

Dear Pippa Whitehouse,

We thank you very much for your comments and insights on the manuscript. We have addressed all your comments one by one and made the necessary modifications in the updated version of the manuscript. Please find below our responses (in blue) to the comments (in black). A marked-up version of the updated manuscript is provided for your convenience below the responses. Throughout our responses, page and line numbers relate to this document. Should you require any further information, please do not hesitate to contact me.

On behalf of all authors,

Vincent Verjans

10 Minor points – page/line numbers relate to version2 of the manuscript (no track changes)

p.1 l.31: most points in this paragraph document why each application is important rather than recommending specific steps for improvement – the text on line 31 is a little misleading since it could be taken to suggest that you will investigate sensitivity to climate conditions in this study

15 We have slightly changed the phrasings in the paragraph to underline the direct role of firm models in these applications. We have also removed the mention to sensitivity to climatic conditions in order to avoid any misunderstanding. However, it should be noted that firm models simulate densification as a function of climate, and their inherent objective is thus to capture the climatic sensitivity of the densification process.

Changes in the text (p.8):

20 "Consequently, uncertainties in modelled densification rates have a direct impact on mass balance estimates, which rely on a correct conversion from measured volume changes to mass changes"

"Model estimates of current and future surface mass balance of the AIS and GrIS are thus dependent on accurate models of firn evolution."

p.2 l.16: check throughout whether it is appropriate use to use 'AIS/GrIS' or 'the AIS/GrIS'

25 Indeed, the previous manuscript was inconsistently alternating between both wordings. We now exclusively use "the AIS/GrIS" when used as a noun (e.g. "locations of the AIS"). We keep the use of "AIS/GrIS" when used as an adjective (e.g. "14 AIS cores").

p.4 l.5-7: text seems a little out of place here, information would fit better at the end of the final paragraph of p.3

30 We moved the paragraph as suggested (p.11 l.1).

p.4 l.25-26: the statement that you use a constant site-specific value is slightly at odds with the statement on lines 29-30 that you add random noise at every model time step. It would be useful to mention the approach used to account for uncertainty earlier in this paragraph

35 We rephrased the paragraph in order to remove the mention to constant values. We also mention our treatment of uncertainty in ρ_0 earlier, as suggested (p.11 l.28):

"At each site, the ρ_0 value is taken in agreement with the shallow densities measured in the corresponding core of the dataset. However, measurements of fresh snow density are highly variable (e.g. Fausto et al., 2018). We account for uncertainty in this parameter by adding normally distributed random noise with standard deviation 25 kg m⁻³ to ρ_0 at every model time step (see Supplementary Information)."

40

p.6 1.9-11: do Li and Zwally (2015) use a different formulation of the equation, or do they just determine different parameter values compared with Li and Zwally (2011)?

The Li and Zwally (2015) model uses a formulation very close, but still different to the Li and Zwally (2011) model. We clarified this aspect in the manuscript (p.13 1.14):

5 " Later, Li and Zwally (2015) developed a densification model calibrated for Antarctic firn. The latter model uses the same governing equations as LZ for c_0 and c_1 but different formulations for β_0 and β_1 (Eq. (6))."

p.6 1.25: to data -> with data

Changed (p.14 1.2)

10

p.7 1.23: several steps are described prior to the mention of figure 2a; please indicate how the text on lines 20-23 relates to steps shown in figure 2. Also, please clarify whether calculations are carried out for both θ_i and θ_i^* on each iteration

The confusion stems from the first iteration (at $i = 0$), which must be executed to start the Random Walk
15 Metropolis algorithm. We have clarified that the first step must compute an initial posterior probability distribution for the algorithm to start because the acceptance step (Fig. 2e) requires a ratio $P(\theta_i^*|Y)/P(\theta_i|Y)$. This first step is executed by using the original models parameter values (from HL, Ar and LZ). This has been included in the description of the algorithm.

We have also emphasised that calculations for θ_i must not be performed at each iteration, since they were
20 performed at a previous stage. As such, the posterior probability value $P(\theta_i|Y)$ is kept in memory and only $P(\theta_i^*|Y)$ must be computed.

We hope that the adjustments made in the text now provide better clarity (p.14 1.25). We also hope that Figure 2 will be next to the relevant text after the typesetting.

25 p.8 1.1: variance -> covariance (as defined on 1.25 of the previous page)

Changed (p.15 1.10)

p.8 1.16: 'a 500 random sample' – rephrase

Changed (p.15 1.25): "from 500 random samples"

30

p.9 1.22: the DML plots are figs. 5g-i

Thank you for pointing this out. The text has been changed (p.16 1.30).

p.9 1.28: 'the better performance at the GrIS evaluation sites...' – make it clear that this text relates to the
35 performance of the original model

Changed (p17. 1.5): "As such, the better performance at the GrIS evaluation sites of the original HL is likely due to its parameterisation being better suited for the particular temperature range corresponding to the conditions of the latter sites."

40 p.10 1.7: it is not clear to me how the LZ dual model was constructed; do you determine different parameters for each ice sheet by dividing the calibration data set, or is the whole formulation of the model different? Refer to Table 2 when quoting results for the LZ dual model

The whole formulation of the model is different. (The formulations are close but not the same). We hope that our changes related to the comment above clarifies this aspect.

Furthermore, we clarified the construction of LZ dual (p.17 l.16): " We compute results at the AIS and GrIS evaluation sites using the Li and Zwally (2015) model for the AIS and the Li and Zwally (2011) model for the GrIS, so that both models are applied to the ice sheet for which they were originally developed. We call this pairing of models LZ dual and evaluate its general performance."

We now refer to Table 2 when stating the RMSE values of LZ dual (p.17 l.21).

p.10: if feasible, it would be useful to include a figure showing the results for LZ dual and IMAU-FDM (e.g. similar to figure 4) in the supplementary information

We have added similar scatter plots for the LZ dual and IMAU models in the Supplementary Information: Figure S6. We inform the reader about this in the main text; in the caption of Figure 4, we added:

"Similar scatter plots for the LZ dual and IMAU results are shown in the Supplementary Information (Fig. S6)."

p.10 l.23: please include information on how uncertainty intervals were constructed in the captions to figures 4 and 5

We added in both captions: "The 95% credible intervals are computed from results of 500 randomly selected parameter combinations from the posterior ensembles of each model (HL, Ar, LZ)."

p.11 l.5: 'indicate a weaker increase...' – weaker than what?

We specified (p.18 l.16): "Our results of stage-1 exponents (a, α) smaller than 1 indicate a weaker increase in densification rates with pressure than assumed in the original versions of Ar and HL."

p.11 l.12: 'The same can be applied...' – not clear what 'The same' refers to

Changed (p.18 l.23): "The difference in sensitivities of stage-1 and stage-2 densification to accumulation also holds in the LZ model (...)"

p.11, l.13: please refer to a figure or table when quoting correlation coefficient values

We now refer to the Figure S5 of the posterior correlation matrices (p.18 l.26).

p.11 l.22: over-sensitivity -> over-sensitivity in Ar

Added (p.18 l.3)

p.12 l.25: what is the reference period? If it is 2000-2017 this should be explicitly stated

The reference periods for both ice sheets are different and the information is provided in Section 2.2. Because this section is in the earlier parts of the manuscript, we agree that it is better to remind the reader about it. We have added a reference to Sect. 2.2 in the text (p.20 l.6): "over the reference period (see Sect. 2.2)."

p.12 l.33: make it clearer that uncertainties in the following sentences are calculated using the CV values quote above/ in table 3

We clarified this point by mentioning our use of CV to estimate uncertainty ranges before providing the uncertainty values (p.20 l.14):

" By using the CV values, we can calculate reasonable uncertainty estimates for cmp_{an} and age_{pc} . For instance, in the dry snow zone of GrIS, simulated compaction anomalies are typically around 20 cm over 2000–2017, and thus come with an uncertainty of the order of ± 4 cm. (...)"

- 5 p.13 1.5: a couple of clarifications needed: (i) what does ‘it’ refer to, and (ii) what does ‘Such numbers’ refer to?
We changed "it" to specify that we refer to the spatial aggregation of uncertainties (p.20 1.20):
"Absolute uncertainty is thus reduced but still critical given the large area of the AIS over which uncertainties are aggregated when mass balance trends are evaluated."
We changed our use of "Such numbers" in accordance with the previous comment (p.20 1.20):
- 10 " The uncertainty ranges calculated from the CV values provide an order of magnitude of errors in firn model outputs that must be accounted for in altimetry-based mass balance assessments and in ice core studies, respectively."

- p.13 1.7: the purpose of the text (paragraph?) starting on this line is initially unclear. For example, it is not clear
- 15 what you mean by ‘the different sensitivities...’. You mention that compaction is sensitive to variability and ‘general increases’ in temperature and accumulation – can you be more explicit about the climate at the two sites, perhaps by including site-specific RACMO2 output in figure 7?
- We changed the starting sentence of the paragraph to relate it more closely to the topic of the study. We explicitly mention that we look at the effect of different models and different parameterisations on firn model output and
- 20 we removed the terms "different sensitivities" which were too vague (p.20 1.23):
"We further investigate how using different models and different parameterisations leads to discrepancies in the modelled compaction. We compute monthly values of compaction anomalies over the 2000–2017 period with the original and MAP models of HL, Ar and LZ (Fig. 7)."
- As suggested, we included the climatic anomalies at the two sites displayed in Figure 7. These mean anomalies
- 25 show both a warming and an increase in accumulation at the sites. We explain the computation of the mean climatic anomalies in the caption:
" Mean climatic anomalies are calculated as a difference between mean climatic values over the period 2000–2017 with respect to the reference period 1960–1979, and based on RACMO2 values."
We considered adding the entire time series of RACMO2 anomalies in accumulation and temperature but after
- 30 due reflection, we preferred not to do so to avoid overloading the figure. In total, the figure would have included the compaction anomaly time series, two climatic anomaly time series and three insets for each site. We believe that adding the mean climatic anomalies to the figure demonstrates that the original HL and the MAP_{HL} models are less sensitive to the general change in climatic conditions. Furthermore, these two models exhibit also much less seasonal variability, which shows that they are also less sensitive to changes in climatic conditions specific to
- 35 each month (e.g. less sensitive to summer months getting warmer).

p.13 1.14: short-scale -> short-timescale
Changed (p.20 1.31)

- 40 p.13 1.26: ‘at most sites...uncertainty intervals do not cover observed DIP values’ – this is an important result but I did not see it stated/quantified anywhere in the main text

We agree that this important result needs to be mentioned in the text and not only in the conclusion. We have added a couple of sentences at the end of the third paragraph of Section 3 (p.16 l.31):

5 " However, at a majority of the evaluation sites, the 95% credible intervals computed for the three models do not include the observed value (Fig. 4). This highlights that the governing equations of the models, which intend to capture densification physics, require improvement, and that parameter calibration in itself cannot overcome this shortcoming. "

p.14 l.12-13: this link takes you to a folder which contains several files that are unrelated to this study, are you able to list a source that just links to the firm data?

10 For the purpose of clarity, we have removed this link and replaced it by the statement (p.21 l.26):

"41 of the 91 firm cores are from the dataset compiled by Matt Spencer 10 (Spencer et al., 2001), which is available upon request."

15 It is important to note that this dataset is only a compilation of firm cores and the rights are not owned by Matt Spencer himself. The dataset is often used in the literature, including in The Cryosphere, but no consistent link is provided for it.

For examples:

<https://doi.org/10.5194/tc-13-845-2019>

<https://doi.org/10.1029/2017JF004597>

20 Tables 1 and S2: some terminology issues for large/small values, e.g. $9 \cdot 10^4$ should be 9×10^4

We have added the multiplicative symbol everywhere necessary in Table 1 and Table S2.

Table 2: explicitly mention RMSE in the caption rather than just 'The errors'

25 Changed: " The Root Mean Squared Errors (RMSE) are calculated with respect to the observations of depth integrated porosity until 15 m depth and until pore close-off."

Figure 1: what is the difference between a circle and a cross?

30 Sorry, we forgot to update the full legend when updating the figure after the first review. The crosses are the evaluation sites and the circles are the calibration sites. Figure 1 has been updated.

Figure 2: in the box titled 'If i is multiple of 100' the second Σ should be Σ_{cov}

Yes, Figure 2 has been updated.

35 Figure 3: please document what the 'posterior samples' are. Do they represent parameters associated with the 500 parameter sets randomly selected from the ensemble of accepted θ ?

We added in the caption: " The posterior samples are 500 randomly selected parameter combinations from the posterior ensembles of each model (HL, Ar, LZ)."

40 Figure 6: mention the difference between the left and right columns in the caption. Also, is it possible to represent AIS and GrIS data points differently, to support statements in main text?

The difference between graphs in the left and right column is now explicitly mentioned in the caption:

"Graphs in the left column display the mean annual temperature on the x-axis and those in the right column display the mean annual accumulation rate."

We have added black contours to the GrIS sites in order to distinguish between AIS and GrIS sites. We hope that this provides better support to the statements in the main text. For example, it shows clearly that the better performance of the original HL with respect to MAP_{HL} is related to the GrIS sites being concentrated in a narrow window of annual mean temperature values.

Figure 7: legend is missing

The legend has been added, sorry about forgetting it in the previous version.

10

Supp Info section S2: the GrIS RMSE value for surface mass balance flux is quoted as 69 mm w.e. in Noël et al. (2019), not 69 m w.e. – check that the units have been correctly converted when applying random noise to the boundary conditions

This was a typo in the text. The value used in the random noise application was indeed 69 mm w.e. The text has been corrected (p.34 l.19 and p.35 l.5).

15

Supp Info section S2: equation S5 contains the term cn , should it be cp ?

Yes, this has been corrected.

20 Supp Info section S2: ‘...must not be iteration specific...’ – needs clarification

We changed the text to clarify our approach (p.35 l.20):

"In contrast to the climatic perturbation, the perturbation in ρ_0 can be specific to each single time step t , and the perturbation thus varies throughout the duration of a firn model simulation. Indeed, it is not unrealistic that a month with anomalously low fresh snow density is immediately followed by a month of anomalously high fresh snow density for example."

25

Supp Info section S2: please include a reference to justify the choice of 25 kg/m³ when defining the perturbation to the fresh snow density values

We have modified the text and included the reference to Reeh et al. (2005) (p.35 l.25):

"We determine surface density values at each site from the firn cores of our dataset, ρ_0^{core} , and we perturb these values based on a standard deviation of 25 kg m⁻³. This value goes in line with a typical window of local variability of 50 kg m⁻³ for ρ_0 (Reeh et al., 2005)."

The relevant sections in Reeh et al. (2005) are Section 4.2 and Figure 2. It should be noted that observations of fresh snow density variability in time but at a same location are scarce. However, it is this value that the perturbations in ρ_0 should represent. In contrast, observations of fresh snow density variability in space are more common. Therefore, we based our choice on the data of Reeh et al. (2005) that show the variability in ρ_0 at sites of similar average temperature. We believe that this variability better suits our interest than an ice sheet scale variability (e.g. Fausto et al., 2018).

35

40 Supp Info section S3: please clarify that ‘original values’ refers to parameter values from the original publications of the HL and Ar models

Changed (p.36 l.4):

"where HL and Ar subscripts denote the values from the original publications of the HL and Ar models, and the mv subscript denotes a modified value of the parameter."

Supp Info section S6: start of second-to-last sentence – clarify what ‘it’ refers to

5 We clarified the subject of the sentence (p.37 1.35):

"As a consequence, the normal approximation results in a slight overestimation of uncertainty and thus conservative estimates of uncertainty."

Supp Info section S7: second sentence should refer to figure S5

10 Thank you, this has been modified (p.38 1.4).

Supp Info, Table S1: Please clarify whether accumulation and temperature values are taken from original publications or RACMO2

We added in the caption: "Values for both \dot{b} and T are computed from the RACMO2 model."

15

We also let you know that a "References" section has been added at the end of the Supplementary Information, with the four references used. Three of these references are also cited in the main manuscript.

Bayesian calibration of firn densification models

Vincent Verjans¹, Amber A. Leeson¹, Christopher Nemeth², C. Max Stevens³, Peter Kuipers Munneke⁴, Brice Noël⁴ and Jan Melchior van Wessem⁴

¹ Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YW, UK.

5 ² Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK

³ Department of Earth and Space Sciences, University of Washington, Seattle, WA, USA

⁴ Institute for Marine and Atmospheric research Utrecht, Utrecht University, Utrecht, the Netherlands

Correspondence to: Vincent Verjans (v.verjans@lancaster.ac.uk)

10 **Abstract.**

Firn densification modelling is key to understanding ice sheet mass balance, ice sheet surface elevation change, and the age difference between ice and the air in enclosed air bubbles. This has resulted in the development of many firn models, all relying to a certain degree on parameter calibration against observed data. We present a novel Bayesian calibration method for these parameters, and apply it to three existing firn models. Using an extensive dataset of firn cores from Greenland and Antarctica, we reach optimal parameter estimates applicable to both ice sheets. We then use these to simulate firn density and evaluate against independent observations. Our simulations show a significant decrease (24 and 56%) in observation-model discrepancy for two models and a smaller increase (15%) for the third. As opposed to current methods, the Bayesian framework allows for robust uncertainty analysis related to parameter values. Based on our results, we review some inherent model assumptions and demonstrate how firn model choice and uncertainties in parameter values cause spread in key model outputs.

20 **1 Introduction**

On the Antarctic and Greenland ice sheets (AIS and GrIS), snow falling at the surface progressively compacts into ice, passing through an intermediary stage called firn. The process of firn densification depends on local conditions, primarily the temperature, the melt rate and the snow accumulation rate, and accurate modelling of densification is key to several applications in glaciology. Firstly, variability in firn densification affects altimetry measurements of ice sheet surface elevation changes. Consequently, ~~it is a large contributor of uncertainty~~ uncertainties in modelled densification rates have a direct impact ~~o~~ in mass balance estimates, which ~~that~~ rely on a correct conversion from measured volume changes to mass changes (Li and Zwally, 2011; McMillan et al., 2016; Shepherd et al., 2019). Errors in the firn related correction can lead to over- or underestimation of mass changes related to surface processes, and to misinterpreting elevation change signals as changes in mass balance and in ice flow dynamics. Secondly, firn models are used to estimate the partitioning of surface meltwater into runoff off the ice sheet, and refreezing within the firn column, which strongly influences mass loss rates (van den Broeke et al., 2016). Model estimates of current and future surface mass balance of the AIS and GrIS are thus dependent on ~~would thus benefit from an improved knowledge of the sensitivity of the densification process to climatic conditions~~ accurate models of

[firn evolution](#). And finally, the densification rate determines the firn age at which air bubbles are trapped in the ice matrix. Knowing this age is crucial for precisely linking samples of past atmospheric composition, which are preserved in these bubbles, to paleo-temperature indicators, which come from the water isotopes in the ice (Buizert et al., 2014).

- 5 Firn densification has been the subject of numerous modelling studies over the last decades (e.g. Herron and Langway, 1980; Goujon et al., 2003; Helsen et al., 2008; Arthern et al., 2010; Ligtenberg et al., 2011; Simonsen et al., 2013; Morris and Wingham, 2014; Kuipers Munneke et al., 2015). However, there is no consensus on the precise formulation that such models should use. Most models adopt a two-stage densification process with the first stage characterising faster densification for firn with density less than a critical value, and then slower densification in the second stage. The firn-model intercomparison of
- 10 Lundin et al. (2017) demonstrated that, even for idealised simulations, inter-model disagreements are large in both stages. Firn compaction is driven by the pressure exerted by the overlying firn layers. Dry firn densification depends on numerous microphysical mechanisms acting at the scale of individual grains, such as grain-boundary sliding, vapour transport, dislocation creep and lattice diffusion (Maeno and Ebinuma, 1983; Alley, 1987; Wilkinson, 1988). Deriving formulations closely describing the densification of firn at the macroscale as a function of these mechanisms is challenging. Consequently, most
- 15 models rely on simplified governing formulations that are calibrated to agree with observations. The final model formulations have usually been tuned to data either from [the AIS](#) (Helsen et al., 2008; Arthern et al., 2010; Ligtenberg et al., 2011) or from [the GrIS](#) (Simonsen et al., 2013; Morris and Wingham, 2014; Kuipers Munneke et al., 2015), consisting of drilled firn cores from which depth-density profiles are measured. However, the calibration of firn densification rates to firn depth-density profiles requires the assumption of a firn layer in steady state. To overcome this limitation, some models have been calibrated
- 20 against other type of data such as strain rate measurements (Arthern et al., 2010; Morris and Wingham, 2014) or annual layering detected by radar reflection (Simonsen et al., 2013), but such measurements remain scarce and do not extend to firn at great depths below the surface. Ultimately, firn model calibration is an inverse problem that relies on using observational data to infer parameter values.
- 25 In this study, we adopt a Bayesian approach in order to address firn model calibration. This provides a rigorous mathematical framework for estimating distributions of the model parameters (Aster et al., 2005; Berliner et al., 2008). Bayesian inversion has been applied in several glaciological studies, and it has been demonstrated that this methodology improves our ability to constrain poorly known factors such as basal topography (Gudmundsson, 2006; Raymond and Gudmundsson, 2009; Brinkerhoff et al., 2016a), basal friction coefficients (Gudmundsson, 2006; Berliner et al., 2008; Raymond and Gudmundsson,
- 30 2009), ice viscosity (Berliner et al., 2008) and the role of the subglacial hydrology systems on ice dynamics (Brinkerhoff et al., 2016b). In the Bayesian framework, model parameters are considered as random variables for which we seek an *a posteriori* probability distribution that captures the probability density over the entire parameter space. This distribution allows not only to identify the most likely parameter combination, but also allows us to set confidence limits on the range of values in each parameter that is statistically reasonable. This enables us to quantify uncertainty in model results, to challenge the assumptions

inherent to the model itself and to assess correlation between different parameters. Calculations rely on Bayes' theorem (see Sect. 2.4 and Eq. (7)), but because of the high-dimensional parameter space and the non-linearity of firm models, solutions cannot be computed in closed form. As such, we apply rigorously designed Monte Carlo methods to approximate the target probability distributions efficiently. By exploiting the complementarity between the Bayesian framework and Monte Carlo techniques, we recalibrate three benchmark firm models and improve our understanding of their associated uncertainty.

2 Data and Methods

2.1 Firm densification data

In order to calibrate three firm densification models, we use observations of firm depth-density profiles from 91 firm cores (see Data Availability and Supplementary Information) located in different climatic conditions on both the GrIS (27 cores) and the AIS (64 cores) (Fig. 1). Using cores from both ice sheets is important since we seek parameter sets that are generally-applicable and not location-specific. We only consider dry densification since meltwater refreezing is poorly represented in firm models and wet-firm compaction is absent (Verjans et al., 2019). As such, we select cores from areas with low mean annual melt (<0.006 m w.e. yr⁻¹) but spanning a broad range of annual average temperatures (-55 to -20°C) and accumulation rates (0.02 to 1.06 m w.e. yr⁻¹). For each core, we use the depth-integrated porosity (*DIP*), also called firm air content. We calculate *DIP* until 15 m depth (*DIP*15, Eq. (1)). For sufficiently deep measurements, we also calculate *DIP**pc*, Eq. (2), taken below 15 m and until pore close-off depth (z_{pc} , where a density of 830 kg m⁻³ is reached). These are the observed quantitative values used for the calibration:

$$DIP15 = \int_0^{15} \frac{\rho_i - \rho}{\rho_i} dz \quad (1)$$

$$DIPpc = \int_{15}^{z_{pc}} \frac{\rho_i - \rho}{\rho_i} dz \quad (2)$$

where z (m) increases downwards, ρ is the density of firm (kg m⁻³) and ρ_i is the density of ice (917 kg m⁻³). In Eq. (2), we consider porosity only below 15 m to avoid dependency between *DIP*15 and *DIP**pc*. We choose to use both *DIP*15 and *DIP**pc* in order to account for first- and second-stage densification. One of the cores has only a single density measurement above 15 m depth and thus its *DIP*15 value is discarded. We note that 48 cores are too shallow to reach z_{pc} and so cores which do reach this depth provide a stronger constraint to the Bayesian inference method. This is sensible because these deep cores carry information about both stages of the densification process.

We use *DIP* as the evaluation metric for the models because of the crucial role of this variable in both surface mass balance modelling and altimetry-based ice sheet mass balance assessments (Ligtenberg et al., 2014). We note that it is commonly used in firm model intercomparison exercises (Lundin et al., 2017; Stevens et al., 2020) and is a quantity of interest for field measurements (Vandecrux et al., 2018). Due to its formulation (Eq. (1) and (2)), *DIP* represents the mean depth-density profile and thus is robust to the presence of individual errors and outliers in density measurements.

~~Observed firn density can be prone to measurement uncertainty, which previous studies point out is about 10%, though it is variable in depth and between measurement techniques employed (Hawley et al., 2008; Conger and McClung, 2009; Proksch et al., 2016). We outline our procedure to account for measurement uncertainty in Sect. 2.4.~~

We separate the dataset into calibration data (69 cores) and independent evaluation data (22 cores). The latter is selected semi-randomly; we ensure that it includes a representative ratio of GrIS-AIS cores and that it covers all climatic conditions, including an outlier of the dataset with high accumulation and temperature (see Supplementary Information). The resulting evaluation data has 8 GrIS and 14 AIS cores; 11 of the 22 cores extend to z_{pc} .

~~Observed firn density can be prone to measurement uncertainty, which previous studies point out is about 10%, though it is variable in depth and between measurement techniques employed (Hawley et al., 2008; Conger and McClung, 2009; Proksch et al., 2016). We outline our procedure to account for measurement uncertainty in Sect. 2.4.~~

2.2 Climate model forcing

At the location of each core, we simulate firn densification under climatic forcing provided by the RACMO2.3p2 regional climate model (RACMO2 hereafter) at 5.5 km horizontal resolution for the GrIS (Noël et al., 2019) and 27 km for the AIS (van Wessem et al., 2018). Each firn model simulation consists of a spin-up by repeating a reference climate until reaching a firn column in equilibrium, which is followed by a transient period until the core-specific date of drilling. The reference climate is taken as the first 20-year period of RACMO2 forcing data (1960-1979 and 1979-1998 for the GrIS and AIS respectively). The number of iterations over the reference period depends on the site-specific accumulation rate and mass of the firn column (mass from surface down to z_{pc}). We ensure that the entire firn column is refreshed during the spin-up but fix the minimum and maximum number of iterations to 10 (200 years spin-up) and 50 (1000 years spin-up). We note that at 33 sites, the core was drilled before the last year of the reference climate and so the transient period is effectively a partial iteration of the spin-up period.

Results of the calibration would depend on the particular climate model used for forcing. We thus propagate uncertainty in modelled climatic conditions into our calibration of firn model parameters by perturbing the temperature and accumulation rates of RACMO2 with normally distributed random noise. Standard deviations of the random perturbations are based on reported errors of RACMO2 (Noël et al., 2019; van Wessem et al., 2018 – see more details in the Supplementary Information). By introducing these perturbations, uncertainty intervals on our parameter values encompass the range of values that would result from using other model-based or observational climatic input.

In addition to the climatic forcing, another surface boundary condition is the fresh snow density, ρ_0 . ~~It is taken as a constant site-specific value. Each~~ At each site, the ρ_0 value is taken in agreement with the shallow densities measured in the corresponding core of the dataset. However, measurements of fresh snow are highly variable (e.g. Fausto et al., 2018). We account for uncertainty in this parameter by adding normally distributed random noise with standard deviation 25 kg m^{-3} to ρ_0 at every model time step (see Supplementary Information). We prefer this approach to the use of available parameterisations

of ρ_0 (Helsen et al., 2008; Kuipers Munneke et al., 2015) to avoid any error in the fresh snow parameterisation to affect the calibration process. ~~Fresh snow density is a poorly constrained boundary condition (e.g. Fausto et al., 2018). We account for uncertainty in this parameter by adding normally distributed random noise with standard deviation 25 kg m^{-3} to ρ_0 at every model time step.~~

5

2.3 Firm densification models

We use the Community Firm Model (Stevens et al., 2020) as the framework of our study because it incorporates the formulations of all three densification models investigated: HL (Herron and Langway, 1980), Ar (Arthern et al., 2010) and LZ (Li and Zwally, 2011). The Robin hypothesis (Robin, 1958) constitutes the fundamental assumption of HL, Ar and LZ. It states that any fractional decrease of the firm porosity, $\frac{\rho_i - \rho}{\rho_i}$, is proportional to an increment in overburden stress. This translates into densification rates depending on a rate coefficient c , assumed different for stage-1 and stage-2 densification:

10

$$\begin{cases} \frac{d\rho}{dt} = c_0 (\rho_i - \rho), & \rho \leq 550 \text{ kg m}^{-3} \\ \frac{d\rho}{dt} = c_1 (\rho_i - \rho), & \rho > 550 \text{ kg m}^{-3} \end{cases} \quad (3)$$

The formulations of the rate coefficients rely on calibration and thus differ between the three models investigated:

15 HL

$$\begin{cases} c_0 = \dot{b}^a k_0^* \exp\left(\frac{-E_0}{RT}\right) \\ c_1 = \dot{b}^b k_1^* \exp\left(\frac{-E_1}{RT}\right) \end{cases} \quad (4)$$

Ar

$$\begin{cases} c_0 = \rho_w \dot{b}^\alpha k_0^{Ar} g \exp\left(\frac{-E_c}{RT} + \frac{E_g}{RT_{av}}\right) \\ c_1 = \rho_w \dot{b}^\beta k_1^{Ar} g \exp\left(\frac{-E_c}{RT} + \frac{E_g}{RT_{av}}\right) \end{cases} \quad (5)$$

LZ

$$\begin{cases} c_0 = \beta_0 l z_a (273.15 - T)^{l z_b} \dot{b} \\ c_1 = \beta_1 l z_a (273.15 - T)^{l z_b} \dot{b} \end{cases} \quad (6)$$

20

$$\text{with } \begin{cases} \beta_0 = l z_{11} + l z_{12} \dot{b} + l z_{13} T_{av} \\ \beta_1 = \beta_0 (l z_{21} + l z_{22} \dot{b} + l z_{23} T_{av})^{-1} \end{cases}$$

where \dot{b} is the accumulation rate (m w.e. yr^{-1}), T the temperature (K), T_{av} the annual mean temperature, R the gas constant, g gravity and ρ_w the water density (1000 kg m^{-3}). All remaining terms are model-specific tuning parameters. For \dot{b} , we use the mean accumulation rate over the lifetime of each specific firm layer because it better approximates the overburden stress than

25

the annual mean (Li and Zwally, 2011). HL and Ar use Arrhenius relationships with activation energies (E terms) capturing temperature sensitivity and exponents characterising the exponential proportionality of the rate coefficients to the accumulation rate. Originally, Herron and Langway (1980) inferred all values from calibration based on 17 firm cores, from which they inferred the values for the six free parameters (Table 1) of HL. In contrast, Arthern et al. (2010) fixed the accumulation exponents in advance ($\alpha = \beta = 1$) and took activation energies (E_c, E_g) from measurements of microscale mechanisms: Nabarro-Herring creep for E_c and grain-growth for E_g . Still, they noted a mismatch with the activation energy fitting their data best. The k_0^{Ar} and k_1^{Ar} parameters were tuned to three measured time series of strain rates collected in relatively warm and high accumulation locations of the AIS. Here, we consider all five $\alpha, \beta, k_0^{Ar}, k_1^{Ar}, E_g$ as free parameters (Table 1) but keep E_c fixed because of its strong correlation with E_g ; our use of monthly model time steps and depth-density profiles as calibration data is not suitable for differentiating effects of $\frac{E_g}{RT_{av}}$ and $\frac{E_c}{RT}$. Equation (6) shows that LZ has eight free parameters (Table 1), all denoted by lz in this paper. In contrast to our approach to Ar, we do not add additional accumulation rate exponents to \dot{b} in Eq. (6) because the dependence of c_0 and c_1 on \dot{b} also involves the coefficients lz_{12} and lz_{22} in the definition of β_0 and β_1 . Li and Zwally (2011) performed their calibration of Eq. (6) against firm cores only from the GrIS. Later, Li and Zwally (2015) proposed a new parameterisation for β_0 and β_1 , but developed a densification model calibrated for Antarctic firm. The latter model uses the same governing equations as LZ for c_0 and c_1 but different formulations for β_0 and β_1 (Eq. (6)). Since one of the goals of this study is to find a densification formulation applicable to firm in both the GrIS and AIS, we choose to apply our calibration method only to the formulations of β_0 and β_1 Eq. (6) specified in Li and Zwally (2011). (Eq. (6)). However, in our results' analysis (Sect. 3), we also consider the performance of the Li and Zwally (2015) model on the AIS cores of our dataset.

20

2.4 Bayesian calibration

In our approach, the free parameters of the firm models are identified as the quantities of interest and we define this parameter set as θ . Hereafter, 'original model values' refers to the values originally attributed by Herron and Langway (1980), Arthern et al. (2010) and Li and Zwally (2011) to their respective sets of free parameters θ . The calibration process relies on Bayes' theorem (Eq. (7)) which allows to update a prior probability distribution $P(\theta)$ for θ based on observed data Y .

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)} \quad (7)$$

We use normal and weakly informative priors centred about the original model values so that the constraint of the prior on $P(\theta|Y)$ is minor (Table 1). As indicated by Morris and Wingham (2014), in HL and Ar, the values of the Arrhenius pre-exponential factors (k_0^*, k_1^*, k_0^{Ar} and k_1^{Ar}) are correlated with their corresponding activation energies (E_0, E_1 and E_g). At a given temperature, a change of the value in the pre-exponential factor can be compensated by adjusting the activation energy to keep the densification rates constant. We express our *a priori* knowledge of these correlations in the prior distributions (see

Supplementary Information). No other pair of parameters in HL, Ar or LZ are clearly correlated *a priori*, but the calibration process captures *a posteriori* correlations by confronting the models ~~to~~ with data. The data Y consists of the observed $DIP15$ and $DIPpc$ values of the calibration data. The marginal likelihood, $P(Y)$, is a constant term independent of θ and does not influence the calibration. We use a normal likelihood function $P(Y|\theta)$, which quantifies the match of the modelled DIP values with the observed:

$$P(Y|\theta) \propto \exp \left[\frac{-1}{2} (X_{15} - Y_{15})^T \Sigma_{15}^{-1} (X_{15} - Y_{15}) - \frac{1}{2} (X_{pc} - Y_{pc})^T \Sigma_{pc}^{-1} (X_{pc} - Y_{pc}) \right] \quad (8)$$

where X_{15} and Y_{15} are vectors containing all modelled and observed values for the calibration data of $DIP15$ respectively, and similarly for X_{pc} and Y_{pc} . We use diagonal covariance matrices Σ_{15} and Σ_{pc} with site-specific variances. The variances determine the spread allowed for the model outputs compared to the observed values and are calculated by taking 10% and 20% margins around $DIP15$ and $DIPpc$ measurements respectively. Allowing for such spread is necessary because multiple causes may lead to model-observation discrepancy such as firm model errors, measurement uncertainties and discrepancies induced by the random perturbations applied to RACMO2 forcing and to ρ_0 . This particular form of the likelihood function assumes independence between model errors in $DIP15$ and in $DIPpc$, which is ensured by our calculation of $DIPpc$ only from 15 m depth to z_{pc} (Eq. (2)). It also assumes normally distributed model errors with respect to the observed values. Both these aspects were verified with preliminary assessments, along with our calculations for the covariance matrices Σ_{15} and Σ_{pc} , as discussed in the Supplementary Information. The posterior distribution $P(\theta|Y)$ gives a probability distribution over the parameter space of a given model conditioned on the calibration data. In our case, with weakly informative priors (Table 1), the distribution $P(\theta|Y)$ is essentially governed by the likelihood function (Eq. (8)). We note here that extreme parameter combinations in the LZ model can lead to negative densification rates. In such cases, we set the modelled DIP values to 0, which leads to extremely low values for the likelihood and for the posterior probability of such parameter sets.

There is no analytical form of $P(\theta|Y)$ and we must investigate the parameter space to generate an ensemble of θ_i approximating $P(\theta|Y)$. Such an investigation is achieved efficiently using Markov Chain Monte Carlo methods. We apply the well-known Random Walk Metropolis (RWM) algorithm (Hastings, 1970) and summarize it in Fig. 2, on which we base the brief following description. A given model (HL, Ar or LZ) starts with ~~a certain parameter set~~ θ_i —the original model parameter values and simulates firm profiles at all the calibration sites. Its $DIP15$ and $DIPpc$ results are compared with observations and the general performance of the model ~~using θ_i is~~ is quantified by the likelihood. From there and with the prior distributions assumed, the posterior ~~probability is~~ probability $P(\theta_i|Y)$ is computed following Eq. (7). At this point, the RWM algorithm starts and the state of the chain ~~is~~ θ_i (Fig. 2a), is set to the original model values and its posterior probability is saved as $P(\theta_i|Y)$. It should be noted that the i subscript designates the iteration number, which is equal to 0 at this initial step. The RWM then proposes a new θ_i^* from a proposal distribution (Fig. 2b). For the latter, we use the symmetric multivariate normal (MVN) distribution which is centred about θ_i . This implies that the random choice of θ_i^* depends only on the current state θ_i and on the proposal

covariance in the MVN, Σ_{prop} , which is discussed below. Using the parameter combination θ_i^* , the model simulates profiles at all calibration sites again (Fig. 2c) and $P(\theta_i^*|Y)$ is computed (Fig. 2d). From there, we either accept or reject the proposed θ_i^* in the ensemble approximating $P(\theta|Y)$. By using the previously computed $P(\theta_i|Y)$, the probability of accepting θ_i^* depends on the ratio $P(\theta_i^*|Y)/P(\theta_i|Y)$ (Fig. 2e). The set saved in the ensemble (Fig. 2g) is θ_i^* if accepted or θ_i if θ_i^* was rejected. The saved set becomes the updated current status for the next iteration θ_{i+1} (Fig. 2a) with its associated posterior probability, $P(\theta_{i+1}|Y)$, and the The algorithm iterates this process and reaches a final posterior distribution over θ . The RWM has the property that the chain will ultimately converge to a stationary distribution that represents the posterior $P(\theta|Y)$. Thus, after a sufficiently high number of iterations of the algorithm, the ensemble of parameter sets is representative of $P(\theta|Y)$. We verify adequate convergence using a number of tests, which are shown in the Supplementary Information. The proposal

10 covariance Σ_{prop} must account for dependence between the different components of θ , i.e. the value of one free parameter can influence the value of another free parameter for the model to reach a good match with the observed data. Σ_{prop} can capture this dependence between parameters and, for optimality, it is updated every given number of iterations (100 in our study) using Eq. (9) (Rosenthal, 2010):

$$\Sigma_{prop} = \frac{2.38^2}{p} \Sigma_{cov} \quad (9)$$

15 where Σ_{cov} is the covariance matrix between the free parameters of the model at this stage of the iterative chain and p is the number of free parameters.

From the posterior probability distributions, we can infer the Maximum a Posteriori (MAP) estimates of each model (MAP_{HL}, MAP_{Ar}, MAP_{LZ}). These are the modes of the multi-dimensional distributions over the space of free parameters and have been identified as the most likely sets by the RWM. The MAP estimates can be compared to the corresponding original model values of the parameters. The posterior distributions additionally incorporate the uncertainty in the parameter values. By performing posterior predictive simulations on the evaluation data, we can assess this remaining uncertainty (Gelman et al., 2013). More specifically, we can assume that a large (500) random sample of the ensemble of accepted θ is representative of the posterior distribution. As such, model results computed with all sets of this sample inform about model performance

25 accounting for uncertainty. Intuitively, a large spread in results from a-500 random samples would indicate a large range of possible sets for the free parameters and thus a high uncertainty in parameter values.

Since there is no analytical form of our posterior distributions, and to facilitate future firm model uncertainty assessments, we can approximate the posterior distributions with MVN distributions whose means and covariances are set to the posterior means and posterior covariance matrices of the calibration. This allows straightforward sampling of random parameter sets

30 instead of relying on posterior samples of the MCMC. We provide information about the normal approximations and assess their validity in the Supplementary Information. Such normal approximations are asymptotically exact and are commonly applied to analytically intractable Bayesian posterior distributions (Gelman et al., 2013).

3 Results

We present the results of the calibration process after 15000 algorithm iterations and compare the MAP and original models' performances against the 22 evaluation cores. We also evaluate the uncertainty of the posterior distributions and compare performances between the different MAP models. All the evaluation simulations are performed without climatic and surface density noise in order to make the evaluation fully deterministic.

For HL and even more so for Ar, the posterior distributions for the parameters demonstrate some strong disagreements with the original values (Figs. 3a, 3b). The 95% credible intervals for each parameter (Table 1) incorporate 95% of the marginal probability density in the posterior. Two original parameter values of HL (a, b) and three of Ar (E_g, α, β) lie in the tails of the posterior distributions (Figs. 3a, 3b) and even outside these intervals in the case of b, E_g, α and β . This indicates that our analysis provides strong evidence against these original values. The strongest disagreements relate to the accumulation exponents of both models (a, b, α, β). In contrast, the original LZ values agree better with the posterior distribution and all lie within the 95% credible intervals (Table 1 and Fig. 3c). The posterior distributions show some strong correlation between certain pairs of parameters (Fig. 3). Notable examples are the pre-exponential factors and their corresponding activation energy in HL and Ar, for which the posterior correlations are even stronger than in the prior distributions. The complete correlation matrices and a detailed analysis of all posterior correlation features are provided in the Supplementary Information.

We use the original models and the MAP estimates to simulate firn profiles at the evaluation sites and we compare *DIP* results with the observed values. This is an effective way to assess possible improvements in parameter estimates reached through our method since the evaluation sites were not used in the calibration process. The match between observations and the model is improved for MAP_{HL} (Fig. 4a) and even more for MAP_{Ar} (Fig. 4b), with the original Ar strongly underestimating *DIP* values. These improvements translate into significantly reduced root mean squared errors (RMSE) in modelled values of both *DIP*15 (-24% for HL and -45% for Ar) and *DIPpc* (-22% and -61%) (Table 2).

For LZ, the relative performance of the MAP_{LZ} model for both *DIP*15 and *DIPpc* is worse (+2% and +24% in RMSE) but differences are of smaller magnitude (Table 2 and Fig. 4c). Parameter values of MAP_{LZ} and the original LZ are closer, which explains more moderate differences in RMSE compared to HL and Ar. Comparing modelled and observed depth-density profiles of evaluation data illustrates the differences in performance visually (e.g. Fig. 5). Profiles of the original models of HL and Ar frequently lie outside the credible intervals of their respective MAP models. In contrast, profiles of MAP_{LZ} and of the original LZ tend to be close together. At the climatic outlier of our evaluation data (DML in Fig. 5), improvements are reached for the three MAP models (Figs. 5e5g, 5f5h, 5i). This demonstrates benefits of this method even at the limits of the calibration range. However, at a majority of the evaluation sites, the 95% credible intervals computed for the three models do not include the observed value (Fig. 4). This highlights that the governing equations of the models, which intend to capture densification physics, require improvement, and that parameter calibration in itself cannot overcome this shortcoming.

Compared to the original HL, MAP_{HL} reaches improvements in $DIP15$ for 12 of the 22 evaluation cores and in $DIPpc$ for 5 of the 11 evaluation cores (Fig. 6a). Generally, MAP_{HL} performs better at AIS sites and worse at GrIS sites. An analysis of the improvement of MAP_{HL} as a function of climatic variables (Fig. 6a) shows that the original HL gives better results in a narrow range of T_{av} : from -30 to -25 °C. As such, the better performance at the GrIS evaluation sites of the original HL is likely due to the original HL its parameterisation being better suited for the particular temperature range corresponding to the conditions of the latter sites. In contrast, MAP_{HL} seems more appropriate for covering a wider range of climatic conditions. For Ar, the original model shows better performance than MAP_{Ar} at few evaluation sites (6 for $DIP15$ and 2 for $DIPpc$) which are only in AIS and confined to low-accumulation conditions (Fig. 6b). This is counterintuitive given that Arthern et al. (2010) tuned the original Ar to measurements from high accumulation sites of the AIS. Finally, the original LZ performs better than MAP_{LZ} at most GrIS sites (Fig. 6c), which is unsurprising given that its original calibration was GrIS-specific. Again, this seems related to the original LZ performing significantly better in the same narrow range of temperatures as for HL. In total, MAP_{LZ} performs better for 10 of the 22 $DIP15$ and 4 of the 11 $DIPpc$ evaluation measurements.

As explained in Sect. 2.3, the original LZ model was developed for GrIS firn only (Li and Zwally, 2011) and later complemented by an AIS-specific model (Li and Zwally, 2015). We compute results at the AIS and GrIS evaluation sites using the Li and Zwally (2015) model for the AIS and the Li and Zwally (2011) model for the GrIS, so that both models are applied to the ice sheet for which they were originally developed. We call this pairing of models LZ dual and evaluate its general performance ~~Using both of these on the evaluation sites of their respective calibration ice sheet, we construct an LZ dual model, which thus really consists of two different models.~~ The RMSE for $DIP15$ of LZ dual is slightly larger (+8 %) than that of MAP_{LZ} and significantly larger (+38 %) for $DIPpc$ (Table 2). We note that the higher RMSE values of LZ dual are strongly affected by its densification scheme performing very poorly at the climatic outlier of the evaluation data, with conditions that are outside of the calibration range of Li and Zwally (2015).

We also compare MAP results with the IMAU firn densification model (IMAU-FDM), which has been used frequently in recent mass balance assessments from altimetry (Pritchard et al., 2012; Babonis et al., 2016; McMillan et al., 2016; Shepherd et al., 2019). IMAU-FDM was developed by adding two tuning parameters to both densification stages of Ar. All four extra-parameters are different for the AIS (Ligtenberg et al., 2011) and for the GrIS (Kuipers Munneke et al., 2015), thus also resulting in two separate models. On the evaluation data, its performance for $DIP15$ is slightly better than MAP_{Ar} and MAP_{LZ} but worse than MAP_{HL} , and its performance for $DIPpc$ is significantly worse than all three MAP models (Table 2).

To assess the uncertainty captured by the Bayesian posterior distributions, we compute results on the evaluation data with the 500 parameter sets randomly selected from each of the three posterior ensembles. For all three models, the average performance of their random sample is similar to the corresponding MAP performance, with a maximum RMSE change of 6% (Table 2).

This demonstrates a low uncertainty in the optimal parameter combinations identified by calibration. Furthermore, the best performing 95th percentile of the random selection allows the construction of the uncertainty intervals shown in Figs 4, 5. Of the original models, LZ reaches the lowest RMSE values. Of all models, MAP_{HL} performs best in *DIP15* and MAP_{Ar} in *DIPpc* (Table 2). MAP_{LZ} performs worse than the other MAP models even when accounting for uncertainty by using the 500-samples random selections (Table 2).

4 Discussion

This calibration method is potentially applicable to models of similar complexity in a broad range of research fields. We exploit it here to investigate the parameter space of HL, Ar and LZ, and to re-estimate optimal parameter values conditioned on observed calibration data; no further complexity is introduced since the number of empirical parameters remains the same. We treat the accumulation exponents of Ar (α, β) as free parameters whereas Arthern et al. (2010) decided to fix their values to 1. Analogous to a and b in HL, these exponents capture the mathematical relationship between densification rates and the accumulation rate, used as a proxy for load increase on any specific firm layer. No physical argument favours a linear proportionality between densification and load increase and any prescribed value for these exponents is a choice of the model designer. Unlike Arthern et al. (2010), Herron and Langway (1980) previously inferred $a = 1$ and $b = 0.5$. Our calibration data shows strong evidence against both these pairs of values; all four are in the extreme tails of the posterior distributions (Fig. 3a, 3b). Our results of stage-1 exponents (a, α) smaller than 1 indicate a weaker increase in densification rates with pressure than assumed in the original versions of Ar and HL. In firm, the load is supported at the contact area between the grains, which increases on average due to grain rearrangement (in stage-1) and grain growth. As such, firm strengthens in time and the actual stress on ice grains increases slower than the total load (Anderson and Benson, 1963). Morris and Wingham (2014) incorporated this by including a temperature-history function, causing slower densification of firm previously exposed to higher temperatures. This is consistent with both grain rearrangement and grain growth because these processes are enhanced at higher temperatures (Alley, 1987; Gow et al., 2004). Lower values of the stage-2 exponents (b, β) illustrate the larger strength of high-density firm with larger contact areas between grains. The same can be applied to difference in sensitivities of stage-1 and stage-2 densification to accumulation also holds in the LZ model, as illustrated by ~~by investigating~~ the posterior correlation between its free parameters. ~~It shows~~ The a positive correlation coefficient ~~(0.74)~~ between the accumulation-related parameters of both stages; lz_{12} and lz_{22} , is significantly positive (0.74, Fig. S5). High values of lz_{12} make β_0 more sensitive to \dot{b} (Eq. (6)). However, β_0 appears in the numerator of the β_1 calculation (Eq. (6)) and higher values of lz_{22} thus moderate the sensitivity of stage-2 densification to \dot{b} . As such, positively correlated lz_{12} and lz_{22} provide further evidence that stage-1 densification rates are more sensitive to accumulation rates. This example demonstrates how posterior correlations provide insights into model behaviour. The posterior correlations of all three models are further discussed in the Supplementary Information.

In the IMAU model introduced in Sect. 3, tuning parameters have been added to Ar in order to reduce its sensitivity to accumulation rates (Ligtenberg et al., 2011; Kuipers Munneke et al., 2015). The calibration method presented in this study detects and adjusts for this over-sensitivity in Ar without the need for more tuning parameters in the governing densification equations. The sensitivity of stage-1 densification to \dot{b} can be computed from the derivative of the rate coefficient:

$$5 \quad \frac{\partial c_0}{\partial \dot{b}} = \rho_w k_0^{Ar} g \exp\left(\frac{-E_c}{RT} + \frac{E_g}{RT_{av}}\right) \alpha \dot{b}^{\alpha-1} \quad (10)$$

Similarly, the derivative $\frac{\partial c_1}{\partial \dot{b}}$ is obtained by replacing k_0^{Ar} and α with k_1^{Ar} and β . Our calibration process strongly favours smaller values of α , β and E_g with respect to the original values (Fig. 3b). We can compare the magnitudes of the derivatives under the original Ar parameterisation and under the MAP parameterisation. The magnitudes vary for particular combinations of T_{av} and \dot{b} . Under all the annual mean climatic regimes of our dataset, the MAP parameters result in a decreased sensitivity
10 of both stage-1 and stage-2 densification rates to \dot{b} .

HL, Ar and LZ only use temperature and accumulation rates as input variables. Other models use additional variables hypothesised to affect densification rates. These include the temperature-history mentioned above (Morris and Wingham, 2014), firm grain size (Arthern et al., 2010), impurity content (Freitag et al., 2013) and a transition region between stage-1 and
15 stage-2 densification (Morris, 2018). Other models are explicitly based on micro-scale deformation mechanisms (Alley, 1987; Arthern and Wingham, 1998; Arnaud et al., 2000). These efforts undoubtedly contribute to progressing towards physically based models. A potential problem with such approaches is overfitting calibration data by adding parameters to model formulations while detailed firm data remain scarce. As long as more firm data is not available to appropriately constrain the role of each variable in model formulations, we favour the use of parsimonious models relying on few input variables. It is
20 noteworthy that MAP_{LZ} , which relies on eight free parameters, performs worse on the evaluation data than MAP_{HL} and MAP_{Ar} with two fewer free parameters. This highlights that gains in model accuracy should rely not only on better calibration of parameters but also on a reconsideration of the governing densification equations. Additionally, firm core data invokes the assumption of a steady-state depth-density profile. As such, parameter calibration poorly captures seasonal climatic effects on densification. Comprehensive datasets of depth-density profiles (Koenig and Montgomery, 2019) are very valuable to model
25 development. Efforts in collecting and publishing strain rate measurements from the field (Hawley and Waddington, 2011; Medley et al., 2015; Morris et al., 2017), and possibly from laboratory experiments (Schleef and Löwe, 2013), can further benefit model calibration and the progress towards more representative equations.

In order to quantify the consequences of our calibration, we investigate two aspects for which firm models are of common use:
30 calculating firm compaction rates and predicting the age of firm at z_{pc} depth, age_{pc} (yr). At every site i of our dataset, we compute the 2000-2017 total compaction anomaly, $cmp_{an,i}$ (m), and the $age_{pc,i}$ value with each of the 500 parameter sets randomly drawn from the posterior ensembles of the three different models (HL, Ar , LZ). This allows evaluation of both

parameter-related and model-related uncertainty. Total compaction anomaly (cmp_{an}) – calculated as the cumulative anomaly in surface elevation change due only to firn compaction changes during the 2000-2017 period with respect to the climatic reference period – is given by:

$$cmp_{an,i} = cmp_{tot,i}^{00-17} - 17cmp_{ref,i}^{yr} \quad (11)$$

5 where cmp_{tot}^{00-17} (m) is the total firn compaction over 2000-2017 and cmp_{ref}^{yr} (m yr⁻¹) is the annual mean compaction over the reference period (see Sect. 2.2). At all sites, we compute the coefficients of variation (CV) for both cmp_{an} and age_{pc} from the 500 simulations with each model, and we average the CVs across all sites. CV is the ratio of the standard deviation to the mean and provides an effective assessment of relative dispersion of model results. Because low mean values of cmp_{an} can inflate its CV, we consider only half of the sites at which the mean computed cmp_{an} is highest. For all three models, the CV values for both cmp_{an} and age_{pc} lie between 5.5 and 7.5% (Table 3). These values give typical uncertainty in firn model output related to uncertain parameter values. Proceeding to the same calculations but using all three models, i.e. an inter-model ensemble of 1500 simulations at each site, gives an overview of the combined parameter- and model-related uncertainty. The CVs are 19.5% for cmp_{an} and 7.5% for age_{pc} , demonstrating larger inter-model disagreement on cmp_{an} calculations (Table 3). By using the CV values, we can calculate reasonable uncertainty estimates for cmp_{an} and age_{pc} . For instance, in the dry
 10 snow zone of GrIS, simulated compaction anomalies are typically around 20 cm over 2000-2017, and thus come with an uncertainty of the order of ± 4 cm. Since pore close-off age here is around 250 years, a reasonable uncertainty range on this value is ± 19 years. In contrast, on the drier AIS, pore close-off age is about 1000 years thus this range increases to ± 75 years. Compaction anomalies hover around 0 cm on most of the dry zone of the AIS because it has not experienced the strong recent surface warming of the GrIS. Absolute uncertainty is thus reduced but still critical given the large area of the AIS over which
 15 it must be integrated uncertainties are aggregated when mass balance trends are evaluated. Such numbers The uncertainty ranges calculated from the CV values provide an order of magnitude of errors in firn model outputs that must be accounted for in altimetry-based mass balance assessments and in ice core studies, respectively.

We further investigate how using different models and different parameterisations leads to discrepancies in the modelled compaction, the different sensitivities in terms of We compute monthly values of compaction anomalies over the 2000-2017
 25 period ~~with~~ the original and MAP models of HL, Ar and LZ (Fig. 7). Ar shows the strongest sensitivity to climatic conditions diverging from these of the reference period; compaction responds strongly to the general increases on GrIS in temperature and accumulation rate, especially in late summer. Due to its lower values for α, β and E_g , MAP_{Ar} exhibits less extreme compaction anomalies than the original Ar and thus less seasonal variability. In sharp contrast to Ar, HL-computed compaction rates remain relatively stable, due to low activation energy values that smooth out the seasonal variability. Firn core
 30 observations provide little information and constraints on seasonal patterns of densification. However, it is noteworthy that MAP_{Ar} and MAP_{LZ} tend to show comparable short-time scale sensitivities (insets in Fig. 7), despite structural differences in the models' governing equations. This might indicate that these models fare relatively well in capturing seasonal fluctuations of densification rates and their sensitivity to climate shifts.

5 Conclusion

We have implemented a Bayesian calibration method to estimate optimal parameter combinations applicable to GrIS and AIS firm for three benchmark firm densification models (HL, Ar, LZ). An extensive dataset of 91 firm cores was separated into calibration and independent evaluation data. Two optimised models (MAP_{HL} , MAP_{Ar}) showed significant improvement against the evaluation data, while MAP_{LZ} reached results close to, but slightly worse, than its original version and inferior to MAP_{HL} and MAP_{Ar} . When compared to other models of greater complexity, the MAP models showed comparable or even improved performances. Furthermore, the Bayesian approach provides a robust way to evaluate the uncertainty related to parameter value choice, which is a major deficiency of current models. By introducing realistic climatic perturbations in the calibration process, the uncertainty intervals obtained account for the effects of an uncertain climatic forcing. However, at most sites where we evaluated, all three models' uncertainty intervals do not cover observed *DIP* values. As such, although model results can be improved by re-calibration methods, model tuning alone is insufficient to reach exact fidelity of firm densification models. The formulation of models' governing equations impacts the remaining errors with respect to observations, which highlights deficiencies in our understanding of dry firm densification. Developing a well-constrained physically detailed model is challenging given the number of mechanisms affecting densification rates and their dependency on microstructural properties of firm, which are difficult to observe. Our study demonstrates that, despite these observational limitations, thorough calibration methods relying only on climatic variables can substantially improve firm model accuracy, and constrain uncertainties.

Author contributions. VV, AL and CN conceived this study. VV performed the development of the calibration method, performed the model experiments and led writing of the manuscript. AL and CN supervised the work. MS developed the Community Firm Model. PKM provided firm core data. BN and JMVW provided the RACMO2 forcing data. All authors provided comments and suggested edits to the manuscript.

Data availability. 41 of the 91 firm cores are from the SUMup dataset (2019 release), which is publicly available from the Arctic Data Center (doi: 10.18739/A26D5PB2S). 41 of the 91 firm cores are from the dataset compiled by Matt Spencer (Spencer et al., 2001), which is ~~available upon request~~ publicly available via the NASA Global Change Master Directory as 'LSSU_PSU_Firm_data' (https://drive.google.com/drive/folders/0B_IQfVYZbeWFMDe4M2ZjOTEtNWVhOS00NjdmLTlxMjctYWZlNmZkMDg2OGFh?hl=en_US). 5 of the 91 firm cores were provided by Joe McConnell and Ellen Mosley-Thompson and are available on request through PKM. 2 of the 91 cores are available via the PANGAEA website (<https://doi.pangaea.de/10.1594/PANGAEA.227732>) and (<https://doi.pangaea.de/10.1594/PANGAEA.615238>). 1 of the 91 cores is available via the NOAA website (<ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/newall/>). 1 of the 91 cores is available via the USAP website (<http://www.usap-dc.org/view/dataset/609215>). All Antarctic RACMO2.3p2 climate data

used are available on request through JMW and yearly climate variables are free to download from the IMAU website (<https://www.projects.science.uu.nl/iceclimate/publications/data/2018/index.php>). All Greenland RACMO2.3p2 climate data used are available on request through BN and yearly SMB and components are free to download (<https://doi.pangaea.de/10.1594/PANGAEA.904428>). The Community Firn Model is available for download on GitHub (<https://github.com/UWGlaciology/CommunityFirnModel>).

Acknowledgements. We thank Lora Koenig and Lynn Montgomery for making the SUMup dataset of firn cores available and easily accessible (Koenig and Montgomery, 2019). Matt Spencer is also acknowledged for publishing a separate dataset of firn cores (Spencer et al., 2001). We thank Joe McConnell and Ellen Mosley-Thompson, supported by the NSF-NASA PARCA Project, for providing additional firn core data (Bales et al., 2009; Banta & McConnell, 2007; McConnell et al. 2002; McConnell et al., 2000; Mosley-Thompson et al., 2001). We thank Malcolm McMillan for his interest in the study and for providing insight into the subject of ice sheet mass balance assessments. VV thanks Elizabeth Morris for pointing out errors in geographical coordinates of some of the firn cores and for her endless interest in firn densification. The Centre for Polar Observation and Modelling is acknowledged for supporting VV in his research. AL and CN research is supported by EPSRC, *A Data Science for the Natural Environment*, EP/R01860X/1. PKM acknowledges support from NESSC (Netherlands Earth System Science Centre). BN was funded by the NWO (Netherlands Organisation for Scientific Research) VENI grant VI.Veni.192.019. We thank all contributors to the development of the CFM who are not authors of this study. All authors thank two anonymous referees for their time and effort in reviewing the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

References

- Alley, R. B.: Firn Densification By Grain-Boundary Sliding : a First Model, *Le J. Phys. Colloq.*, 48(C1), C1-249-C1-256, doi:10.1051/jphyscol:1987135, 1987.
- Anderson, D. L. and Benson, C. S.: The densification and diagenesis of snow, in W. D. Kingery, ed. *Ice and snow: properties, processes, and applications*, pp. 391–411, MIT Press, Cambridge, MA., 1963.
- 5 Arnaud, L., Gay, M., Barnola, J.-M. J. M. and Duval, P.: Physical modeling of the densification of snow / firn and ice in the upper part of polar ice sheets, in *Physics of Ice Core Records*, edited by T. Hondoh, pp. 285–305, Hokkaido University Press, Sapporo, Japan., 2000.
- Arthern, R. J. and Wingham, D. J.: The Natural Fluctuations of Firn Densification and Their Effect on the Geodetic Determination of Ice Sheet Mass Balance, *Clim. Change*, 40(3–4), 605–624, doi:10.1023/A:1005320713306, 1998.
- 10 Arthern, R. J., Vaughan, D. G., Rankin, A. M., Mulvaney, R. and Thomas, E. R.: In situ measurements of Antarctic snow compaction compared with predictions of models, *J. Geophys. Res. Earth Surf.*, 115, 1–12, doi:10.1029/2009JF001306, 2010.
- Aster, R. C., Borchers, B. and Clifford, H. T.: *Parameter estimation and inverse problems*, Elsevier, Amsterdam., 2005.
- 15 Babonis, G. S., Csatho, B. and Schenk, T.: Mass balance changes and ice dynamics of Greenland and Antarctic ice sheets from laser altimetry, *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 41(July), 481–487, doi:10.5194/isprsarchives-XLI-B8-481-2016, 2016.
- Berliner, L. M., Jezek, K., Cressie, N., Kim, Y., Lam, C. Q. and Van Der Veen, C. J.: Modeling dynamic controls on ice streams: A Bayesian statistical approach, *J. Glaciol.*, 54(187), 705–714, 2008.
- 20 Brinkerhoff, D. J., Aschwanden, A. and Truffer, M.: Bayesian Inference of Subglacial Topography Using Mass Conservation, *Front. Earth Sci.*, 4(February), 1–15, doi:10.3389/feart.2016.00008, 2016a.
- Brinkerhoff, D. J., Meyer, C. R., Bueler, E., Truffer, M. and Bartholomaus, T. C.: Inversion of a glacier hydrology model, *Ann. Glaciol.*, 57(72), 84–95, doi:10.1017/aog.2016.3, 2016b.
- van den Broeke, M. R., Enderlin, E. M., Howat, I. M., Kuipers Munneke, P., Noël, B. P. Y., van de Berg, W. J., van Meijgaard, E. and Wouters, B.: On the recent contribution of the Greenland ice sheet to sea level change, *The Cryosphere*, 10, 1933–1946, doi:10.5194/tc-10-1933-2016, 2016.
- 25 Buizert, C., Gkinis, V., Severinghaus, J. P., He, F., Lecavalier, B. S., Kindler, P., Leuenberger, M., Carlson, A. E., Vinther, B., Masson-Delmotte, V., White, J. W. C., Liu, Z., Otto-Bliesner, B. and Brook, E. J.: Greenland temperature response to climate forcing during the last deglaciation, *Science (80-.)*, 345(6201), 1177–1180, doi:10.1126/science.1254961, 2014.
- 30 Conger, S. M. and McClung, D.: Instruments and Methods: Comparison of density cutters for snow profile observations, *J. Glaciol.*, 55, 163–169, doi:10.3189/002214309788609038, 2009.
- Fausto, R. S., Box, J. E., Vandecrux, B., van As, D., Steffen, K., MacFerrin, M., Machguth H., Colgan W., Koenig L. S., McGrath D., Charalampidis, C., and Braithwaite, R. J.: A Snow Density Dataset for Improving Surface Boundary Conditions in Greenland Ice Sheet Firn Modeling, *Front. Earth Sci.*, 6, 51 pp., doi:10.3389/feart.2018.00051, 2018.
- 35 Freitag, J., Kipfstuhl, S., Laepple, T. and Wilhelms, F.: Impurity-controlled densification: a new model for stratified polar firn, *J. Glaciol.*, 59(218), 1163–1169, doi:10.3189/2013jog13j042, 2013.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D.: *Bayesian Data Analysis*, Third Edit., CRC Press Taylor & Francis Group, Boca Raton., 2013.
- 40 Goujon, C., Barnola, J.-M. and Ritz, C.: Modeling the densification of polar firn including heat diffusion: Application to close-off characteristics and gas isotopic fractionation for Antarctica and Greenland sites, *J. Geophys. Res. Atmos.*, 108, n/a-n/a, doi:10.1029/2002JD003319, 2003.
- Gow, A. J., Meese, D. A. and Bialas, R. W.: Accumulation variability, density profiles and crystal growth trends in ITASE firn and ice cores from West Antarctica, *Ann. Glaciol.*, 39, 101–109, doi:10.3189/172756404781814690, 2004.
- 45 Gudmundsson, G. H.: Estimating basal properties of glaciers from surface measurements, in Knight, P.G., ed. *Glacier science and environmental change*. Oxford, Blackwell, pp. 415–417., 2006.
- Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57(1), 97–109, 1970.
- Hawley, R., Brandt, O., Morris, E., Kohler, J., Shepherd, A. and Wingham, D.: Techniques for measuring high-resolution

- firn density profiles: Case study from Kongsvegen, Svalbard. *J. Glaciol.*, 54(186), 463–468, doi:10.3189/002214308785837020, 2008.
- Hawley, R. L. and Waddington, E. D.: Instruments and Methods in situ measurements of firn compaction profiles using borehole optical stratigraphy, *J. Glaciol.*, 57(202), 289–294, doi:10.3189/002214311796405889, 2011.
- 5 Helsen, M. M., van den Broeke, M. R., van de Wal, R. S. W., van de Berg, W. J., van Meijgaard, E., Davis, C. H., Li, Y. and Goodwin, I.: Elevation changes in antarctica mainly determined by accumulation variability, *Science* (80-.), 320, 1626–1629, doi:10.1126/science.1153894, 2008.
- Herron, M. and Langway, C.: Firn densification: an empirical model, *J. Glaciol.*, 25(93), 373–385, <https://doi.org/10.3189/S0022143000015239>, 1980.
- 10 Koenig, L. and Montgomery, L.: Surface mass balance and snow depth on sea ice working group (SUMup) snow density subdataset, Greenland and Antarctica, 1950–2018, , doi:10.18739/A26D5PB2S, 2019.
- Kuipers Munneke, P., Ligtenberg, S. R. M., Noël, B. P. Y., Howat, I. M., Box, J. E., Mosley-Thompson, E., McConnell, J. R., Steffen, K., Harper, J. T., Das, S. B. and van den Broeke, M. R.: Elevation change of the Greenland ice sheet due to surface mass balance and firn processes, 1960–2013, *The Cryosphere*, 9, 3541–3580, doi:10.5194/tcd-9-3541-2015, 2015.
- 15 Li, J. and Zwally, H. J.: Modeling of firn compaction for estimating ice-sheet mass change from observed ice-sheet elevation change, *Ann. Glaciol.*, 52, 1–7, doi:10.3189/172756411799096321, 2011.
- Li, J. and Zwally, H. J.: Response times of ice-sheet surface heights to changes in the rate of Antarctic firn compaction caused by accumulation and temperature variations, *J. Glaciol.*, 61, 1037–1047, doi:10.3189/2015JoG14J182, 2015.
- 20 Ligtenberg, S. R. M., Helsen, M. M. and Van Den Broeke, M. R.: An improved semi-empirical model for the densification of Antarctic firn, *The Cryosphere*, 5, 809–819, doi:10.5194/tc-5-809-2011, 2011.
- Ligtenberg, S. R. M., Kuipers Munneke, P., and van den Broeke, M. R.: Present and future variations in Antarctic firn air content, *The Cryosphere*, 8, 1711–1723, doi:10.5194/tc-8-1711-2014, 2014.
- Lundin, J. M. D., Stevens, C. M., Arthern, R., Buizert, C., Orsi, A., Ligtenberg, S. R. M., Simonsen, S. B., Cummings, E., 25 Essery, R., Leahy, W., Harris, P., Helsen, M. M. and Waddington, E. D.: Firn Model Intercomparison Experiment (FirnMICE), *J. Glaciol.*, 63, 401–422, doi:10.1017/jog.2016.114, 2017.
- Maeno, N. and Ebinuma, T.: Pressure sintering of ice and its implication to the densification of snow at polar glaciers and ice sheets, *J. Phys. Chem.*, 87(21), 4103–4110, doi:10.1021/j100244a023, 1983.
- McMillan, M., Leeson, A., Shepherd, A., Briggs, K., Armitage, T., Hogg, A., Kuipers Munneke, P., van den Broeke, M., 30 Noël, B., van de Berg, W. J., Ligtenberg, S., Horwath, M., Groh, A., Muir, A. and Gilbert, L.: A high-resolution record of Greenland mass balance, *Geophys. Res. Lett.*, 43, 7002–7010, doi:10.1002/2016GL069666, 2016.
- Medley, B., Ligtenberg, S. R. M., Joughin, I., Van Den Broeke, M. R., Gogineni, S. and Nowicki, S.: Antarctic firn compaction rates from repeat-track airborne radar data: I. Methods, *Ann. Glaciol.*, 56(70), 155–166, doi:10.3189/2015AoG70A203, 2015.
- 35 Morris, E. M.: Modeling Dry-Snow Densification without Abrupt Transition, *Geosciences*, 8(12), 464, doi:10.3390/geosciences8120464, 2018.
- Morris, E. M. and Wingham, D. J.: Densification of polar snow: Measurements, modeling, and implications for altimetry, *J. Geophys. Res. Earth Surf.*, 119, 349–365, doi:10.1002/2013JF002898, 2014.
- Morris, E. M., Mulvaney, R., Arthern, R. J., Davies, D., Gurney, R. J., Lambert, P., De Rydt, J., Smith, A. M., Tuckwell, R. 40 J. and Winstrup, M.: Snow Densification and Recent Accumulation Along the iSTAR Traverse, Pine Island Glacier, Antarctica, *J. Geophys. Res. Earth Surf.*, 122, 2284–2301, doi:10.1002/2017JF004357, 2017.
- Noël, B., van de Berg, W. J., van Wessem, J. M., van Meijgaard, E., van As, D., Lenaerts, J. T. M., Lhermitte, S., Kuipers Munneke, P., Smeets, C. J. P. P., van Ulf, L. H., van de Wal, R. S. W., and van den Broeke, M. R.: Modelling the climate and surface mass balance of polar ice sheets using RACMO2 –Part 1: Greenland (1958–2016), *The Cryosphere*, 12, 811–831, <https://doi.org/10.5194/tc-12-811-2018>, 2018.
- 45 Noël, B., van de Berg, W. J., Lhermitte, S. and van den Broeke, M. R.: Rapid ablation zone expansion amplifies north Greenland mass loss, *Sci. Adv.*, 5(9), eaaw0123, doi:10.1126/sciadv.aaw0123, 2019.
- Pritchard, H. D., Ligtenberg, S. R. M., Fricker, H. A., Vaughan, D. G., Van Den Broeke, M. R. and Padman, L.: Antarctic ice-sheet loss driven by basal melting of ice shelves, *Nature*, 484(7395), 502–505, doi:10.1038/nature10968, 2012.
- 50 Proksch, M., Rutter, N., Fierz, C., and Schneebeli, M.: Intercomparison of snow density measurements: bias, precision, and

vertical resolution, *The Cryosphere*, 10, 371–384, doi:10.5194/tc10-371-2016, 2016.

- Raymond, M. J. and Gudmundsson, G. H.: Estimating basal properties of ice streams from surface measurements: A non-linear Bayesian inverse approach applied to synthetic data, *The Cryosphere*, 3(2), 265–278, doi:10.5194/tc-3-265-2009, 2009.
- 5 Robin, G. d. Q.: *Glaciology III: Seismic shooting and related investigations*, in *Norwegian-British-Swedish Antarctic Expedition, 1949–52, Scientific Results*, vol. 5, Norsk Polarinstitt, Oslo., 1958.
- Rosenthal, J.: *Optimal Proposal Distributions and Adaptive MCMC*, (1), 1–25, doi:10.1201/b10905-5, 2010.
- Schleef, S. and Löwe, H.: X-ray microtomography analysis of isothermal densification of new snow under external mechanical stress, *J. Glaciol.*, 59(214), 233–243, doi:10.3189/2013JoG12J076, 2013.
- 10 Shepherd, A., Gilbert, L., Muir, A. S., Konrad, H., McMillan, M., Slater, T., Briggs, K., Sundal, V., Hogg, A. and Engdahl, E.: Trends in Antarctic Ice Sheet elevation and mass. *Geophysical Research Letters*, 46, 8174–8183, doi:10.1029/2019GL082182, 2019.
- Simonsen, S. B., Stenseng, L., Adalgeirsdóttir, G., Fausto, R. S., Hvidberg, C. S. and Lucas-Picher, P.: Assessing a multilayered dynamic firn-compaction model for Greenland with ASIRAS radar measurements, *J. Glaciol.*, 59, 545–558, doi:10.3189/2013JoG12J158, 2013.
- 15 Spencer, M. K., Alley, R. B. and Creyts, T. T.: Preliminary firn-densification model with 38-site dataset, *J. Glaciol.*, 47, 671–676, <https://doi.org/10.3189/172756501781831765>, 2001.
- Stevens, C. M., Verjans, V., Lundin, J. M. D., Kahle, E. C., Horlings, A. N., Horlings, B. I., and Waddington, E. D.: The Community Firn Model (CFM) v1.0, *Geosci. Model Dev. Discuss.*, <https://doi.org/10.5194/gmd-2019-361>, in review, 2020.
- 20 Verjans, V., Leeson, A. A., Max Stevens, C., MacFerrin, M., Noël, B. and Van Den Broeke, M. R.: Development of physically based liquid water schemes for Greenland firn-densification models, *The Cryosphere*, 13(7), 1819–1842, doi:10.5194/tc-13-1819-2019, 2019.
- van den Broeke, M. R., Enderlin, E. M., Howat, I. M., Kuipers Munneke, P., Noël, B. P. Y., van de Berg, W. J., van Meijgaard, E. and Wouters, B.: On the recent contribution of the Greenland ice sheet to sea level change, *The Cryosphere*, 10, 1933–1946, doi:10.5194/tc-10-1933-2016, 2016.
- 25 van Wessem, J. M., Jan Van De Berg, W., Noël, B. P. Y., Van Meijgaard, E., Amory, C., Birnbaum, G., Jakobs, C. L., Krüger, K., Lenaerts, J. T. M., Lhermitte, S., Ligtenberg, S. R. M., Medley, B., Reijmer, C. H., Van Tricht, K., Trusel, L. D., Van Uft, L. H., Wouters, B., Wuite, J. and Van Den Broeke, M. R.: Modelling the climate and surface mass balance of polar ice sheets using RACMO2 - Part 2: Antarctica (1979–2016), *The Cryosphere*, 12(4), 1479–1498, doi:10.5194/tc-12-1479-2018, 2018.
- 30 Vandecrux, B., MacFerrin, M., Machguth, H., Colgan, W. T., van As, D., Heilig, A., Stevens, C. M., Charalampidis, C., Fausto, R. S., Morris, E. M., Mosley-Thompson, E., Koenig, L., Montgomery, L. N., Miège, C., Simonsen, S. B., Ingeman-Nielsen, T., and Box, J. E.: Firn data compilation reveals widespread decrease of firn air content in western Greenland, *The Cryosphere*, 13, 845–859, doi:10.5194/tc-13-845-2019, 2019.
- 35 Wilkinson, D.: A pressure-sintering model for the densification of polar firn and glacier ice, *J. Glaciol.*, 34(116), 40–45, doi:10.3189/S0022143000009047, 1988.

40

| Parameter | Value in original model | Prior distribution | MAP | 95 % Credible interval |
|-----------------------------------|-------------------------|---------------------------------|---------|------------------------|
| k_0^* [m w.e. $^{-\alpha}$] | 11 | $N(11, 100)$ | 17.4 | 7.58; 28.4 |
| k_1^* [m w.e. $^{-b}$] | 575 | $N(575, 9 \times 9 \cdot 10^4)$ | 524 | 260; 1060 |
| E_0 [J mol $^{-1}$] | 10 160 | $N(10160, 4 \times 10^6)$ | 10 840 | 9 000; 12 290 |
| E_1 [J mol $^{-1}$] | 21 400 | $N(21400, 4 \times 10^6)$ | 20 800 | 18 900; 22 300 |
| a [/] | 1 | $N(1, 0.4)$ | 0.91 | 0.74; 1.02 |
| b [/] | 0.5 | $N(0.5, 0.4)$ | 0.63 | 0.54; 0.78 |
| k_0^{Ar} [m w.e. $^{-\alpha}$] | 0.07 | $N(0.07, 4.9 \times 10^{-3})$ | 0.077 | 0.046; 0.137 |
| k_1^{Ar} [m w.e. $^{-\beta}$] | 0.03 | $N(0.03, 9 \times 10^{-4})$ | 0.025 | 0.015; 0.048 |
| E_c [J mol $^{-1}$] | 60 000 | Fixed: 60000 | / | / |
| E_g [J mol $^{-1}$] | 42 400 | $N(42400, 16 \times 10^6)$ | 40 900 | 39 700; 42 000 |
| α [/] | 1 | $N(1, 0.4)$ | 0.80 | 0.66; 0.89 |
| β [/] | 1 | $N(1, 0.4)$ | 0.68 | 0.59; 0.81 |
| lz_a | 8.36 | $N(8.36, 36)$ | 7.31 | 3.93; 12.82 |
| lz_b | -2.061 | $N(-2.061, 2)$ | -2.124 | -2.319; -1.896 |
| lz_{11} | -9.788 | $N(-9.788, 36)$ | -14.710 | -20.839; -5.469 |
| lz_{12} | 8.996 | $N(8.996, 36)$ | 7.269 | 2.680; 17.724 |
| lz_{13} | -0.6165 | $N(-0.6165, 1)$ | -1.019 | -1.389; -0.509 |
| lz_{21} | -2.0178 | $N(-2.0178, 2)$ | -1.513 | -2.970; -0.258 |
| lz_{22} | 8.4043 | $N(8.4043, 36)$ | 6.0203 | 4.911; 12.942 |
| lz_{23} | -0.0932 | $N(-0.0932, 0.25)$ | -0.0913 | -0.133; -0.0460 |

Table 1. Information for the free parameters of HL (top), Ar (middle) and LZ (low). $N(x, y)$ designates a normal distribution of mean x and variance y . The variances in the prior distributions are taken to generate weakly informative distributions. Some prior correlation is prescribed for the pairs (k_0^*, E_0) , (k_1^*, E_1) , (k_0^{Ar}, E_g) , (k_1^{Ar}, E_g) and (k_0^{Ar}, k_1^{Ar}) (see Supplementary Information). MAP estimates and credible intervals are results from the calibration process.

5

| Model | RMSE (<i>DIP15</i>) [m] | RMSE (<i>DIPpc</i>) [m] |
|----------------------|---------------------------|---------------------------|
| HL original | 0.503 | 2.395 |
| HL MAP | 0.382 | 1.862 |
| HL 500 random sample | 0.396 | 1.899 |
| Ar original | 0.772 | 4.566 |
| Ar MAP | 0.426 | 1.780 |
| Ar 500 random sample | 0.448 | 1.889 |
| LZ original | 0.452 | 1.812 |
| LZ dual | 0.505 | 3.883 |
| LZ MAP | 0.463 | 2.392 |
| LZ 500 random sample | 0.486 | 2.296 |
| IMAU-FDM | 0.418 | 2.681 |

Table 2. Model results on the evaluation data. The ~~errors~~ Root Mean Squared Errors (RMSE) are calculated with respect to the observations of depth integrated porosity until 15 m depth and until pore close-off.

| Coefficient of Variation | HL | Ar | LZ | Combined (HL, Ar, LZ) |
|--------------------------|------|------|------|-----------------------|
| cmp_{an} | 5.8% | 5.8% | 6.5% | 19.5% |
| age_{pc} | 6.5% | 5.8% | 7.5% | 7.5% |

Table 3. Coefficients of variation for the 2000-2017 cumulative compaction anomaly (cmp_{an}) and firn age at pore close-off depth (age_{pc}).

- 5 Values are computed from results of 500 randomly selected parameter combinations from the posterior ensembles of each model (HL, Ar, LZ). Coefficients of variation are averaged across all sites of the dataset.

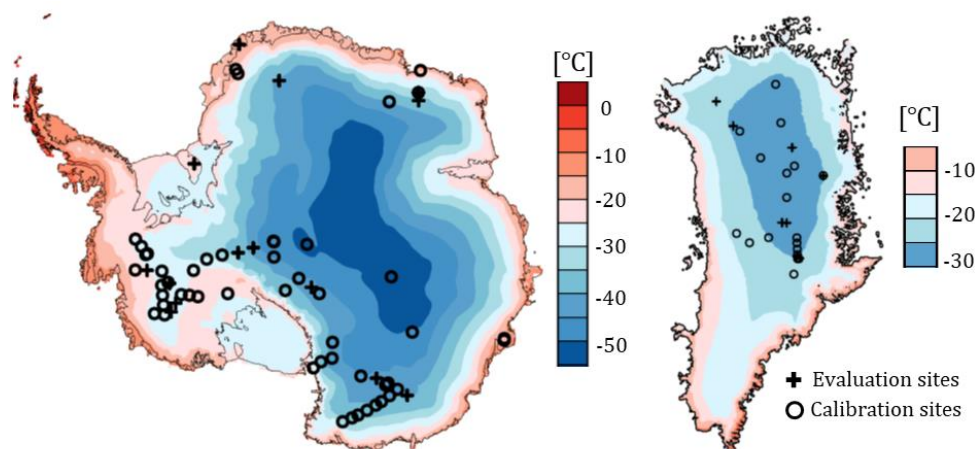


Figure 1. Maps of Antarctic (left) and Greenland (right) ice sheets. Background is mean annual air temperature as modelled by RACMO2.

- 10 Note the different colour scales.

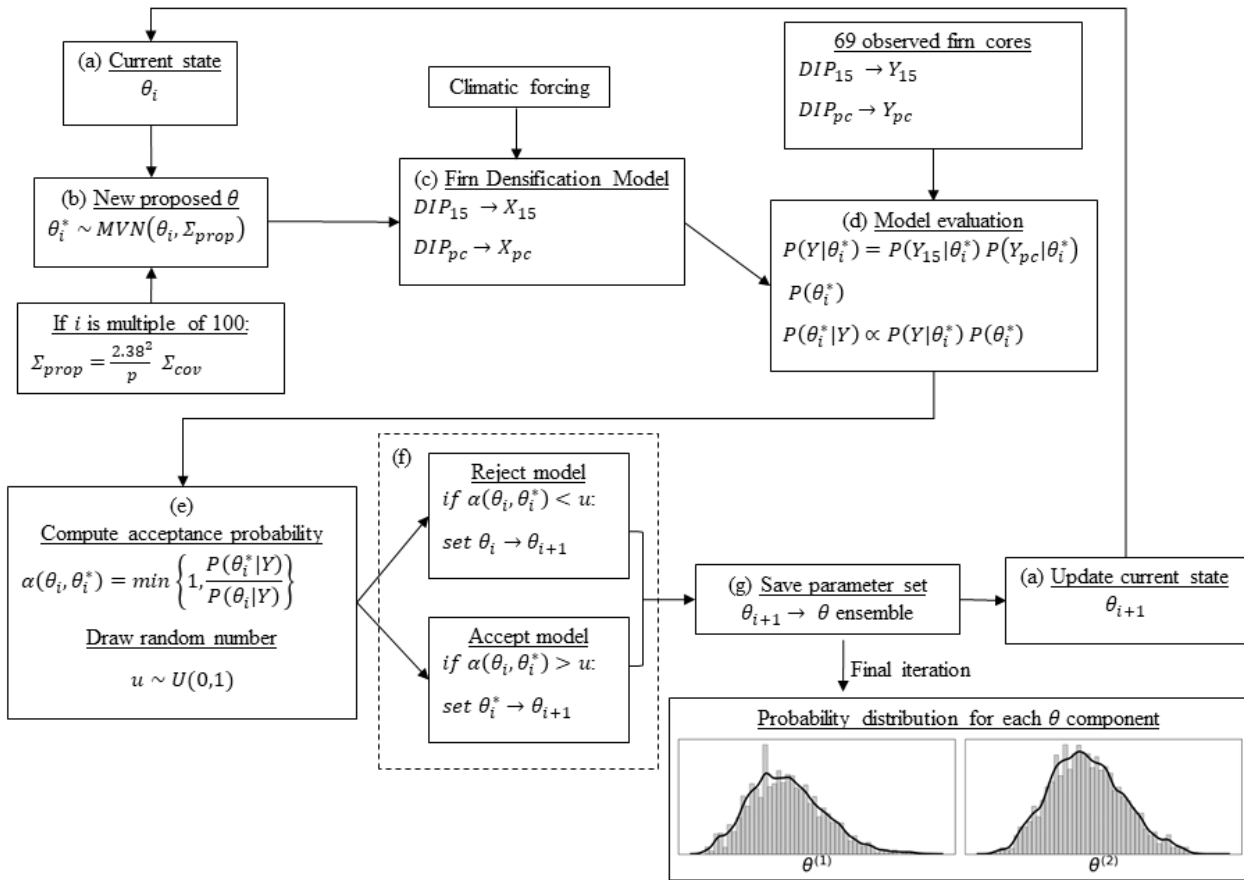


Figure 2. Implementation of the Random Walk Metropolis algorithm. θ represents a parameter combination of any given firm densification model investigated.

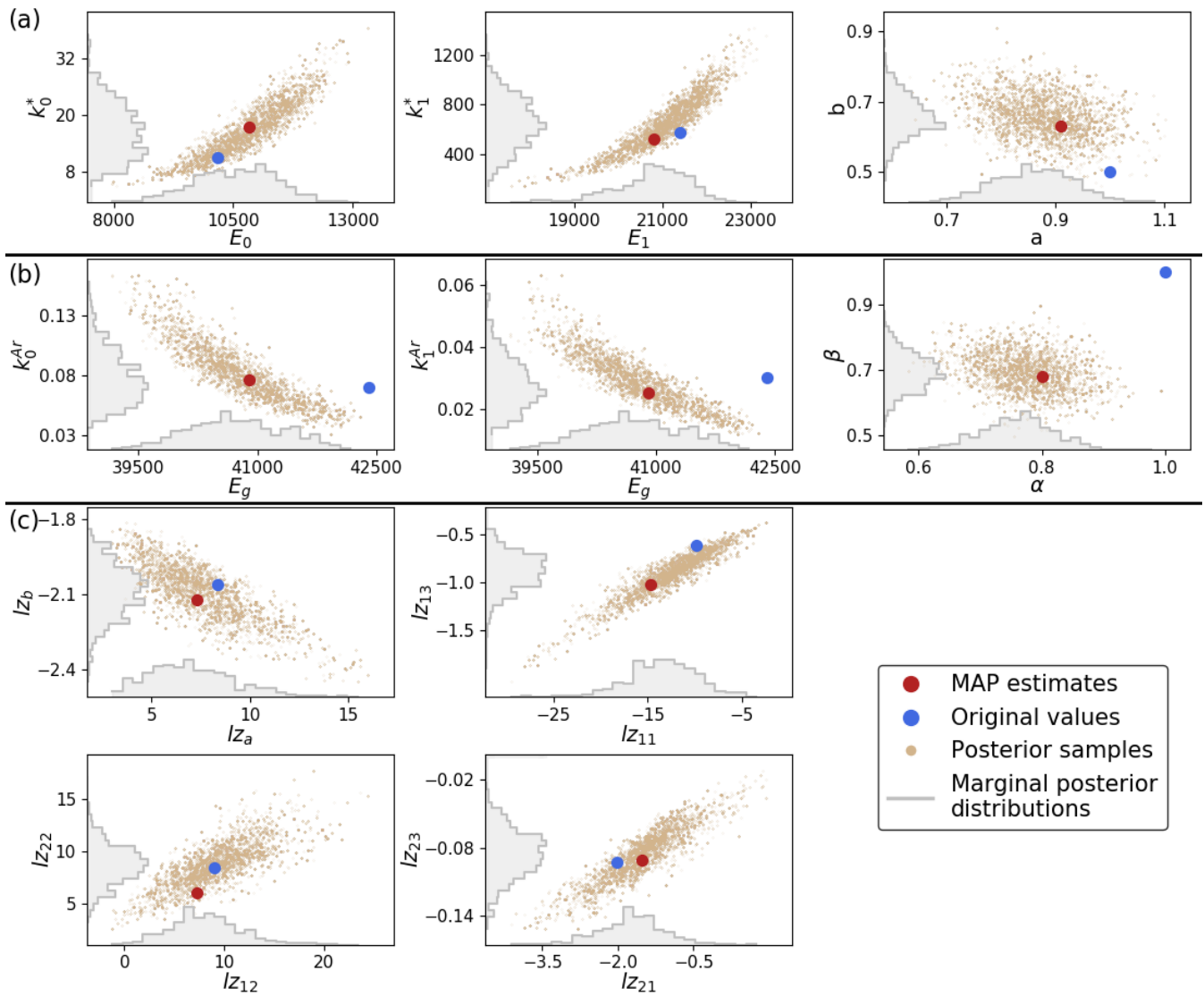


Figure 3. Posterior probability distributions, shown for pairs of parameters, for (a) HL, (b) Ar, (c) LZ. Where possible, correlated parameters share the same graph (see Supplementary Information for full correlation matrices). The posterior samples are 500 randomly selected parameter combinations from the posterior ensembles of each model (HL, Ar, LZ).

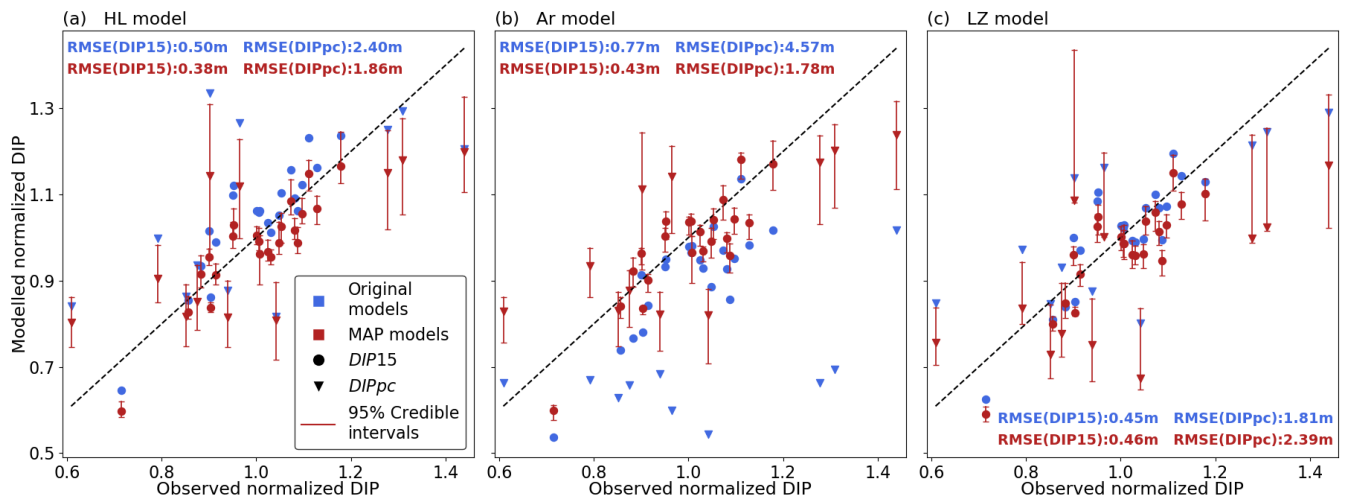


Figure 4. Comparison of evaluation data *DIP* with model results. The 95% credible intervals are computed from results of 500 randomly selected parameter combinations from the posterior ensembles of each model (HL, Ar, LZ). Similar scatter plots for the LZ dual and IMAU results are shown in the Supplementary Information (Fig. S6).

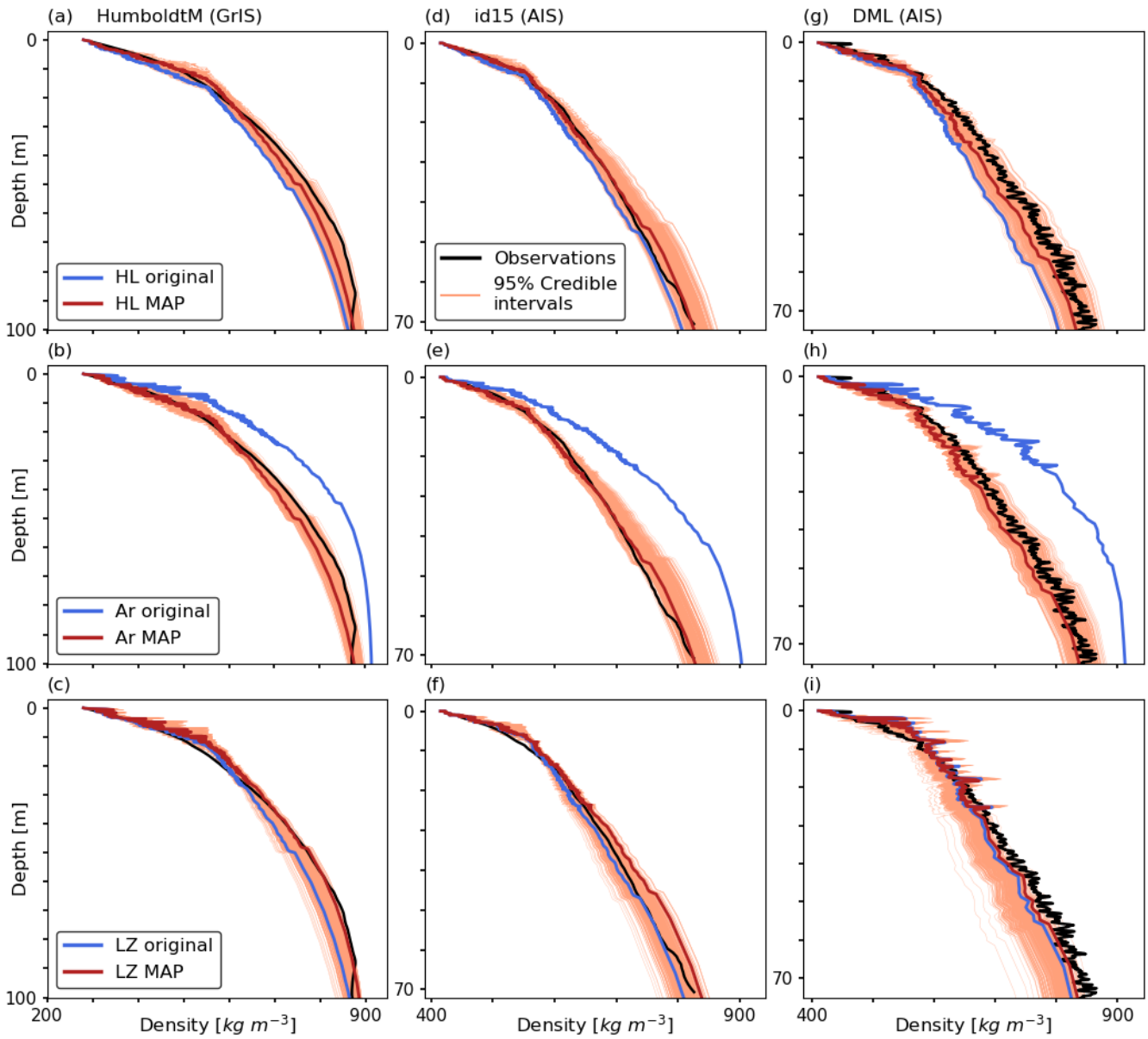


Figure 5. Depth-density profiles at three evaluation sites. DML is a climatic outlier of our dataset with particularly high temperatures and accumulation rates. The 95% credible intervals are computed from results of 500 randomly selected parameter combinations from the posterior ensembles of each model (HL, Ar, LZ).

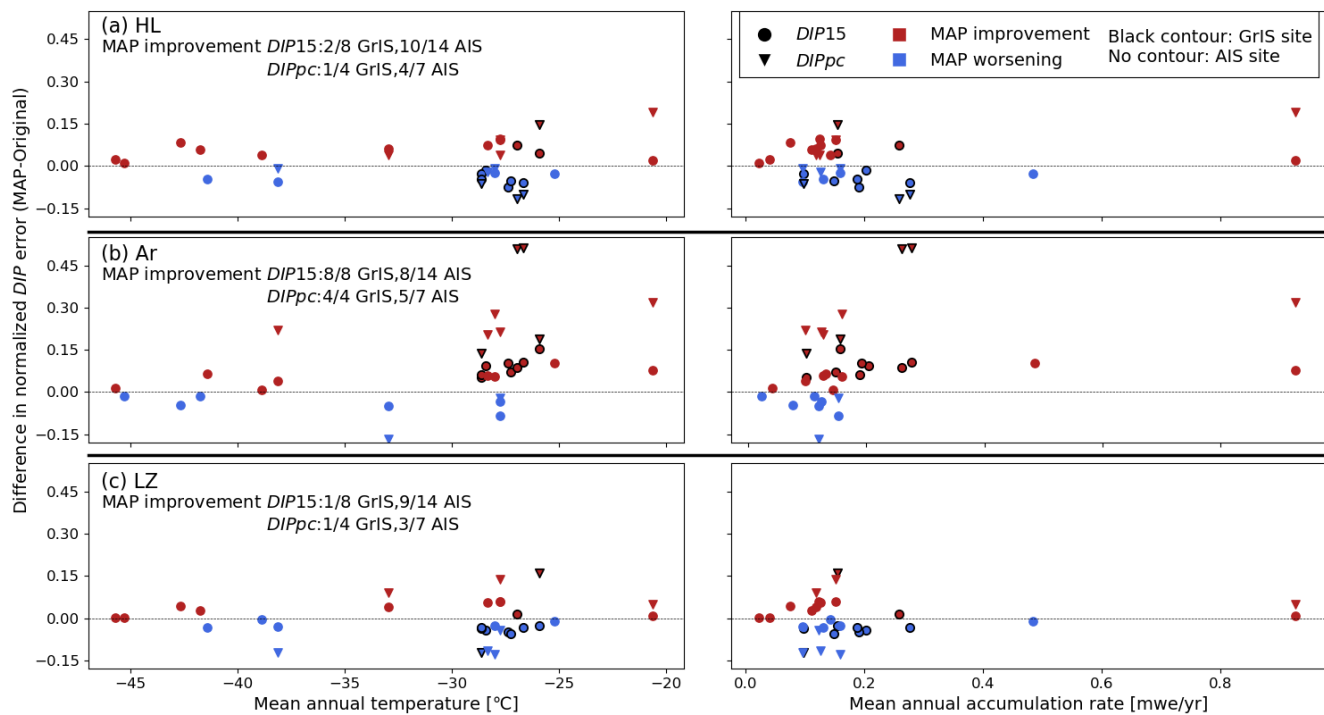


Figure 6. Improvements of the MAP models with respect to the original models for the evaluation data. The ratios indicate the ratios of cores for which an improvement is achieved by the corresponding MAP. Graphs in the left column display the mean annual temperature on the x-axis and those in the right column display the mean annual accumulation rate.

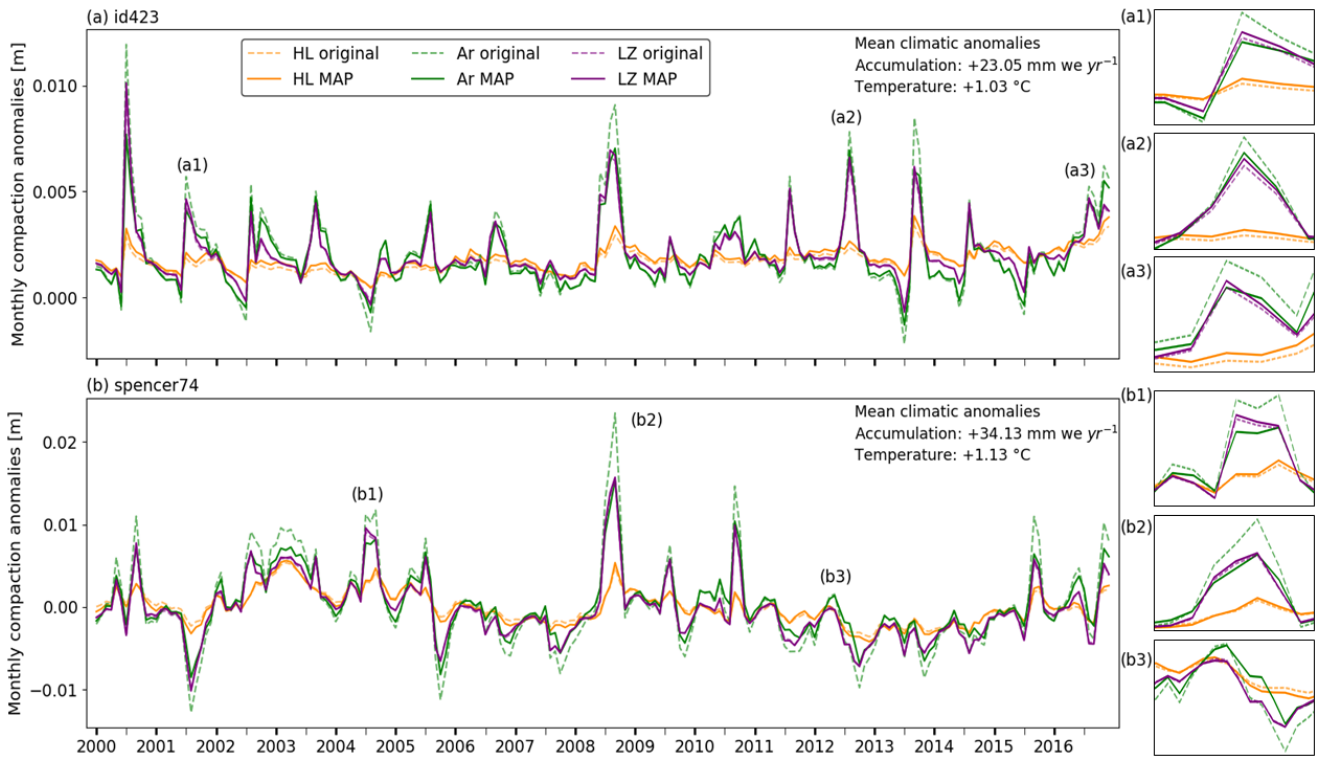


Figure 7. Monthly time-series of compaction anomalies at two sites on the GrIS. Insets show details for particular intervals of the time-series. Mean climatic anomalies are calculated as a difference between mean climatic values over the period 2000-2017 with respect to the reference period 1960-1979, and based on RACMO2 values.

S1 Separation between the calibration and the evaluation data

The 91 sites of the dataset span a broad range of temperature and accumulation rate conditions (Table S1 and Fig. S1). As explained in the main text, our objective is to select the evaluation data (22 cores) randomly but still making it representative of (i) all climatic conditions and (ii) the ratio of GrIS to AIS sites of the dataset. We
5 separate the 91 observed cores in three tiers of lowest, middle and highest T_{av} and we select randomly 7 cores in each tier for the evaluation data. We repeat this random selection until 5 to 10 out of the 21 cores are from GrIS, with the remainder from AIS. Finally, our dataset includes two sites that are climatic outliers with respect to the others (DML and spencer4 in Table S1) with high T_{av} and \dot{b} values (Figure S1). We select randomly one of these
10 for the evaluation data. Proceeding to the selection based on \dot{b} rather than T_{av} would be similar given the strong correlation between both variables.

S2 Application of random noise in the boundary conditions

In order to let uncertainty in RACMO2 output affect the calibration process, we perturb the temperature and accumulation time series that serve as climatic forcing for the firm models. At each iteration (a round of
15 simulations with a given parameter set at all the calibration sites) and for each individual calibration site, we randomly draw an individual climatic perturbation value c_p from a standard Normal distribution (Eq. (S1)). As such, every calibration site has a specific c_p value, which changes at each iteration. We use observed statistics of RACMO2 errors in temperature and Surface Mass Balance to determine the perturbation. For GrIS, Noël et al. (2019) report RMSE values with respect to field observations for temperature and surface mass balance flux of
20 2.1 K and 69 mm w.e. yr^{-1} respectively (in their Supplementary Material). Each monthly value of the RACMO2 time series is therefore perturbed by the corresponding RMSE value scaled by c_p (Eq. (S2), (S3), (S4)).

We favour this approach rather than drawing a different random perturbation at each time step. The latter method would cause perturbations of opposite signs to occur on a very short timescale, which would result in unrealistic short term climatic variability (e.g. a very warm perturbation could be immediately followed by a very cold
25 perturbation in the next month). Also, using the same c_p value to quantify the magnitude of the perturbation for temperature and accumulation preserves the strong correlation between these variables. Warm (cold) temperature perturbations coincide with high (low) accumulation perturbations, which keeps our random perturbations physically plausible.

30 The part of the total accumulation perturbation attributed to each monthly time step is weighted by the actual accumulation at that time step. This attributes larger absolute noise in accumulation to high-accumulation months and lower absolute noise to low-accumulation months (Eq. (S3), (S4)).

35 Our approach is summarized in Eq. (S1), (S2), (S3) and (S4). These equations are applied at all iterations of the calibration process.

$$c_p \sim N(0,1) \quad \text{at all calibration sites} \quad (\text{S1})$$

$$T_t^* = T_t + c_p \sigma_T \quad \text{at all } t \quad (\text{S2})$$

$$\dot{b}_{tot}^* = n_{yr} c_p \sigma_{SMB} \quad (\text{S3})$$

$$40 \quad \dot{b}_t^* = \dot{b}_t + \dot{b}_{tot}^* \frac{\dot{b}_t}{\sum_t \dot{b}_t} \quad (\text{S4})$$

where T_t and \dot{b}_t are temperature and accumulation rate as computed by RACMO2 at time step t and the * superscript denotes the perturbed quantity. n_{yr} is the total number of years in a given simulation, \dot{b}_{tot}^* is the total accumulation perturbation applied for that simulation and σ_T and σ_{SMB} are the temperature and surface mass balance flux RMSE values (as mentioned above, $\sigma_T = 2.1$ K and $\sigma_{SMB} = 69$ mm w.e. yr⁻¹ for GrIS). Note that by using a RMSE value on the surface mass balance flux, we overestimate uncertainty because the observed RMSE is mostly driven by errors in melt amounts which do not apply at the sites of our dataset, all from the dry snow zone area. For AIS, we apply the exact same process for perturbing the temperature variables. We use the RMSE value reported by van Wessem et al. (2018) and set $\sigma_T = 1.3$ K. The accumulation conditions of AIS forces the use of a slightly different method for perturbing the accumulation rate. In terms of magnitude, RACMO2 errors are much larger in coastal areas, where accumulation rates are high. In contrast, in the dry interior of the ice sheet where most of the cores of our dataset come from, the magnitude of RACMO2 errors is small due to low accumulation rates. As such, applying noise based on the ice sheet wide RMSE value would result in noise signals larger than actual accumulation values at most of our dry sites. We thus use the average RACMO surface mass balance bias of 5% (van Wessem et al., 2018) as a proxy for one standard deviation. For AIS, Eq. (S3) and (S4) are replaced by Eq. (S5).

$$\dot{b}_t^* = \dot{b}_t + 0.05 c_{p\#} \dot{b}_t \quad (S5)$$

As explained in Sect. 2.2, we also let uncertainty in fresh snow density, ρ_0 , affect the calibration process by applying random perturbations to each ρ_0^t . In contrast to the climatic perturbation, the perturbation in ρ_0 ~~must not be iteration specific but~~ can be specific to each single time step t , and the perturbation thus varies throughout the duration of a firm model simulation. Indeed, it is not unrealistic that a month with anomalously low fresh snow density is immediately followed by a month of anomalously high fresh snow density for example. We determine surface density values at each site from the firm cores of our dataset, ρ_0^{core} , and we perturb these values based on a standard deviation of 25 kg m⁻³. This value goes in line with a typical window of local variability of 50 kg m⁻³ for ρ_0 (Reeh et al., 2005). As such, adding noise to ρ_0 simplifies to Eq. (S6).

$$\rho_{0,t}^* \sim N(\rho_0^{core}, 25) \quad (S6)$$

We emphasize that the aim of this study is not to conduct a complete sensitivity analysis of firm densification to climatic forcing and to fresh snow density. The objective of the perturbations is to let reasonable estimates of errors in those fields to be accounted for in the calibration process.

S3 Prior correlations in HL and Ar

The Arrhenius form of HL and Ar (Eq. (4) and (5)) allows us to include some correlation in the prior distributions over the parameters of these models. The values of the Arrhenius pre-exponential factors (k_0^* , k_1^* , k_0^{Ar} and k_1^{Ar}) are correlated with their corresponding activation energies (E_0 , E_1 and E_g). For any given constant temperature, modelled densification rates, $\frac{d\rho}{dt}$, can be kept constant despite a change in the pre-exponential factor if the corresponding activation energy is changed accordingly and vice versa. As such, changes in these parameters can potentially compensate in the calculation of DIP values and of the likelihood function (Eq. (8)).

By enforcing constant $\frac{d\rho}{dt}$, exact compensation is ensured by the following equalities:

$$k_{0,mv}^* = k_{0,HL}^* \exp\left(\frac{E_{0,mv} - E_{0,HL}}{RT}\right) \quad (S7)$$

$$k_{1,mv}^* = k_{1,HL}^* \exp\left(\frac{E_{1,mv} - E_{1,HL}}{R T}\right) \quad (\text{S8})$$

$$k_{0,mv}^{Ar} = k_{0,Ar}^{Ar} \exp\left(\frac{E_{g,Ar} - E_{g,mv}}{R T}\right) \quad (\text{S9})$$

$$k_{1,mv}^{Ar} = k_{1,Ar}^{Ar} \exp\left(\frac{E_{g,Ar} - E_{g,mv}}{R T}\right) \quad (\text{S10})$$

where *HL* and *Ar* subscripts denote the ~~original~~ values from the original publications of the HL and Ar models, and the *mv* subscript denotes a modified value of the parameter. Firstly, we generate 10000 random values of temperature *T* in the range of annual mean temperatures covered by our dataset. Secondly, for each random temperature, we generate random values of $E_{0,mv}$, $E_{1,mv}$ and $E_{g,mv}$ in an interval of $\pm 500 \text{ J mol}^{-1}$ around the original values. Thirdly, we calculate the corresponding values in the pre-exponential factors from Eq. (S7), (S8), (S9) and (S10). This results in 10000 pairs of $(k_{0,mv}^*, E_{0,mv})$, $(k_{1,mv}^*, E_{1,mv})$, $(k_{0,mv}^{Ar}, E_{g,mv})$ and $(k_{1,mv}^{Ar}, E_{g,mv})$, from which we calculate correlation coefficients. The absolute values of all four correlation coefficients lie in the interval [0.75; 0.78]. We decide to fix all prior correlation coefficients to -0.75 (HL parameters, negatively correlated) and 0.75 (Ar parameters, positively correlated). The process described necessarily results in perfectly correlated $k_{0,mv}^{Ar}$ and $k_{1,mv}^{Ar}$. We also set the prior correlation between these parameters to 0.75. We emphasize here that any other pair of *a priori* uncorrelated parameters can certainly be correlated *a posteriori* if the calibration process identifies such quantitative behaviour when the observed data is considered. This is highlighted and further discussed in Sect. S7.

S4 The likelihood function, Eq. (8)

The covariance matrices Σ_{15} and Σ_{pc} that appear in Eq. (8) are diagonal matrices with the site-specific variances on the diagonal. At each site, we take 10% of the observed *DIP*₁₅ and 20% of the observed *DIP*_{pc} as the standard deviation, and the variance value is the square of the standard deviation. We take the higher value of 20% for *DIP*_{pc} because density errors propagate in firm models. Equation (3) shows that densification rates depend on the density value itself, resulting in error propagation through time. As such, if a model shows an offset compared to observations at 15 m depth, it is likely to show an even stronger offset at z_{pc} . Taking a higher variance alleviates the strength of this effect on the likelihood calculations by allowing a larger spread of model results compared to observed *DIP*_{pc} values.

The form of Eq. (8) corresponds to a normal likelihood function. This assumes that model *DIP* results are normally distributed around the observed values. To support this assumption, we conducted a preliminary verification of errors in *DIP*₁₅ ($X_{15} - Y_{15}$) and *DIP*_{pc} ($X_{pc} - Y_{pc}$) computed with the three original models (HL, Ar, LZ) on the entire dataset. We compute a basic Kolmogorov-Smirnov test for both sets of errors: residuals in *DIP*₁₅ and in *DIP*_{pc}. The resulting p-values are very large: 0.94 and 0.86 respectively. The distributions of these errors are thus in line with a normal distribution. We show the Quantiles-Quantiles plots for both sets of residuals in Figure S2. As explained in the main text, the form of Eq. (8) also assumes independence between errors in *DIP*₁₅ and *DIP*_{pc}, which is the reason why *DIP*_{pc} is calculated only from depths below 15 m. As such, observations-model discrepancies are essentially governed by parameter values of stage-1 densification for *DIP*₁₅ and by parameter values of stage-2 densification for *DIP*_{pc}, with little interaction between both. The same preliminary verification as mentioned above allows us to evaluate the correlation between *DIP*₁₅ and *DIP*_{pc} errors for all three original models on the entire dataset. This yields correlation coefficients of 0.13, 0.60 and -0.01 for the original models HL, Ar and LZ respectively. **S5 Convergence diagnostics**

For convergence of the RWM algorithm, the chain must traverse between the peaks of the target posterior distribution multiple times. Simply examining the trace of the RWM algorithm for each parameter provides an

effective way to verify this criterion. The trace is the history of accepted parameter values over the entire chain. We show this sampling history in Fig. S3. The fuzzy appearance for each parameter of each model indicates an efficient exploration of the parameter space as the samples from RWM algorithm oscillate around the posterior mode.

- 5 In addition to this, we compute the Gelman-Rubin statistic, which provides a numerical test for convergence (Gelman et al., 2013). The motivation behind this test is that if each chain (run independently) converges to the same posterior distribution, then the variances within each chain should be approximately the same. For each model (HL, Ar, LZ), we launch three different chains from different initial parameter values. For each parameter of each model, we calculate the mean within sample variance W :

$$10 \quad W = \frac{s_1^2 + s_2^2 + s_3^2}{3} \quad (\text{S1})$$

where s^2 denotes the variance of an individual chain. We then calculate the between sample variance:

$$B = \frac{n}{3-1} \sum_{i=1}^3 (\bar{\theta}_i - \bar{\bar{\theta}})^2 \quad (\text{S2})$$

where n denotes the number of iterations within each chain, $\bar{\theta}_i$ the mean parameter value within each chain and $\bar{\bar{\theta}}$ is the mean of $(\bar{\theta}_1, \bar{\theta}_2, \bar{\theta}_3)$. From there, the estimate of the variance of the posterior distribution is given by:

$$15 \quad \hat{\sigma}^2 = \frac{n-1}{n} W + \frac{1}{n} B \quad (\text{S3})$$

And the Gelman-Rubin statistic is defined as:

$$R = \sqrt{\frac{\hat{\sigma}^2}{W}} \quad (\text{S4})$$

- Large values of R indicate that estimates of θ values between the different chains are significantly different. With more iterations, the chains progressively converge to the same stationary distributions and the estimates of θ become similar, resulting in values of R close to 1. We reach $R < 1.1$ for all parameters, which proves adequate convergence (Gelman et al., 2013). Two parameters of the LZ model needed a larger number of iterations to reach $R < 1.1$.

S6 Normal approximation to the posterior

- The ensembles of parameter combinations obtained for each model provide large samples, representative of the posterior probability distributions over their respective parameter space. The most efficient way to assess parameter-related uncertainty is to run a model with a high number of random parameter combinations from these ensembles, which is demonstrated in Sect. 3. However, this means that for any firm modelling study, access must be easy to such posterior ensembles or an MCMC algorithm must be re-executed. To circumvent these practical difficulties, it is approximately correct to sample random parameter combinations from a multivariate normal distribution centred about the mean of the posterior ensemble and with covariance matrix set to the posterior ensemble covariance matrix. This is commonly referred to as a normal approximation to the posterior (Gelman et al., 2013). Table S2 provides both the posterior mean and posterior covariance for the HL, Ar and LZ models. We assess how random samples from the normal approximations compare to samples from the posterior ensembles in Fig. S4. Posterior samples and the normal approximations are very similar, with correlations only slightly less well captured in the tails of the distributions. ~~It~~ As a consequence, the normal approximation results in a slight overestimation of uncertainty and thus conservative estimates of uncertainty. This has been confirmed by additional model simulations with values sampled from the normal approximations (not shown).

S7 Posterior correlation between parameters

The joint posterior distributions for the parameters of each model also allow us to analyse the models' internal structure, i.e. the correlation between their different parameters. The full correlation matrices are given in Fig. [S4S5](#). In HL, the strongest correlation coefficients r are unsurprisingly found for the pairs of pre-exponential factor and activation energy governing densification in stage-1 (k_0^* and E_0) and in stage-2 (k_1^* and E_1) with r of 0.91 and 0.92 respectively. Higher activation energies (E_0 and E_1) imply stronger thermal barriers to the densification process and thus slower densification, and the pre-exponential factors (k_0^* and k_1^*) counter-balance the effect to still match observed *DIP* values. In the same way, the activation energies are negatively correlated with their respective accumulation rate exponent (a and b), but more moderately (r values of approximately -0.5). The negative correlation of -0.28 between a and b themselves might be linked to the density at 15 m being the lower boundary and the upper boundary condition for the calculation of *DIP*₁₅ and *DIP*_{pc} respectively. Higher values of a tend to cause lighter firm at 15 m depth. Lower E_0 values compensate for this effect on *DIP*₁₅ because the shallow firm densifies faster due to its greater sensitivity to temperature variations. The lighter 15 m depth density also affects *DIP*_{pc}, and lower values of b compensate for this by enhancing the densification rate, which explains the negative correlation between a and b . In Ar, the interpretation is more challenging due to the use of a same activation energy in both stages. There is a strong correlation between the activation energy E_g and both pre-exponential factors k_0^{Ar} ($r = -0.89$) and k_1^{Ar} ($r = -0.90$), for the same reason as in HL. As such, this induces a strong positive correlation between the latter parameters ($r = 0.76$). The negative correlation between α and k_1^{Ar} ($r = -0.41$) is more surprising because these parameters apply to different stages, but it reveals an interesting pattern. Higher temperatures raise densification rates at warmer sites, where accumulation rates are also higher thus further amplifying this effect. Higher accumulation rates nevertheless cause light recently deposited firm to be buried rapidly, which may cause lower density firm governed by the fast stage-1 densification to extend below 15 m. To avoid underestimation of *DIP*_{pc} at such sites, stage-1 densification rates must remain low enough but there is no possibility for adjusting a stage-1 specific activation energy. Lower α values generate this effect while only marginally affecting densification at colder low-accumulation sites. Thus, high k_1^{Ar} without a complementary lower α would cause *DIP*_{pc} underestimation at warm and high accumulation sites. We note here that, through the calibration process, the data enhanced the prior correlations we assigned in the HL and Ar models. Analysis of correlation coefficients in LZ is less straightforward because its governing equations, Eq. (6), are less interpretable than the plain Arrhenius relationship of HL and Ar. Still, we highlight some correlated pairs of parameters. As could be expected from Eq. (6), lz_a and lz_b are negatively correlated ($r = -0.80$). Also, the independent term of stage-1 densification lz_{11} is strongly correlated with the corresponding temperature-related parameter (lz_{13} , $r = 0.94$). The same is valid for stage-2 densification between lz_{21} and lz_{23} ($r = 0.90$). The positive correlation between lz_{12} and lz_{22} ($r = 0.74$) is discussed in the main text.

Tables

| Site | Lat | Lon | Core depth [m] | Year | Mean \dot{b} [m w.e. yr ⁻¹] | Mean T [°C] | ρ_0 [kg/m ³] | DIP15 [m] | Var DIP15 [m ²] | DIP _{pc} [m] | Var DIP _{pc} [m ²] |
|-------------|-------|--------|----------------|------|-------------------------------------------|-------------|-------------------------------|-----------|-----------------------------|-----------------------|-----------------------------------------|
| EGRIP | 75.63 | -35.98 | 20.1 | 2017 | 0.113 | -29.0 | 285 | 7.816 | 0.611 | / | / |
| Summit * | 72.58 | -38.47 | 22.1 | 2017 | 0.205 | -28.4 | 330 | 7.500 | 0.562 | / | / |
| id359 | 73.94 | -37.63 | 102.4 | 1993 | 0.124 | -28.8 | 240 | 6.708 | 0.450 | 11.456 | 5.250 |
| id369 | 75.00 | -30.00 | 19.9 | 1997 | 0.135 | -27.6 | 335 | 7.454 | 0.556 | / | / |
| id373 | 75.25 | -37.62 | 100.8 | 1993 | 0.106 | -29.5 | 275 | 7.826 | 0.612 | 12.372 | 6.123 |
| id385 | 76.00 | -43.49 | 109.8 | 1995 | 0.124 | -29.3 | 315 | 7.857 | 0.617 | 13.186 | 6.955 |
| id423 * | 76.62 | -36.40 | 143.2 | 1993 | 0.093 | -29.1 | 310 | 7.716 | 0.595 | 10.666 | 4.550 |
| id514 | 77.25 | -49.22 | 119.6 | 1995 | 0.163 | -28.3 | 300 | 7.575 | 0.574 | 13.217 | 6.987 |
| id531 * | 77.45 | -51.06 | 75.0 | 2009 | 0.198 | -27.4 | 320 | 7.434 | 0.553 | / | / |
| id534 | 80.00 | -41.14 | 96.0 | 1994 | 0.105 | -28.4 | 335 | 7.811 | 0.610 | 11.345 | 5.148 |
| Basin8 | 69.80 | -36.49 | 29.8 | 2003 | 0.350 | -25.6 | 300 | 7.396 | 0.547 | / | / |
| D2 | 71.80 | -46.34 | 101.3 | 2003 | 0.421 | -23.4 | 370 | 7.051 | 0.497 | 14.097 | 7.949 |
| D4 | 71.39 | -43.94 | 143.9 | 2003 | 0.390 | -24.6 | 300 | 7.394 | 0.547 | 12.770 | 6.523 |
| HumboldtM * | 78.47 | -56.98 | 141.9 | 1995 | 0.384 | -24.8 | 280 | 8.062 | 0.650 | 10.947 | 4.794 |
| NASAE1 * | 74.98 | -29.97 | 19.9 | 1997 | 0.135 | -27.6 | 340 | 7.394 | 0.547 | / | / |
| spencer6 * | 72.57 | -37.62 | 82.3 | 1994 | 0.176 | -29.0 | 360 | 4.889 | 0.239 | / | / |
| spencer16 | 71.75 | -40.75 | 15.0 | 1954 | 0.289 | -27.0 | 340 | 7.216 | 0.521 | / | / |
| spencer17 | 77.95 | -39.18 | 60.0 | 1973 | 0.080 | -29.3 | 300 | 5.002 | 0.250 | 7.781 | 2.421 |
| spencer66 * | 70.75 | -35.96 | 109.0 | 1987 | 0.247 | -27.3 | 300 | 7.340 | 0.539 | 14.852 | 8.823 |
| spencer67 | 70.63 | -35.83 | 128.6 | 1988 | 0.262 | -27.0 | 325 | 7.098 | 0.504 | 14.114 | 7.968 |
| spencer68 * | 70.65 | -37.48 | 105.6 | 1988 | 0.263 | -26.9 | 325 | 7.172 | 0.514 | 14.505 | 8.416 |
| spencer69 | 70.67 | -38.79 | 24.8 | 1988 | 0.252 | -27.1 | 305 | 7.184 | 0.516 | / | / |
| spencer70 | 70.64 | -39.62 | 100.1 | 1988 | 0.262 | -27.0 | 290 | 6.772 | 0.459 | 14.026 | 7.869 |

| | | | | | | | | | | | |
|-----------|--------|---------|-------|------|-------|-------|-----|-------|-------|--------|--------|
| spencer71 | 71.76 | -35.87 | 77.8 | 1988 | 0.203 | -28.2 | 275 | 7.043 | 0.496 | 13.094 | 6.858 |
| spencer72 | 71.48 | -35.88 | 25.7 | 1988 | 0.207 | -28.0 | 330 | 7.223 | 0.522 | / | / |
| spencer73 | 71.15 | -35.85 | 70.8 | 1988 | 0.214 | -27.7 | 340 | 7.230 | 0.523 | / | / |
| spencer74 | 70.85 | -35.85 | 26.2 | 1988 | 0.264 | -26.9 | 330 | 7.087 | 0.502 | / | / |
| SouthPole | -90.00 | 0.00 | 122.9 | 2001 | 0.055 | -47.8 | 325 | 7.613 | 0.580 | 22.312 | 19.913 |
| Newall | -77.58 | 162.50 | 111.1 | 1989 | 0.043 | -31.2 | 305 | 7.160 | 0.513 | 4.132 | 0.683 |
| Berkner * | -79.61 | -45.72 | 178.2 | 1995 | 0.124 | -28.3 | 345 | 6.255 | 0.391 | 9.658 | 3.731 |
| DML * | -71.41 | -9.92 | 78.2 | 2007 | 0.902 | -20.6 | 410 | 6.037 | 0.364 | 10.228 | 4.185 |
| id9 | -76.77 | -101.74 | 111.6 | 2012 | 0.313 | -24.7 | 470 | 6.194 | 0.384 | 12.119 | 5.875 |
| id10 | -76.95 | -121.22 | 62.0 | 2011 | 0.213 | -28.4 | 355 | 6.947 | 0.483 | / | / |
| id11 | -77.06 | -89.14 | 114.5 | 2001 | 0.346 | -26.5 | 415 | 5.879 | 0.346 | 11.201 | 5.019 |
| id12 | -77.61 | -92.25 | 67.8 | 2001 | 0.301 | -27.8 | 350 | 6.019 | 0.362 | / | / |
| id13 | -77.68 | -124.00 | 59.3 | 2000 | 0.155 | -28.2 | 350 | 6.411 | 0.411 | / | / |
| id14 | -77.76 | 153.38 | 97.1 | 2006 | 0.048 | -44.6 | 360 | 6.833 | 0.467 | 17.516 | 12.272 |
| id15 * | -77.84 | -102.91 | 70.7 | 2001 | 0.486 | -25.1 | 415 | 5.853 | 0.343 | / | / |
| id17 | -77.88 | 158.46 | 98.5 | 2006 | 0.058 | -41.1 | 350 | 6.419 | 0.412 | 11.687 | 5.464 |
| id18 | -77.96 | -95.96 | 57.4 | 2010 | 0.354 | -28.0 | 335 | 6.752 | 0.456 | / | / |
| id19 | -78.08 | -120.08 | 57.8 | 2000 | 0.171 | -27.7 | 315 | 6.253 | 0.391 | / | / |
| id20 | -78.12 | -95.65 | 70.5 | 2001 | 0.324 | -27.7 | 385 | 6.265 | 0.393 | / | / |
| id22 * | -78.33 | -124.48 | 59.9 | 2000 | 0.152 | -27.7 | 285 | 6.509 | 0.424 | 8.989 | 3.232 |
| id24 | -78.43 | -115.92 | 59.8 | 2000 | 0.318 | -27.8 | 390 | 6.295 | 0.396 | / | / |
| id26 | -78.73 | -111.50 | 60.7 | 2000 | 0.329 | -27.8 | 350 | 6.427 | 0.413 | / | / |
| id28 | -79.04 | 149.68 | 100.1 | 2006 | 0.040 | -44.6 | 405 | 6.703 | 0.449 | 15.584 | 9.714 |
| id29 * | -79.13 | -122.27 | 63.1 | 2000 | 0.127 | -27.8 | 300 | 6.507 | 0.423 | 9.926 | 3.941 |
| id30 | -79.16 | -104.97 | 72.7 | 2001 | 0.306 | -28.7 | 400 | 5.921 | 0.351 | / | / |
| id33 | -79.38 | -111.24 | 104.8 | 2000 | 0.239 | -28.2 | 370 | 6.159 | 0.379 | 12.943 | 6.701 |
| id35 * | -79.48 | -112.09 | 160.0 | 2011 | 0.162 | -28.0 | 460 | 6.181 | 0.382 | 11.824 | 5.592 |
| id39 | -80.62 | -122.63 | 57.5 | 1999 | 0.094 | -25.9 | 370 | 6.253 | 0.391 | / | / |

| | | | | | | | | | | | |
|-------------|--------|---------|-------|-------|-------|-------|-----|-------|-------|--------|--------|
| id43 | -81.20 | -126.17 | 48.3 | 1999 | 0.070 | -24.5 | 325 | 6.268 | 0.393 | 4.975 | 0.990 |
| id46 | -82.00 | -110.01 | 62.2 | 2002 | 0.180 | -27.8 | 340 | 6.161 | 0.380 | / | / |
| id48 | -83.50 | -104.99 | 61.7 | 2002 | 0.220 | -31.0 | 360 | 6.098 | 0.372 | / | / |
| id49 * | -84.40 | 140.63 | 50.1 | 2007 | 0.023 | -45.4 | 340 | 6.886 | 0.474 | / | / |
| id50 | -85.00 | -105.00 | 44.9 | 2002 | 0.157 | -36.3 | 360 | 6.422 | 0.412 | / | / |
| id51 | -85.78 | 145.72 | 41.7 | 2007 | 0.033 | -46.1 | 310 | 6.767 | 0.458 | / | / |
| id52 * | -86.50 | -107.99 | 71.6 | 2002 | 0.147 | -38.8 | 340 | 6.882 | 0.474 | / | / |
| id53 | -86.84 | 95.31 | 20.8 | 2003 | 0.042 | -53.3 | 355 | 6.535 | 0.427 | / | / |
| id54 * | -88.00 | -107.98 | 54.1 | 2002 | 0.133 | -41.4 | 355 | 7.009 | 0.491 | / | / |
| id55 | -88.51 | 178.53 | 99.3 | 2007 | 0.081 | -48.2 | 320 | 6.880 | 0.473 | / | / |
| id56 | -89.93 | 144.39 | 139.5 | 2002 | 0.080 | -48.6 | 345 | 6.319 | 0.399 | 25.046 | 25.092 |
| spencer1 | -80.00 | -120.00 | 307.0 | 1968 | 0.120 | -27.2 | 350 | 6.987 | 0.488 | 10.314 | 4.255 |
| spencer4 | -66.72 | 113.18 | 200.9 | 1989 | 1.060 | -22.0 | 380 | 7.848 | 0.616 | 12.847 | 6.602 |
| spencer5 | -74.50 | 123.17 | 49.5 | 1980 | 0.037 | -51.8 | 345 | 8.262 | 0.683 | / | / |
| spencer7 | -85.25 | 166.50 | 79.9 | 19997 | 0.028 | -39.7 | 305 | 7.003 | 0.490 | 8.202 | 2.691 |
| spencer8 | -66.77 | 112.80 | 180.0 | 1997 | 0.488 | -22.7 | 385 | 7.385 | 0.545 | 10.640 | 4.528 |
| spencer22 | -73.60 | -12.43 | 25.5 | 1996 | 0.220 | -22.5 | 380 | 3.920 | 0.154 | / | / |
| spencer25 | -74.02 | -12.02 | 26.5 | 1996 | 0.171 | -30.7 | 390 | 5.412 | 0.293 | / | / |
| spencer29 * | -75.00 | 2.00 | 20.6 | 1996 | 0.072 | -42.9 | 320 | 7.602 | 0.578 | / | / |
| spencer33 | -70.68 | 44.32 | 123.5 | 1978 | 0.114 | -33.1 | 385 | 6.385 | 0.408 | 7.022 | 1.972 |
| spencer34 * | -70.68 | 44.32 | 109.0 | 1978 | 0.114 | -33.1 | 375 | 6.161 | 0.380 | 6.909 | 1.909 |
| spencer61 | -73.10 | 39.75 | 99.7 | 1978 | 0.069 | -42.3 | 360 | 7.005 | 0.491 | 16.245 | 10.556 |
| spencer62 * | -71.18 | 45.97 | 100.2 | 1997 | 0.091 | -38.2 | 395 | 7.049 | 0.497 | 16.344 | 10.686 |
| spencer76 | -90.00 | 0.00 | 122.1 | 1997 | 0.055 | -47.8 | 360 | 4.906 | 0.241 | 25.586 | 26.185 |
| spencer77 | -75.00 | 147.00 | 15.8 | 1961 | 0.042 | -46.1 | 385 | 7.184 | 0.516 | / | / |
| spencer78 * | -74.00 | 143.00 | 16.0 | 1961 | 0.043 | -45.5 | 375 | 7.205 | 0.519 | / | / |
| spencer79 | -73.00 | 142.00 | 15.7 | 1961 | 0.057 | -44.0 | 325 | 7.148 | 0.511 | / | / |
| spencer80 | -73.00 | 141.00 | 16.0 | 1961 | 0.057 | -44.0 | 355 | 6.876 | 0.473 | / | / |

| | | | | | | | | | | | |
|-------------|--------|---------|-------|------|-------|-------|-----|-------|-------|--------|--------|
| spencer81 | -72.00 | 140.00 | 16.9 | 1961 | 0.080 | -42.7 | 335 | 6.936 | 0.481 | / | / |
| spencer82 * | -71.00 | 139.00 | 15.6 | 1961 | 0.120 | -41.6 | 375 | 6.848 | 0.469 | / | / |
| spencer83 | -72.00 | 143.00 | 15.7 | 1961 | 0.087 | -41.3 | 405 | 6.796 | 0.462 | / | / |
| spencer84 | -72.00 | 146.00 | 16.2 | 1961 | 0.086 | -40.9 | 410 | 6.876 | 0.473 | / | / |
| spencer85 | -72.00 | 148.00 | 15.9 | 1961 | 0.096 | -40.2 | 360 | 6.745 | 0.455 | / | / |
| spencer86 | -72.00 | 151.00 | 15.8 | 1961 | 0.103 | -39.7 | 400 | 6.963 | 0.485 | / | / |
| spencer87 | -72.00 | 154.00 | 15.9 | 1961 | 0.130 | -38.0 | 355 | 6.430 | 0.414 | / | / |
| spencer88 | -72.00 | 156.00 | 15.7 | 1961 | 0.130 | -37.6 | 395 | 7.050 | 0.497 | / | / |
| spencer89 | -72.00 | 159.00 | 15.7 | 1961 | 0.115 | -35.7 | 370 | 6.665 | 0.444 | / | / |
| spencer90 | -83.47 | 138.80 | 340.5 | 1994 | 0.020 | -45.2 | 420 | / | / | 10.046 | 4.037 |
| spencer91 | -83.47 | -138.80 | 47.0 | 1987 | 0.058 | -27.1 | 295 | 7.037 | 0.495 | 3.530 | 0.499 |
| spencer92 | -78.47 | 106.80 | 179.3 | 1996 | 0.022 | -54.6 | 360 | 8.790 | 0.773 | 20.368 | 16.594 |

Table S1. The 91 firn core dataset used in this study. * symbols indicate the core is part of the evaluation data. Lat and Lon designate latitude and longitude respectively. Year indicates the year of drilling of the core. \dot{b} is the accumulation rate. T is the temperature. Values for both \dot{b} and T are computed from the RACMO2 model. ρ_0 is the surface density boundary condition that was derived individually for each core by extrapolating density measurements until the surface (random noise is added to ρ_0 as discussed in Sect. S2). Var designates the site-specific variance used for the terms of Σ_{15} and Σ_{pc} (see Text S4 for their calculation). The core spencer90 has only a single density measurement above 15 m depth and its *DIP15* is discarded.

| | Parameters | Posterior mean | Posterior covariance matrix | | | | | | |
|----|--------------------------------------------------------------------|-------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|
| HL | $k_0^*, k_1^*, E_0, E_1, a, b$ | $[16.7, 649, 10760, 21000, 0.88, 0.66]$ | $\begin{bmatrix} 34.4 & 40.2 & 4500 & 324 & -0.0685 & -0.0195 \\ 40.2 & 44000 & 618 & 161000 & 1.087 & -3.670 \\ 4502 & 618 & 710000 & 7080 & -29.95 & 1.94 \\ 324 & 1610000 & 7080 & 694000 & 7.86 & -27.51 \\ -0.0685 & 1.087 & -29.95 & 7.86 & 0.0051 & -0.0012 \\ -0.0195 & -3.670 & 1.94 & -27.51 & -0.0012 & 0.0036 \end{bmatrix}$ | | | | | | |
| Ar | $k_0^{Ar}, k_1^{Ar}, E_g, \alpha, \beta$ | $[0.080, 0.028, 40900, 0.78, 0.69]$ | $\begin{bmatrix} 5.62 \times 10^{-4} & 1.55 \times 10^{-4} & -12.66 & 9.65 \times 10^{-5} & -3.23 \times 10^{-4} \\ 1.55 \times 10^{-4} & 7.41 \times 10^{-5} & -4.64 & -2.04 \times 10^{-4} & 1.05 \times 10^{-4} \\ -12.66 & -4.64 & 360000 & 11.0 & 4.67 \\ 9.65 \times 10^{-5} & -2.04 \times 10^{-4} & 11.0 & 0.00330 & -0.00101 \\ -3.23 \times 10^{-4} & 1.05 \times 10^{-4} & 4.67 & -0.00101 & 0.00312 \end{bmatrix}$ | | | | | | |
| LZ | $lz_a, lz_b, lz_{11}, lz_{12}, lz_{13}, lz_{21}, lz_{22}, lz_{23}$ | $[7.56, -2.091, -14.71, 7.269, -1.019, -1.513, 6.0203, -0.09127]$ | $\begin{bmatrix} 5.27 & -0.198 & -1.20 & -1.68 & -0.0239 & 0.00553 & -0.0606 & 0.00413 \\ -0.198 & 0.0116 & 0.218 & -0.0612 & 0.0134 & -0.0158 & -0.00229 & -7.37 \times 10^{-4} \\ -1.20 & 0.218 & 14.6 & -3.96 & 0.801 & 0.368 & 0.354 & 0.0129 \\ -1.68 & -0.0612 & -3.96 & 13.3 & -0.309 & -0.0850 & 5.40 & 0.0166 \\ -0.0239 & 0.0134 & 0.801 & -0.309 & -0.0502 & -0.0173 & 0.0252 & -4.42 \times 10^{-4} \\ 0.00553 & -0.0158 & 0.368 & -0.0850 & -0.0173 & 0.446 & -0.429 & 0.0131 \\ -0.0606 & -0.00229 & 0.354 & 5.40 & 0.0252 & -0.429 & 3.94 & -2.59 \times 10^{-4} \\ 0.00413 & -7.37 \times 10^{-4} & 0.0129 & 0.0166 & -4.42 \times 10^{-4} & 0.0131 & -2.59 \times 10^{-4} & 4.80 \times 10^{-4} \end{bmatrix}$ | | | | | | |

Table S2. The posterior means and covariance matrices for the free parameters of HL, Ar and LZ. These statistics can be used to generate random parameter combinations following a normal approximation.

Figures

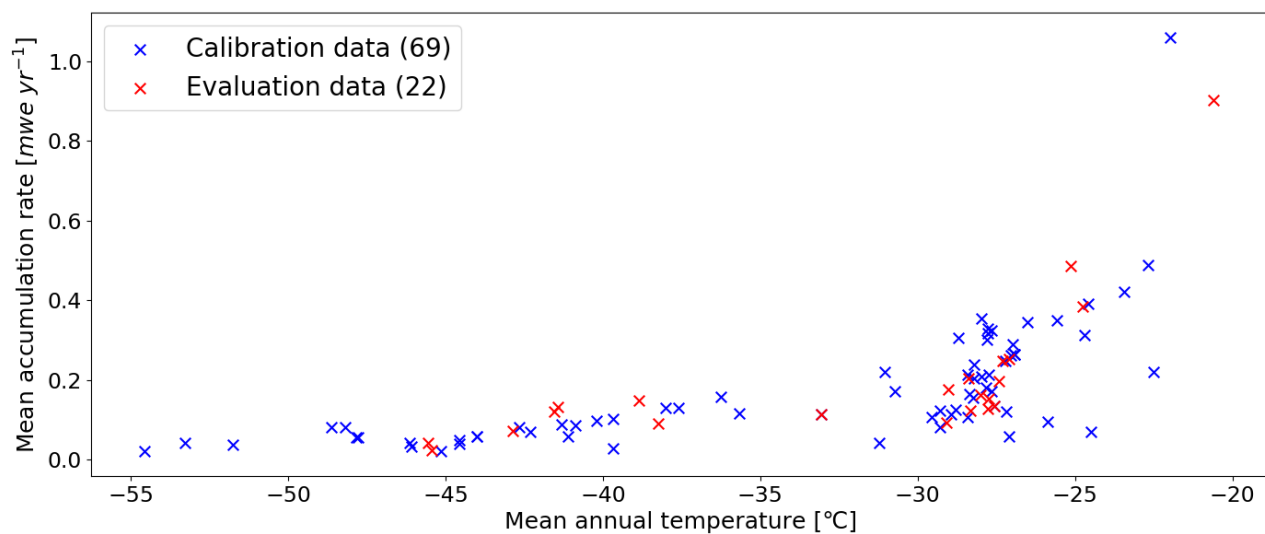
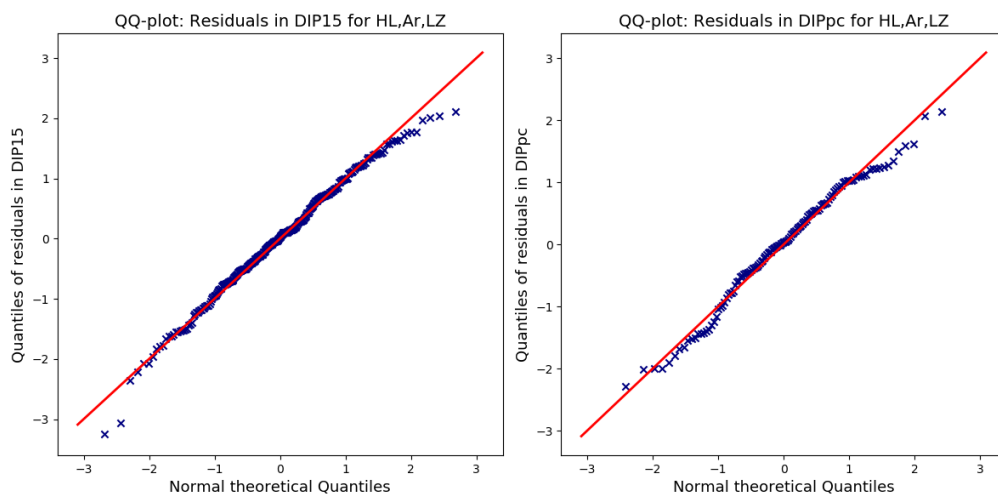


Figure S1. Climatic conditions at the 91 sites of the dataset



5

Figure S2. Quantiles-Quantiles plots for the errors of the three original models (HL, Ar, LZ) computed on the entire dataset. The alignment of the points along the red line informs about the fit to a normal distribution.

10

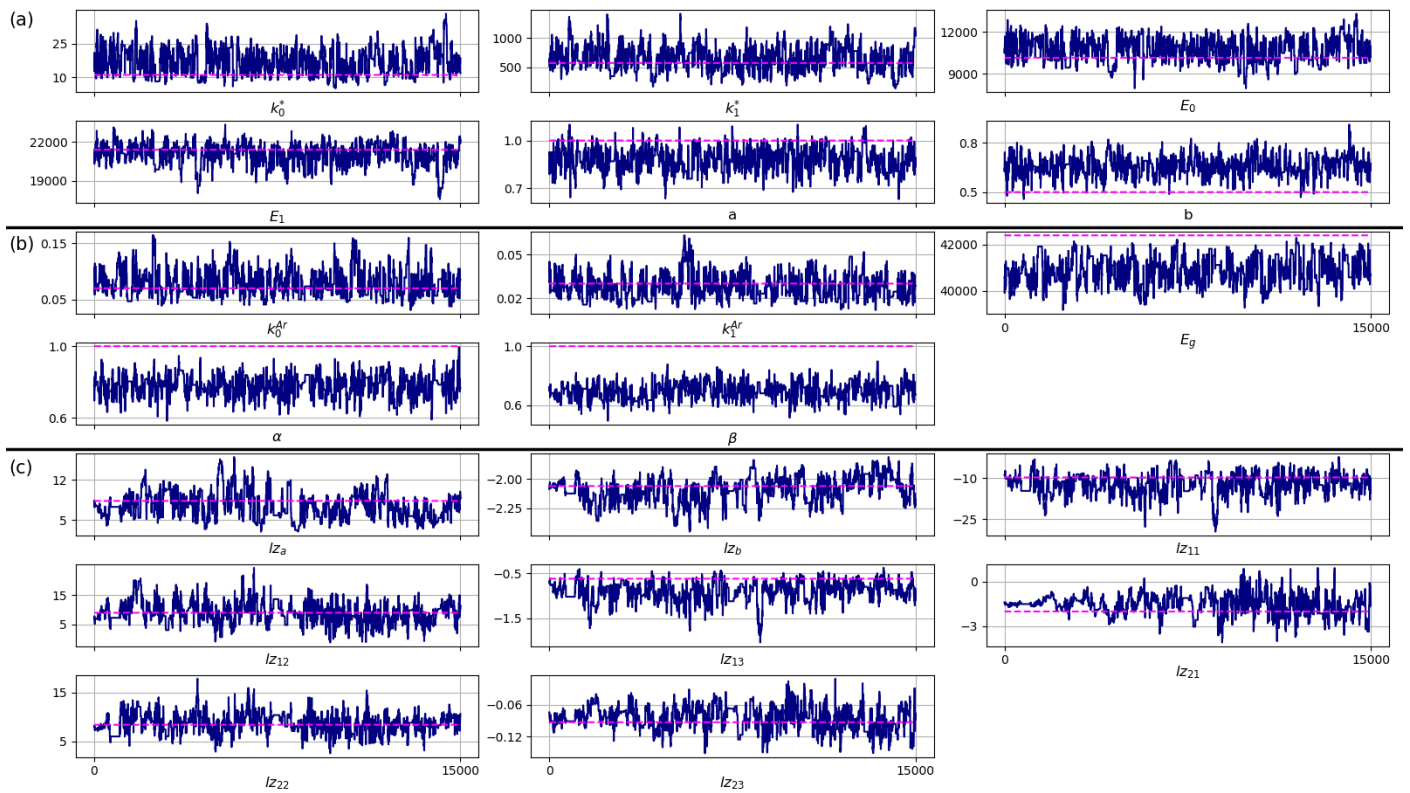


Figure S3. Sampling chains of each parameter for (a) HL, (b) Ar, (c) LZ. The x-axis displays the iteration number, the y-axis displays the parameter value. The dashed pink line shows the value of the original model, which is also the starting point of each chain.

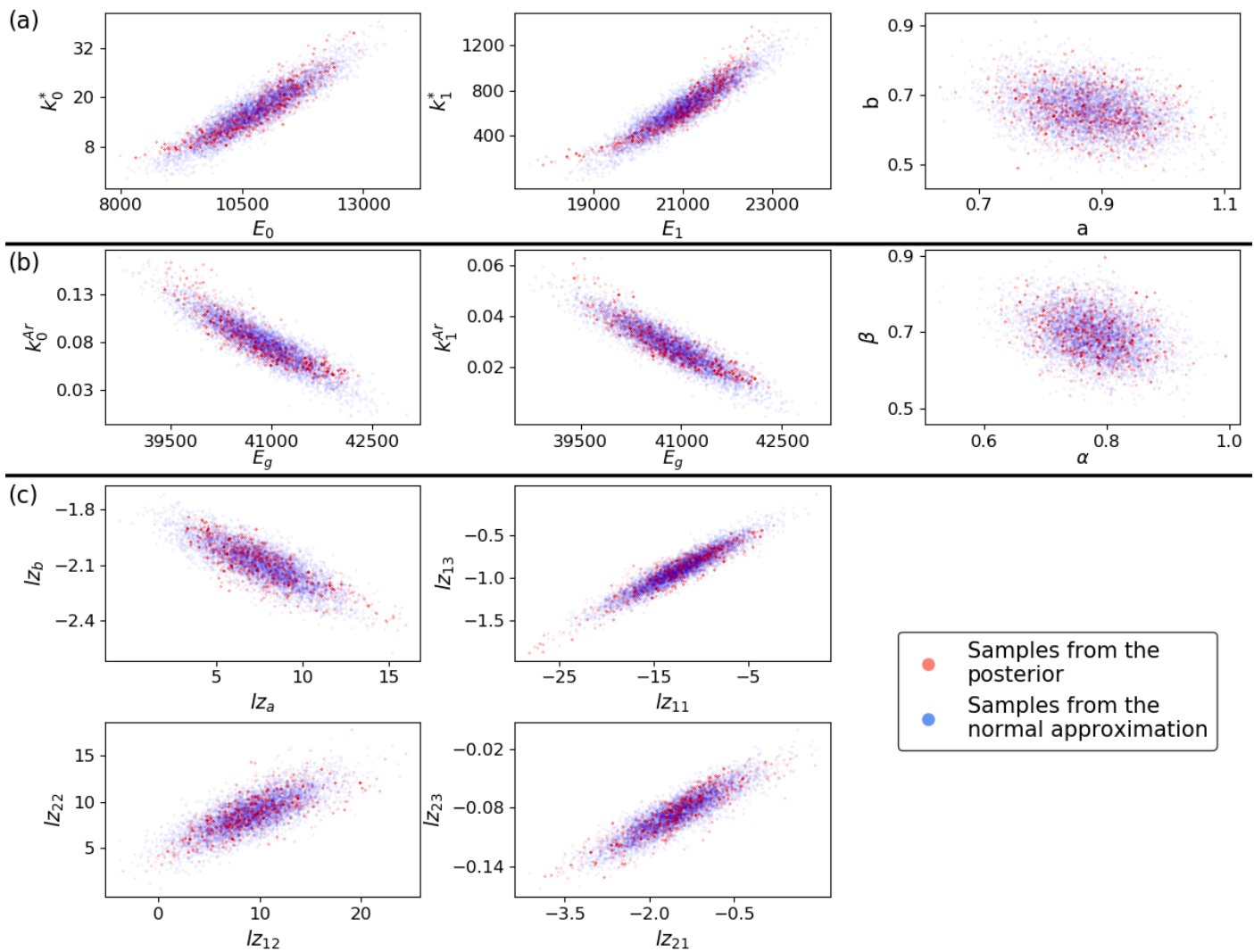


Figure S4. Evaluation of the normal approximations to the posterior distributions for (a) HL, (b) Ar, (c) LZ. Where possible, correlated parameters share a same graph.

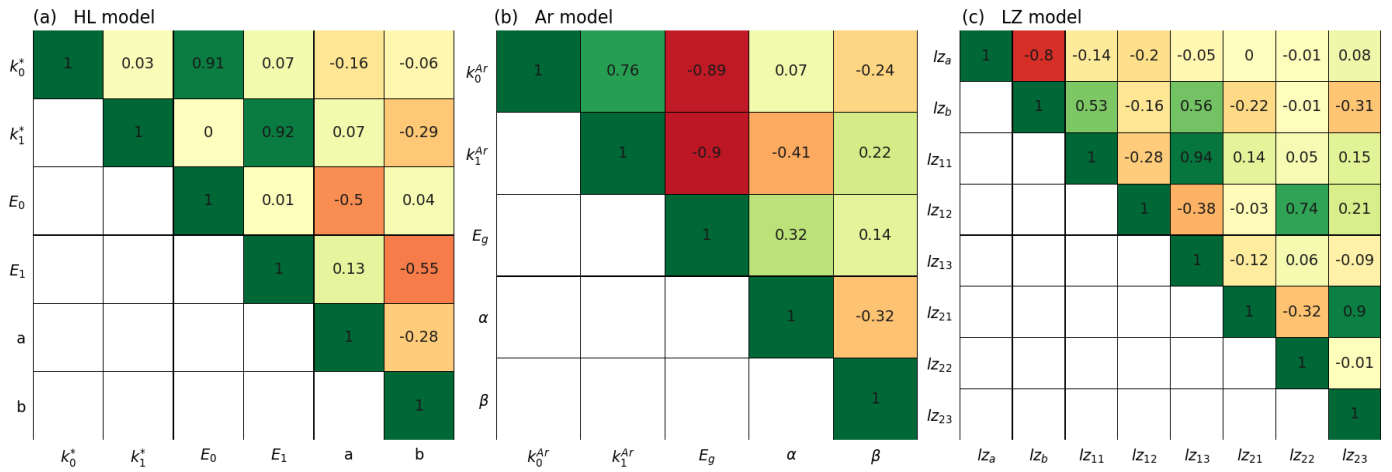


Figure S5. Posterior correlation matrices.

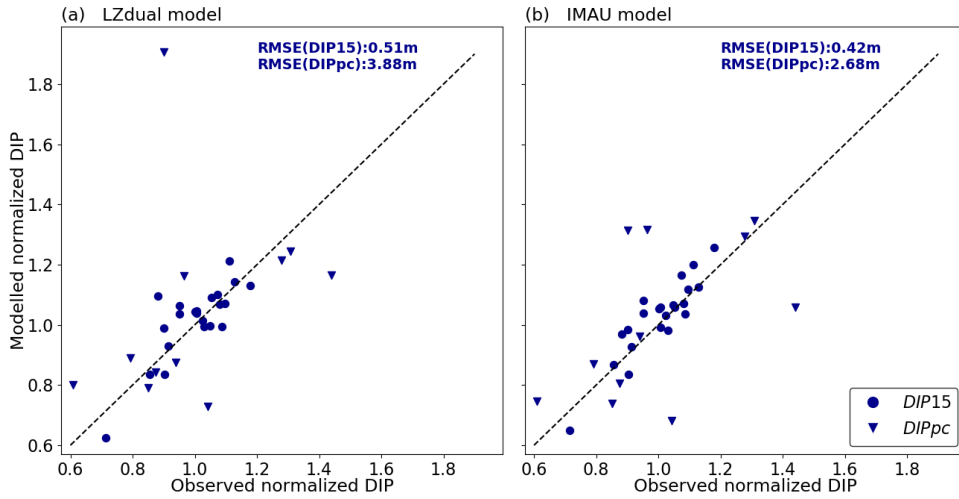


Figure S6. Comparison of evaluation data *DIP* with model results for the LZ dual and IMAU models.

References

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D.: Bayesian Data Analysis, Third Edit., CRC Press Taylor & Francis Group, Boca Raton., 2013.

5 Noël, B., van de Berg, W. J., Lhermitte, S. and van den Broeke, M. R.: Rapid ablation zone expansion amplifies north Greenland mass loss, Sci. Adv., 5(9), eaaw0123, doi:10.1126/sciadv.aaw0123, 2019.

Reeh, N., Fisher, D. A., Koerner, R. M. and Clausen, H. B.: An empirical firndensification model comprising ice lenses, in Annals of Glaciology, vol. 42, pp. 101–106., 2005.

10 van Wessem, J. M., Jan Van De Berg, W., Noël, B. P. Y., Van Meijgaard, E., Amory, C., Birnbaum, G., Jakobs, C. L., Krüger, K., Lenaerts, J. T. M., Lhermitte, S., Ligtenberg, S. R. M., Medley, B., Reijmer, C. H., Van Tricht, K., Trusel, L. D., Van Uft, L. H., Wouters, B., Wuite, J. and Van Den Broeke, M. R.: Modelling the climate and surface mass balance of polar ice sheets using RACMO2 - Part 2: Antarctica (1979-2016), The Cryosphere, 12(4), 1479–1498, doi:10.5194/tc-12-1479-2018, 2018.

15