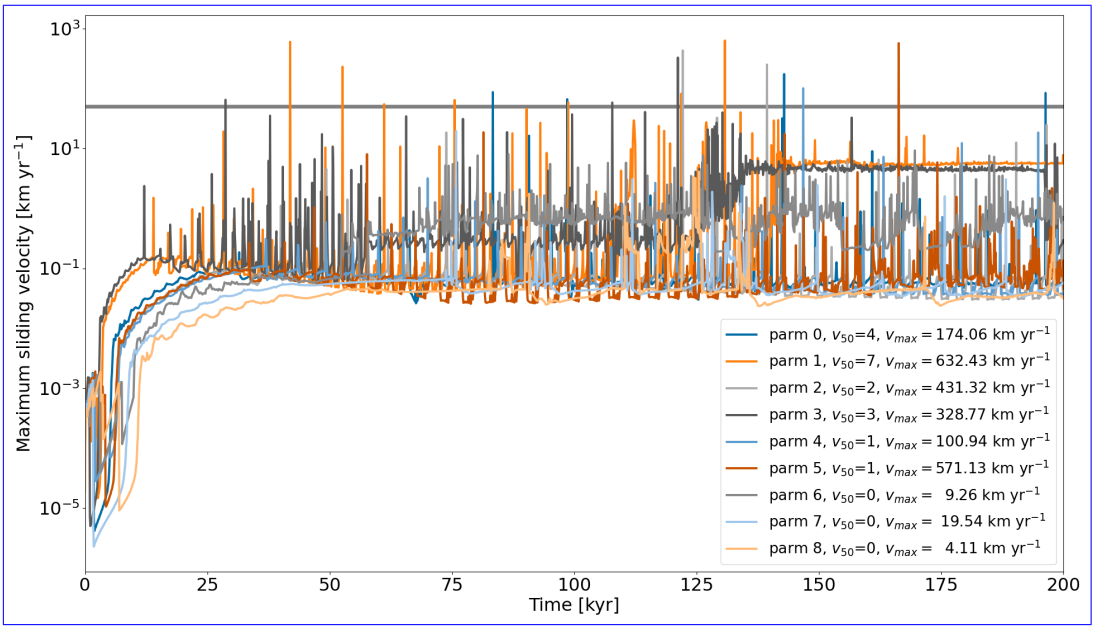


Figure S11. Maximum sliding velocity (max(max(abs($\boldsymbol{u}$)), max(abs($\boldsymbol{v}$)))) at each time step (100 yr interval) within the whole model domain for all ~~10~~ 9 parameter vectors using PISM without an upper limit for the SSA velocity. The black horizontal line marks 50 km yr^{-1} and v_{50} indicates the number of time steps exceeding this velocity. v_{max} is the highest maximum sliding velocity in a run.

~~respectively). Therefore, the only significant differences in model response between the full- and subensemble occur when the bed thermal model is removed.~~

~~Differences in the percentage change of the mean event characteristics (between comparison and base setup) for the full 10 run PISM ensemble and a subensemble including only the 3 runs with a maximum basal sliding velocity $v_{max} < 50$ km yr^{-1}. A positive difference indicates a larger change for the subensemble. The percentages in the titles of each subplot represent the differences in the event characteristics of the base runs. For example, the mean number of events in the subensemble is ~ 83 % smaller than in the full ensemble. The shaded pink regions mark the combined noise estimates described in the text and listed in the last row of Tab. **??**. The different colors resemble different model setups. The x-axis is logarithmic.~~

~~Without a bed thermal model, both ensembles show an increase in the number of events (28 vs. 163 % for the full- and subensemble, respectively). The large difference between the two ensembles is due to a low number of events across all 3 subensemble runs (maximum of 8 events per run for the subensemble (base setup) compared to 46 for the full ensemble). In fact, 6 out of 7 runs with $v_{max} \geq 50$ km yr^{-1} show a decreasing number of events when removing the bed thermal model. Therefore, the larger model response in the number of events for the subensemble resembles the differences in the base model runs (black numbers in subtitles of Fig. **??**) rather than a numerical solver instability. Even larger differences in the event characteristics of the base runs occur for the mean period and duration.~~

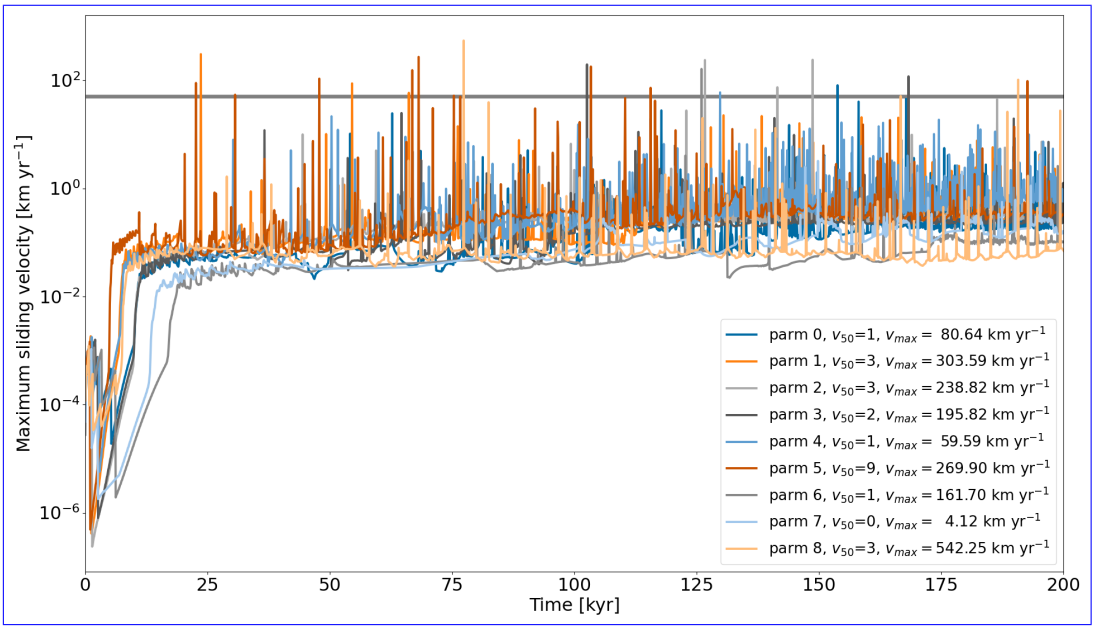


Figure S12. Maximum sliding velocity (max(max(abs($\boldsymbol{u}$)), max(abs($\boldsymbol{v}$)))) at each time step (100 yr interval) within the whole model domain for ~~10~~ 9 parameter vectors with till friction angles between 5 and 10° and values of $C_c = 0.2$, $e_0 = 0.6$, and ~~$\delta = 0.01$~~ $\delta_e = 0.01$ using PISM without an upper limit for the SSA velocity. The black horizontal line marks 50 km yr^{-1} and v_{50} indicates the number of time steps exceeding this velocity. v_{max} is the highest maximum sliding velocity in a run.

~~The difference in mean ice volume change between the base runs of the two ensembles is only −0.01 %. Both ensembles show an increase in mean ice volume change when removing the bed thermal model, but the effect is stronger when including the high-velocity runs (283 vs. 374 %).~~

~~Based on the above analysis, the model response is qualitatively similar for the full- and subensemble, but differences can occur when the mean values of the base runs are far apart, e.g., for the mean period when removing the bed thermal model (75 vs. −24 % for the full- and subensemble, respectively). However, this represents a physically different response rather than a numerical instability. Since the maximum sliding velocity of 50 km yr^{-1} is only exceeded for maximum 7 time steps per run (Fig. S11), all runs are included in the analysis.~~

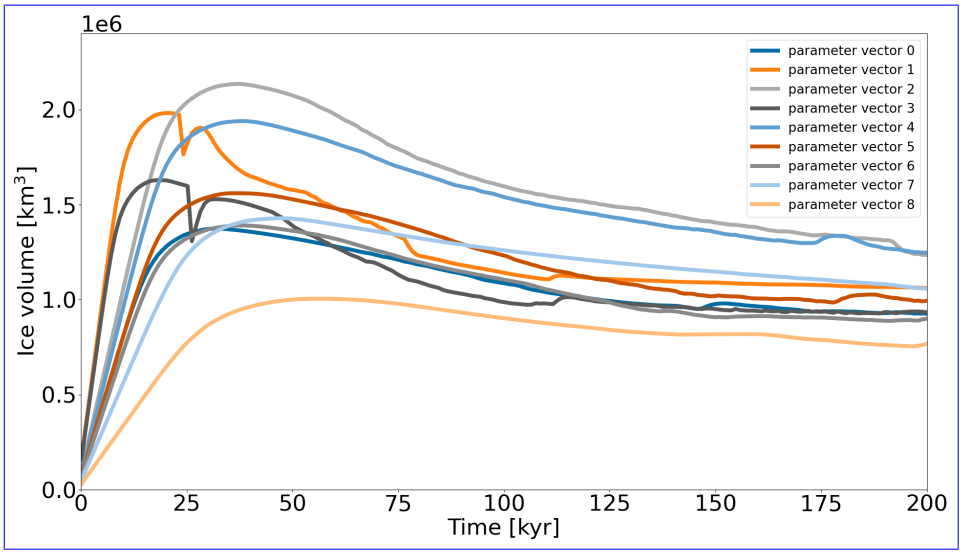


Figure S13. Ice volume in the eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait for all ~~10~~ 9 PISM parameter vectors when using the ~~base~~ reference setup but a soft bed till friction angle of 1°.~~Note that the run for parameter vector 9 crashed (sudden drop to 0 km^3 ice volume).~~

S3 ~~GSM - temporal resolution of output time series~~

~~Pseudo-Hudson Strait ice volume of a GSM model run with different output time steps. The horizontal grid resolution is 3.125 km.~~

S3 ~~PISM - temporal resolution of output time series~~

~~Ice volume in the surge-affected area (eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait) of a PISM model run with different output time steps. The horizontal grid resolution is 25 km.~~

S3 ~~PISM - Ice volume - pseudo-Hudson Strait vs. surge-affected area~~

S2.1 Ice volume - pseudo-Hudson Strait vs. surge-affected area

During a surge, ice from the pseudo-Hudson Bay and areas surrounding the pseudo-Hudson Strait is rapidly transported into the mostly ice-free pseudo-Hudson Strait. Consequently, the ice sheet extends further to the East (increasingly stronger melting), covering almost the entire pseudo-Hudson Strait area. Due to the complex interaction between ice transport and melting area, times of minimum ice volume over the area most affected by the pseudo-Hudson Strait surge (eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait, e.g., Fig. 5 and video 06 of Hank (2023)) correspond to maxima in the pseudo-Hudson Strait ice volume for most surges (grey lines in Fig. S14). However, some ice volume minima do not align with a maximum of the pseudo-Hudson Strait ice volume (red lines in Fig. S14). This inconsistency hampers the detection of surges when using the pseudo-Hudson Strait ice volume and can lead to flawed statistics. To avoid this issue, we use the ice volume in the surge-affected area, for which surges appear as minima, for all PISM results. A comparison between PISM results based on the pseudo-Hudson Strait and the surge-affected area is shown in Fig. S15. Note that we only consider the eastern half of the pseudo-Hudson Bay because some runs also show surges on the Western side of the ice sheet (e.g., 50 km run in video 09 of Hank (2023)).

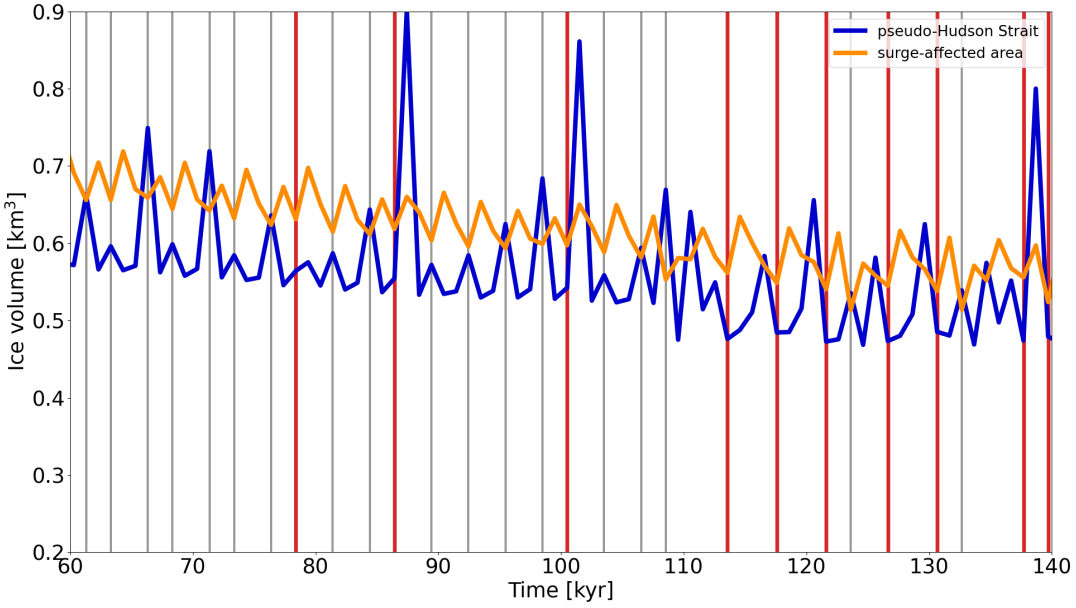


Figure S14. Normalized pseudo-Hudson Strait and surge-affected area (eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait) ice volume for parameter vector 1 using PISM. For most surges~~events~~, the minimum ice volume over the surge-affected area aligns with a maxima in the pseudo-Hudson Strait ice volume (grey lines). This is, however, not true for all ~~surge events~~ surges (thick red lines) and can lead to flawed statistics. See also video 06 of Hank (2023).

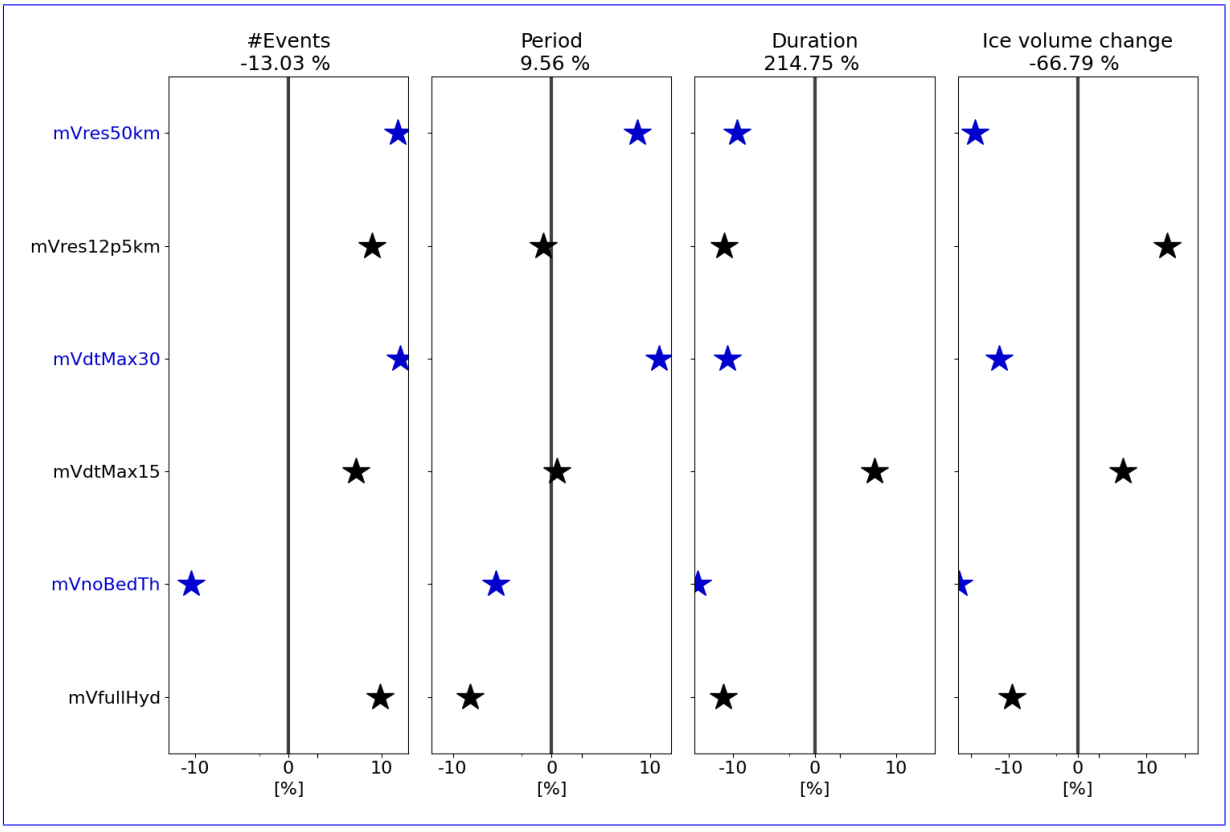


Figure S15. Differences in the percentage change of the mean ~~event~~ surge characteristics (between comparison and ~~base~~ reference setup) when using the ice volume of the surge-affected area (eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait) compared to only the pseudo-Hudson Strait ice volume. A positive difference indicates a larger change for the analysis based on the pseudo-Hudson Strait ice volume. The different colors were added for visual alignment of the individual model setups, the stars are the ensemble mean differences, and the horizontal bars represent the ensemble standard deviations. The percentages in the titles of each subplot represent the differences in the ~~event~~ surge characteristics of the ~~base~~ reference runs. For example, the mean number of ~~events~~ surges based on the pseudo-Hudson Strait ice volume is ~~$\sim 2.0\ \%$~~ $\sim 13\ \%$ smaller than for the ice volume of the surge-affected area. The different colors resemble different model setups. Note that the surge threshold is ~~$40^4\ \mathrm{km}^3$~~ $4 \cdot 10^4\ \mathrm{km}^3$ when using the surge-affected area ice volume and $0.5 \cdot 10^4\ \mathrm{km}^3$ for the pseudo-Hudson Strait ice volume ($\sim 5\ \%$ of mean ice volume across all runs). The x-axis is logarithmic.

S3 ~~Event characteristics~~Temporal resolution of output time series

S3.1 GSM

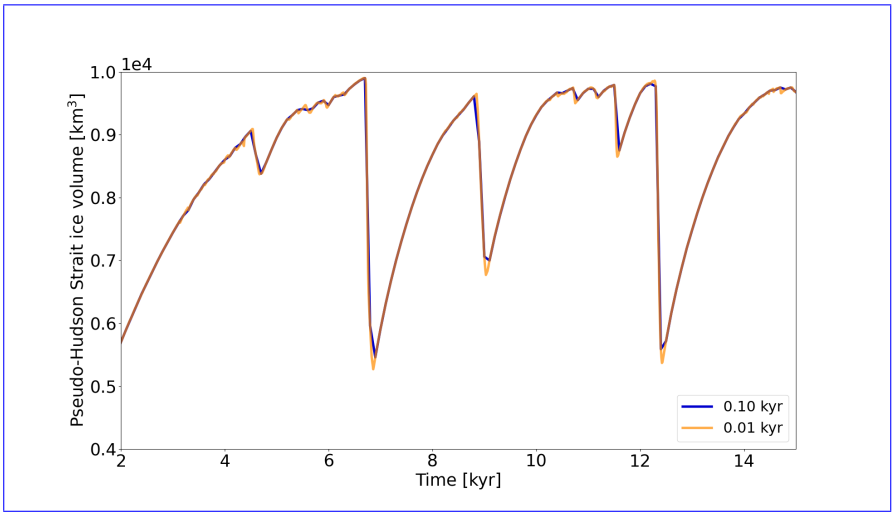


Figure S16. Pseudo-Hudson Strait ice volume of a GSM model run with ~~visual illustration of the event characteristics used to compare~~ different ~~model setups~~output time steps. The horizontal grid resolution is 3.125 km.

S3.2 PISM

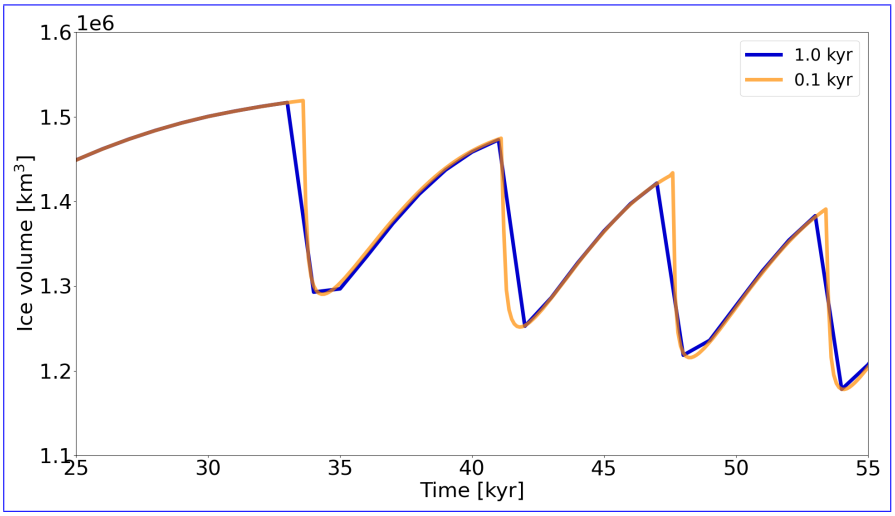


Figure S17. Ice volume in the surge-affected area (eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait) of a PISM model run (parameter vector 5) with different output time steps. The horizontal grid resolution is 25 km.

S4 RMSE and mean bias

The RMSE and mean bias values presented throughout the paper are calculated according to the following equations

$$rmse = \sqrt{\frac{\sum_{t=0}^{t_{max}} (x_t - b_t)^2}{N}} \cdot \frac{100}{b_m}, \text{ and} \tag{S3}$$

$$mean\ bias = \frac{\sum_{t=0}^{t_{max}} (x_t - b_t)}{N} \cdot \frac{100}{b_m}, \tag{S4}$$

~~,~~where x_t and b_t are the (pseudo-Hudson Strait) ice volume values at time t of the comparison setup and ~~base~~ reference setup, respectively. t_{max} is the maximum time, N the number of time steps, and b_m the mean of the ~~base~~ reference setup time series. These values are then averaged over all 5 parameter vectors. Crashed runs are excluded from the averaging process.

S5 Comparison between different model setups

The analysis to compare the different model setups follows

1. run 1 parameter vector with the ~~base setup (Tab.~~reference setup (Table 2)
2. calculate the ~~event~~ surge characteristics for this ~~base~~ reference run
3. re-run the same parameter vector for one of the comparison setups (Sec. 2.1.3 and 2.2.4)
4. calculate the ~~event~~ surge characteristics for the comparison run
5. calculate the differences in ~~event~~ surge characteristics between the ~~base~~ reference run and comparison run expressed as percentage ~~deviations from the base~~ differences from the reference run (positive for increase compared to the value of the ~~base~~ reference run)
6. repeat steps 1) to 5) for all parameter vectors (5 for the GSM, ~~10~~ 9 for PISM)
7. take the average of all percentage ~~deviations for each event~~ differences for each surge characteristic

Percentage ~~deviations~~ differences for crashed comparison runs are not considered for the final average and runs with less than 2 ~~events~~ surges require special treatment. In these cases, the period is set to a NaN value, leading to a NaN difference between that particular run and the corresponding ~~base~~ reference run. We use Numpys *numpy.nanmean()* and *numpy.nanstd()* to ignore these NaN values when averaging over all parameter vectors. Similarly, all ~~event~~ surge characteristics except for the number of ~~events~~ surges are set to NaN values for runs with no ~~events~~ surges at all.

S6 ~~GSM - 12.5 km noise estimate~~Minimum numerical error estimates

S6.1 GSM

S6.1.1 Minimum numerical error estimates at 12.5 km

Metric	original 12.5 km runs	stricter numerical convergence [% difference]	stricter numerical convergence with increased maximum iterations [% difference]
#~~Events~~ Surges	81 ± 42	2.3 ± 8.5	$\mathbf{2.8 \pm 9.2}$
mean period	2.3 ± 0.8 kyr	~~-2.1 ± 7.5~~ $\mathbf{-2.1 \pm 7.5}$	-1.5 ± 9.3
mean duration	0.6 ± 0.2 kyr	-1.4 ± 9.4	$\mathbf{2.6 \pm 14.2}$
mean pseudo-Hudson Strait ice volume change	$2.2 \pm 1.1 \cdot 10^3$ km^3	~~20.9 ± 53.0~~ $\mathbf{20.9 \pm 53.0}$	-5.1 ± 12.2

Table S2. Percentage differences of ~~event~~ surge characteristics between GSM runs with regular and stricter numerical convergence and increased maximum iterations for the ice dynamics loops at 12.5 km (except first column). The values represent the average of 5 parameter vectors. No runs crashed and all runs had more than 1 surge~~event~~. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above. The bold numbers mark the largest MNEE for each surge characteristic.

S6.1.2 Adding surface temperature noise

Metric	~~base~~ reference setup	± 0.1°C noise	± 0.5°C noise ~~implicit coupling~~
#~~Events~~ Surges	180 ± 100	-4.0 ± 4.3	-4.1 ± 7.0 ~~1.1 ± 4.9~~
mean period	1.1 ± 0.5 kyr	4.8 ± 5.3	3.8 ± 6.8 ~~-0.3 ± 5.3~~
mean duration	0.3 ± 0.1 kyr	1.3 ± 4.4	0.9 ± 4.3 ~~-12.7 ± 9.5~~
mean pseudo-Hudson Strait ice volume change	$1.7 \pm 0.2 \cdot 10^3$ km^3	0.9 ± 4.1	2.1 ± 5.5 ~~-25.1 ± 18.7~~
RMSE	-	8.0 ± 2.5	7.8 ± 2.1 ~~7.3 ± 2.5~~
Mean Bias	-	-0.1 ± 0.2	0.1 ± 0.0

Table S3. Percentage differences of surge characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM reference setup for two different amplitudes of surface temperature noise (except first column). No runs crashed and all runs had more than 1 surge. The first 20 kyr of each run are treated as a spin-up interval for the surge characteristics (not the RMSE and mean bias).

S6.1.3 Implicit thermodynamics/ice dynamics coupling

Metric	reference setup	implicit coupling
#Surges	180 ± 100	1.1 ± 4.9
mean period	1.1 ± 0.5 kyr	-0.3 ± 5.3
mean duration	0.3 ± 0.1 kyr	-12.7 ± 9.5
mean pseudo-Hudson Strait ice volume change	$1.7 \pm 0.2 \cdot 10^3$ km^3	-25.1 ± 18.7
RMSE	-	7.3 ± 2.5
Mean Bias	-	1.8 ± 1.5

Table S4. Percentage differences of ~~event~~ surge characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM ~~base~~ reference setup for ~~two different amplitudes of surface temperature noise and~~ implicit coupling between the thermodynamics and ice dynamics in the GSM (except first column). No runs crashed and all runs had more than 1 surge~~event~~. The first 20 kyr of each run are treated as a spin-up interval for the ~~event~~ surge characteristics (not the RMSE and mean bias).

S7 ~~PISM - Noise estimate~~

190 ### S6.1 PISM

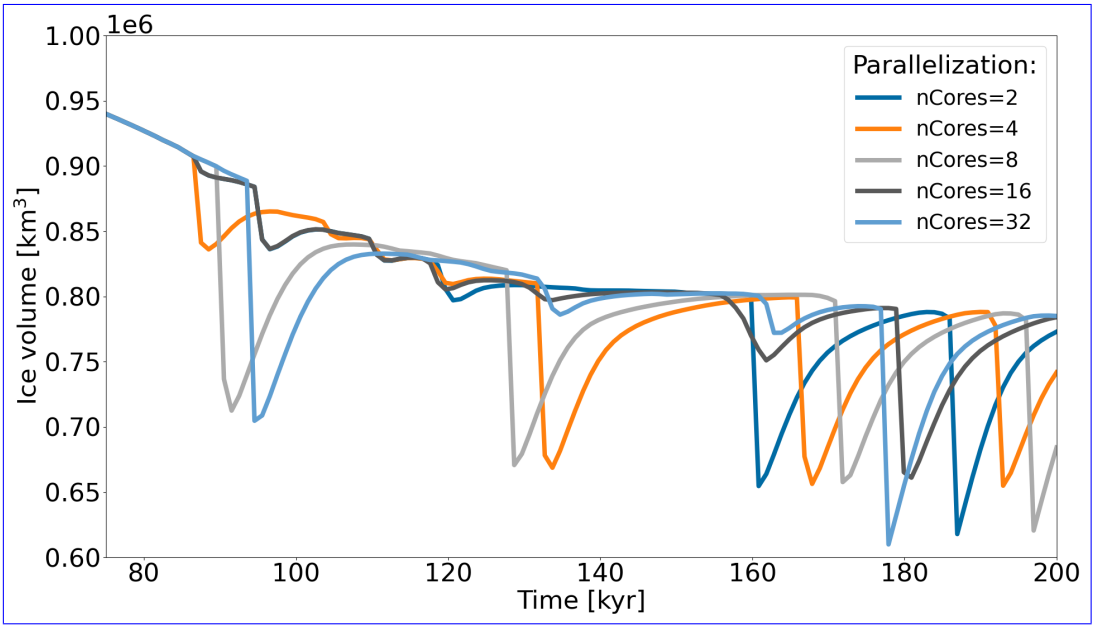


Figure S18. Ice volume in the eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait for parameter vector ~~1~~ 8 and different ~~number~~ numbers of cores/processes using the ~~Parallel Ice Sheet Model (~~PISM~~)~~.

S7 ~~GSM - Smooth sediment transition zone and non-flat topography~~

S6.0.1 Relative tolerances

Setup	number of surges	mean period	mean duration	mean ice volume change	nC	nS1
25 km reference setup	35 ± 25	10 ± 10 kyr	3 ± 2 kyr	$1.1 \pm 0.3 \cdot 10^5$ km^3	0	0
nCores= 8, KSP= 10^{-10}	-10.7 ± 21.6	-1.8 ± 5.6	11.2 ± 32.8	5.3 ± 11.2	0	1
nCores= 16, KSP= 10^{-10}	-4.7 ± 22.4	-2.7 ± 4.8	0.1 ± 9.3	-0.4 ± 5.1	0	1
nCores= 32, KSP= 10^{-10}	-10.5 ± 28.7	-3.7 ± 6.2	-1.2 ± 12.4	1.2 ± 7.2	4	1
nCores= 8, KSP= 10^{-10}, PIC= 10^{-7}	8.3 ± 11.8	-28.6 ± 17.4	-17.1 ± 35.2	9.0 ± 21.5	6	0
nCores= 16, KSP= 10^{-10}, PIC= 10^{-7}	98.1 ± 135.9	-22.5 ± 10.1	-14.4 ± 24.6	15.7 ± 54.4	5	0
nCores= 32, KSP= 10^{-10}, PIC= 10^{-7}	83.8 ± 125.1	-18.2 ± 11.6	-13.9 ± 25.7	23.2 ± 56.4	5	0

Table S5. Percentage differences of surge characteristics compared to the PISM reference setup with different numbers of cores and adjusted relative tolerances for the Picard iteration in the calculation of the vertically-averaged effective viscosity (PIC, default is 10^{-4}) and the Krylov linear solver used at each Picard iteration (KSP, default is 10^{-7} (except first row). The values represent the average of 9 parameter vectors. Crashed runs (nC) are not considered and runs with just one surge (nS1) are ignored when calculating the change in mean period. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above. Note that more than 50 % of all runs with KSP= 10^{-10} and PIC= 10^{-7} did not finish within the time limit set by the computational cluster and are considered as crashed runs (nC). A direct comparison of runs with these tolerances can be found in Fig. S19. Note that all test runs without preconditioning (removes processor-number-dependence of results) crashed during the spin-up phase and long before the first surge occurs.

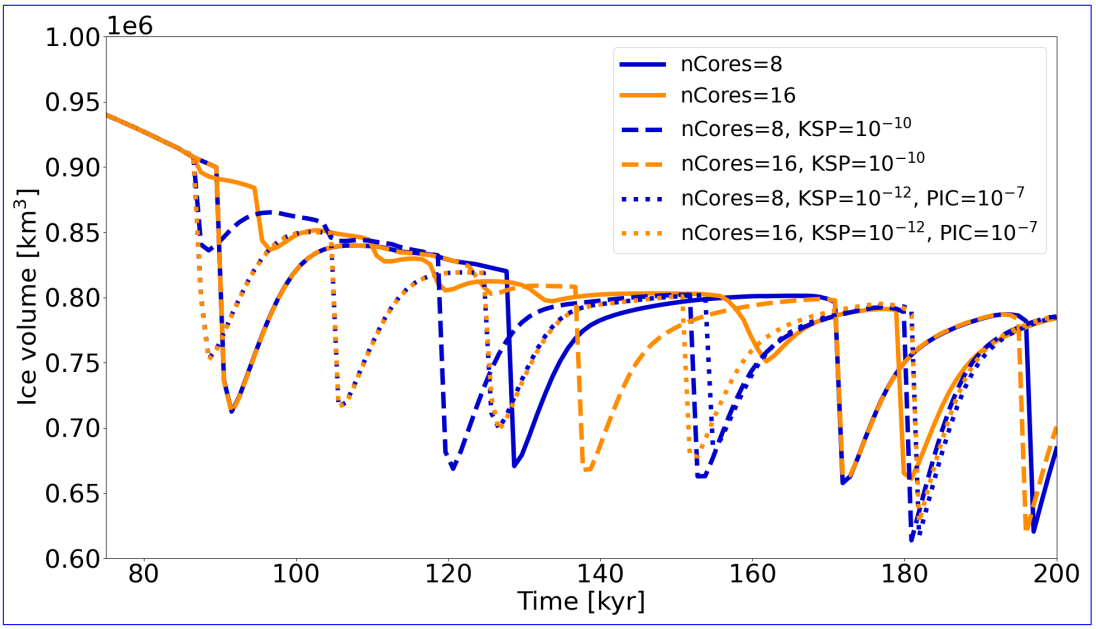


Figure S19. ~~Pseudo-Hudson Strait ice~~ Ice volume ~~for parameter vector 0 with and without a 200 m deep topography~~ in the eastern half of the pseudo-Hudson Bay and ~~Hudson~~ the pseudo-Hudson Strait ~~region~~ for parameter vector 8 and different number of cores/processes using the ~~GSM. In runs~~ PISM with ~~a non-flat topography,~~ different relative tolerances for the ~~initial glaciation is delayed because~~ Picard iteration in the ~~pseudo-Hudson Strait topography~~ calculation of the vertically-averaged effective viscosity (PIC, default is ~~below sea level. The horizontal grid resolution~~ 10^{-4}) and the Krylov linear solver used at each Picard iteration (KSP, default is ~~3.125 km~~10^{-7}.

~~Comparing the runs for parameter vector 1 more closely shows that when using a non-flat topography, the surges now start and propagate at the southernmost and northernmost end of the~~

S6.0.2 Adding surface temperature noise

Metric	reference setup	±0.1°C noise	±0.5°C noise
#Surges	35 ± 25	-12.4 ± 24.2	-12.0 ± 25.2
mean period	10 ± 10 kyr	-5.6 ± 8.0	-4.0 ± 8.8
mean duration	3 ± 2 kyr	11.5 ± 37.9	2.0 ± 12.9
mean pseudo-Hudson Strait ~~, where the topography is deepest and begins to slope upwards. Additionally, the surges tend to propagate faster and extend further to the West and in North-South direction than without topography (e.g., 8.0 to 8.3 kyr in the bottom row of video 07 of Hank (2023)) . This is mainly due to warmer basal conditions in the transition zone and Hudson Bay region before the start of the surge (200 m bed depression increases the heat generation at the bed (video 08 of Hank (2023)) which, in turn, increases the average basal temperature with respect to the pressure melting point). An interesting effect of the 200 m deep Hudson Strait and 500 m deep ocean is that the pressure melting point is first reached further inland and not at the eastern end of the~~ ice volume change	$1.1 \pm 0.3 \cdot 10^5$ km^3	1.9 ± 15.9	2.5 ± 8.1
RMSE	-	4.1 ± 3.5	4.3 ± 2.6
Mean Bias	-	-0.8 ± 1.9	0.1 ± 0.3
nS1	0	1	1

Table S6. Percentage differences of surge characteristics, ice volume RMSE and mean bias compared to the PISM reference setup for two different amplitudes of surface temperature noise (except first column). No runs crashed and all runs showed at least 1 surge. Runs with just one surge (nS1) are ignored when caluclating the change in mean period. The first 20 kyr of each run are treated as a spin-up interval for the surge characteristics (not the RMSE and mean bias).

S7 Sensitivity experiments with a significant effect

S7.1 Bed thermal model

Metric	
number of surges	
mean period	
mean duration	
mean pseudo-Hudson Strait ~~, as is the case for a flat topography (e.g., 7.8 to 8.1 kyr in the bottom row of video 07 of Hank (2023)).~~ ice volume change	
RMSE	
Mean Bias	

Table S7. Percentage differences (except first column) of surge characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM reference setup for runs with only one bed thermal layer (20 m deep). No runs crashed and all runs had more than 1 surge. The first 20 kyr of each run are treated as a spin-up interval for the surge characteristics (not the RMSE and mean bias).

S8 ~~GSM - Bed thermal model~~

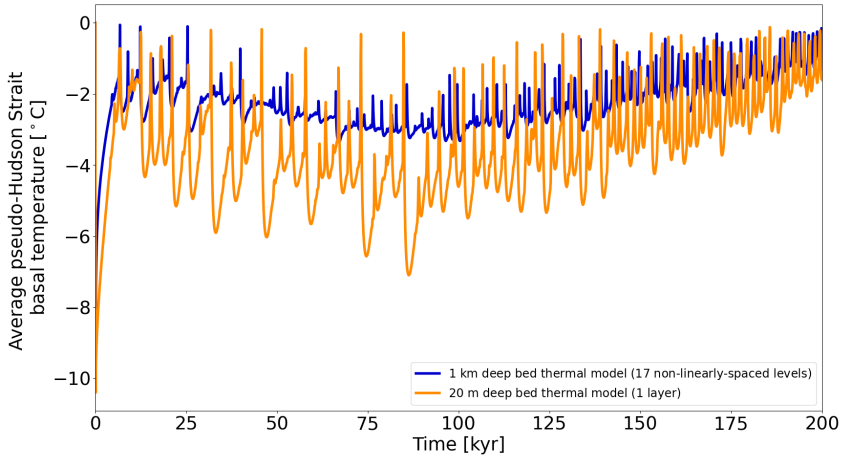


Figure S20. Average pseudo-Hudson Strait basal ice temperature with respect to the pressure melting point for parameter vector 1 with a 20 m and 1 km deep bed thermal model (17 non-linearly-spaced levels) using the GSM. The horizontal grid resolution is 3.125 km.

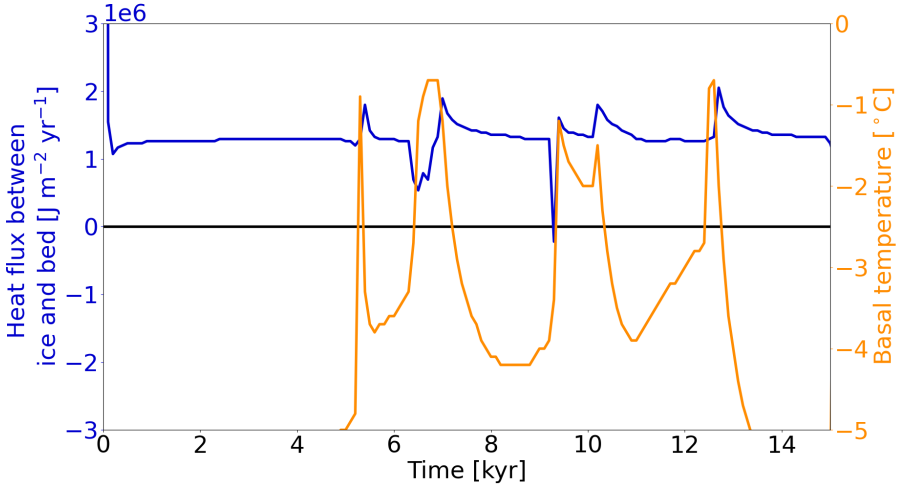


Figure S21. Heat flux at the base of the ice sheet (positive from bed into ice) and basal ice temperature for a grid cell in the center of the pseudo-Hudson Strait (grid cell center at $x = 376.5625$ km and $y = 248.4375$ km, white star in Fig. 1) and parameter vector 1 with only one bed thermal layer (20 m deep) using the GSM. The horizontal grid resolution is 3.125 km.

S8 ~~PISM - Bed thermal model~~

Metric	reference setup	no bed thermal model
number of surges	35 ± 25	20.7 ± 140.5
mean period	10 ± 10 kyr	79.9 ± 80.0
mean duration	3 ± 2 kyr	69.8 ± 60.4
mean ice volume change	$1.1 \pm 0.3 \cdot 10^5$ km^3	395.8 ± 240.5
RMSE	-	36.0 ± 5.3
Mean Bias	-	-27.1 ± 5.6

Table S8. Percentage differences of surge characteristics, ice volume RMSE and mean bias compared to the PISM reference setup for runs without a bed thermal model (except first column). No runs crashed and all runs had more than 1 surge. The first 20 kyr of each run are treated as a spin-up interval for the surge characteristics (not the RMSE and mean bias).

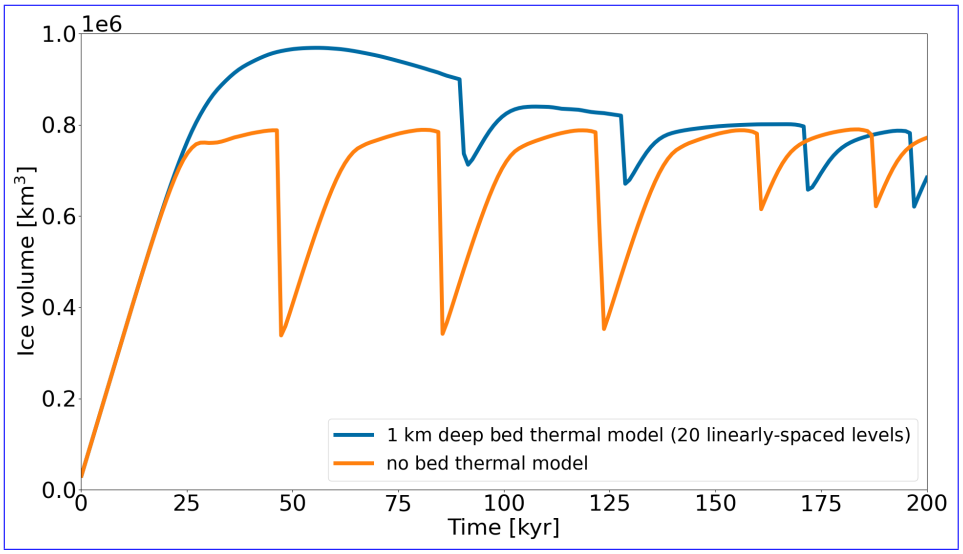


Figure S22. Ice volume in the eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait for parameter vector 8 with and without the 1 km deep (20 linearly-spaced levels) bed thermal model using the ~~Parallel Ice Sheet Model (~~PISM~~)~~. The horizontal grid resolution is 25 km.

~~S8 GSM - Weight of adjacent minimum basal temperature~~

S7.1 Basal temperature at the grid cell interface

~~S8 GSM - Basal temperature ramps at different resolutions~~

S7.1 Basal temperature ramps at different resolutions

To simplify the comparison of different temperature ramps, we calculate a single value score based on all ~~event~~ surge characteristics. The calculation steps are as follows.

1. calculate the absolute values for all ~~event~~ surge characteristic means
2. calculate the average across all ramps for all characteristics (means and standard deviations separately, total of 4 means and 4 standard deviations)
3. for each ramp, divide all ~~event~~ surge characteristics by their corresponding average
4. sum the values for all ~~event~~ surge characteristics (separately for mean and std)

Metric	~~base setup~~ reference setup (TpmTrans)	$W_{Tb,\min} = 0.0$ TpmInt	$W_{Tb,\min} = 1.0$ TpmInt, upw
~~#Events~~ nC	0	0	0
nS0	0	4	1
nS1	0	0	1
number of surges	180 ± 100	~~-9.6 ± 6.5~~ -96.9 ± 6.3	~~-3.7 ± 7.8~~ -90.2 ± 15.4
mean period	1.1 ± 0.5 kyr	~~14.7 ± 13.5~~ 106.2 ± 0.0	~~3.0 ± 0.8~~ 1645.4 ± 2136.8
mean duration	0.3 ± 0.1 kyr	~~5.1 ± 4.9~~ -15.9 ± 0.0	~~-2.6 ± 3.3~~ 11.1 ± 17.4
mean pseudo-Hudson Strait ice volume change	$1.7 \pm 0.2 \cdot 10^3$ km^3	~~-1.9 ± 4.0~~ -66.2 ± 0.0	~~4.0 ± 6.7~~ -60.4 ± 6.5
RMSE	-	~~7.8 ± 2.5~~ 7.4 ± 2.4	~~8.0 ± 2.5~~ 9.4 ± 2.6
Mean Bias	-	~~-0.1 ± 0.1~~ 4.0 ± 1.6	~~0.3 ± 0.1~~ 6.7 ± 2.4

Table S9. Percentage differences (except first column) of ~~event~~ surge characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM ~~base~~ reference setup ~~($W_{Tb,\min} = 0.5$)~~ for different ~~weights of~~ approaches to calculate the ~~adjacent minimum~~ basal temperature ~~for~~ at the ~~basal sliding temperature ramp~~ grid cell interface (~~except first column~~Sec. 3.3.2). ~~No~~ Crashed runs ~~crashed~~ (nC) are not considered and ~~all~~ runs ~~had more than~~ without surges (nS0) only contribute to the change in surge number. Runs with only 1 surge ~~event~~(nS1) are excluded from the calculation of the mean period. The first 20 kyr of each run are treated as a spin-up interval for the ~~event~~ surge characteristics (not the RMSE and mean bias).

The above calculation combines the 4 ~~event~~ surge characteristics to a single value for the mean and standard deviation of each ramp. We keep separate values for the mean and standard deviation since the two metrics contain different information. Smaller values indicate a better agreement with the 3.125 km ~~base~~ reference setup.

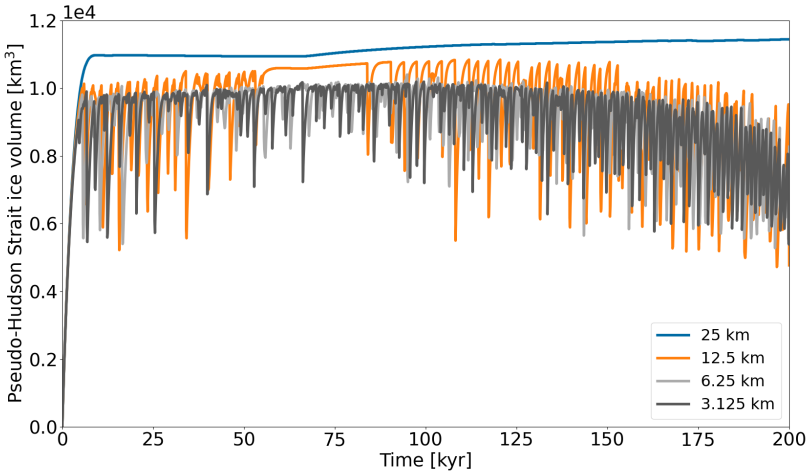


Figure S23. Pseudo-Hudson Strait ice volume for parameter vector 1 and different horizontal grid resolutions using the GSM. A constant temperature ramp with $T_{ramp} = 0.0625°$C and $T_{exp} = 28$ is used for all horizontal grid resolutions (magenta line in Fig. 2).

Metric	#~~Events~~ Surges	mean period	mean duration	mean pseudo-Hudson Strait ice volume change	RMSE	Mean Bias
$T_{exp} = 5, T_{ramp} = 1$	-63.5 ± 17.1	121.7 ± 29.5	300.0 ± 116.6	95.5 ± 39.5	21.8 ± 4.8	-17.8 ± 5.6
$T_{exp} = 15, T_{ramp} = 1$	-39.0 ± 10.2	64.6 ± 22.8	179.2 ± 117.3	51.5 ± 35.3	17.4 ± 3.7	-11.2 ± 4.6
$T_{exp} = 28, T_{ramp} = 0.5$	-17.1 ± 7.1	28.6 ± 21.0	64.0 ± 54.1	18.7 ± 12.4	10.0 ± 3.3	-3.5 ± 3.2
$T_{exp} = 5, T_{ramp} = 0.0625$	-9.5 ± 5.1	16.5 ± 12.4	14.9 ± 12.0	3.9 ± 2.9	8.1 ± 2.4	-0.8 ± 0.6
$T_{exp} = 10, T_{ramp} = 0.0625$	-9.3 ± 5.0	10.1 ± 5.0	8.8 ± 7.0	3.4 ± 4.0	8.0 ± 2.4	-0.4 ± 0.3
$T_{exp} = 28, T_{ramp} = 0.125$	-4.6 ± 6.4	3.2 ± 4.2	4.4 ± 5.4	0.3 ± 2.1	7.9 ± 2.2	-0.3 ± 0.2
$T_{exp} = 14, T_{ramp} = 0.0625$	-7.1 ± 5.1	9.3 ± 7.4	7.3 ± 7.5	2.8 ± 3.4	7.8 ± 1.9	-0.2 ± 0.1
$T_{exp} = 15, T_{ramp} = 0.0625$	-4.9 ± 4.7	8.4 ± 10.5	4.8 ± 4.4	0.3 ± 6.3	7.8 ± 2.0	-0.2 ± 0.1
$T_{exp} = 20, T_{ramp} = 0.0625$	-3.0 ± 4.7	2.0 ± 3.9	-0.1 ± 2.9	1.3 ± 4.2	7.9 ± 2.4	-0.1 ± 0.1
$T_{exp} = 25, T_{ramp} = 0.0625$	-1.2 ± 3.5	4.1 ± 7.7	0.5 ± 1.1	-1.5 ± 3.0	7.8 ± 2.4	-0.0 ± 0.1
3.125 km reference setup	$\mathbf{180 \pm 100}$	$\mathbf{1.1 \pm 0.5}$ **kyr**	$\mathbf{0.3 \pm 0.1}$ **kyr**	$\mathbf{1.7 \pm 0.2 \cdot 10^3}$ **km^3**	-	-
$T_{exp} = 30, T_{ramp} = 0.0625$	-2.4 ± 3.6	2.4 ± 3.9	-0.1 ± 2.9	-0.4 ± 2.4	7.9 ± 2.2	0.0 ± 0.1
$T_{exp} = 35, T_{ramp} = 0.0625$	-2.6 ± 4.7	2.6 ± 4.8	0.5 ± 4.3	-0.6 ± 4.1	7.9 ± 2.3	0.1 ± 0.2
$T_{exp} = 45, T_{ramp} = 0.0625$	-1.3 ± 4.8	1.8 ± 4.4	-0.1 ± 1.6	-1.6 ± 4.1	7.8 ± 2.2	0.1 ± 0.1
$T_{exp} = 56, T_{ramp} = 0.0625$	-1.7 ± 4.7	1.3 ± 4.7	-3.2 ± 2.1	-0.4 ± 5.2	7.7 ± 2.1	0.2 ± 0.0
$T_{exp} = 28, T_{ramp} = 0.03125$	-0.8 ± 4.9	3.2 ± 8.0	-2.3 ± 3.1	-0.2 ± 3.8	7.8 ± 2.3	0.2 ± 0.1

Table S10. Percentage differences of ~~event~~ surge characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM ~~base~~ reference setup ($T_{ramp} = 0.0625$, $T_{exp} = 28$) for different basal temperature ramps (except for ~~base~~ reference setup). The ramps are sorted from widest (first row) to sharpest (last row, see Fig. S24). The bold ~~base~~ reference values in the middle of the table ~~seperate~~ separate the ramps that are wider (above) and sharper (below) than the ~~base~~ reference setup. No runs crashed and all runs had more than 1 surge~~event~~. The first 20 kyr of each run are treated as a spin-up interval for the ~~event~~ surge characteristics (not the RMSE and mean bias).

ramp	score-mean	score-std	sum of scores
res= 25 km, $T_{exp} = 5$, $T_{ramp} = 0.5$	**0.85**	**2.92**	**3.77**
res= 25 km, $T_{exp} = 15$, $T_{ramp} = 1$	1.08	3.53	4.61
res= 25 km, $T_{exp} = 10$, $T_{ramp} = 0.5$	1.58	3.88	5.46
res= 12.5 km, $T_{exp} = 10$, $T_{ramp} = 0.25$	3.87	4.11	7.98
res= 12.5 km, $T_{exp} = 20$, $T_{ramp} = 0.25$	3.04	2.55	**5.59**
res= 12.5 km, $T_{exp} = 25$, $T_{ramp} = 0.25$	3.43	3.31	6.74
res= 12.5 km, $T_{exp} = 28$, $T_{ramp} = 0.25$	**2.93**	2.72	5.65
res= 12.5 km, $T_{exp} = 30$, $T_{ramp} = 0.25$	3.54	**2.45**	5.99
res= 12.5 km, $T_{exp} = 35$, $T_{ramp} = 0.25$	3.30	2.98	6.28
res= 12.5 km, $T_{exp} = 45$, $T_{ramp} = 0.25$	3.36	3.17	6.54
res= 12.5 km, $T_{exp} = 28$, $T_{ramp} = 0.0625$	3.40	2.87	6.27
res= 6.25 km, $T_{exp} = 10$, $T_{ramp} = 0.125$	2.73	1.69	4.42
res= 6.25 km, $T_{exp} = 15$, $T_{ramp} = 0.125$	2.13	**1.17**	**3.30**
res= 6.25 km, $T_{exp} = 20$, $T_{ramp} = 0.125$	1.92	2.98	4.90
res= 6.25 km, $T_{exp} = 25$, $T_{ramp} = 0.125$	2.09	2.35	4.44
res= 6.25 km, $T_{exp} = 28$, $T_{ramp} = 0.125$	2.05	2.10	4.15
res= 6.25 km, $T_{exp} = 30$, $T_{ramp} = 0.125$	1.95	1.63	3.58
res= 6.25 km, $T_{exp} = 35$, $T_{ramp} = 0.125$	1.94	1.66	3.60
res= 6.25 km, $T_{exp} = 45$, $T_{ramp} = 0.125$	**1.70**	2.67	4.37
res= 6.25 km, $T_{exp} = 28$, $T_{ramp} = 0.0625$	1.80	2.56	4.36

Table S11. Single value scores for the mean and standard deviation of the basal temperature ramps. The temperature ramps are shown in Fig. S25. A total of 12, 13, and 13 ramps were tested at 25 km, 12.5 km, and 6.25 km horizontal grid resolution, respectively. Note that ramps whose sum (score-mean + score-std) differ by more than 50 % from the minimum sum at the corresponding resolution are not listed here. The minimum scores for the mean, standard deviation, and sum at each resolution are marked as bold numbers. No runs crashed and all runs had more than 1 ~~event~~surge. Note that the sum of scores can be slightly off due to rounding (± 0.01).

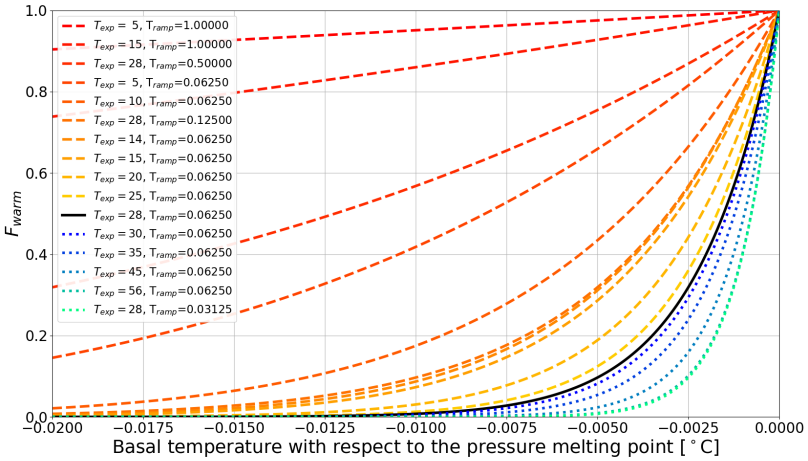


Figure S24. Temperature ramps for different values of T_{ramp} and T_{exp}. The black solid line shows the ramp used for the 3.125 km horizontal grid resolution ~~base~~ reference setup ($T_{ramp} = 0.0625$, $T_{exp} = 28$). The solid and dotted lines show ramps that are wider and sharper than the ~~base~~ reference setup, respectively. The depicted temperature ramps are the same as the ones listed in Fig. 9 and ~~Tab.~~Table S10.

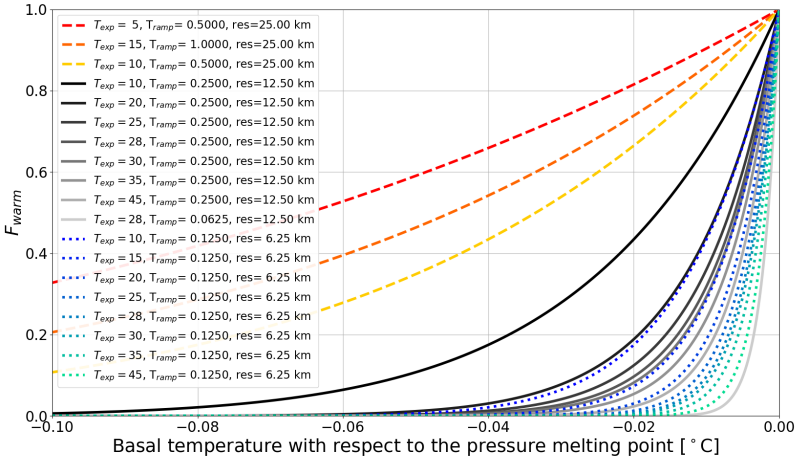


Figure S25. Shown are the temperature ramps listed in ~~Tab.~~Table S11 at 25 km (solid lines), 12.5 km (dashed lines), and 6.25 km horizontal grid resolution (dotted lines).

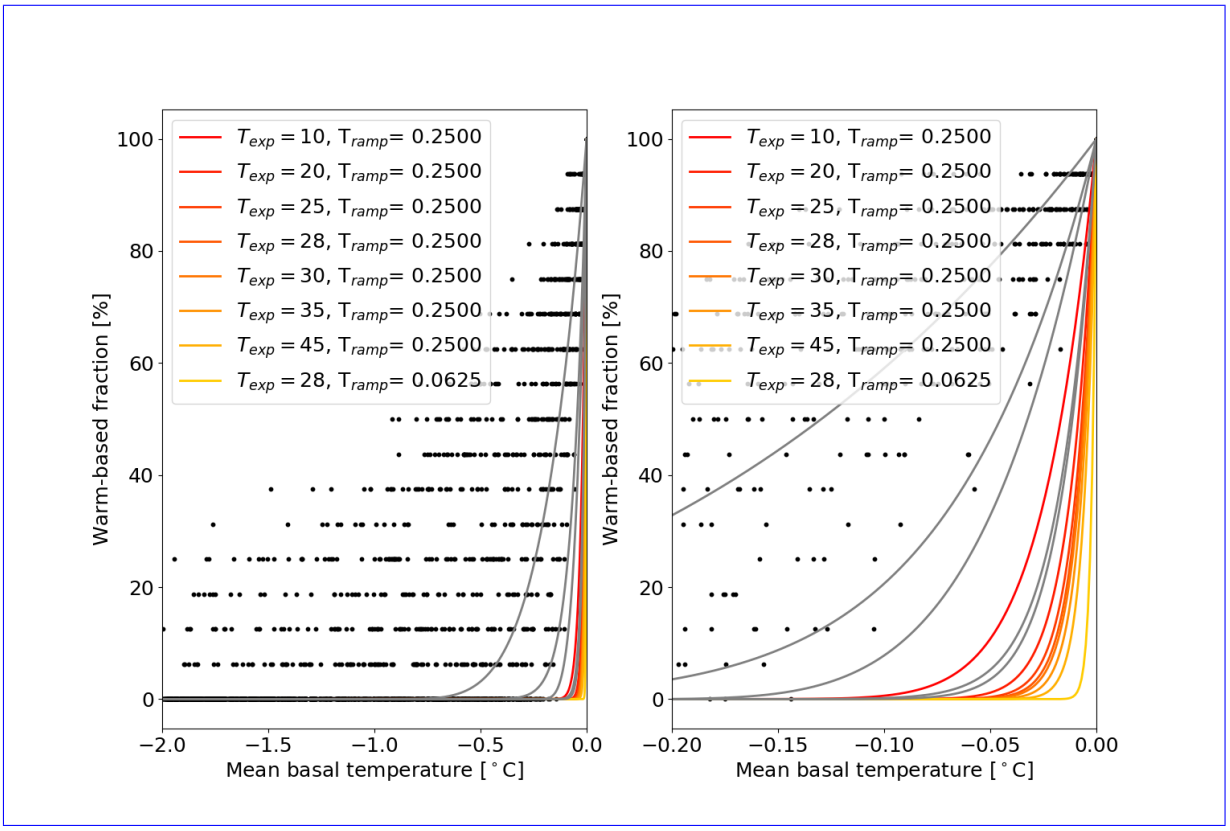


Figure S26. Warm-based fraction (basal temperature with respect to the pressure melting point at 0 °C) vs. mean basal temperature with respect to the pressure melting point when upscaling a 3.125 km run to 12.5 km horizontal grid resolution including all 5 parameter vectors using the GSM. Only grid cells within the pseudo-Hudson Strait and time steps within the surges of the 10 kyr after the first surge are considered. The restriction to the 10 kyr after the first surge for these experiments is set by storage limitations due to the high temporal resolution of the model output fields (10 yr). The colored ramps correspond to the 12.5 km horizontal grid resolution basal temperature ramps in ~~Tab.~~Table S11 and the gray lines show all other ramps that were tested at this resolution.

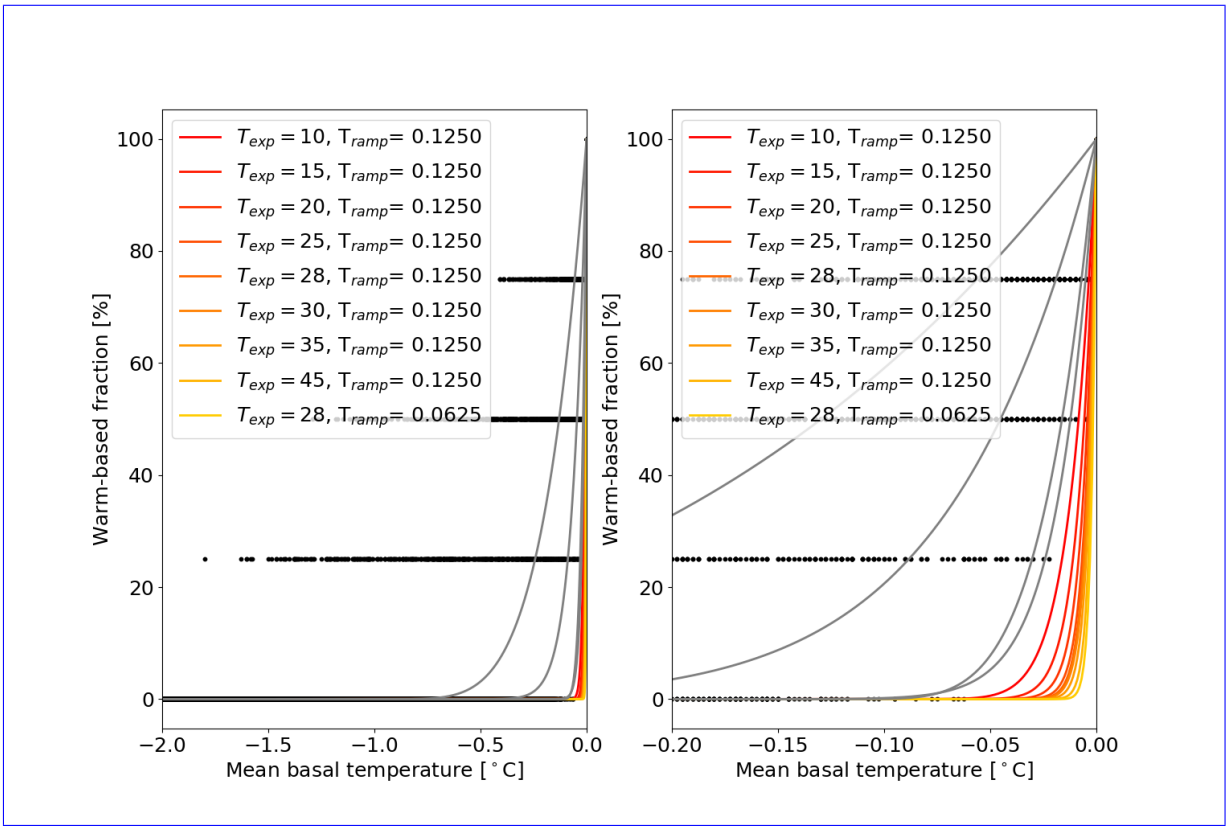


Figure S27. Warm-based fraction (basal temperature with respect to the pressure melting point at $0\ ^{\circ}\mathrm{C}$) vs. mean basal temperature with respect to the pressure melting point when upscaling a $3.125\ \mathrm{km}$ run to $6.25\ \mathrm{km}$ horizontal grid resolution including all 5 parameter vectors using the GSM. The colored ramps correspond to the $6.25\ \mathrm{km}$ horizontal grid resolution basal temperature ramps in ~~Tab.~~Table S11 and the gray lines show all other ramps that were tested at this resolution. Otherwise same as Fig. S26.

S7.2 Smooth sediment transition zone and non-flat topography

Metric	reference setup (abrupt transition) reference values	3.125 km wide transition	25 km wide transition	3.125 km wide transition with HB/HS topography	25 km wide transition with HB/HS topography
number of surges	180 ± 100	-4.2 ± 8.9	1.0 ± 11.4	36.3 ± 17.3	19.9 ± 22.6
mean period	1.1 ± 0.5 kyr	4.6 ± 9.2	-0.4 ± 10.4	2.2 ± 48.0	14.5 ± 45.1
mean duration	0.3 ± 0.1 kyr	2.7 ± 3.6	7.2 ± 4.4	10.2 ± 17.5	24.3 ± 9.0
mean pseudo-Hudson Strait ice volume change	$1.7 \pm 0.2 \cdot 10^3$ km^3	0.2 ± 4.8	-1.7 ± 4.1	8.4 ± 10.0	17.3 ± 15.6
RMSE	-	7.9 ± 2.3	8.0 ± 2.2	11.2 ± 1.8	12.2 ± 2.0
Mean Bias	-	0.0 ± 0.2	-0.6 ± 0.5	-6.2 ± 1.9	-6.6 ± 2.2

Table S12. Percentage differences of surge characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM reference setup for runs with a smooth transition between hard bedrock and soft sediment, and runs with a pseudo-Hudson Bay/Hudson Strait (HB/HS) topography (except first column). No runs crashed and all runs had more than 1 surge. The first 20 kyr of each run are treated as a spin-up interval for the surge characteristics (except for the RMSE and mean bias).

Comparing the runs for parameter vector 1 more closely shows that when using a non-flat topography, the surges now start and propagate at the southernmost and northernmost end of the pseudo-Hudson Strait, where the topography is deepest and begins to slope upwards. Additionally, the surges tend to propagate faster and extend further to the West and in North-South direction than without topography (e.g., 8.0 to 8.3 kyr in the bottom row of video 07 of Hank (2023)). This is mainly due to warmer basal conditions in the transition zone and Hudson Bay region before the start of the surge (200 m bed depression increases the heat generation at the bed (video 08 of Hank (2023)) which, in turn, increases the average basal temperature with respect to the pressure melting point). An interesting effect of the 200 m deep Hudson Strait and 500 m deep ocean is that the pressure melting point is first reached further inland and not at the eastern end of the pseudo-Hudson Strait, as is the case for a flat topography (e.g., 7.8 to 8.1 kyr in the bottom row of video 07 of Hank (2023)).

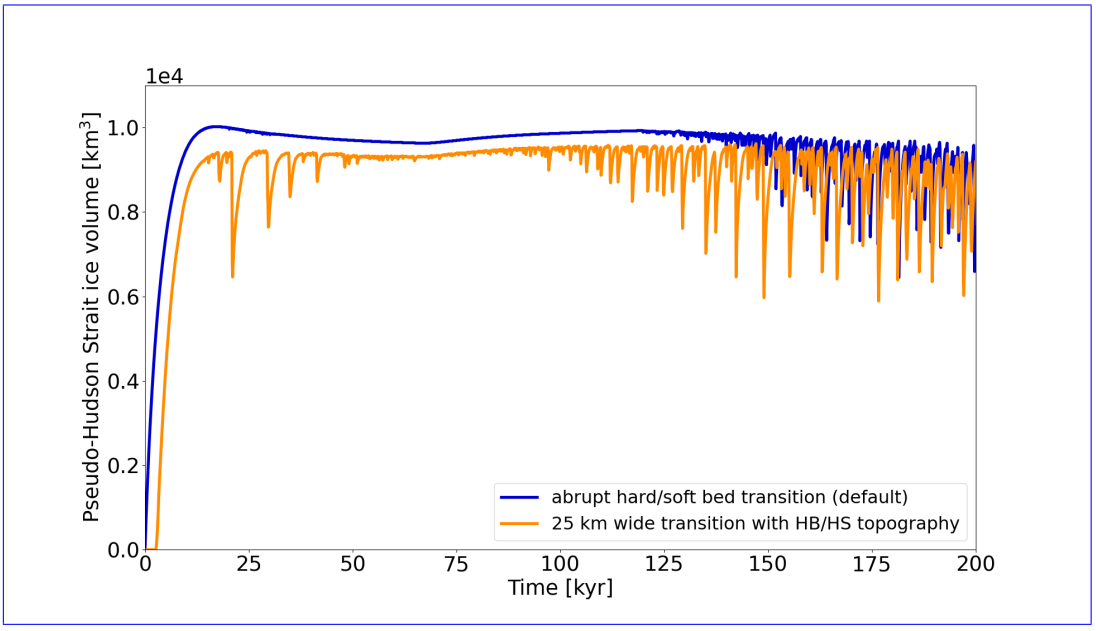


Figure S28. Pseudo-Hudson Strait ice volume for parameter vector 0 with and without a 200 m deep topography in the pseudo-Hudson Bay and Hudson Strait region using the GSM. In runs with a non-flat topography, the initial glaciation is delayed because the pseudo-Hudson Strait topography is below sea level. The horizontal grid resolution is 3.125 km.

Metric	reference setup (smooth transition, e.g., Fig. S8) reference values	abrupt transition	smooth transition with HB/HS topography
number of surges	35 ± 25	-4.8 ± 13.4	-0.1 ± 190
mean period	10 ± 10 kyr	7.4 ± 24.5	79.0 ± 64.1
mean duration	3 ± 2 kyr	21.5 ± 59.4	56.2 ± 53.4
mean pseudo-Hudson Strait ice volume change	$1.1 \pm 0.3 \cdot 10^5$ km^3	14.1 ± 37.8	389.5 ± 169.8
RMSE	-	4.4 ± 2.5	37.7 ± 2.5
Mean Bias	-	-0.2 ± 0.3	-30.4 ± 1.4
nS0	0	0	3

Table S13. Percentage differences of surge characteristics, ice volume RMSE and mean bias compared to the PISM reference setup for runs with an abrupt transition between hard bedrock and soft sediment, and runs with a pseudo-Hudson Bay/Hudson Strait (HB/HS) topography (except first column). No runs crashed and runs without surges (nS0) only contribute to the change in surge number. The first 20 kyr of each run are treated as a spin-up interval for the surge characteristics (except for the RMSE and mean bias).

S7.3 Basal hydrology

Metric	no hydrology	local hydrology	no hydrology, double C_{rmu}	no hydrology, double C_{fslid}
number of surges	180 ± 100	-3.8 ± 23.8	-9.5 ± 3.9	-3.0 ± 8.8
mean period	1.1 ± 0.5 kyr	17.4 ± 44.9	12.4 ± 4.1	4.5 ± 10.3
mean duration	0.3 ± 0.1 kyr	11.6 ± 19.1	3.1 ± 5.6	2.3 ± 3.5
mean ice volume change	$1.7 \pm 0.2 \cdot 10^3$ km^3	20.2 ± 44.7	10.5 ± 5.9	-0.9 ± 5.8
RMSE	-	8.7 ± 2.6	8.5 ± 2.7	7.8 ± 2.2
Mean Bias	-	-0.9 ± 0.8	-0.4 ± 0.4	-0.1 ± 0.1

Table S14. Percentage differences of surge characteristics, ice volume RMSE and mean bias of GSM runs with a local basal hydrology model compared to runs without sub-glacial hydrology in percent (except first column). Additionally shown are the changes in surge characteristics when doubling the values of the soft and hard bed sliding coefficient (C_{rmu} and C_{fslid} in Table 1, respectively). No runs crashed and all runs had more than 1 surge. The first 20 kyr of each run are treated as a spin-up interval for the surge characteristics (not the RMSE and mean bias).

S8 ~~GSM~~ Sensitivity experiments without a significant effect

S8.1 Weight of adjacent minimum basal temperature

Depending on the location of the adjacent minimum grid cell center basal temperature, either the ice flow (when the adjacent minimum basal temperature is downstream) or upstream propagation of the surge should be affected (decreasing basal interface temperature with increasing weight). For the large-scale surges, the adjacent minimum basal temperature is almost exclusively located upstream (e.g., video 02 of Hank (2023)). Changing the weight of the adjacent minimum basal temperature, therefore, affects the surge propagation rather than blocking parts of the ice flow.

Here we compare the effect of three different weights on the GSM surge characteristics (Eq. (S2)): no consideration of adjacent minimum basal temperature ($W_{Tb,\mathrm{min}} = 0.0$), basal temperature at the interface depends to 50 % on the adjacent minimum basal temperature at the grid cell center (reference setup, $W_{Tb,\mathrm{min}} = 0.5$), and basal temperature at the interface is equal to the adjacent minimum basal temperature at the grid cell center ($W_{Tb,\mathrm{min}} = 1.0$).

The surge cycling response to changes in $W_{Tb,\mathrm{min}}$ is not coherent (Table S15). For instance, the mean surge period increases for both $W_{Tb,\mathrm{min}} = 0.$ and ~~time step dependence~~$W_{Tb,\mathrm{min}} = 1.0$ compared to the reference $W_{Tb,\mathrm{min}} = 0.5$. However, standard deviations are large, indicating a different model response for different parameter vectors.

S8.2 Different approaches to basal hydrology

Metric	reference setup	$W_{Tb,\min} = 0.0$	$W_{Tb,\min} = 1.0$
#Surges	180 ± 100	-9.6 ± 6.5	-3.7 ± 7.8
mean period	1.1 ± 0.5 kyr	14.7 ± 13.5	3.0 ± 0.8
mean duration	0.3 ± 0.1 kyr	5.1 ± 4.9	-2.6 ± 3.3
mean pseudo-Hudson Strait ice volume change	$1.7 \pm 0.2 \cdot 10^3$ km^3	-1.9 ± 4.0	4.0 ± 6.7
RMSE	- ~~Resolution~~	7.8 ± 2.5	8.0 ± 2.5
Mean Bias	-	-0.1 ± 0.1	0.3 ± 0.1

Table S15. Percentage differences of surge characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM reference setup ($W_{Tb,\min} = 0.5$) for different weights of the adjacent minimum basal temperature for the basal sliding temperature ramp (except first column). No runs crashed and all runs had more than 1 surge. The first 20 kyr of each run are treated as a spin-up interval for the surge characteristics (not the RMSE and mean bias).

Here we compare the effects on surge characteristics when using a horizontal transport model in PISM instead of a simple local basal hydrology. In general, PISM experiments with a mass-conserving horizontal transport hydrology model yield similar results to the local hydrology model (Fig. 7 and Table S16). The mean duration, period, and ice volume change increase (11 %, 10 %, and 7 %, respectively), while the number of surges decreases (5 %). These differences are on the same level as the MNEEs (Table 6) and show large standard deviations, indicating a different model response for different parameter vectors. The ice volume RMSE and mean bias are also small (+3.9 % and −0.1 %, respectively).

Metric	local hydrology	horizontal transport
number of surges	35 ± 25	-4.6 ± 14.5
mean period	10 ± 10 kyr	10.8 ± 27.8
mean duration	3 ± 2 kyr	10.5 ± 35.4
mean ice volume change	$1.1 \pm 0.3 \cdot 10^5$ km^3	6.8 ± 17.9
RMSE	-	3.9 ± 2.5
Mean Bias	-	-0.1 ± 0.3

Table S16. Percentage differences of surge characteristics, ice volume (eastern half of pseudo-Hudson Bay and the pseudo-Hudson Strait) RMSE and mean bias of PISM runs with a mass-conserving horizontal transport hydrology model compared to the local hydrology model (except first column). No runs crashed and all runs had more than 1 surge. The first 20 kyr of each run are treated as a spin-up interval for the surge characteristics (not the RMSE and mean bias).

S8.3 Basal hydrology instead of basal temperature ramp as the primary smoothing mechanism

We examine the effects of a local basal hydrology as main smoothing mechanism for basal sliding (compared to a basal temperature ramp) by using a very sharp ramp ($T_{ramp} = 0.001$, $T_{exp} = 28$), minimizing the smoothing effect of the basal temperature ramp. The change in surge characteristics between runs with local basal hydrology and the sharp temperature ramp and the GSM reference setup is similar (maximum difference of 3 %; compare Table S14 and S17) to the runs with local basal hydrology and the reference basal temperature ramp ($T_{ramp} = 0.0625$, $T_{exp} = 28$), indicating that the local basal hydrology is the primary smoothing mechanism in both cases. The differences in the change of surge characteristics between the reference and the steeper ramp are smaller than the MNEEs, preventing further analysis.

Metric	reference setup	steeper ramp ($T_{ramp} = 0.001$, $T_{exp} = 28$), local hydrology
#Surges	180 ± 100	-3.8 ± 24.6
mean period	1.1 ± 0.5 kyr	16.0 ± 42.0
mean duration	0.3 ± 0.1 kyr	8.7 ± 17.0
mean pseudo-Hudson Strait ice volume change	$1.7 \pm 0.2 \cdot 10^3$ km^3	21.5 ± 43.3
RMSE	-	8.9 ± 3.2
Mean Bias	-	-0.6 ± 0.9

Table S17. Percentage differences of surge characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM reference setup with local basal hydrology instead of a basal temperature ramp as the primary smoothing mechanism (except first column). No runs crashed and all runs had more than 1 surge. The first 20 kyr of each run are treated as a spin-up interval for the surge characteristics (not the RMSE and mean bias).

S9 Convergence study

S9.1 GSM convergence study without basal hydrology

Analyzing individual GSM parameter vectors in detail shows that some discrepancies prevail even when using a resolution-dependent temperature ramp. In the case of parameter vector 1, for example, surges do still not occur for the coldest temperatures (Fig. S29). Note the asymmetry in termination and onset of surge cyclicity ($\Delta t_1 < \Delta t_2$). For increasing temperatures after the minimum surface temperature $T_{\min} = -15°\mathrm{C}$ at $t_{\min} = 66.7$ kyr, the first surge occurs at a surface temperature slightly higher than the initial temperature T_{init}, for which oscillations occur. The difference between Δt_1 and Δt_2 is ~ 25 kyr and closely resembles the lag of the average pseudo-Hudson Strait basal temperature with respect to the pressure melting point behind the surface temperature changes. For example, the minimum average pseudo-Hudson Strait basal temperature with respect to the pressure melting point ($T_{\mathrm{bpm}} = -3.2°\mathrm{C}$) occurs 23 kyr after the minimum surface temperature (not shown). The period without oscillations in the 25 km run corresponds to a period of somewhat smaller and less frequent oscillations in the higher resolution runs (Fig. 12).

Setup	number of surges	mean period	mean duration	mean pseudo-Hudson Strait ice volume change	nS0
3.125 km reference setup	180 ± 100	1.1 ± 0.5 kyr	0.3 ± 0.1 kyr	$1.7 \pm 0.2 \cdot 10^3$ km^3	0
25 km, constant ramp	-95.1 ± 7.4	942.3 ± 517.70	300.0 ± 172.22	95.9 ± 52.6	3
25 km, resolution-dependent ramp	-78.1 ± 18.2	414.5 ± 309.0	119.5 ± 17.6	91.9 ± 23.6	1
25 km, $T_{ramp} = 0.5, T_{exp} = 5$	-15.9 ± 20.4	29.7 ± 24.6	43.8 ± 36.6	3.5 ± 18.7	0
12.5 km, constant ramp	-59.2 ± 16.5	129.0 ± 41.8	90.3 ± 17.9	50.3 ± 76.5	0
12.5 km, resolution-dependent ramp, also minimum mean score	-56.5 ± 15.1	115.7 ± 46.8	101.1 ± 20.5	33.0 ± 66.3	0
6.25 km, constant ramp	-24.2 ± 13.1	36.4 ± 20.9	24.8 ± 8.5	14.9 ± 14.2	0
6.25 km, resolution-dependent ramp	-27.9 ± 9.9	42.2 ± 18.9	32.1 ± 6.3	15.9 ± 12.3	0
6.25 km, $T_{ramp} = 0.125$, $T_{exp} = 45$	-25.3 ± 13.6	37.9 ± 26.7	28.2 ± 7.0	9.8 ± 11.6	0
0.5 year maximum time step	-4.4 ± 4.5	5.4 ± 4.8	2.0 ± 2.4	-0.5 ± 5.6	0
0.25 year maximum time step	-1.8 ± 3.2	2.6 ± 4.4	-0.1 ± 3.4	-0.1 ± 3.4	0

Table S18. Percentage differences of surge characteristics compared to the 3.125 km GSM reference setup (except first row). The values represent the average of 5 parameter vectors. No runs crashed and runs without surges (nS0) only contribute to the change in surge numbers. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above. The resolution-dependent ramps ($T_{exp} = 28$) and constant ramp (black line, $T_{ramp} = 0.0625$, $T_{exp} = 28$) are shown in Fig. 2. The third ramp listed for each resolution is the ramp with the smallest mean score (Table S11).

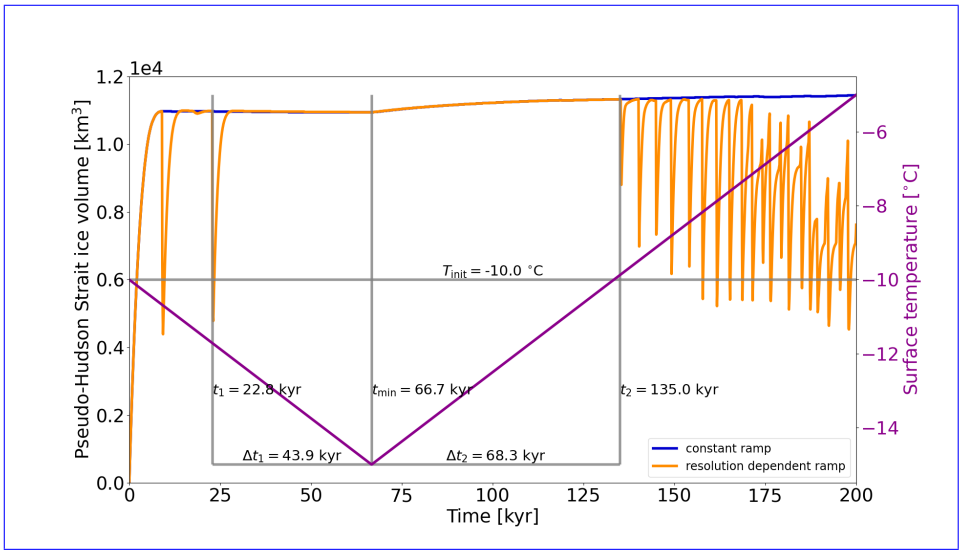


Figure S29. Pseudo-Hudson Strait ice volume for parameter vector 1 and different basal temperature ramps using the GSM (constant ramp: $T_{ramp} = 0.0625°\text{C}$ and $T_{exp} = 28$; resolution-dependent ramp: $T_{ramp} = 0.5°\text{C}$ and $T_{exp} = 28$, see Fig. 2). The right axis shows the surface temperature when ignoring the lapse rate dependency. t_{min}, t_1, and t_2 mark the time of the minimum surface temperature, the start of the last surge before t_{min}, and the start of the first surge after t_{min}, respectively. Δt_1 and Δt_2 represent the time difference between t_{min} and t_1 and t_2, respectively. T_{init} indicates the surface temperature at the beginning of the run.

Setup	mean RMSE	mean Bias
25 km, constant ramp	17.1 ± 4.7	14.6 ± 4.4
12.5 km, constant ramp	10.3 ± 2.5	4.0 ± 1.1
6.25 km, constant ramp	8.7 ± 2.2	0.4 ± 0.5
25 km, resolution-dependent ramp	15.3 ± 3.0	10.3 ± 2.5
12.5 km, resolution-dependent ramp	10.3 ± 2.8	3.0 ± 2.3
6.25 km, resolution-dependent ramp	8.5 ± 2.3	0.2 ± 0.6
25 km, $T_{ramp} = 0.5$, $T_{exp} = 5$	12.8 ± 2.4	6.4 ± 1.7
12.5 km, $T_{ramp} = 0.25$, $T_{exp} = 28$, same as resolution-dependent ramp	10.3 ± 2.8	3.0 ± 2.3
6.25 km, $T_{ramp} = 0.125$, $T_{exp} = 45$	8.5 ± 2.0	0.6 ± 0.5
0.5 year maximum time step	7.7 ± 2.4	0.0 ± 0.1
0.25 year maximum time step	8.0 ± 2.5	0.0 ± 0.1

Table S19. Pseudo-Hudson Strait ice volume RMSE and mean bias compared to the 3.125 km GSM ~~base~~ reference setup in percent. The values represent the average of 5 parameter vectors. No runs crashed and the entire 200 kyr run time is used (no spin-up interval).

S10 ~~GSM - Basal hydrology instead of basal temperature ramp as the primary smoothing mechanism~~

~~Metric base setup steeper ramp ($T_{ramp} = 0.001$, $T_{exp} = 28$), local hydrology #Events 180 ± 100 -3.8 ± 24.6 mean period 1.1 ± 0.5 kyr 16.0 ± 42.0 mean duration 0.3 ± 0.1 kyr 8.7 ± 17.0~~

S9.1 GSM convergence study with basal hydrology

Based on the results without basal hydrology (Sec. 3.4.1), 5 basal temperature ramps ($T_{exp} = [5, 10, 15, 20, 28]$) with a resolution-dependent T_{ramp} (Eq. (9)) are tested for all resolutions. As it is unclear which basal temperature ramp should be used at the highest horizontal grid resolution (3.125 km), we test two different ramps ($T_{exp} = [5, 28]$). The experiments that yield the smallest differences in surge characteristics (smallest mean score in Table S20 and S21) compared to the corresponding 3.125 km reference runs (bold rows) are presented in Table S22.

Similar to the results without a basal hydrology model, the smallest differences in surge characteristics (except the mean pseudo-Hudson Strait ice volume change~~$1.7 \pm 0.2 \cdot 10^3$ km^3 21.5 ± 43.3RMSE - 8.9 ± 3.2 Mean Bias - -0.6 ± 0.9 Percentage differences of event characteristics, pseudo-Hudson Strait ice volume RMSE and mean bias compared to the GSM base setup with local basal hydrology instead of a basal temperature ramp as the primary smoothing mechanism (except first column). No runs crashed and all runs had more than 1 surge event. The first 20 kyr of each run are treated as a spin-up interval for the event characteristics (not the RMSE and mean bias).~~) occur for the coarsest horizontal grid resolution (25 km, Table S22). This likely indicates that the optimal ramps at 12.5 and 6.25 km horizontal grid resolution have not been found.

S10 ~~GSM - Resolution scaling with basal hydrology~~

In general, the resolution-dependent ramp with $T_{exp} = 5$ leads to the smallest differences between coarse and high-resolution runs. The differences in surge characteristics are significantly smaller than for a resolution-dependent temperature ramp without local basal hydrology (except for the mean pseudo-Hudson Strait ice volume change, Table S18 vs. S22), further underlining the importance of the basal hydrology.

~~Setup mean RMSE mean Bias 25 km, $T_{ramp} = 0.5$, $T_{exp} = 5$ 14.3 ± 3.2 6.0 ± 0.9~~ Except for 12.5 km ~~, $T_{ramp} = 0.25$, $T_{exp} = 5$ 11.2 ± 4.2 0.6 ± 2.4 6.25 km, $T_{ramp} = 0.125$, $T_{exp} = 28$ 10.0 ± 3.0 0.5 ± 0.6 25 km, $T_{ramp} = 0.5$, $T_{exp} = 5$ 14.5 ± 3.2 6.8 ± 0.4~~ horizontal grid resolution, the resolution-dependent ramp with $T_{exp} = 5$ yields a self-consistent response across all resolutions. At 12.5 km, ~~$T_{ramp} = 0.25$, $T_{exp} = 10$ 11.7 ± 4.2 1.6 ± 2.5~~ the next closest exponent ($T_{exp} = 10$) has the minimum mean score. However, given that there is no single best ramp across all resolutions, we assess different ramps as to whether differences are within inferred MNEEs (DWINS). To this end, we calculate the differences between the ramp with the minimum mean score and all other ramps at each resolution and for all surge characteristics (Table S20 and S21). We rule out ramps for which the differences exceed the maximum MNEEs (maximum of Table 5 and S2) for more than one surge characteristic (DWINS failures).

Under these criteria and when using $T_{exp} = 5$ at 3.125 km horizontal grid resolution, the resolution-dependent ramp with $T_{exp} = 10$ remains within the DWINS ensemble for all resolutions (Table S21). The results for $T_{exp} = 28$ at 3.125 km horizontal grid resolution do not yield a single ramp that remains within the DWINS ensemble at all resolutions (Table S20). However, except for 6.25 km, ~~$T_{ramp} = 0.125$, $T_{exp} = 5$ 10.1 ± 1.8 0.6 ± 0.8 Resolution scaling of pseudo-Hudson Strait ice volume RMSE and mean bias with local basal hydrology in percent. The three upper ramps are compared to the 3.125 km GSM setup with $T_{exp} = 28$, the lower three to $T_{exp} = 5$. The values represent the average of 5 parameter vectors. No runs crashed and the entire 200 kyr run time is used (no spin-up interval).~~ for which the differences between the tested basal temperature ramps are the smallest, $T_{exp} = 5$ yields the minimum mean-score.

The pseudo-Hudson Strait ice volume RMSE and mean bias show convergence (smaller differences) for both 3.125 km horizontal grid resolution setups (Table S23).

ramp	score-mean	score-std	sum of scores	~~DWINN~~ DWINS failures
res= 25 km, $T_{exp} = 5$, $T_{ramp} = 0.5$	**0.99**	4.31	**5.31**	0
res= 25 km, $T_{exp} = 10$, $T_{ramp} = 0.5$	1.44	5.29	6.74	2
res= 25 km, $T_{exp} = 15$, $T_{ramp} = 0.5$	4.80	**3.05**	7.85	4
res= 25 km, $T_{exp} = 20$, $T_{ramp} = 0.5$	5.65	3.47	9.11	4
res= 25 km, $T_{exp} = 28$, $T_{ramp} = 0.5$	7.11	3.88	11.00	4
res= 12.5 km, $T_{exp} = 5$, $T_{ramp} = 0.25$	**3.69**	4.60	8.29	0
res= 12.5 km, $T_{exp} = 10$, $T_{ramp} = 0.25$	3.81	5.07	8.88	2
res= 12.5 km, $T_{exp} = 15$, $T_{ramp} = 0.25$	3.82	4.11	7.93	2
res= 12.5 km, $T_{exp} = 20$, $T_{ramp} = 0.25$	4.21	3.42	7.63	3
res= 12.5 km, $T_{exp} = 28$, $T_{ramp} = 0.25$	4.47	**2.81**	**7.28**	4
res= 6.25 km, $T_{exp} = 5$, $T_{ramp} = 0.125$	4.03	4.29	8.33	3
res= 6.25 km, $T_{exp} = 10$, $T_{ramp} = 0.125$	3.94	**3.76**	7.70	3
res= 6.25 km, $T_{exp} = 15$, $T_{ramp} = 0.125$	4.65	3.90	8.55	1
res= 6.25 km, $T_{exp} = 20$, $T_{ramp} = 0.125$	3.79	3.82	**7.60**	1
res= 6.25 km, $T_{exp} = 28$, $T_{ramp} = 0.125$	**3.59**	4.23	7.82	0

Table S20. Single value scores for the mean and standard deviation of the basal temperature ramps and the number of ~~DWINN~~ DWINS failures (maximum 4) for a resolution-dependent ~~base~~ reference temperature ramp with $T_{exp} = 28$ in the GSM. The minimum scores for the mean, standard deviation, and sum at each resolution are marked as bold numbers. At = 25 km, 1 run crashed for $T_{exp} = 10$ and 1 run showed no ~~events~~ surges for $T_{exp} = [15, 20, 28]$. Note that the sum of scores can be slightly off due to rounding (± 0.01).

ramp	score-mean	score-std	sum of scores	~~DWINN~~ DWINS failures
res= 25 km, $T_{exp} = 5$, $T_{ramp} = 0.5$	**0.84**	3.91	**4.75**	0
res= 25 km, $T_{exp} = 10$, $T_{ramp} = 0.5$	1.21	5.04	6.25	1
res= 25 km, $T_{exp} = 15$, $T_{ramp} = 0.5$	4.89	**3.29**	8.18	4
res= 25 km, $T_{exp} = 20$, $T_{ramp} = 0.5$	5.76	3.63	9.40	4
res= 25 km, $T_{exp} = 28$, $T_{ramp} = 0.5$	7.30	4.13	11.43	4
res= 12.5 km, $T_{exp} = 5$, $T_{ramp} = 0.25$	3.97	4.49	8.45	2
res= 12.5 km, $T_{exp} = 10$, $T_{ramp} = 0.25$	**3.77**	4.60	8.37	0
res= 12.5 km, $T_{exp} = 15$, $T_{ramp} = 0.25$	3.79	4.13	7.93	1
res= 12.5 km, $T_{exp} = 20$, $T_{ramp} = 0.25$	4.10	3.50	**7.59**	1
res= 12.5 km, $T_{exp} = 28$, $T_{ramp} = 0.25$	4.37	**3.28**	7.65	2
res= 6.25 km, $T_{exp} = 5$, $T_{ramp} = 0.125$	**3.53**	4.44	7.97	0
res= 6.25 km, $T_{exp} = 10$, $T_{ramp} = 0.125$	4.27	3.77	8.04	0
res= 6.25 km, $T_{exp} = 15$, $T_{ramp} = 0.125$	4.59	3.82	8.42	1
res= 6.25 km, $T_{exp} = 20$, $T_{ramp} = 0.125$	3.91	**3.64**	**7.55**	1
res= 6.25 km, $T_{exp} = 28$, $T_{ramp} = 0.125$	3.70	4.33	8.03	3

Table S21. Single value scores for the mean and standard deviation of the basal temperature ramps and the number of ~~DWINN~~ DWINS failures (maximum 4) for a resolution-dependent ~~base~~ reference temperature ramp with $T_{exp} = 5$ in the GSM. The minimum scores for the mean, standard deviation, and sum at each resolution are marked as bold numbers. At $= 25$ km, 1 run crashed for $T_{exp} = 10$ and 1 run showed no ~~events~~ surges for $T_{exp} = [15, 20, 28]$. Note that the sum of scores can be slightly off due to rounding (± 0.01).

S10 ~~PISM - Resolution and time step dependence~~

Setup	number of surges	mean period	mean duration	mean pseudo-Hudson Strait ice volume change
3.125 km, $T_{ramp} = 0.0625$, $T_{exp} = 28$	**197 ± 131**	**1.5 ± 1.1 kyr**	**0.3 ± 0.2 kyr**	**$2.0 \pm 0.7 \cdot 10^3$ km^3**
25 km, $T_{ramp} = 0.5$, $T_{exp} = 5$	9.7 ± 59.9	15.5 ± 42.3	24.3 ± 36.1	13.6 ± 46.7
12.5 km, $T_{ramp} = 0.25$, $T_{exp} = 5$	-36.1 ± 17.6	68.0 ± 49.8	97.1 ± 60.3	3.0 ± 26.4
6.25 km, $T_{ramp} = 0.125$, $T_{exp} = 28$	-13.2 ± 31.1	27.0 ± 40.6	25.7 ± 25.2	5.6 ± 27.5
3.125 km, $T_{ramp} = 0.0625$, $T_{exp} = 5$	**190 ± 118**	**1.3 ± 0.7 kyr**	**0.3 ± 0.2 kyr**	**$1.8 \pm 0.4 \cdot 10^3$ km^3**
25 km, $T_{ramp} = 0.5$, $T_{exp} = 5$	-2.4 ± 35.8	16.1 ± 31.4	20.7 ± 30.3	14.3 ± 35.8
12.5 km, $T_{ramp} = 0.25$, $T_{exp} = 10$	-37.7 ± 12.1	61.7 ± 44.1	63.4 ± 34.8	20.5 ± 39.0
6.25 km, $T_{ramp} = 0.125$, $T_{exp} = 5$	-25.6 ± 13.9	37.8 ± 23.8	41.1 ± 21.3	0.3 ± 19.8

Table S22. Percentage differences (except bold rows) of surge characteristics compared to the 3.125 km GSM setups with local basal hydrology (bold rows, $T_{exp} = [5, 28]$) for the ramps with the smallest mean score (analysis steps described in Sec. S7.1). The values represent the average of 5 parameter vectors. No runs crashed and all runs had more than 1 surge. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above.

Setup	mean RMSE	mean Bias
25 km, $T_{ramp} = 0.5$, $T_{exp} = 5$	14.3 ± 3.2	6.0 ± 0.9
12.5 km, $T_{ramp} = 0.25$, $T_{exp} = 5$	11.2 ± 4.2	0.6 ± 2.4
6.25 km, $T_{ramp} = 0.125$, $T_{exp} = 28$	10.0 ± 3.0	0.5 ± 0.6
25 km, $T_{ramp} = 0.5$, $T_{exp} = 5$	14.5 ± 3.2	6.8 ± 0.4
12.5 km, $T_{ramp} = 0.25$, $T_{exp} = 10$	11.7 ± 4.2	1.6 ± 2.5
6.25 km, $T_{ramp} = 0.125$, $T_{exp} = 5$	10.1 ± 1.8	0.6 ± 0.8

Table S23. Resolution scaling of pseudo-Hudson Strait ice volume RMSE and mean bias with local basal hydrology in percent. The three upper ramps are compared to the 3.125 km GSM setup with $T_{exp} = 28$, the lower three to $T_{exp} = 5$. The values represent the average of 5 parameter vectors. No runs crashed and the entire 200 kyr run time is used (no spin-up interval).

S9.1 GSM convergence study with active SSA everywhere (no basal hydrology)

Setup	number of surges	mean period	mean duration	mean pseudo-Hudson Strait ice volume change	nS0
3.125 km reference setup	216 ± 146	1.3 ± 0.8 kyr	0.3 ± 0.1 kyr	$1.4 \pm 0.6 \cdot 10^3$ km^3	0
25 km	-76.9 ± 17.9	432.2 ± 384.3	151.7 ± 49.0	163.8 ± 65.9	1
12.5 km	-61.3 ± 23.8	179.8 ± 127.1	154.3 ± 79.3	40.1 ± 54.4	0
6.25 km	-46.2 ± 11.6	66.7 ± 12.9	59.8 ± 19.5	75.1 ± 42.9	0

Table S24. Percentage differences of surge characteristics compared to the 3.125 km GSM reference setup (except first row) with a resolution-dependent basal temperature ramp ($T_{exp} = 28$, Fig. 2) and active SSA everywhere. The values represent the average of 5 parameter vectors. No runs crashed and runs without surges (nS0) only contribute to the change in surge numbers. The first 20 kyr of each run are treated as a spin-up interval and are not considered in the above.

S9.2 PISM convergence study

Similar to the results presented for the GSM (Sec. S9.1 and S9.1), analyzing individual parameter vectors for PISM shows significant differences in surge behavior for different horizontal grid resolutions. Parameter vector ~~1~~ 8 at 25 km horizontal grid resolution, for example, ~~first transitions from high-frequency small-scale oscillations to low-frequency oscillations with larger amplitudes (110 kyr) before it remains at a constant ice volume (starting at 150 kyr,~~ only shows 4 oscillations (Fig. S30~~and video 09 of Hank (2023)~~). In contrast, ~~high-frequency oscillations are present until the end of the run at~~ more oscillations occur for both the 12.5 km and 50 km horizontal grid resolution ~~. The~~ run. Additionally, most of the 50 km ~~run also shows oscillations for the entire run period, but the ice volume at 200 kyr is roughly $0.3 \cdot 10^6$ km^3 (27.5 %)larger than the corresponding value at 12.5 km.~~

~~Ice Volume in the eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait for parameter vector 1 and different horizontal grid resolutions using the Parallel Ice Sheet Model (PISM). See also video 09 of Hank (2023).~~ surges transport ice toward the West, whereas the 25 and 12.5 km runs almost exclusively surge through the pseudo-Hudson Strait (video 09 of Hank (2023)).

Setup	number of surges	mean period	mean duration	mean ice volume change	nC	nS0	nS1
12.5 km reference setup	**22 ± 19**	**9 ± 6 kyr**	**3 ± 1 kyr**	**$1.3 \pm 0.2 \cdot 10^5$ km^3**	**4**	**1**	**0**
50 km	4.1 ± 46.0	15.3 ± 47.4	11.9 ± 33.7	30.6 ± 39.6	0	0	1
25 km	-28.3 ± 12.0	46.5 ± 31.9	6.3 ± 13.8	4.0 ± 20.3	0	0	0
25 km reference setup	**35 ± 25**	**10 ± 10 kyr**	**3 ± 2 kyr**	**$1.1 \pm 0.3 \cdot 10^5$ km^3**	**0**	**0**	**0**
0.5 year maximum time step	-15.3 ± 34.1	-4.4 ± 7.4	-1.5 ± 7.5	3.8 ± 10.5	0	1	0
0.25 year maximum time step	2.8 ± 14.9	1.4 ± 8.6	3.4 ± 14.3	5.5 ± 11.9	0	0	0

Table S25. ~~Ice volume in the eastern half~~ Percentage differences (except bold rows) of ~~the pseudo-Hudson Bay~~ PISM surge characteristics due to different horizontal grid resolutions and maximum time steps. Note that the ~~pseudo-Hudson Strait~~ 12.5 km (highest resolution tested) is used as a reference for the grid resolution convergence study, whereas the 25 km setup is used for the maximum time step experiments. The values represent the average of 9 parameter ~~vector 1 and different~~ vectors for the maximum time ~~steps using~~ step experiments and only 5 for the ~~Parallel Ice Sheet Model~~ resolution convergence study. 4 of the 12.5 km runs crashed after $\sim$ 50 kyr because they hit the run-time limit on the computational cluster (~~PISM~~7 days) and 1 12.5 km run does not show a surge (nS0). Crashed runs (nC) are not considered and runs without surges in the comparison setup only contribute to the change in surge numbers. Runs without surges in the reference setup are not considered. The ~~horizontal grid resolution is 25 km~~first 20 kyr of each run are treated as a spin-up interval and are not considered in the above.

Setup	nC	mean RMSE	mean Bias
50 km	~~4~~ 0	~~12.9 ± 4.1~~ 11.1 ± 2.6	~~9.0 ± 5.5~~ 6.5 ± 4.1
25 km	0	~~8.5 ± 2.0~~ 7.4 ± 1.4	~~3.3 ± 1.4~~ 3.7 ± 0.8
0.5 year maximum time step	0	~~4.7 ± 3.0~~ 3.7 ± 2.8	~~-0.1 ± 2.4~~ -0.1 ± 0.4
0.25 year maximum time step	0	~~3.8 ± 3.1~~ 3.5 ± 2.3	~~-0.7 ± 2.3~~ -0.1 ± 0.2

Table S26. Ice volume RMSE and mean bias (in percent) due to different horizontal grid resolutions and maximum time steps. Note that the 12.5 km (highest resolution tested~~, one crashed run~~) is used as ~~base~~ a reference for the grid resolution convergence study, whereas the 25 km setup is used for the maximum time step experiments. The values represent the average of ~~10~~ 9 parameter vectors for the maximum time step experiments and only 4 for the resolution convergence study. 4 of the 12.5 km runs crashed after $\sim$ 50 kyr because they hit the run-time limit on the computational cluster (7 days) and 1 12.5 km run does not show a surge (nS0). Crashed runs (nC) are not considered. The entire 200 kyr run time is used (no spin-up interval).

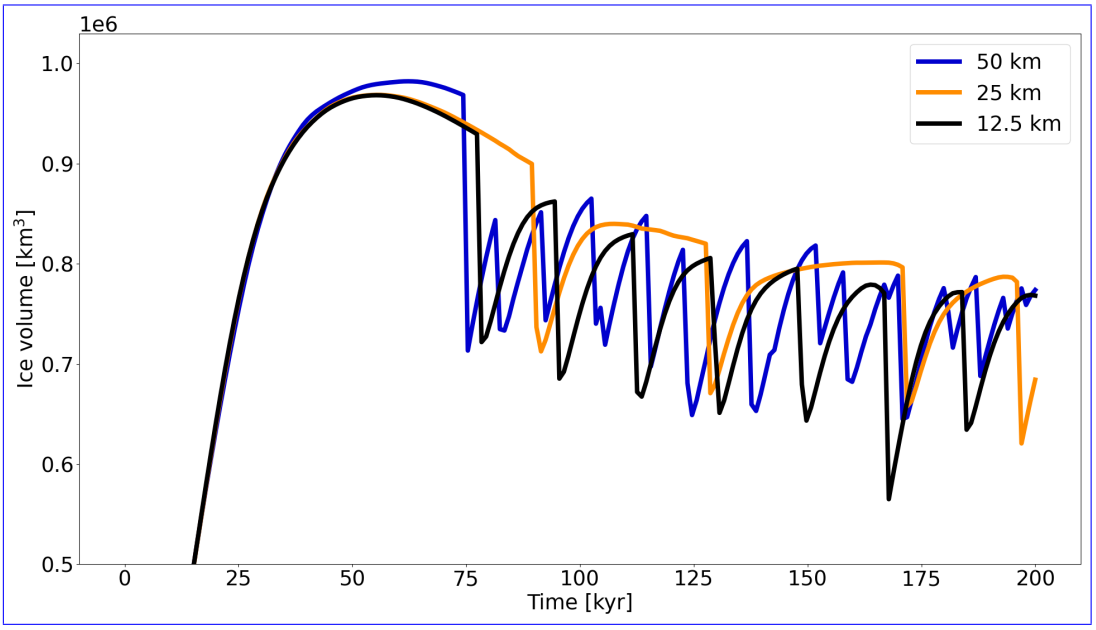


Figure S30. Ice Volume in the eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait for parameter vector 8 and different horizontal grid resolutions using the PISM. See also video 09 of Hank (2023).

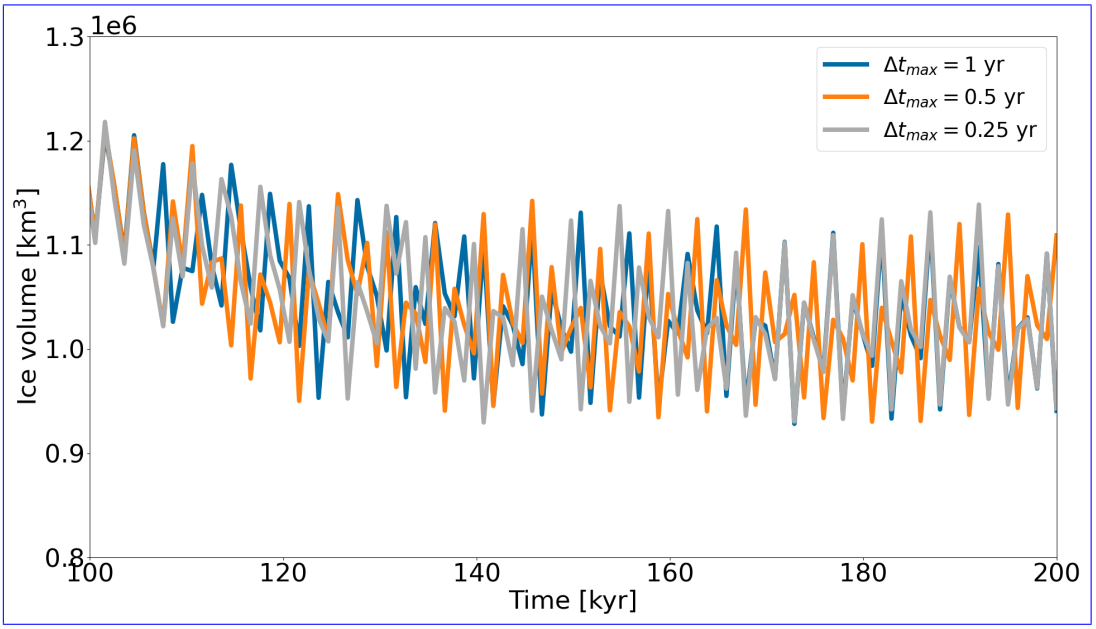


Figure S31. Ice volume in the eastern half of the pseudo-Hudson Bay and the pseudo-Hudson Strait for parameter vector 1 and different maximum time steps using the PISM. The horizontal grid resolution is 25 km.

References

Drew, M. and Tarasov, L.: Surging of a Hudson Strait Scale Ice Stream: Subglacial hydrology matters but the process details don't, The Cryosphere Discussions, 2022, 1–41, https://doi.org/10.5194/tc-2022-226, 2022.

Hank, K.: Supplementary material for "Numerical issues in modeling ice stream surge cycling", https://doi.org/10.5281/zenodo.7905404, 2023.

K.M. Cuffey and W.S.B. Paterson.: The Physics of Glaciers, Butterworth-Heinemann/Elsevier, Burlington, MA, 4th edn., 2010.