Lancaster Environment Centre

Lancaster University

Addressing uncertainty in firn densification models for applications on the Greenland and Antarctic ice sheets.

Thesis submitted by
Vincent Verjans
in March 2021

for the degree of Doctor of Philosophy in Environmental Science

Abstract

Ice sheets form from snow which has been compacted into ice over many years. As snow transitions into ice, it passes through an intermediate stage known as firn. Both the Greenland and Antarctic ice sheets are covered by a firn layer that can be up to ~150 m thick, with lighter firn closer to the surface and denser firn at depth. Firn models aim at representing densification processes as a function of climate (e.g. snowfall and temperature), and model simulations can be performed on large spatial scales, e.g. for ice sheet mass balance assessments, or at individual locations, e.g. for analyses of ice cores. Firn densification models are empirical models, thus relying on calibration with observational data. In this context, there are several sources of inter-model discrepancies: calibration methodology, calibration data, level of model complexity, physical processes included and their parameterisation, and climatic input forcing. For all these reasons, uncertainty about firn densification results is large, yet difficult to evaluate. This thesis aims to better quantify uncertainty from these aspects, and their impact on firn model output.

In Chapter 1, a new model implementation for simulating meltwater percolation in firn models is presented and evaluated. Representing meltwater infiltration and the subsequent effects on firn densification and temperature properties remains a primary source of uncertainty in firn models. This study provides a novel and physics-based approach to represent the meltwater flow process and its interplay with firn densification.

Uncertainty associated with model parameterisation and with the calibration process is addressed in Chapter 2. Using a large dataset of firn cores, a Bayesian calibration framework is designed and used in order to re-estimate model parameter values. This statistical approach further allows to quantify parametric uncertainties and their direct impact on modelled densification rates.

Chapter 3 is the first study to thoroughly evaluate uncertainty in firn thickness changes at the ice sheet scale, the East Antarctic ice sheet more specifically. Using statistical emulation of firn densification models, a large ensemble of model simulations that accounts for the different uncertainty sources in firn model simulations is constructed. Statistical analysis of the results evaluates spatial patterns of the resulting total uncertainty and quantifies the contributions of the different uncertainty sources. Finally, this study demonstrates the direct impact of such firn model uncertainties on the interpretation of altimetry-based mass balance assessment in East Antarctica.

Together, these chapters bring novel firn physics, statistical techniques, and uncertainty evaluation methods into firn modelling. They contribute to firn model accuracy and robust estimation of uncertainties, both being crucial to several important glaciological applications.

Acknowledgments

I am grateful to my primary supervisor, Amber Leeson. Amber has guided my early steps in the world of Science and has taught me a lot about turning ideas into research. I thank her also for giving me a lot of freedom in my research. She went even further than that by prompting my interest in a variety of subjects. Thanks to her, I could learn about several topics and I could develop genuine scientific curiosity. Whatever the future holds, I will never forget her role in my scientific journey.

I also thank my secondary supervisors, who helped me at different stages of my thesis. Mal McMillan shared with me his experience and knowledge about glaciology and about scientific rigour. Our discussions were sometimes lengthy and about what could be considered as minor details, but they were always interesting and very instructive. Chris Nemeth was patient enough to teach me about statistics and their application. He contributed significantly to my scientific skills, whatever the p-value. Finally, the experience and insights from Keith Beven were very helpful in the early stages of my PhD.

The contribution of Max definitely merits a special recognition. He has been an astounding modelling mentor. With hindsight, collaborating on the CFM development with him was among the aspects of my PhD that I enjoyed the most. Our discussions about new codes, and most often about code debugging, were challenging and exciting at the same time. Max has played an important role in my enthusiasm for glaciological modelling.

I want to thank my close family for supporting me through my entire PhD. It was also my first experience living far away from them, and that has posed some unprecedented challenges. While there have been ups and downs during these 3.5 years, our relation means a lot to me. I particularly thank my mum. Every single day, her support, advice and encouragements are essential to me.

I have sincerely appreciated the company of friends and colleagues who have shared the PhD experience with me, and I would like to thank two of them in particular: Antoine and Margherita. Despite the distance, it has been important to have them by my side.

Finally, I thank the many scientists, most of them glaciologists, that contributed to this crucial step of my scientific life. This includes the Lancaster glaciology group, researchers that I met at conferences, my co-authors, and reviewers. I believe that our work has a common goal, and this should not be forgotten.

"In questions of science the authority of a thousand is not worth the humble reasoning of a single individual." – Galileo Galilei (1612)

"Science and everyday life cannot and should not be separated." – Rosalind Franklin

"The best thing about firn modelling is that Herron and Langway works. The worst thing about firn modelling is that Herron and Langway works." – Anonymous firn modeller

Table of contents

List of publications

Publications included in this thesis

Official statements of authorship are provided in Appendix D.

I.

Verjans, V., Leeson, A., Stevens, C. M., 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, 1819-1842, https://doi.org/10.5194/tc-13-1819-2019, 2019.

This publication was submitted on the 24th of January 2019 and published on the 9th of July 2019 in the journal The Cryosphere.

VV and AL conceived this study. VV developed the new firn model schemes for liquid water, performed model experiments, analysed the results and led writing the manuscript. AL supervised the work and contributed extensively to writing of the manuscript. MS developed the original firn model and contributed to the development of the model schemes and model experiments. MMF provided the firn core data. BN and MvdB provided the climatic forcing data. All authors provided comments and suggested edits to the manuscript.

II.

Verjans, V., Leeson, A. A., Nemeth, C., Stevens, C. M., Kuipers Munneke, P., Noël, B., and van Wessem, J. M.: Bayesian calibration of firn densification models, The Cryosphere, 14, 3017–3032, https://doi.org/10.5194/tc-14-3017-2020, 2020.

This publication was submitted on the 15th of November 2019 and published on the 15th of September 2020 in the journal The Cryosphere.

VV, AL and CN conceived this study. VV developed the statistical calibration method, performed the model experiments, analysed the results and led writing the manuscript. AL and CN supervised the work. AL assisted in writing the manuscript. MS developed the original firn model. PKM provided firn core data. BN and JMvW provided the climatic forcing data. All authors provided comments and suggested edits to the manuscript.

III.

Verjans, V., Leeson, A. A., McMillan, M., Stevens, C. M., van Wessem, J. M., van de Berg, W. J., van den Broeke, M. R., Kittel, C., Amory, C., Fettweis, X., Hansen, N., Boberg, F., and Mottram, R.: Uncertainty in East Antarctic firn thickness constrained using a model ensemble approach, Geophysical Research Letters, https://doi.org/10.1029/2020GL092060, 2021.

This publication was submitted on the 11th of December 2020 and published on the 15th of March 2021 in the journal Geophysical Research Letters.

VV conceived this study with guidance from AL and MMM. VV developed the statistical emulator, performed the model experiments, analysed the results and led writing of the manuscript. AL and MMM supervised the work and assisted writing the manuscript. MS developed the original firn model. JMvW, WJvdB, MvdB, CK, CA, XF, NH, FB and RM provided the climatic forcing data. MvdB, CK, CA, XF, NH and FB provided comments and suggested edits to the manuscript.

Publications not included in this thesis, but published during the PhD project

IV.

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, 2020.

VV contributed to the development of the "Community Firn Model" and assisted writing the manuscript.

V.

Vandecrux, B., Mottram, R., Langen, P. L., Fausto, R. S., Olesen, M., Stevens, C. M., **Verjans, V.**, Leeson, A., Ligtenberg, S., Kuipers Munneke, P., Marchenko, S., van Pelt, W., Meyer, C. R., Simonsen, S. B., Heilig, A., Samimi, S., Marshall, S., Machguth, H., MacFerrin, M., Niwano, M., Miller, O., Voss, C. I., and Box, J. E.: The firn meltwater Retention Model Intercomparison Project (RetMIP): evaluation of nine firn models at four weather station sites on the Greenland ice sheet, The Cryosphere, 14, 3785–3810, https://doi.org/10.5194/tc-14-3785-2020, 2020.

VV led the contribution of the "Community Firn Model" to the model intercomparison by performing the model experiments. VV provided comments and suggested edits to the manuscript.

List of tables

List of figures

0. General introduction

0.1 The polar ice sheets

The cryosphere is one of the major dynamic components of the Earth system, interacting with the atmosphere, the ocean, the lithosphere and the biosphere. The Greenland and Antarctic ice sheets (GrIS and AIS, respectively) are the largest ice reservoirs of the cryosphere: combined, they account for more than 99% of all ice on Earth, and for ~68% of all freshwater (Shiklomanov and Rodda, 2003). Paleoclimatic records show that variations in the size and extent of ice sheets depend on global climatic changes (Alley et al., 2005). In turn, ice sheets also have a profound impact on the global Earth system. For example, they reflect incoming solar radiation, release freshwater into the oceans, dictate global sea level and change the land surface elevation. Since the early 1990's, the combined loss of ice from the GrIS and AIS is ~6600 billion tons, which corresponds to a global sea level contribution of +18 mm (Shepherd et al., 2018, 2020). This has many implications for the other components of the Earth system and for societies worldwide, compounded by the increasing rate of mass loss (Meredith et al., 2019). For these reasons, understanding the dynamical changes of the GrIS and AIS in the context of current climatic changes is a major scientific challenge. This includes a finer process-based understanding of ice sheet mass balance, as well as an improved ability to perform and interpret measurements of ice sheet changes.

0.2 Definition of firn

On the GrIS and AIS, the area where snow accumulation exceeds melting is defined as the accumulation zone, and it spans the majority of both ice sheets. In the accumulation zone, snow layers are progressively buried under more recently deposited snow, which causes compaction of the material. Snow is ultimately transformed into ice, and this process is the necessary condition for the inception and progressive growth of ice sheets. Freshly fallen snow is very loosely compacted and can have density values as low as 50 kg m^{-3} (Cuffey and Paterson, 2010). Progressive densification reduces the air content and ice is defined as the stage where interconnections between pores in the solid material are closed. This pore close-off density is ~830 kg m^{-3}, but fully densified ice has typical density values of ~917 kg m^{-3}. Firn is commonly defined as the intermediate stage of transformation between freshly fallen snow and ice (Cuffey and Paterson, 2010), and thus its density can span a large range of values (e.g. Fig. 0.1).

Throughout densification, at the microstructural level, ice crystals transform due to their mutual displacement, to changes in their size and shape, and to internal deformation. In addition to possessing a large range of density values, firn also therefore has highly variable microstructural properties, and

thus the mechanisms involved in the densification process vary accordingly. In the absence of meltwater influence, dry firn densification is commonly separated in distinct stages (Fig. 0.1), differing by their governing mechanisms (Maeno and Ebinuma, 1983). In the first stage of densification, rounding and packing of the ice particles, known as grain boundary sliding (Alley, 1987), dominate until firn reaches a critical density of ~550 kg m^{-3}. At densities higher than this critical density, sintering becomes the dominant process. Sintering of ice particles occurs through several transport mechanisms: volume diffusion, surface diffusion, grain boundary diffusion, sublimation, and plastic deformation (Blackford, 2007). Different mechanisms dominate at different stages of the densification process, but they all act simultaneously. The transition from the second to the third stage of densification was suggested by Maeno and Ebinuma (1983). From this second critical density (~730 kg m^{-3}), remaining air is confined to thin channels along the intersections of grain boundaries. The last critical density corresponds to the pore close-off density. In the fourth stage, further slow densification results from the compression of air bubbles by plastic deformation of ice, also called ice creep. The manifestation of these four stages is displayed in Figure 0.1, showing that the relationship between density and overburden pressure is non-linear and varies through each of these stages. Parts of the accumulation zones of the Greenland and Antarctic ice sheets are affected by surface melt in summer, and so in these areas wet firn densification must also be considered. This includes refreezing of meltwater – a key process affecting firn densification. Meltwater can percolate through the firn, be retained by capillary forces and refreeze at depths where the firn has subfreezing temperature conditions. When this melt-refreezing process occurs, it efficiently contributes to firn densification and raises firn temperature through latent heat release.

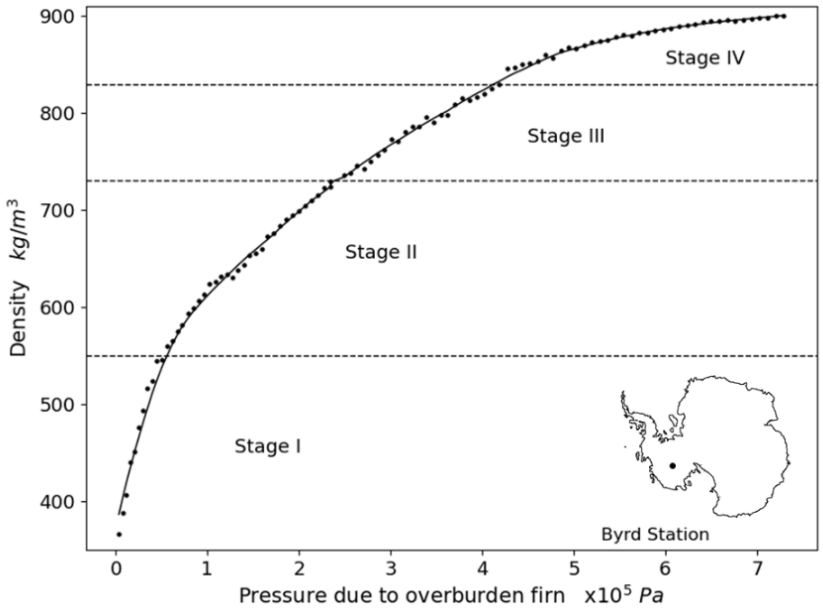


Figure 0.1. Firn density – overburden pressure relationship illustrating the four different stages of dry firn densification, which are governed by different mechanisms. Data are from Gow (1968) and was collected at Byrd Station, Antarctica, in the dry snow zone.

The speed of firn densification varies strongly and depends on climatic conditions. Densification rates have an Arrhenius-type dependence on temperature (Arthern et al., 2010), with some studies supporting a temperature-dependence of the activation energy (Zwally and Li, 2002; Scapozza and Bartelt, 2003). The rate of snowfall determines the overburden pressure exerted on the firn column. Consequently, seasonal, multi-annual and long-term changes in climate affect the GrIS and AIS firn thickness, density and air content properties.

The climatic sensitivity of firn densification also results in spatial differences of densification rates, associated with local climatic patterns. Additionally, complex variabilities in firn microstructural properties affect the several mechanisms that govern densification. As such, modelling firn densification with fidelity at the scale of ice sheets, which span areas of ~1.7×10^6 (GrIS) and ~14×10^6 (AIS) km^2, represents a major challenge.

0.3 Importance of firn modelling

Firn densification modelling has several important applications in glaciology. First, fluctuations in firn thickness affect surface elevation of the ice sheets. The advent of polar altimetry satellite missions in the early 1990's opened the way for regular measurements of ice sheet elevation changes (e.g. Shepherd et al., 2018; 2020). These measurements are subsequently converted to a mass change in order to evaluate the mass balance of the ice sheets and thus their contribution to global sea level. This volume to mass conversion relies on accurate partitioning of the elevation change signal between firn densification variability and changes in ice flow below the firn column. Accounting for variability in firn densification rates, and short-term fluctuations in firn thickness, is therefore crucial to ensure the accuracy of altimetry-based mass balance assessments, and for identifying long-term trends in mass changes (Morris and Wingham, 2015). Secondly, in areas affected by surface melt, pore space within the firn provides a substantial reservoir for the retention of meltwater. Infiltration and refreezing of surface meltwater in the porous and cold underlying firn column mean that firn has the capacity to delay the runoff of meltwater to the ocean. This process acts as a buffer on the contribution to sea-level rise from increases in surface melt and plays an important role in ice sheet Surface Mass Balance (SMB). This buffer effect is especially important on the GrIS, where summer melt affects ~64% of the total area, and this percentage of melt-affected areas has been increasing over recent decades (Tedesco and Fettweis, 2020). Model studies show that ~40% of the GrIS surface meltwater is estimated to refreeze in firn, but that this buffer capacity may plummet as firn porosity becomes saturated due to increasing melt rates (van Angelen et al., 2013; van den Broeke et al., 2016). Finally, accurate estimations of firn densification rates are important for paleoclimate analyses based on ice cores (Goujon et al., 2003). Due to the porosity of firn, air bubbles become isolated from the

atmosphere only once the pore close-off density is reached. Past air composition is inferred from air bubbles trapped in ice cores, while past climatic conditions are inferred from isotopic signals in the surrounding ice matrix. The time needed for deposited snow to reach pore close-off density implies that ice can be significantly older than air bubbles entrapped at the same vertical level. This age lag of the gas depends strongly on firn densification rates: it is typically several centuries but can be as high as ~4000 years in the cold and dry interior of the AIS (Loulergue et al., 2007). As such, accurate firn models are required to link chronologies of past changes in temperature and atmospheric composition with precision.

These aspects have motivated research into firn model development over the past 40 years (e.g. Herron and Langway, 1980; Barnola et al., 1991; Arthern and Wingham 1998; Arthern et al., 2010; Li and Zwally, 2011; Ligtenberg et al., 2011; Morris and Wingham, 2014). Approaches to firn modelling can be separated into two main categories. First, the microscale approach seeks to express densification through governing equations that represent all the microphysical processes at play (Alley 1987; Arthern and Wingham 1998; Fourteau et al., 2020). Due to the many complexities associated with this approach, most models have instead focused on a macroscale approach. This entails formulating bulk densification equations, which are assumed to encompass all the relevant mechanisms. The macroscale approach incorporates tunable model parameters, and thus relies on calibration of these parameters against field measurements. Firn strain rates, defined as thinning of the firn column, can provide direct observations of firn compaction rates. But continuous strain measurements are logistically difficult to acquire and thus relatively scarce. A common method for firn model calibration is to use measured depth-density profiles derived from firn cores instead. However this approach implies the assumption that the firn column is in steady-state (Sorge's Law – Bader, 1954). By making this assumption, time derivatives can be substituted for space derivatives, hence linking depth-density profiles to densification rates.

0.4 Existing firn models

The Herron and Langway (1980) model is considered a benchmark in firn modelling and is the foundation for many later developments. In this model, densification results from the progressive compression of firn, and thus densification rates are linked to strain rates:

$$\dot{\varepsilon} = \frac{-1}{\rho}\frac{d\rho}{dt} \tag{0.1}$$

where ρ is firn density (kg m^{-3}) and $\dot{\varepsilon}$ is the volumetric strain rate (s^{-1}). The strain rate is assumed to be a function of firn density with a certain proportionality coefficient, or rate coefficient, to be determined:

$$\dot{\varepsilon} = \frac{-1}{\rho}\, c\, (\rho_i - \rho) \tag{0.2}$$

where c is the rate coefficient (s^{-1}) and ρ_i is ice density (917 kg m^{-3}). The dependence of strain rates on firn density illustrates that compaction must account for structural changes of firn as density increases. The rate coefficient captures the contributions to strain rates from all factors not dependent on the firn density itself. In the Herron and Langway (1980) model, the rate coefficient only depends on temperature and on the rate of increase in overburden stress, approximated by the snow accumulation rate.

The prevalent underpinning of the macroscale approach is to express the rate coefficient through an Arrhenius-type equation to model densification rates; Herron and Langway (1980) were the first to propose its form:

$$c = k\ \dot{b}^{a} \exp\left(\frac{-E_c}{R\ T}\right) \quad (0.3)$$

Expressing the governing equation for densification rates by substitution:

$$\frac{d\rho}{dt} = (\rho_i - \rho)\ k\ \dot{b}^{a} \exp\left(\frac{-E_c}{R\ T}\right) \quad (0.4)$$

where R is the universal gas constant (8.314 J K^{-1} mol^{-1}), T is temperature (K) and $\dot{b}$ is the snow accumulation rate (m w.e. yr^{-1}). The factor k, the exponent a and the activation energy E_c (J mol^{-1}) are considered as calibration parameters. This framework captures the strong sensitivity of densification rates to temperature, quantified by the activation energy. It also indirectly accounts for overburden pressure by using the accumulation rate as a proxy for the rate at which the weight of the material overlying any given firn layer increases. The exponent a characterises whether this relation is linear or nonlinear. Its value was determined by comparing measured rate coefficients at sites of similar temperature but different accumulation rates.

Herron and Langway (1980) calibrated their model with 17 firn cores from GrIS and AIS. They suggested a critical density of ρ = 550 kg m^{-3} at which the values of the three calibration parameters change, hence separating the modelled densification process into two different stages. This model form has served as the basis for most of the firn densification models since then. Model developments have primarily focused on refining the parameter estimates and on adding additional terms to the governing equation to approximate effects of other factors such as snow type and impurity content (e.g. Arthern et al., 2010; Freitag et al., 2013; Morris and Wingham, 2014). The use of the climatic variables has also progressed from using only mean annual accumulation and surface temperature to calculating the overburden pressure (Li and Zwally, 2011) and computing the firn vertical temperature profile by coupling the densification equation to the heat diffusion equation (Goujon et al., 2003). Importantly, these advancements relax the steady-state assumption and thus mean that firn densification models can be used dynamically. Models have generally been calibrated to special regions of interest, and most separate GrIS and AIS (e.g. Li and Zwally, 2011; Ligtenberg et al., 2011; Kuipers Munneke et al., 2015b; Li and Zwally, 2015).

Most firn densification models are numerical, finite-difference and one-dimensional vertical models, in which each firn layer in depth is treated as a separate volume. Lagrangian discretisation is most common (e.g. Kuipers Munneke et al., 2015b; Stevens et al., 2020), but some models take an Eulerian approach (e.g. Colin and Meyer, 2017; Langen et al., 2017). In the former, individual firn layers are modelled and are progressively buried in the firn column. In the latter, firn layers of fixed depth and/or mass constitute the model domain. At the surface boundary, models require climatic forcing which will dictate the evolution of firn density in depth and in time (e.g. Eq. 0.4). Ideally the forcing comes from weather station data, because they provide, in theory, the most accurate observations of local meteorological variables. But on the ice sheets such stations are sparse in space, have limited coverage in time, can present data gaps and sometimes require corrections. For paleoclimatic studies, climatic fields determined from ice-core data provide multi-millennial forcing at a temporally coarse resolution (e.g. Buizert et al., 2015). However, ice sheet wide studies aimed at determining recent elevation changes attributable to firn or estimating meltwater retention within the firn column require both large-scale and temporally fine climatic forcing. These applications thus rely on Regional Climate Models (RCM) or atmospheric reanalysis products to provide continuous time series of temperature, snow accumulation and snow melt covering the entire ice sheets.

Recent advances in ice sheet studies have highlighted the importance of percolation, retention and refreezing of meltwater for several reasons (van As et al., 2016a; Vandecrux et al., 2020). First, the extent of the GrIS affected by seasonal surface melt is increasing (Tedesco and Fettweis, 2020). As such, deep firn layers with high retention capabilities have newly become exposed to meltwater infiltration – increasing the importance of refreezing in GrIS mass balance calculations. Second, perennial firn aquifers have been discovered in some areas of the GrIS accumulation zone (Foster et al., 2014; Miège et al., 2016). These perennial liquid water bodies deep in the firn column could well store substantial amounts of meltwater (Koenig et al., 2014) and could influence ice flow dynamics if they drain through vertical crevasses towards the bed of the ice sheet (Poinar et al., 2019). Third, water retention in firn on ice shelves is thought to be a key factor in ice shelf stability. Saturation of the firn can trigger the process of hydrofracturing, whereby water-filled crevasses develop and decrease the mechanical stability of the ice shelf (Kuipers Munneke et al., 2014b). Progressive firn saturation is hypothesised to have been influential in the notorious 2002 collapse of the Larsen B ice shelf in Antarctica (Leeson et al., 2020), and there is recent evidence of perennial firn aquifers on ice shelves (Montgomery et al., 2020). For these reasons, it is now clear that a better understanding of the physics of meltwater behaviour in firn, and models that adequately capture this behaviour, are needed.

0.5 Uncertainty in firn modelling and contribution of this work

Addressing model uncertainty is a vast and continuous process. In academic research uncertainty is traditionally defined in two different ways (Beven, 2009). First, it can be perceived as the existence of a set of possible alternatives of which only one is valid. The limitation of this approach is that any element must be either in- or outside of the set of alternatives. Second, in statistics, probability theory allows to express uncertainty quantitatively by allocating some measure ranging between 0 and 1 to any alternative. The latter approach usually treats uncertainty as originating from randomness, called aleatory uncertainty. But uncertainty can also arise from lack of information and knowledge, referred to as epistemic uncertainty. Both aleatory and epistemic uncertainty arise in environmental modelling.

In the case of firn densification, aleatory uncertainty can be associated with different snow crystal structures as they fall from the sky, randomness in the agency of the crystals once they settle and the intrinsic randomness of climatic variability, for example. Such uncertainties are inherent to firn densification and are thus irreducible by nature, but they should be quantified as accurately as possible. To date, epistemic uncertainty likely contributes more to uncertainty in firn model output. Different sources of epistemic uncertainty, as identified by Refsgaard et al. (2007), and which are directly applicable to firn densification modelling are:

- Input uncertainty, i.e. the uncertainty inherent to the climatic time series provided by RCMs, weather station data or paleoclimatic records that are used to force the firn model. It also includes uncertainty in the snow density at the onset of the densification process.
- Model structure uncertainty. This is likely the most pressing issue in firn modelling. It originates from our incomplete understanding of firn compaction physics and associated processes such as meltwater percolation. One example of this is that a simple Arrhenius relation is not fully representative of the true physics governing densification, but is a simple and useful approximation.
- Parameter uncertainty relates to the exact values of model parameters being unconstrained. This is particularly relevant to firn models because their development relies on different calibration techniques but also on different type of calibration data and on data collected at different locations and in different climates.
- Model technical uncertainty arises from the computer implementation of the model. It includes the space and time resolution of model simulations, as well as coding practices of the model developer and numerical approximations from the software.

Ultimately, all these sources of uncertainty contribute to firn model output uncertainty.

While the importance of firn densification model accuracy is well-established, little effort has actually been dedicated to address uncertainty in and from firn models to date. Many altimetry studies do not

consider firn thickness and firn density uncertainties, and others approximate these using temporal variability in RCM time series of snowfall and in modelled firn densities (McMillan et al., 2016; Shepherd et al., 2019). Ice core studies typically account for uncertainty in the climatic forcing and in the exact value of the pore close-off density, but not in the firn model output itself (Capron et al., 2013). And, despite our lack of understanding of meltwater behaviour in firn, ice sheet wide SMB studies rely on a single model output as a best estimate for meltwater refreezing. This approach is questionable because it does not account for the well-known limitations of meltwater schemes in firn models and for the large ranges of possible parameter values in the schemes, which are poorly constrained. For uncertainty calculations, they use average discrepancies of the RCM accumulation and melt fields with respect to observations but neglect spatial variability in these values and uncertainty in the firn model (van den Broeke et al., 2016). These shortcomings underline how much progress must be achieved by the firn modelling community to satisfy better scientific rigour. In particular, the numerous sources of uncertainty in firn model output need to be addressed, as illustrated above.

The firn model intercomparison experiment of Lundin et al. (2017) was an important first step in uncertainty assessment. They used synthetic steady-state climatic forcing followed by a step change in either accumulation or temperature to evaluate differences in both steady and transient responses from eight firn densification models. Analysing key metrics such as depth integrated porosity (DIP, also referred to as firn air content – FAC) and pore close-off depth, they showed significant disagreement between models in steady-state conditions. Importantly, the disagreement was further amplified in the transient responses, despite the simplicity of the step change. This likely has major implications with respect to firn model agreement in simulations forced with highly variable time series of polar climate.

This thesis focuses on specific issues related to uncertainty in firn densification models. It addresses structural uncertainty by improving model representation of firn processes. Also, parametric uncertainties are investigated to better constrain the empirical nature of firn models. Finally, it focuses on quantifying firn model output uncertainty and its implications. The research performed for this thesis has been continuously associated with the development of the Community Firn Model (CFM, Stevens et al., 2020). The CFM is an open source model that facilitates the applicability of firn modelling for various research interests among the glaciology community. Improving the model and extending its applicability have been an integral part of this thesis. The CFM also contributes to reduce technical uncertainties as defined above: the model code is openly available and can thus be accessed, used and verified by all scientists.

In order to address structural uncertainty associated with representation of meltwater behaviour, Chapter 1 describes the implementation of advanced liquid water flow schemes to simulate meltwater percolation and refreezing. The newly implemented schemes are based on hydrological theory of liquid water flow in porous media. Model results are compared to simple empirical schemes and to firn observations from the percolation zone of the GrIS in order to assess their benefits and shortcomings. The findings demonstrate that model results in melt areas are sensitive not only to the meltwater scheme, but also to the parameterisation of firn compaction equations. In this light, further focus is dedicated to firn model calibration in Chapter 2. The latter is an in-depth evaluation of parametric uncertainty in three benchmark firn models. Bayesian statistical methods are used to re-calibrate the firn models and to better constrain parameter uncertainty. This statistical application makes use of an extensive set of firn cores from both the AIS and the GrIS in order to overcome the location-specific aspect of model parametrisations. It also incorporates input and aleatory uncertainties by accounting for stochastic climatic perturbations in the calibration. The results demonstrate that, despite being calibrated with the same method and dataset, inter-model disagreements remain significant. This motivates large-scale intercomparisons of model results to evaluate how such disagreements translate into firn model output uncertainty for ice sheet wide applications. With this aim, Chapter 3 investigates implications of firn model output uncertainty on altimetry-based mass balance assessments of the East Antarctic ice sheet (EAIS). Such assessments on the EAIS are particularly sensitive to uncertainty and fluctuations in firn thickness. By applying statistical emulation of firn densification models, an ensemble of model scenarios of 1992-2017 firn thickness change is developed. This ensemble approach allows to evaluate uncertainty related to climatic forcing, boundary conditions on surface snow density and choice of firn densification model. Ultimately, the ensemble quantifies modelled firn thickness change, estimates the associated uncertainty, and provides a direct basis of comparison for satellite measurements of surface elevation changes.

1. Development of physically based liquid water schemes for Greenland firn-densification models

1.1 Abstract

As surface melt is increasing on the Greenland ice sheet (GrIS), quantifying the retention capacity of the firn layer is critical to link meltwater production to meltwater runoff. Firn-densification models have so far relied on empirical approaches to account for the percolation-refreezing process, and more physically based representations of liquid water flow might bring improvements to model performance. Here we implement three types of water percolation schemes into the Community Firn Model: the bucket approach, the Richards Equation in a single-domain and the Richards Equation in a dual-domain, which accounts for partitioning between matrix and fast preferential flow. We investigate their impact on firn densification at four locations on the GrIS and compare model results with observations. We find that for all of the flow schemes, significant discrepancies remain with respect to observed firn density, particularly the density variability in depth, and that inter-model differences are large (porosity of the upper 15 m firn varies by up to 47 %). The simple bucket scheme is as efficient in replicating observed density profiles as the single-domain Richards Equation and the most physically detailed dual-domain scheme does not necessarily reach best agreement with observed data. However, we find that the implementation of preferential flow simulates ice layer formation more reliably and allows for deeper percolation. We also find that the firn model is more sensitive to the choice of densification scheme than to the choice of water percolation scheme. The disagreements with observations and the spread in model results demonstrate that progress towards an accurate description of water flow in firn is necessary. The numerous uncertainties about firn structure (e.g. grain size and shape, presence of ice layers) and about its hydraulic properties, as well as the one dimensionality of firn models render the implementation of physically based percolation schemes difficult. Additionally, the performance of firn models is still affected by the various effects affecting the densification process such as microstructural effects, wet snow metamorphism and temperature sensitivity when meltwater is present.

1.2 Introduction

Estimating the properties of the firn layer – and how it evolves under a warming climate – is a critical step in measuring the ice sheets' contribution to sea level rise, yet it remains one of the key sources of uncertainty in present assessments (McMillan et al., 2016). Accurate estimates of firn thickness and density are required for the conversion of space-borne measurements of volume change into mass change (e.g. McMillan et al., 2016; Shepherd et al., 2018). Also, assessments of the Greenland and Antarctic ice sheets' contribution to sea level require information on firn density and spatial distribution in order to calculate meltwater retention potential and the capacity of firn to buffer the

flow of meltwater to the ocean (Harper et al., 2012; Machguth et al., 2016; van den Broeke et al., 2016). Surface melting has become more widespread and intense on the Greenland Ice Sheet (GrIS), with annual total melt rates rising by 11.4 Gt yr^{-2} between 1991 and 2015 (van Angelen et al., 2014; van den Broeke et al., 2016). This melt water percolates into the firn layer, where it can refreeze, run off, or remain liquid in temperate firn. Refreezing of liquid water in firn, known as internal accumulation, has an impact both on ice-sheet mass balance and on heat fluxes from the surface to the ice sheet (van Pelt et al., 2012). As such, understanding physical processes in firn, including in particular the transport of liquid water, is becoming increasingly important in order to accurately constrain and predict the mass balance of the GrIS (van den Broeke et al., 2016).

Densification of dry firn is typically modelled as a function of near-surface air temperature and accumulation (Herron and Langway, 1980; Arthern et al., 2010; Li and Zwally, 2011; Simonsen et al., 2013; Morris and Wingham, 2014; Kuipers Munneke et al., 2015a,b). If applied to wet firn, these models are often modified to include a simplified representation of liquid water percolation, the bucket scheme, which assumes flow and refreezing through the firn column occur in a single time step (Simonsen et al., 2013; Kuipers Munneke et al., 2015b; Steger et al., 2017a). Observations have shown that, in reality, liquid water transport in firn is characterised by flow patterns that are heterogeneous in space and time (Pfeffer and Humphrey, 1996; Humphrey et al., 2012). Incorporation of liquid water schemes representing such flow patterns would enable models to better represent the transport of mass and heat through the firn; these schemes might also improve modelled densification in wet firn conditions (Kuipers Munneke et al., 2015a; van As et al., 2016a; Meyer and Hewitt, 2017). Liquid water flow however, is a complex function of several properties and processes that are difficult to constrain by observations and, as a corollary, are difficult to represent in these models (e.g. presence of impermeable ice layers, snow hydraulic properties, grain-size, lateral runoff). The infiltration of water through firn can be partitioned between the progressive advance of a uniform wetting front through the pores, called matrix flow, and fast, localised, preferential flow (Colbeck, 1979; Waldner et al., 2004; Katsushima et al., 2013). This dual-nature of water flow has been reported in observations of the firn layer of the GrIS, where preferential flow pathways come in the form of discrete vertical conduits and are crucial to effectively transport surface meltwater in deep subfreezing firn (Pfeffer and Humphrey, 1996; Parry et al., 2007; Humphrey et al., 2012; Cox et al., 2015). The detection of Perennial Firn Aquifers (PFA), in which large amounts (140 Gt) of liquid water is stored year round in deep firn, further emphasises the importance of firn hydrology on the GrIS mass balance (Forster et al., 2014; Koenig et al., 2014). Snow modellers have developed liquid water schemes based on the Richards Equation (RE) to simulate matrix flow (Hirashima et al., 2010; Wever et al., 2014; D'Amboise et al., 2017). The RE is a continuity equation describing water flow in unsaturated porous media and is widely used in hydrological models. Recently, a preferential flow scheme has been included in the SNOWPACK model to account for heterogeneous percolation (Wever et al.,

2016). Until now however, such developments have not been implemented in firn-densification models.

In this study, we describe and compare liquid water schemes of different levels of physical complexity from snow models, and we apply these in combination with firn-densification models in order to evaluate the impact of the treatment of liquid water flow on modelled firn densification and temperature. We use the Community Firn Model (CFM) as the modelling framework for our study; the CFM is able to simulate numerous physical processes in firn and includes a large choice of governing formulations for densification. We use the common bucket approach and also develop schemes for liquid water flow in firn following physically based advances in snow models (Wever et al., 2014, 2016; D'Amboise et al., 2017).

We simulate liquid water flow and firn densification starting from 1980 at four sites on the GrIS: DYE-2, NASA-SE, KAN-U and a PFA site (Fig. 1.1). These sites were chosen because they are collocated with recently drilled firn cores which allow a direct comparison of model results with observations. By comparing simulated firn densification to observations at these sites, we investigate the sensitivity of the system to the choice of liquid water flow scheme and the sensitivity of the flow schemes to various parameterisations of firn structural properties. Finally, we perform simulations with a range of firn-densification formulae and assess the relative importance of the choice of liquid water flow scheme to the choice of the underlying densification equation.

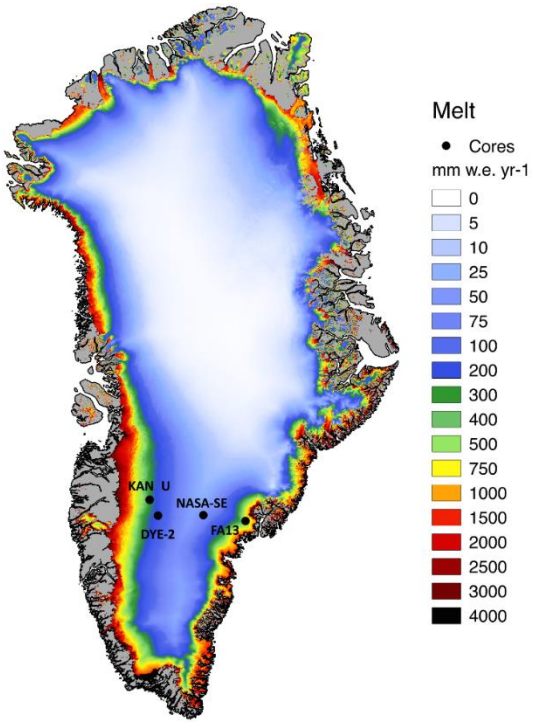


Figure 1.1. Study sites locations and mean annual melt rates (1958-2017) from RACMO2.3p2

1.3 Firn Model and Data

In this study we use and further develop the CFM, an open-source firn-densification modelling framework. We refer the reader to Stevens (2018) for details and briefly summarise the main characteristics here. The CFM is one-dimensional and works in a Lagrangian framework; it is forced at its upper boundary by observed or modelled values for accumulation, surface temperature, surface density, rain, and snow melt. The CFM includes many of the commonly used dry firn-densification schemes (e.g. Herron and Langway, 1980; Helsen et al., 2008; Arthern et al., 2010; Li and Zwally, 2011; Morris and Wingham, 2014; Kuipers Munneke et al., 2015b). We refer the reader to the original publications for details on the different densification schemes and briefly outline the expressions used in our simulations in this section.

1.3.1 Firn-densification model

As a base case, we use the firn-densification formulation implemented in the snow model CROCUS (Vionnet et al., 2012), Eq. (1.1). It has previously been used in model studies of firn densification on the GrIS and also on polar ice caps (Gascon et al., 2014; Langen et al., 2017). This model is formulated so that densification is based on the overburden stress:

$$\frac{d\rho}{dt} = \rho \, \frac{\sigma}{\eta} \tag{1.1}$$

where ρ is the density of the firn (kg m^{-3}), σ is the stress due to weight of the upper layers (kg m^{-1} s^{-2}) and η is the snow viscosity (kg m^{-1} s^{-1}) following the parameterisation:

$$\eta = f_1 f_2 \eta_0 \frac{\rho}{c_\eta} \exp\left[a_\eta (T_0 - T) + b_\eta \rho\right] \tag{1.2}$$

where η_0 = 7.62237 kg s^{-1} m^{-1}, a_η = 0.1 K^{-1}, b_η= 0.023 m^3 kg^{-1}, T is the firn temperature (K) and T_0 = 273.15 K. The parameter c_η is set to 358 kg m^{-3} as suggested by van Kampenhout et al. (2017) when using Eq. (1.1) for polar firn. There are two additional correction factors, f_1 and f_2 , depending on firn microstructural properties. The factor f_1 accounts for the presence of liquid water:

$$f_1 = \frac{1}{1+60\,\theta} \tag{1.3}$$

where θ is the volumetric water content (m^3 m^{-3}). In this study, we neglect the change in snow viscosity for grain-sizes smaller than 0.3 mm by keeping the constant value f_2 = 4 (after Langen et al., 2017; van Kampenhout et al., 2017).

Several firn-densification equations have been derived and calibrated for the GrIS specifically. We favoured the use of Eq. (1.1) as our base case because (i) most of these calibrated schemes were developed for dry firn densification whereas the CROCUS formulation accounts for the presence of liquid water explicitly, (ii) applying a percolation scheme in a stress-based densification model rather

than in an accumulation-rate-based model ensures that the redistribution of mass associated with percolation will affect the densification appropriately and (iii) the CROCUS densification scheme is currently used by the regional climate model MAR and by the Earth System Model CESM to quantify firn densification on the GrIS (Fettweis et al., 2017; van Kampenhout et al., 2017).

1.3.2 Climatic forcing

To force the model at its upper boundary we use three-hourly skin temperature, melt, snowfall, rain and sublimation fields simulated by the latest version of the RACMO2 regional climate model (RACMO2.3p2, Noël et al., 2018). This model has a 5.5 km horizontal resolution grid and has been explicitly adapted for use over the polar ice sheets. Discrepancies between the climatic forcing and the real climatic history can bias the firn models' results. For the areas of our study sites (Sect. 1.3.4), we provide detailed statistical comparisons between RACMO2.3p2 output and available observations (see Appendix A1). We further refer to Noël et al. (2018) for a more detailed discussion about the performance of RACMO2.3p2 on the GrIS scale and related uncertainties. Additionally, Ligtenberg et al. (2018) have demonstrated the impact of recent developments in RACMO2.3p2 on firn modelling, mostly yielding improvements in modelled densification.

If the solid input rate (snowfall – sublimation) is negative over a time step, the CFM treats it as a corresponding mass loss in the surface layer; liquid water is evaporated before solid mass gets sublimated. The temperature of a newly accumulated snow layer is defined as the skin temperature at that time step. Deep temperatures in the model are thus mostly determined by the mean surface temperature applied during the spin-up process (Sect. 1.3.5) together with latent heat release through refreezing. We use a Neumann boundary condition for the temperature at the bottom of the domain and use a 250 m deep column to account for the large thermal mass of the ice sheet during the transient run.

In addition to latent heat release due to refreezing, the CFM accounts for heat conduction through the different layers to determine the temperature profile. In accordance with previous firn modelling studies (Kuipers Munneke et al., 2015b; Steger et al., 2017a), we make the firn conductivity, k_s, a function of density following Anderson (1976):

$$k_s = 0.021 + 2.5 \left(\frac{\rho}{1000}\right)^2 \quad (1.4)$$

This formulation for k_s assumes a simple relation between thermal conductivity and density, but we note that alternative formulations have been derived, some including a dependence on microstructural parameters (e.g. Yen, 1981; Löwe et al., 2013). Another boundary condition is the density of every fresh snow layer deposited at the surface. To reduce the sources of possible uncertainties, we simply

use a constant and site-specific surface density according to the surface value of the drilled firn cores instead of a parameterised formulation (see Sect. 1.3.4).

1.3.3 Grain-size

The temporal evolution of grain-size in firn is poorly understood and observational constraints are scarce. However, the grain-size is a key variable for the RE, and the flow schemes used in this study thus require an initial grain-size and a grain-growth rate. For the former, we use the empirical formulation of Linow et al. (2012) derived from observations of snow samples from Antarctica and Greenland:

$$r_0 = (b_0 + b_1(T_{av} - T_0) + b_2\, \dot{b}\, \frac{\rho_i}{\rho_w}) \quad (1.5)$$

where ρ_i is the ice density (917 kg m^{-3}), ρ_w that of liquid water (1000 kg m^{-3}), $\dot{b}$ is the mean annual accumulation rate (m w.e. yr^{-1}), T_{av} the mean annual surface temperature (K) and b_0, b_1 and b_2 are calibration parameters taking the values 0.781 m, 0.0085 m K^{-1} and -0.279 yr (m w.e.)$^{-1}$ respectively. For grain growth rate, the relationship proposed by Katsushima et al. (2009) is applied:

$$\frac{dr}{dt} = \frac{1}{8\, r^2\, 10^9}\; min\left[\frac{2}{\pi}\left(1.28 \times 10^{-8} + 4.22 \times 10^{-10}\left(\theta_{weight,\%}\right)^3\right), 6.94 \times 10^{-8}\right] \quad (1.6)$$

where r is the grain radius (m) and $\theta_{weight,\%}$ is the mass liquid water content expressed in percent and is thus related to θ (see Eq. (1.3)):

$$\theta_{weight,\%} = 100\, \frac{\theta\, \rho_w(\rho_i - \rho)}{\rho_i\, \rho} \quad (1.7)$$

Equation (1.6) combines a wet snow metamorphism formula and a higher limit of growth rate of ice particles, both derived from laboratory measurements.

To study the sensitivity of the model to the grain-size implementation, we also use an alternative option based on the approach for West-Antarctic firn of Arthern et al. (2010) and supported by observations of Gow et al. (2004); the grain radius in newly deposited layers (r_0) has the constant value of 0.1 mm and the grain growth rate is formulated as:

$$\frac{dr}{dt} = \frac{1}{2\, r}\; k_g \exp\left(\frac{-E_g}{RT}\right) \quad (1.8)$$

where E_g is the activation energy for grain growth (42.4 kJ mol^{-1}), R is the gas constant (8.314 J mol^{-1} K^{-1}) and k_g a parameter that takes the value 1.3 10^{-7} m^2 s^{-1}. Note that Eq. (1.8) does not take the impact of liquid water presence on firn metamorphism into account.

1.3.4 Study sites

We perform simulations at four study sites in the percolation zone of the GrIS where the availability of well-documented firn cores allows for model-observations comparisons: NASA-SE, DYE-2, KAN-U and FA13 (perennial Firn Aquifer) (Fig. 1.1). NASA-SE (66.48ºN, 42.50ºW, 2372 m a.s.l.) is located in the upper part of the percolation zone with a mean annual temperature of -20°C and

relatively low melt rates (50 mm yr^{-1}). DYE-2 (66.48ºN, 46.28ºW, 2126 m a.s.l.) is a slightly warmer site (T_{av} = -18°C), and melt is about three times greater than at NASA-SE (150 mm yr^{-1}). KAN-U (67.00ºN, 47.03ºW, 1838 m a.s.l.) is near the equilibrium line altitude and has warmer temperatures (T_{av} = -8°C) and significant melting (280 mm yr^{-1}). FA13 (66.18ºN, 39.04ºW, 1563 m a.s.l., T_{av} = -13°C) is a location known to contain a firn aquifer (Foster et al., 2014). The persistence of deep saturated layers year round is due to the coupling of high melt rates (587 mm w.e. yr^{-1}) with high accumulation rates (1002 mm w.e. yr^{-1}) (Kuipers Munneke et al., 2014a). Multiple firn core density data exist, a large part of which is available in the SUMup dataset (Montgomery et al., 2018). The selection of these four particular sites is motivated by their variety in climatic and glaciological conditions: a cold site with low melt rates, a cold site with high melt rates, a site close to the equilibrium line with substantial refreezing and a site with the presence of a firn aquifer. We perform transient firn-model simulations for each site until the date that a core was drilled. The cores at NASA-SE and DYE-2 were drilled in spring of the years 2016 and 2017 respectively, as part of the FirnCover project. The cores at KAN-U (Machguth et al., 2016) and at FA13 (Koenig et al., 2014) were drilled in spring 2013. The firn temperature measurements were given in the sources mentioned for the density data, except at KAN-U for which it comes from the collocated automatic weather station overseen by the Programme for Monitoring of the Greenland Ice Sheet (PROMICE) (van As et al., 2016b). The fixed surface densities for DYE-2, NASA-SE, KAN-U and FA13 are 325, 240, 325 and 365 kg m^{-3} respectively and were taken in accordance with the surface density of the drilled cores.

1.3.5 Spin-up and domain definition

In order to simulate the evolution of the firn layer in time, we start the transient simulation from an initial state in equilibrium with a reference climate. In accordance with previous GrIS firn studies (Kuipers Munneke et al., 2015b; Steger et al., 2017a), we take the 1960-1979 climate as reference climate because it predates the onset of the general warming of Greenland and the subsequent increase in surface melt. We iterate over the reference climate until 70 m w.e. of snow has been accumulated, which ensures the entire firn column is refreshed. The number of iterations over the reference climate is thus site-specific. This spin-up process starts from an analytical solution for the density profile (Herron and Langway, 1980) with temperatures corrected to account for latent heat release by refreezing (Reeh, 2008). During the spin-up process we use the simple bucket approach, and the more advanced flow schemes, detailed in the next section, are turned on only at the end of the spin-up for the transient simulation. This is because using the advanced schemes over long periods is computationally expensive. The domain on which the flow calculations are applied is a subset of the entire CFM domain; this sub-domain is defined each time the flow routine is called in the transient run. The bottom of the sub-domain is defined as the depth below which all layers have density higher than the pore close-off value (830 kg m^{-3}), because infiltration of liquid water becomes negligible at this point. The thickness of the layers deposited in every three hourly time step determines the vertical

resolution, and we apply a merging process only to individual layers less than 2 cm thick (see Appendix A2.8).

1.4 Liquid Water schemes

The water flow schemes are added to the dry-densification model detailed in Sect. 1.3.1 and are thus also effectively one-dimensional, representing no lateral exchange of heat and mass although lateral runoff is used as a mass sink. In this section, we present the three different flow schemes that we implement in the CFM: (1) the Bucket method (BK), (2) a single-domain Richards Equation scheme (R1M) and (3) a dual-permeability Richards Equation scheme (DPM). Because of its robustness and ease of implementation, BK is the current 'state-of-the-art' in firn-densification models that are interactively coupled to regional climate models. R1M is used in several stand-alone snow models to describe water flow (Hirashima et al., 2010; Wever et al., 2014; D'Amboise et al., 2017), and DPM is entirely based on the scheme implemented in the snow model SNOWPACK (Wever et al., 2016), where dual-permeability means that separate domains for matrix flow and preferential flow coexist with liquid water exchanged between these domains.

1.4.1 Bucket model

The bucket percolation scheme is commonly used to account for the vertical transport of meltwater in firn models, though the precise form of its implementation is variable. Each layer in the model can refreeze meltwater according to its 'cold content', i.e. the energy required to raise the temperature of the layer to the melting point. Starting from the surface, the meltwater may percolate through successive layers, thus allowing for refreezing at depth. Meltwater is progressively depleted due to refreezing and retention according to each layers' water-holding capacity, which is the part of the water that is stored in some of the available pore space and not subject to vertical transfer. The water-holding capacity acts as an approximation of the effect of capillary forces on water retention. Percolation proceeds until all the meltwater is stored (refrozen or retained) or until it reaches a layer with a density exceeding the impermeability threshold (780-830 kg m^{-3}), at which point all the water in excess is instantly treated as lateral runoff. The BK thus requires two parameters: the water-holding capacity and the impermeability threshold. We test two possibilities for the former and three for the latter. The water-holding capacity can be prescribed by the calculations of Coléou and Lesaffre (1998) for the mass proportion of water in a firn layer, W_w:

$$W_w = 0.057 \, \frac{\rho_i - \rho}{\rho} \tag{1.9}$$

This mass proportion is then converted to the water-holding capacity, θ_h:

$$\theta_h = \frac{W_w}{(1 - W_w)} \, \frac{\rho \, \rho_i}{\rho_w (\rho_i - \rho)} \tag{1.10}$$

Using constant values of the water-holding capacity is also common practice (Reijmer et al., 2012; Steger et al., 2017a). Our base case scenario uses a fixed θ_h at 0.02, or 2% of the pore space available

for liquid water retention. This low value assumes effective downward percolation and is meant to account for vertical preferential flow (Reijmer et al., 2012). For that reason, we consider this as a good basis for comparison with the DPM that explicitly accounts for such flow.
We test three values for the impermeability threshold; these were selected in accordance with Gregory et al. (2014), who tested firn permeability of Antarctic samples in a lab and reported that impermeability can occur over density values ranging from 780 kg m^{-3} to 840 kg m^{-3}. We thus take our three test values to be 780 kg m^{-3}, 810 kg m^{-3} and 830 kg m^{-3}, respectively the lower bound and middle of this range and a commonly-used value of pore close-off density.

1.4.2 Richards Equation

Vertical movement of water in a variably saturated porous medium can be described by the one-dimensional version of the RE:

$$\frac{\partial \theta}{\partial t} - \frac{\partial}{\partial z}\left[K(\theta)\left(\frac{\partial h}{\partial z} + 1\right)\right] = 0 \quad (1.11)$$

where K is the hydraulic conductivity (m s^{-1}), h is the pressure head (m of water), z is the vertical coordinate (m, taken positive downwards) and θ is as defined in Eq. (1.3). The +1 term accounts for the effect of gravity. The RE is an equation expressing the mass conservation law and Darcy's law and it includes the 'suction head', i.e. the suction force exerted at the surface of individual grains.

A water-retention curve describes the relationship between θ and h required by Eq. (1.11). We use the van Genuchten (1980) model which is typically applied in studies of liquid water flow through snow (Jordan, 1995; Hirashima et al., 2014; Wever et al., 2014; D'Amboise et al., 2017):

$$\theta = \theta_r + (\theta_{sat} - \theta_r)\frac{(1+(\alpha|h|^n)^{-m}}{Sc} \quad (1.12)$$

where θ_r is the residual water content (m^3 m^{-3}), θ_{sat} is the volumetric liquid water content at saturation (m^3 m^{-3}). Sc is a correction coefficient following Wever et al. (2014). The parameters α, n and m are tuning coefficients, with α being related to the maximum pore size and n and m being related to the pore size distribution. These three parameters, referred to as the van Genuchten parameters, are specific to the modelled porous medium and for snow; a common approach is to use the parameterisation developed by Yamaguchi et al. (2012) in a laboratory study:

$$\alpha = 4.4 \times 10^6 \left(\frac{\rho}{2r}\right)^{-0.98} \quad (1.13)$$

$$n = 1 + 2.7 \times 10^{-3} \left(\frac{\rho}{2r}\right)^{0.61} \quad (1.14)$$

$$m = 1 - \frac{1}{n} \quad (1.15)$$

Yamaguchi et al. (2012) measured the water-retention curve for a range of grain radii (0.025 to 2.9 mm) and densities (361 to 636 kg m^{-3}) in different snow samples by using a gravity drainage column method.

The porosity is the part of the volume not occupied by the solid matrix and, in the case of firn, is defined as:

$$P = 1 - \frac{\rho}{\rho_i} \quad (1.16)$$

The volumetric liquid water content at saturation is proportional to the porosity (Wever et al., 2014):

$$\theta_{sat} = P\ \frac{\rho_i}{\rho_w} \quad (1.17)$$

Note that water is not assumed to fill the entire pore space in saturated conditions and the correction factor $\frac{\rho_i}{\rho_w}$ included in Eq. (1.17) accounts for the required space to allow the liquid water to freeze. This correction factor agrees with the finding of Yamaguchi et al. (2010) that trapped irreducible air still occupies 10 % of the porosity in water-saturated snow.

The parameter θ_{sat} thus represents the pore space available for liquid water and from there we can define the effective saturation as:

$$Se = \frac{\theta - \theta_r}{\theta_{sat} - \theta_r} \quad (1.18)$$

and Se must be bounded between 0 and 1. In completely dry layers, a zero effective saturation would lead to infinite values in the head pressure calculation and thus, we use a numerical adjustment to avoid this happening (see Appendix A2.3). The residual water content θ_r is defined as the amount of liquid water that cannot be removed by gravity as it is held by capillary tension at the surface of the solid grains. Following Yamaguchi et al. (2010), a constant value of θ_r = 0.02 can be taken but in case of refreezing, θ can approach zero and θ_r must be adjusted accordingly. We take θ_r following a piecewise function:

$$\theta_r = \min\left[0.02, 0.9\ \theta\right] \quad (1.19)$$

The numerical requirement of an effective saturation value strictly greater than zero causes the persistence of very low flow rates, even for liquid water contents close to the residual water content because gravitational flow persists as Se>0. Over long time periods, layers cannot hold any residual water content and eventually dry out under the effect of gravity. By taking the coefficient 0.9 in Eq. (1.19) instead of 0.75 used in snow models (Wever et al., 2014; D'Amboise et al., 2017), we partially reduce this effect because this lowers the effective saturation value Se for any value of volumetric water content θ approaching zero.

The hydraulic conductivity ($K(\theta)$) is the ability of the fluid to flow through the porous medium under a certain hydraulic gradient dependent on pressure head and gravity. Thus, $K(\theta)$ depends on the effective saturation and on the properties of both the porous medium and the fluid; fluid flow is enhanced in highly saturated layers. The hydraulic conductivity is described by the van Genuchten-Mualem model (Mualem, 1976; van Genuchten, 1980):

$$K(\theta) = Ksat\ Se^{1/2}\left[1-\left(1-Se^{\frac{1}{m}}\right)^{m}\right]^{2} \quad (1.20)$$

where $Ksat$ is the hydraulic conductivity in saturated conditions (Se = 1). For the case of water flow through snow, it has been inferred using three-dimensional images of the microstructure by Calonne et al. (2012) as:

$$Ksat = 3.0\,r^{2}\exp\left(-0.013\,\rho\right)\left(\frac{g\,\rho_w}{\mu}\right) \quad (1.21)$$

Where g is the gravitational acceleration (9.8 m s^{-2}) and μ is 0.001792 kg m^{-1} s^{-1}, the dynamic viscosity of liquid water at 273.15 K. Equation (21) shows that simulated water flow is faster in layers with coarser grains and lower densities. These conditions correspond to cases where the connectivity between the pore spaces is high. With respect to the hydraulic conductivity parameterisation, we additionally modify the permeability of ice layers. The hydraulic conductivity of any layer with density exceeding the impermeability threshold is set to zero, rendering it impermeable to incoming flow and leading to the ponding of water on top of that layer and subsequent enhancement of runoff rates (Sect. 3.4.2). This RE implementation completely describes R1M and provides the basis of DPM, further detailed in the next section.

Details of the numerical implementations that are required to maintain stability and to improve computational efficiency for the RE calculations are discussed in Appendix A.

1.4.3 Dual-permeability model

Physical models of preferential flow in snow are still scarce (Hirashima et al., 2014; Wever et al., 2016). In this section, we explain how the SNOWPACK dual-permeability model (Wever et al., 2016) is implemented in the CFM. The firn column is separated into two domains and water flow in both is governed by the RE (Sect. 3.2). We define F as the pore space allocated to the preferential flow domain and accordingly 1-F as the pore space for the matrix flow domain. Wever et al. (2016) used a grain-size dependence for F, but their regression was performed on only four data points measured in idealised snow laboratory conditions (Katsushima et al., 2013). The experimental grain-sizes ranged from 0.1 to 0.8 mm and the water input from 480 to 550 mm per day, which is not representative of firn conditions in Greenland (Figs. 1.1 and 1.2). Moreover, due to the typical grain-size ranges in firn (Gow et al., 2004; Lyapustin et al., 2009), the model would regularly be forced to use for F the minimal value for numerical stability implemented in SNOWPACK. To deal with this uncertain parameter but still retain fidelity with respect to the SNOWPACK implementation, we favour the use of a constant value based on observations in natural snow. Marsh and Woo (1984) and Williams et al. (2010) reported that rapid flow paths occupy respectively 22% and 5% to 30% of the area and we thus fix the value F = 0.2. The extension of the preferential flow area within the snowpack is very likely to be a function of grain-size and meltwater influx and these dependencies are still uncertain (Avanzi et

al., 2016). The value of F thus determines the value of the saturated liquid water content θ_{sat} in both domains and instead of Eq. (1.17), we write:

$$\begin{cases} \theta_{sat,m} = (1-F)P\frac{\rho_i}{\rho_w} \\ \theta_{sat,p} = F\,P\frac{\rho_i}{\rho_w} \end{cases} \quad (1.22)$$

where from hereon, the subscripts m and p stand for matrix and preferential flow domain respectively. Equation (22) shows that the volumetric water content in the preferential flow domain is smaller than that in the matrix flow domain. All the input of meltwater is added to the matrix flow domain. For the regulation of the exchange of water between domains, we also closely follow the transfer processes of SNOWPACK (Wever et al., 2016) which are executed at the same 15-minute time step. We briefly summarize the transfer processes below.

Water from the matrix flow domain can enter the preferential flow domain of the layer below if the pressure head in the layer reaches the water entry suction, h_{we}, of the underlying layer. The parameter can be expressed as (Katsushima et al., 2013; Hirashima et al., 2014; Wever et al., 2016):

$$h_{we} = 0.0437(2r)^{-1} + 0.01074 \quad (1.23)$$

The amount of water transferred into the preferential flow domain equals the amount of water in excess of h_{we}. If after the transfer, Se in the matrix flow domain still exceeds Se in the preferential flow domain of the underlying layer, their respective Se are equalised by transferring the appropriate amount of water from the overlying matrix flow domain to the underlying preferential flow domain. In addition, in every individual firn layer where Se in the matrix flow domain exceeds Se in the preferential flow domain, matrix and preferential Se are equalised by transferring water from the matrix flow domain to the preferential flow domain. This serves to avoid the presence of horizontal pressure gradients in wet snow.

Water can flow from the preferential flow domain to the matrix domain by two processes. The first process is when the saturation in the preferential flow domain exceeds a threshold value Θ. Wever et al. (2016) determined Θ by tuning its value to best match observations. When this threshold is reached, the amount of water corresponding to the cold content of the layer flows back into the matrix domain. If there is still water in excess of the threshold in the preferential flow domain, saturation in both domains is set equal to one another. The second process simulates the heat flow from the preferential flow domain (at the melting point) to the colder surrounding matrix domain. Instead of transferring sensible heat, this process allows liquid water and its inherent latent heat to be exchanged to account for a theoretical heat flow, Q, and thus approximating Fourier's law:

$$Q = k_s \frac{(T-T_0)}{\left(\sqrt{\frac{1+F}{2\pi}} - \sqrt{\frac{F}{\pi}}\right)} \quad (1.24)$$

This formulation assumes a linear horizontal temperature gradient in the matrix and a circular shape of the preferential flow path's perimeter. From Eq. (1.24), the corresponding water transfer is calculated as:

$$\Delta\theta_{p\to m} = \frac{2\,N\sqrt{\pi F}\,Q\,\Delta t_{15}}{L_f\,\rho_w} \quad (1.25)$$

where Δt_{15} is the 15 minutes time step (s), L_f is the specific latent heat of fusion (335 500 J kg^{-1}) and N is a tuning parameter representing the number of preferential flow paths per square meter (m^{-2}). In their study, Wever et al. (2016) arrived at a best parameter set for Θ and N of 0.1 and 0 m^{-2} based on comparisons of ice-layer occurrence and runoff amounts with observations in alpine snowpacks. Note that the use of a null value for N is implausible in our case of firn-column simulations. Indeed, this would imply that liquid water would persist and flow deeper in the preferential flow domain in saturation conditions below the Θ value until the bottom of a subfreezing firn column, which can be up to 70 meters thick in some areas of the GrIS. Therefore, we use the smallest non-zero value of N tested by Wever et al. (2016) and the parameters Θ and N are fixed to 0.1 and 0.2 m^{-2} respectively.

The hydraulic conductivity of ice layers is not artificially set to zero in the preferential flow domain as it is in the matrix flow domain. Preferential flow thus provides a way for water to flow through an ice layer, reproducing observations that ice layers are not totally impermeable barriers and can lead to localised piping events (Marsh and Woo, 1984; Pfeffer and Humphrey, 1998; Williams et al., 2010; Sommers et al., 2017). An exception for this is the bottom of the domain: as preferential flow is stopped at the last layer, it does not percolate through the solid ice.

1.4.4 Additional processes in the single- and dual-domain schemes

1.4.4a Refreezing process

In R1M and DPM a 'cold content' is calculated for every firn layer, similarly to BK (Sect. 3.1) and refreezing in both flow schemes is executed at the 15 minutes time step (equivalent to the time step of the transfer processes of DPM, Sect. 3.3).

When refreezing occurs, every layer freezes the maximum of its liquid water content that its cold content allows. For numerical reasons, refreezing cannot dry out a layer completely; instead, a very low value of liquid water remains in every layer (see Appendix A2.3). The refrozen water densifies the firn layer and modifies its hydraulic properties. The remaining liquid water is still subject to flow and infiltrates deeper into the firn column.

In DPM, refreezing is restricted to the matrix flow domain (see Appendix A2.7). In the preferential flow domain, liquid water can percolate through cold layers, as has been observed in field studies on the GrIS (e.g. Pfeffer and Humphrey, 1996; Humphrey et al., 2012). For this liquid water to refreeze,

it first has to be transferred back to the matrix flow domain. Preferential flow thus provides a way for liquid water to bypass cold firn layers and subsequently to infiltrate deeper layers.

1.4.4b Aquifer development and lateral runoff

In R1M and DPM, lateral runoff in the firn column is simulated using the parameterisation of Zuo and Oerlemans (1996):

$$\frac{dRu}{dt} = \frac{L_{excess}}{\tau_{Ru}} \quad (1.26)$$

$$\tau_{Ru} = c_1 + c_2 \exp(-c_3 S) \quad (1.27)$$

where Ru is the amount of meltwater that runs off (m), L_{excess} is the excess of liquid water amount with respect to the residual water content (m) and τ_{Ru} is a characteristic runoff time (s). The constants c_1, c_2 and c_3 are parameters derived by comparison with observations by Zuo and Oerlemans (1996) for the GrIS and S is the surface slope. The meltwater input is immediately treated as lateral runoff if the surface layer is an impermeable ice layer or if it is saturated.

Equation (26) leads to the complete drainage of a layer with a zero slope in only 26 days, precluding the formation and persistence of perennial firn aquifers. Therefore, we don't apply Eq. (1.26) in the layers at the bottom of the firn column. All the inflow of water reaching this section is added to the aquifer, allowing the model to progressively fill the pore space of the bottom layers in the firn column with meltwater.

Firn aquifers are known to be affected by drainage mechanisms not represented in the model, for example via crevasses (Poinar et al., 2017), and possibly hydrofracture and rapid drainage events (Koenig et al., 2014). Miller et al. (2018) found discharge rates within the firn aquifer to be 4.3 x 10^{-6} m s^{-1} by borehole dilution tests in the field. We tested this approach in our model by applying this value as a constant discharge rate for aquifers formed in our simulations. We found however that, using this approach, an aquifer was not sustained; suggesting that such discharge rates must be dependent on the total amount of water within the aquifer and are likely temporally variable. To account for drainage processes and yet allow the formation of an aquifer, we therefore limited the amount of water stored in the firn aquifer to 1.65 m w.e., i.e. the water level measured in the field by Koenig et al. (2014). In firn aquifers forming at the bottom of the firn column, the saturation in both domains is equalised and the model does not perform flow calculation in this lowest part of the domain (see Appendix A2.5).

1.4.5 Investigating model sensitivity

In Sections 2 and 3, we highlight several factors influencing BK, R1M and DPM. For each of the schemes, we analyse results generated using three possible impermeability thresholds: 780 kg m^{-3}

(ip780), 810 kg m^{-3} (ip810) and 830 kg m^{-3} (ip830). This provides a way to compare the sensitivity of the simple BK and of the physically based schemes (R1M and DPM) to a common parameter. For BK, we try two different formulations of the water-holding capacity: constant at 0.02 (wh02) and according to the parameterisation of Coléou and Lesaffre (1998), Eq. (1.9) (whCL). For R1M and for DPM, we test two different grain-size implementations: the Linow et al. (2012) surface grain-size calculation, Eq. (1.5), coupled to the Katsushima et al. (2009) grain growth rate, Eq. (1.6) (grLK), and the grain-size implementation of Arthern et al. (2010), Eq. (1.8) (grA). It is important to examine model sensitivity to the grain-size variable as almost all the hydraulic parameters of the RE depend on it. The different sensitivity tests are summarised in Table 1.1.

Bucket Model (BK)					Single-domain RE (R1M)					Dual-Permeability RE (DPM)				
Impermeability threshold (ip)			Water-holding capacity (wh)		Impermeability threshold (ip)			Grain-size formulation (gr)		Impermeability threshold (ip)			Grain-size formulation (gr)	
780	810	830	0.02	CL	780	810	830	LK	A	780	810	830	LK	A

Table 1.1 Summary of the sensitivity tests

1.5 Results

In this section, we describe and discuss the model performance at each of the four sites tested (DYE-2, NASA-SE, KAN-U and FA13). We begin by comparing BK, R1M and DPM in a base case parameterisation: BK wh02 ip810, R1M grLK ip810 and DPM grLK ip810 respectively. Then, we perform various tests to investigate the sensitivity of the flow schemes to variations in their parameter values. We refer to ice layers as layers with a density value exceeding the impermeability threshold in the model, and to liquid water input as the total of meltwater and rain influx. The DPM approach features two tuning parameters, N and Θ. Model results and depth-density profiles were found to be weakly sensitive to the value of N and Θ and so we omit consideration of these from the remainder of our study. Results of simulations and observations are inter-compared based on the firn air content (FAC; the depth integrated porosity in a firn column) over the top 15 m of firn and the temperature at 10 m depth. Comparing the modelled FAC and 10 m depth temperature values with observed data depicts the ability of the tested models to reproduce the bulk condition of the upper firn column. We also qualitatively assess the degree to which the models form a 'realistic' ice layer distribution and depth-density profile. One would not expect simulated values of either to match observations precisely given the high spatial variability of firn structure (Marchenko et al., 2017), but it is indicative of the models' performance in reproducing heterogeneity in firn density.

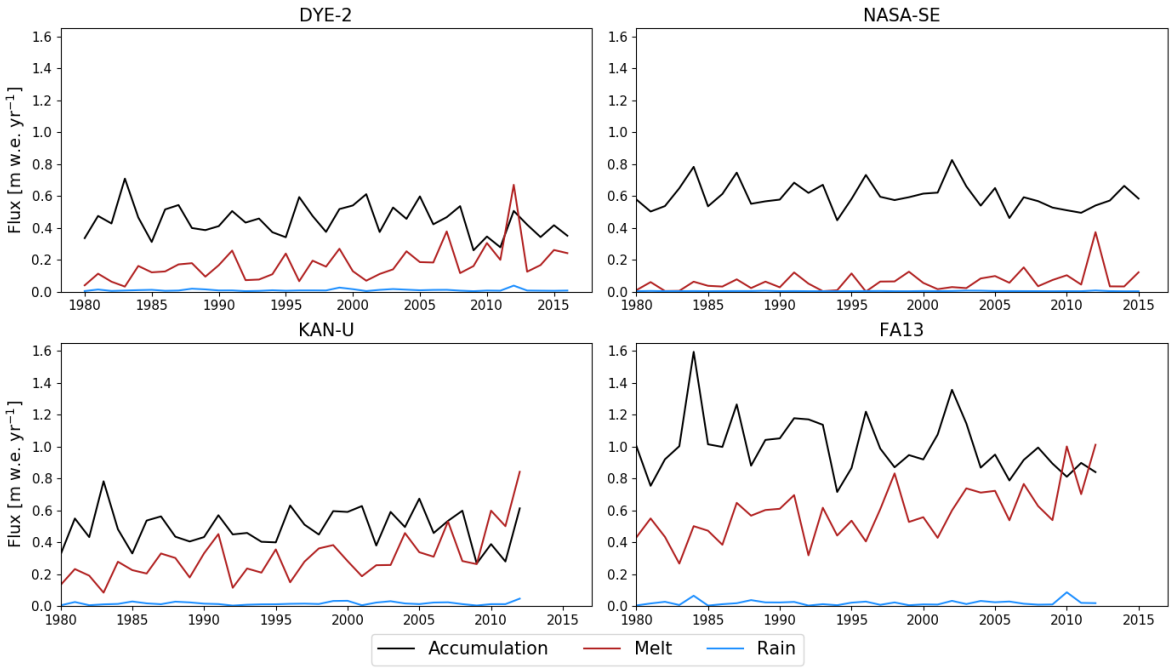


Figure 1.2. Annual surface mass fluxes from RACMO2.3p2 at the study sites (1980-drilling date)

1.5.1 DYE-2

DYE-2 has a typical liquid water input between 0.1 and 0.3 m w.e. yr^{-1} (Fig. 1.2), which is moderate in the context of our study sites. The extreme melt year of 2012 (Nghiem et al., 2012) is an exception, with an estimated input of more than 0.7 m w.e. Using BK, almost all of this meltwater refreezes locally and runoff is close to zero (Table 1.2) until the 2012 summer when ice layers ($\rho \geq 810$ kg m^{-3}) start forming in the top 2 m (Fig. 1.3a). Runoff increases in the subsequent years because meltwater reaches these ice layers. In R1M and DPM, small amounts of runoff occur between 1980 and 2011 due to the lateral runoff implementation, Eq. (1.26). Beginning in summer 2004, some ice layers start to form in R1M (Fig. 1.3b) due to the refreezing of water held close to the surface by capillary forces. Over the 2012 summer, surface layers are progressively melted, bringing ice layers closer to the surface. The ponding and refreezing of water on the top ice layer allows it to thicken. This then acts as an impermeable barrier to vertical percolation from 2012 onwards, resulting in a more than sixfold increase in runoff (Table 1.2). In contrast, runoff remains low in DPM, in which several ice layers form in the upper firn as early as summer 1996 (Fig. 1.3c). These ice layers generally form deeper than 2 m due to more effective water transfer from the near-surface to lower layers; preferential flow provides a path for ponding meltwater in the matrix flow domain to bypass ice layers and continue to percolate vertically, thus maintaining low runoff amounts. Preferential flow brings part of the 2012 meltwater to depths greater than 12 m. For each flow scheme, the modelled FAC underestimates the observed value by 4-16%. This can partly be attributed to the tendency of the CROCUS scheme to slightly overestimate densification rates in the upper part of polar firn (Gascon et al., 2014). FAC is

underestimated more strongly in DPM (16 %) than in BK and R1M (4 %) because in DPM the deeper firn is not isolated from surface meltwater percolation (Table 1.2).

	Refreezing / Inflow (1980-2011)	Refreezing / Inflow (2012-2016)	Runoff / Inflow (1980-2011)	Runoff / Inflow (2012-2016)	Top 15 m FAC [m] (anomaly vs observations)	T 10m [K] (anomaly vs observations [K])
BK (wh02 ip810)	0.96	0.67	0.01	0.31	5.01 (-4 %)	260.88 (+0.21)
R1M (grLK ip810)	0.91	0.63	0.05	0.35	4.99 (-4 %)	260.27 (-0.40)
DPM (grLK ip810)	0.95	0.95	0.02	0.03	4.38 (-16 %)	263.39 (+2.72)
Observations	/	/	/	/	5.21	260.67
DPM (grLK ip780)	0.95	0.96	0.02	0.02	4.39 (-16 %)	263.45 (+2.78)
BK (wh02 ip780)	0.96	0.60	0.02	0.38	5.15 (-1 %)	260.68 (+0.01)
BK (whCL ip810)	0.92	0.62	0.05	0.36	5.21 (+0 %)	259.40 (-1.27)
BK (wh02 ip830)	0.97	0.68	0.00	0.31	4.96 (-5 %)	260.98 (+0.31)
R1M (grA ip810)	0.83	0.59	0.14	0.38	5.19 (-0 %)	259.64 (-1.03)
DPM (grA ip810)	0.93	0.93	0.04	0.05	4.40 (-16 %)	263.08 (+2.41)
HL DPM (grLK ip810)	0.95	0.96	0.02	0.02	4.16 (-20 %)	263.78 (+3.11)
KM DPM (grLK ip810)	0.95	0.95	0.02	0.02	3.35 (-36 %)	262.74 (+2.07)
HL R1M (grLK ip810)	0.90	0.72	0.07	0.26	4.65 (-11 %)	260.41 (-0.26)
KM R1M (grLK ip810)	0.91	0.75	0.06	0.23	4.00 (-23 %)	260.26 (-0.41)

Table 1.2. Model outputs at DYE-2 site. / indicates no data.

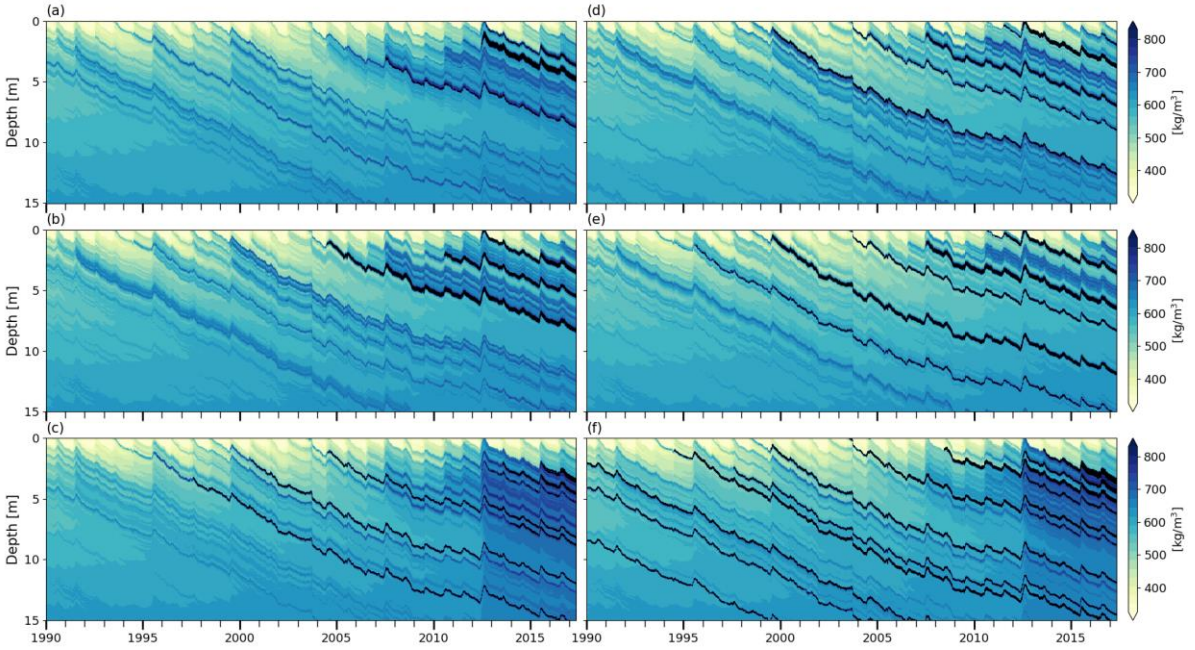


Figure 1.3. Modelled firn density at DYE-2 (a) BK wh02 ip810, (b) R1M grLK ip810, (c) DPM grLK ip810, (d) BK whCL ip810, (e) R1M grA ip810, (f) DPM grA ip810, black indicates solid ice.

Modelled density profiles using each flow scheme are compared with observations (Fig. 1.4a). Mean density is reproduced reasonably well using each of the three flow schemes, but no configuration is able to qualitatively reproduce the strong variability in density observed. For example, numerous

high-density layers separated by much lower density intervals are clear in the observations. Regardless of the flow scheme, only a few ice layers are formed in the model and these tend to be confined to the upper 6 m, which has been affected by the higher melting rates of the recent years. In older firn deposited under lower-melt conditions, the number of density peaks and their amplitude is underestimated even more strongly. Several ice layers are observed in the 10 – 20 m depth range; only DPM simulates the presence of ice layers here.

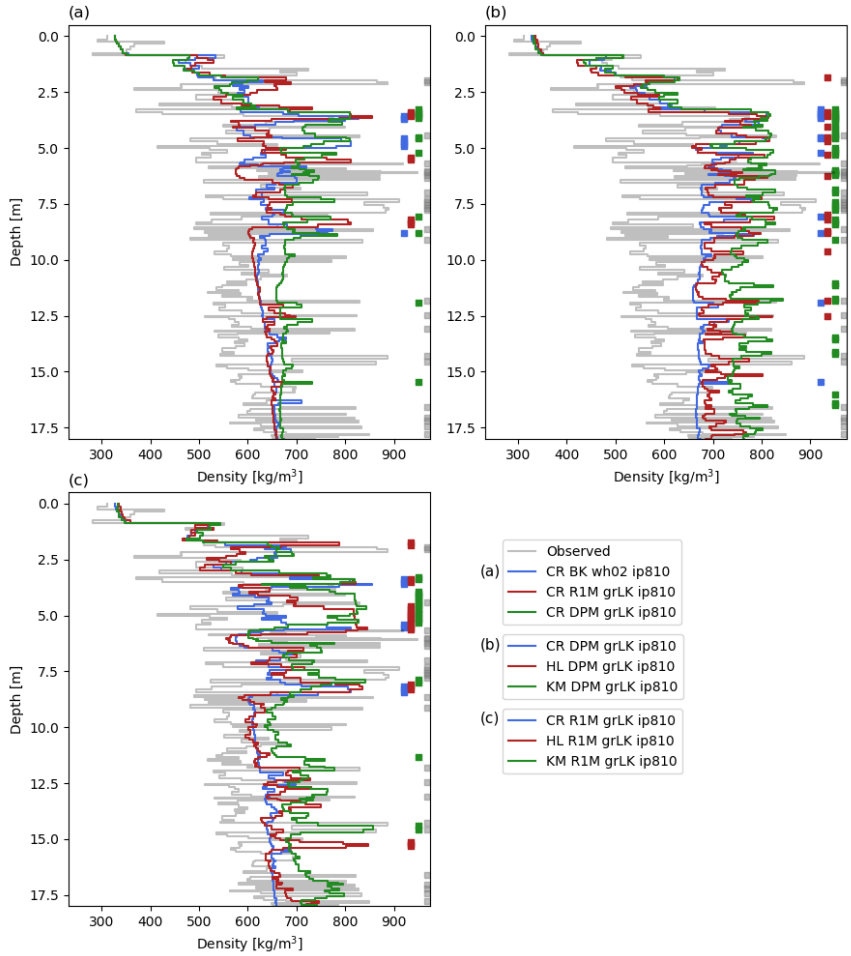


Figure 1.4. Measured and modelled depth-density profiles at DYE-2 on 11/05/2017. Thick vertical lines show ice layers. The modelled densities are averaged at the vertical resolution of the drilled core. CR: CROCUS, HL: Herron and Langway, KM: Kuipers Munneke.

The three flow schemes lead to significantly different firn thermal conditions. The temperature at 10 m depth of BK and R1M agree well with observations (+0.2 and -0.4 K). In contrast, 10 m temperature is strongly overestimated in DPM (+2.7 K) because it allows percolation at depth, subsequent refreezing, and latent heat release. The summer 2012 percolation raises the 10 m depth temperature to within a few degrees of melting using DPM. Since the DPM method seems to exaggerate deep percolation, we tested a lower impermeability threshold (DPM grLK ip780) which should favour the formation of shallow ice layers, the ponding of water in the matrix flow domain,

more lateral runoff and colder temperatures at depth. The ice layers do form slightly earlier in the melt seasons but not noticeably shallower than in DPM ip810. The partitioning between runoff and refreezing is barely affected and the 10 m temperature bias remains (Table 1.2).

The BK method gives a density profile closer to R1M than to DPM. In order to mimic the behaviour of DPM we increase the impermeability threshold in BK (BK wh02 ip830) to make it more effective in transporting water vertically; however, model results are only weakly affected by this change (Table 1.2). We also modify the water-holding capacity in BK according to the parameterisation of Coléou and Lesaffre (1998) (BK whCL ip810) which allows more water to be retained in the low-density layers close to the surface. Ice layers appear earlier in the simulation and at shallower depths (Fig. 1.3d). This increases the amount of runoff in BK whCL ip810 with respect to BK wh02 ip810 (+4 % of the water input over the entirety of the transient model run); however, in the surface layers, where high amounts of water are retained, refreezing dominates. As a result, much less water percolates to the deeper firn and there is less refreezing and latent heat release. All of this leads to a significantly higher FAC (+4 %) and colder 10 m temperature (-1.5 K) relative to BK wh02 ip810.

For models based on the RE (R1M and DPM), we test sensitivity to grain-size by implementing a parameterisation for grain growth based on Arthern et al. (2010) (denoted grA). Using this parameterisation, grain-sizes tend to be smaller, and so more water tends to be retained and refrozen close to the surface due to stronger capillary forces. Compared to the R1M grLK ip810 experiment, the R1M grA ip810 causes formation of ice layers earlier in the simulation (beginning in 1996) and shallower in the firn column (Fig. 1.3e), favouring water ponding and subsequent runoff (+7 % of the water input over the entirety of the transient model run). Stronger capillarity also means that saturation is higher for percolation to occur, which in turn increases the simulated runoff since more water is in excess of the residual water content. The enhanced runoff and shallower percolation lead to a higher FAC (+4 %) and a colder 10 m temperature (-0.6 K). In DPM, the flow and refreezing patterns are also altered by the grain-size formulation: DPM grA ip810 produces ice layers much earlier (beginning in summer 1981), at shallower depths and in larger numbers (Fig. 1.3f). Runoff is however only slightly increased (+2%). The FAC remains similar to DPM grLK ip810, but the 10 m temperature is 0.3 K lower and the warm bias is thus reduced (an 11 % decrease) (Table 1.2).

Finally, we investigate differences in the depth-density profiles simulated at DYE-2 attributed to different firn-densification formulations in contrast to those observed due to the use of different flow schemes. We first choose to apply the DPM grLK ip810 flow scheme with the additional firn-densification formulations of Herron and Langway (1980) (HL) and of Kuipers Munneke et al. (2015b) (KM), both calibrated for GrIS firn. The HL and KM models are forced with the same three-hourly climatic forcing as the CROCUS (base case) model. The FAC (-5 %), 10 m temperature (+0.4

K) and mean density profile (Fig. 1.4b) predicted by the HL densification model agree reasonably well with that predicted by CROCUS, although HL predicts greater density variability due to its stronger dependence on the annual temperature cycle. In contrast, the KM model predicts much higher densification rates and thus greater densities, with several thick ice layers in the 3-8 m depth range, some exceeding a meter thickness. This results in a much lower FAC value compared to the CROCUS model (-24 %) and in this case, differences between flow schemes are small with respect to the choice of the densification formulation. Since the warm bias of DPM can cause temperature-dependent densification formulations to overestimate densities, we also compare the three densification formulations coupled to R1M grLK ip810 (Fig. 1.4c). Similar to the results using the DPM flow scheme, the HL profile agrees reasonably well with the CROCUS model (FAC value is -7 %) but predicts that a meter-thick ice layer formed at 5 m depth (Fig. 1.4c) during the 2012 summer. Discrepancies between CROCUS and KM are only slightly reduced using R1M; for example, the FAC predicted by KM is 20% less than that predicted by CROCUS. This can be attributed to greater densities at depth (>8 m) and to much higher densities in the depth range 3-5 m. The latter corresponds to the layers affected by meltwater refreezing and considerable latent heat release in the 2012 summer.

1.5.2 NASA-SE

NASA-SE is a site characterised by high accumulation rates, ranging between 0.5 and 0.8 m w.e. yr^{-1}, and low rates of liquid water input, typically between 0.01 and 0.15 m w.e. yr^{-1} (Fig. 1.2). Under these conditions, abundant pore space and cold content are available for prompt refreezing of the summer meltwater so one would expect a smaller sensitivity of the model to the flow scheme applied. In BK, no runoff is produced over the entire simulation (Table 1.3) since refreezing of small amounts of melt does not lead to the formation of impermeable ice layers. R1M and DPM have very low runoff amounts with a small spike in the summer of 2012 when there was 0.38 m w.e. of liquid water input. No ice layer forms in the top 15 m of the firn column using any of the liquid water schemes, in agreement with the observed core (Fig. 1.5a). Changing the impermeability threshold results in identical model results since no layer exceeds the lowest possible value in the depth range where water percolates. The three water-transport schemes predict a similar FAC; they all underestimate the observed value by approximately 3% (Table 1.3). This is because the mean firn density is well-captured by the model but somewhat overestimated at depths greater than 8 m (Fig. 1.5a). R1M simulates a single density peak at 8 m depth (Fig. 1.5a), corresponding to the 2012 summer meltwater percolation, due to capillary forces effectively retaining the relatively high meltwater volume produced in that year close to the surface and exposing it to delayed refreezing once these layers cool below the freezing point. DPM also produces a density peak (albeit a much smaller one) at a similar depth, and more-effective downward percolation results in a uniform increase in density over the next 3 m. Finally, BK also produces a small density peak; however, this is at a greater depth of 9 m since it

assumes water flow to be instantaneous in a time step and the major part of the refreezing occurs as water reaches deeper cold layers. Again, none of the percolation schemes capture the observed variability in density. Also, despite the low melt/accumulation ratio, the three percolation schemes overestimate the 10 m temperature by 1.4-2.2 K (Table 1.3).

	Refreezing / Inflow (1980-2011)	Refreezing / Inflow (2012-2015)	Runoff / Inflow (1980-2011)	Runoff / Inflow (2012-2015)	Top 15 m FAC [m] (anomaly vs observations)	T 10m [K] (anomaly vs observations [K])
BK (wh02 ip810)	0.97	0.97	0.00	0.00	6.78 (-3 %)	257.91 (+1.94)
R1M (grLK ip810)	0.94	0.89	0.02	0.08	6.78 (-3 %)	257.39 (+1.42)
DPM (grLK ip810)	0.95	0.94	0.01	0.03	6.77 (-3 %)	258.18 (+2.21)
Observations	/	/	/	/	6.98	255.97
BK (whCL ip810)	0.95	0.94	0.00	0.02	6.81 (-2 %)	256.97 (+1.00)
R1M (grA ip810)	0.92	0.83	0.04	0.13	6.83 (-2 %)	256.99 (+1.02)
DPM (grA ip810)	0.95	0.92	0.01	0.04	6.78 (-3 %)	258.11 (+2.14)
HL R1M (grLK ip810)	0.93	0.86	0.04	0.11	8.13 (+17 %)	258.19 (+1.22)
KM R1M (grLK ip810)	0.93	0.86	0.03	0.10	7.53 (+8 %)	257.74 (+1.77)

Table 1.3. Model outputs at NASA-SE site. / indicates no data.

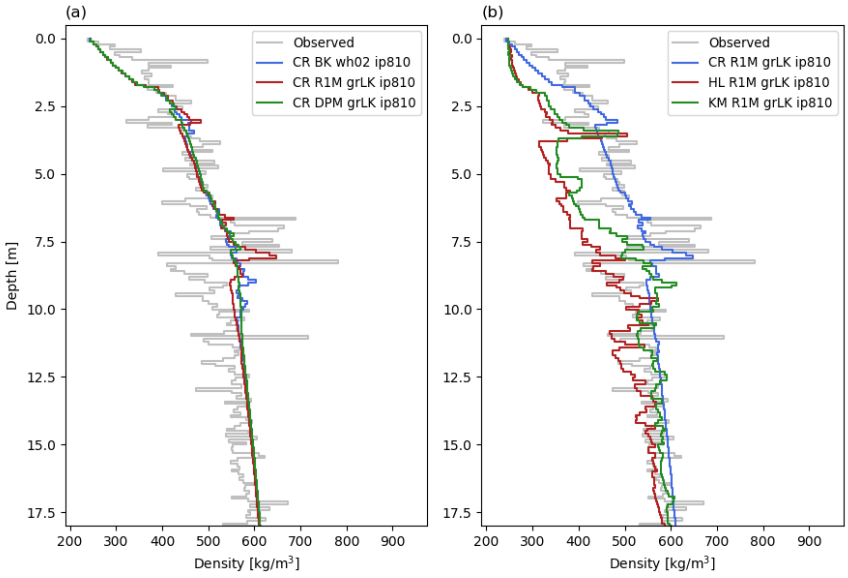


Figure 1.5. Measured and modelled depth-density profiles at NASA-SE on 04/05/2016. The modelled densities are averaged at the vertical resolution of the drilled core. CR: CROCUS, HL: Herron and Langway, KM: Kuipers Munneke.

Increasing the water-holding capacity in BK (BK whCL ip810) leads to a minor increase in the FAC (< 1%) and a 0.9 K cooling of the 10 m temperature, because the surface layers have a relatively low density (surface boundary condition of 240 kg m^{-3} at this site) and thus retain high amounts of water with the whCL parameterisation (Table 1.3). The R1M and the DPM density profiles are weakly sensitive to a change in the grain-size formulation from grLK to grA (Table 1.3). This is due to the

small meltwater amounts with meltwater refreezing only slightly closer to the surface because of the stronger capillarity retention in the grA models. However, we note that simply changing the grain-size formulation in R1M from grLK to grA leads to a 0.4 K colder 10 m temperature and thus decreases the bias with respect to observations by 28 % (Table 1.3).

We used the R1M grLK ip810 model with the HL and the KM densification formulations in order to prevent the DPM's warm bias from skewing the modelled densification rates. As expected in this relatively dry site, the modelled profiles are much more sensitive to the dry-densification than to the percolation scheme (Table 1.3 and Fig. 1.5a and b). The maximal difference in FAC between the three densification formulations tested is 20% compared to less than 1 % between the three flow schemes and their possible parameterisations. In contrast with the DYE-2 simulations, the CROCUS model predicts the fastest densification and thus the lowest FAC. HL and KM predict 20% and 11% greater FAC than CROCUS, respectively, and CROCUS is in closest agreement with the observations, underestimating FAC by just 3%.

1.5.3 KAN-U

KAN-U is a high-melt site with an average melt rate over the 1980-2013 period of 0.33 m w.e. yr^{-1}, and in the last three years of our simulation (2010-2013), the RCM calculates annual melt exceeding annual accumulation (Fig. 1.2). Since surface temperatures are relatively high (annual mean around -8 °C), refreezing of the summer meltwater depletes the cold content over large depth ranges. Beginning in summer 1990 in the BK simulation, some ice layers are present in the depth range 3-8 m (Fig. 1.6a), allowing part of the meltwater to runoff and impeding percolation to greater depths. At the start of 2012, there is a thick ice layer in the upper 4 m and another one forms at the surface during the summer. As a result, refreezing is constrained to the uppermost firn layers and a large part of the water input runs off (Table 1.4). In R1M, the high water content and the almost-continuous presence of ice layers in the upper 5 m from summer 1986 onwards (Fig. 1.6b) cause relatively high runoff rates throughout the simulation (28 % of the water input over the entirety of the transient model run). As in the BK simulation, runoff is particularly high in 2012 due to ice layers impeding vertical percolation below 1 m (Table 1.4). In the DPM simulation, the preferential flow mechanism leads to the formation of multiple ice layers in the depth range 4-10 m from 1987 onwards (Fig. 1.6c). Runoff rates remain low but there is a notable increase in 2012. This is due to the formation of ice layers close to the surface, which allows ponding of water in the matrix flow domain. The preferential flow domain is unable to accommodate all the ponding water, and part of it is treated as lateral runoff (Eq. (1.26)). While matrix flow typically remains constrained to the upper 5 m (Fig. 1.7a), the recent (2010 to 2012) high-melt summers cause preferential flow to reach much greater depths (e.g. up to 35 m in the 2012 summer (Fig. 1.7b)). Since preferential flow can transfer water below ice layers, the refreezing process can fill the pore space available at depth, leading to substantial thickening of the

ice layers. As a result, the FAC is much smaller in the DPM simulation than in the BK (-39 %) and the R1M (-35 %) simulations, in which runoff limits the amount of meltwater refreezing.

	Refreezing / Inflow (1980-2011)	Refreezing / Inflow (2012)	Runoff / Inflow (1980-2011)	Runoff / Inflow (2012)	Top 15 m FAC [m] (anomaly vs observations)	T 10m [K] (anomaly vs observations [K])
BK (wh02 ip810)	0.81	0.18	0.17	0.81	3.92 (+59%)	263.93 (-1.73)
R1M (grLK ip810)	0.74	0.20	0.23	0.79	3.69 (+50 %)	263.09 (-2.57)
DPM (grLK ip810)	0.91	0.77	0.07	0.23	2.40 (-2 %)	270.18 (+4.52)
Observations	/	/	/	/	2.46	265.66
DPM (grLK ip780)	0.91	0.75	0.07	0.25	2.77 (+13 %)	268.63 (+2.97)
DPM (grLK ip 830)	0.90	0.82	0.07	0.18	2.18 (-11%)	271.31 (+5.65)
BK (wh02 ip780)	0.78	0.19	0.21	0.81	4.11 (+67 %)	263.46 (-2.20)
BK (whCL ip810)	0.78	0.22	0.19	0.78	4.05 (+65 %)	262.23 (-3.43)
BK (wh02 ip830)	0.83	0.21	0.15	0.79	3.61 (+47 %)	264.69 (-0.97)
R1M (grLK ip830)	0.76	0.20	0.22	0.79	3.64 (+48 %)	263.21 (-2.45)
R1M (grA ip810)	0.66	0.26	0.31	0.73	4.08 (+66 %)	262.41 (-3.25)
DPM (grA ip810)	0.87	0.69	0.10	0.30	2.36 (-4 %)	270.28 (+4.62)
HL R1M (grLK ip810)	0.74	0.18	0.23	0.82	3.36 (+37 %)	263.34 (-2.32)
KM R1M (grLK ip810)	0.74	0.20	0.23	0.80	2.70 (+10%)	262.82 (-2.84)

Table 1.4. Model outputs at KAN-U site. / indicates no data.

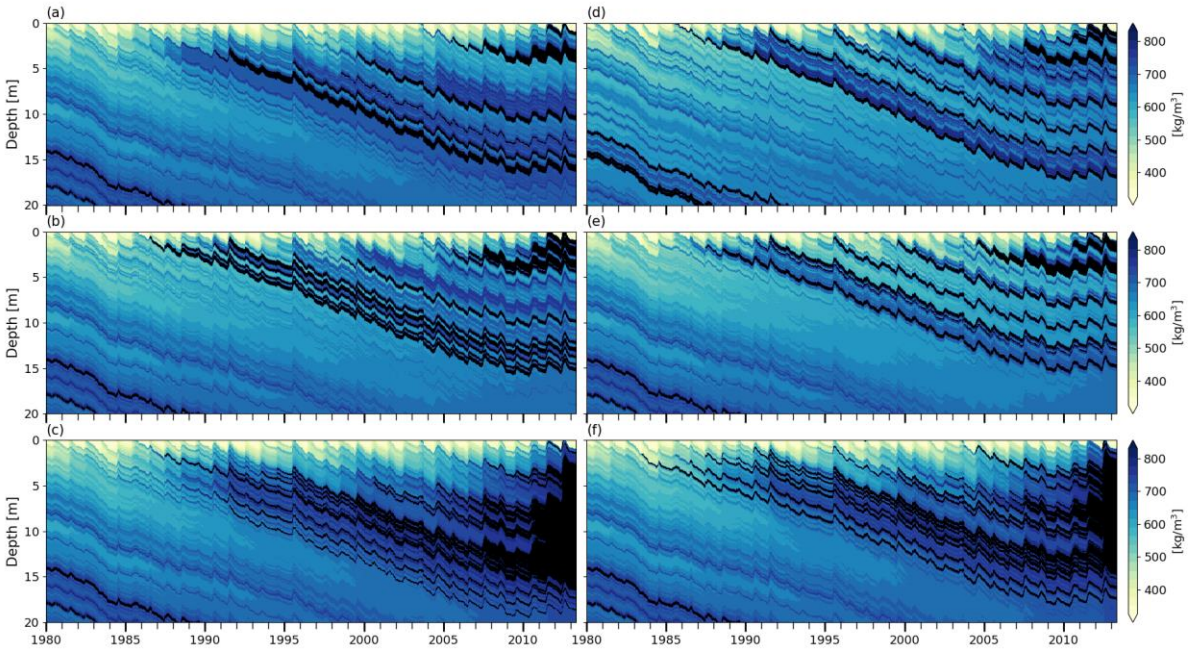


Figure 1.6. Modelled firn density at KAN-U. (a) BK wh02 ip810, (b) R1M grLK ip810, (c) DPM grLK ip810, (d) BK whCL ip810, (e) R1M grA ip810, (f) DPM grA ip810, black indicates ice layers

The observations reveal a thick, almost continuous ice slab over the depth range of 1-7 m (Fig. 1.8a). Below it, the density is more variable but remains generally high causing a low FAC (Table 1.4). Both the BK and the R1M simulation significantly overestimate the FAC (+59% and +50 %). In contrast,

the average FAC of the DPM simulation is very close to the observed value (-2 %), however, the DPM density profile shows an almost continuous ice slab from 3 to 17 m depth (Fig. 1.8a) and does not reproduce the lower density intervals observed. This demonstrates an important limitation of the liquid water schemes: since water cannot be retained in layers exceeding the impermeability threshold, these layers can only further densify by the dry densification mechanism and not by water refreezing. The overestimation of the ice slab thickness in the DPM profile is thus compensated by the underestimation of its density, which leads to the good agreement with the observed FAC value. BK reproduces the presence of the ice slab at 1 m depth, but it underestimates its thickness and simulated a thick (2 m) low-density region (Fig. 1.8a). Below the observed ice slab, the agreement with observed average density is reasonable but variability in density is underestimated. Despite also underestimating the thickness of the ice slab, the R1M profile agrees better with the observed density profile: it produces only two thin, low-density layers in the slab, and more high density peaks and ice layers below 7 m which is in better agreement with the observed density variability.

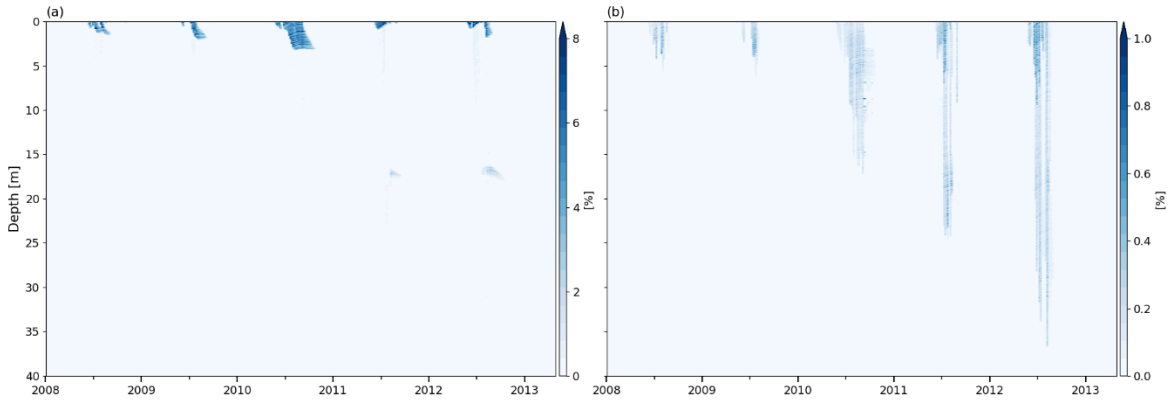


Figure 1.7. Volumetric water content at KAN-U site for DPM grLK ip810 in (a) Matrix flow domain, (b) Preferential flow domain, note difference in scales

With respect to the 10 m temperature, the BK method is biased cold but gives results in reasonable agreement with the observations (-1.7 K). This bias is more pronounced in R1M (-2.6 K). In contrast, DPM largely overestimates the 10 m temperature (+4.5 K), as a result of its overestimation of percolation and subsequent refreezing at depth.

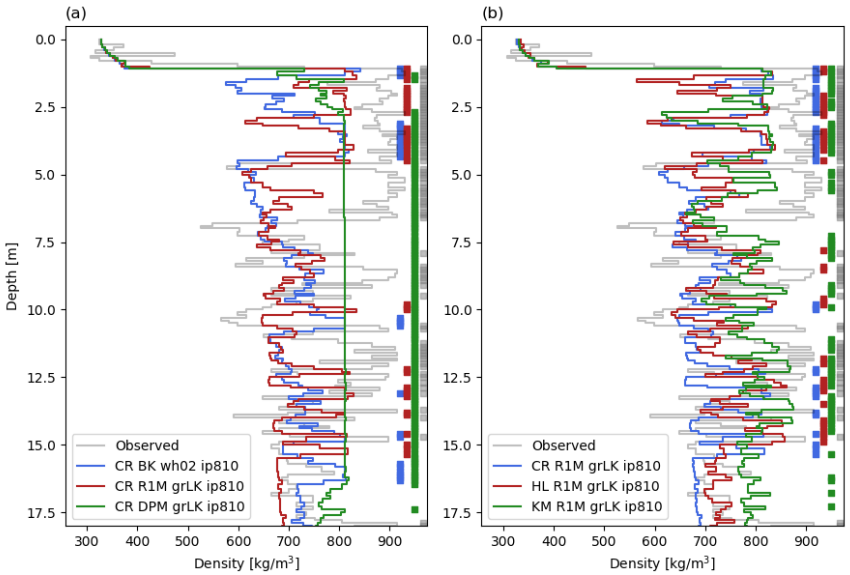


Figure 1.8. Measured and modelled depth-density profiles at KAN-U on 28/04/2013. Thick vertical lines show ice layers. The modelled densities are averaged at the vertical resolution of the drilled core. CR: CROCUS, HL: Herron and Langway, KM: Kuipers Munneke.

Changing the impermeability threshold for DPM (DPM wh02 ip780 and ip830) does not alter the pattern of the modelled depth-density profile, but the corresponding changes to the density of the ice slab has an impact on the FAC (+15 % for ip780 and -9 % for ip830). Other factors further affect the FAC: runoff rates slightly decrease with higher impermeability thresholds (Table 1.4) and the mass of the ice layers increases the overburden stress on the firn column below, increasing the densification rate. In addition, higher (lower) impermeability thresholds lead to warmer (colder) 10 m temperatures (+1.1 K for ip830 and -1.6 K for ip780), due to enhanced latent heat release. Compared to BK wh02 ip810, decreasing the impermeability threshold (BK wh02 ip780) leads to formation of ice layers in earlier years and closer to the surface and thus more runoff (+3 % of the water input over the entirety of the transient model run), which in turn increases the FAC (+5 %) and decreases the 10 m temperature (-0.5 K). Increasing the threshold (BK wh02 ip830) has the opposite effect (-8 % for the FAC and +0.8 K for the 10 m temperature compared to BK wh02 ip810). If we instead allow for a greater water-holding capacity (BK whCL ip810), the partitioning between runoff and refreezing remains very similar (Table 1.4). However, the FAC and the 10 m temperature are changed (+3 % and -1.7 K compared to BK wh02 ip810). The lower temperature is due to latent heat release from refreezing being more concentrated in the surface layers (Fig. 1.6d). The formation of ice layers earlier in the year and at shallower depths allows part of the underlying firn to remain free of refreezing, which increases the FAC. Furthermore, colder temperatures cause a higher firn viscosity thus decreasing the densification rates. Since the R1M formulation both overestimates the FAC and underestimates the 10 m temperature, we test an increase in its impermeability threshold (R1M grLK

ip830), allowing for deeper percolation. Both the decrease in FAC (-1 %) and increase in 10 m temperature (+0.1 K) compared to R1M grLK ip810 are minor.

With the grA formulation in DPM (DPM grA ip810), water is more efficiently transferred vertically through the preferential flow domain, which causes an increase in the number of ice layers formed during the simulation (Fig. 1.6f), a slight decrease in FAC (-2 %) and a slight increase in the 10 m temperature (+0.1 K) relative to DPM grLK ip810. The nearly-continuous ice slab, which extends to 17 m depth below the final winter accumulation, explains the weak sensitivity of the final FAC and 10 m temperature values of DPM to grain-size. In contrast, applying the grA formulation in R1M (R1M grA ip810) leads to a considerable increase in FAC (+11 %) and a decrease in 10 m temperature (-0.7 K) compared to R1M grLK ip810. This is due to higher water content during percolation events and, especially in the most recent years of our simulation, refreezing and ice-layer formation at shallower depths (Fig. 1.6e). This increases the runoff and isolates the deeper firn from meltwater percolation. As in the cases of DYE-2 and NASA-SE, the change in FAC due to different grain-size formulations in R1M is greater than the change due to switching from BK to R1M (Table 1.4).

The modelled depth-density profiles also differ according to the densification formulation used (Fig. 1.8b). We compare the different densification formulations using R1M grLK ip810, thus avoiding the effect of the strong temperature bias of DPM on the densification process. Densification in KM is sensitive to high firn temperatures, and it predicts the highest densities: it produces the highest density values in the ice slab range, the most ice layers below the ice slab and the lowest FAC value (-27 % compared to the CROCUS formulation). HL behaves in a similar way to CROCUS in the upper 5 m, apart from a much lower density interval in the 2-2.5 m depth range. In deeper firn, densities simulated using HL tend to lie between those simulated using KM and CROCUS, and its FAC difference with the CROCUS (-9 %) is less than that of KM. The DPM scheme simulates a depth-density profile of an ice slab over a 14 m range, which is in stark contrast with BK and R1M. Apart from this, the choice of the densification formulation has a greater influence on the model than the choice of liquid water scheme and of any of their respective parameterisations presented here, in spite of the high water input at this site.

1.5.4 FA13

The FA13 site is representative of conditions in the southeast part of the GrIS; it has both high accumulation and high melt rates (mean 1980-2012 rates of 1.09 and 0.64 m w.e. yr^{-1} respectively, Fig. 1.2). This favours the insulation of summer percolating meltwater from winter atmospheric temperatures, typically leading to the formation of PFAs (Kuipers Munneke et al., 2014a). Here, the initial conditions and the spin-up process cause the deep firn to be close to the melting point at the start of the transient run.

The warm firn, combined with the high water influx, allows liquid water to reach greater depths than at the other sites in all three flow schemes. Additionally, the firn – ice transition depth becomes important in the FA13 simulations. The observed core shows that the 810 kg m^{-3} density is reached and maintained from 24 m depth. The CROCUS densification scheme predicts that this density horizon occurs at 60 m depth. Since CROCUS has been developed for seasonal snow, the densification at high overburden stress is probably not well captured by the model (Stevens, 2018). Because of this, we base our simulations for FA13 on the HL densification model, which predicts this transition depth to be around 21 m.

The total refreezing rates are similar for the three flow schemes (Table 1.5). Since the deep firn is close to the melting point, the total refreezing amounts are essentially determined by the cold content provided in winter and the precise behaviour of the percolation has a minor impact. However, variability of refreezing with depth differs between schemes, which leads to differences in the 15 m FAC values (Table 1.5) and in the modelled depth-density profiles (Fig. 1.9a, b and c and Fig. 1.10a). FAC is consistently underestimated (-23 to -30 %) because firn density is overestimated above 10 m. R1M and DPM overestimate density most strongly with FAC values 9 and 10 % smaller than BK respectively, and both schemes simulate the presence of a thick ice layer in the upper 10 m of the firn, which is not observed in the core. The BK model produces only a single thin ice layer in the 10 upper meters (0.2 m thick at 9 m depth), which is in good agreement with the observations (showing a single thin ice layer at 7.5 m depth). Below 10 m, the modelled densities are generally in better agreement and all the schemes produce several ice layers (Fig. 1.10a and Fig. 1.9a, b and c).

	Refreezing / Inflow (1980-2011)	Refreezing / Inflow (2012)	Runoff / Inflow (1980-2011)	Runoff / Inflow (2012)	Top 15 m FAC [m] (anomaly vs observations)	T 10m [K] (anomaly vs observations [K])	Remaining water [m]
BK (HL wh02 ip810)	0.55	0.28	0.45	0.73	3.82 (-23 %)	271.75 (+0.10)	0
R1M (HL grLK ip810)	0.50	0.28	0.49	0.71	3.47 (-30 %)	270.94 (-0.71)	0
DPM (HL grLK ip810)	0.51	0.32	0.38	0.70	3.45 (-30 %) of which 0.81 m of water	273.15 (+1.5)	1.53
Observations	/	/	/	/	4.96	271.65	1.65
BK (HL whCL ip810)	0.60	0.23	0.38	0.81	3.38 (-32 %)	270.99 (-0.66)	0.09
R1M (HL grA ip810)	0.46	0.30	0.52	0.69	3.60 (-27 %)	269.77 (-1.88)	0
DPM (HL grA ip810)	0.51	0.36	0.39	0.70	3.36 (-32 %) of which 0.83 m of water	273.15 (+1.5)	1.48
DPM (HL grLK ip830)	0.51	0.29	0.39	0.70	3.57 (-28 %) of which 0.38 m of water	272.13 (+0.48)	1.64
CROCUS R1M (grLK ip810)	0.49	0.31	0.49	0.68	3.94 (-21%)	271.46 (-0.19)	0
KM R1M (grLK ip810)	0.51	0.37	0.45	0.99	2.29 (-54 %)	270.90 (-0.75)	0

Table 1.5. Model outputs at FA13 site. / indicates no data.

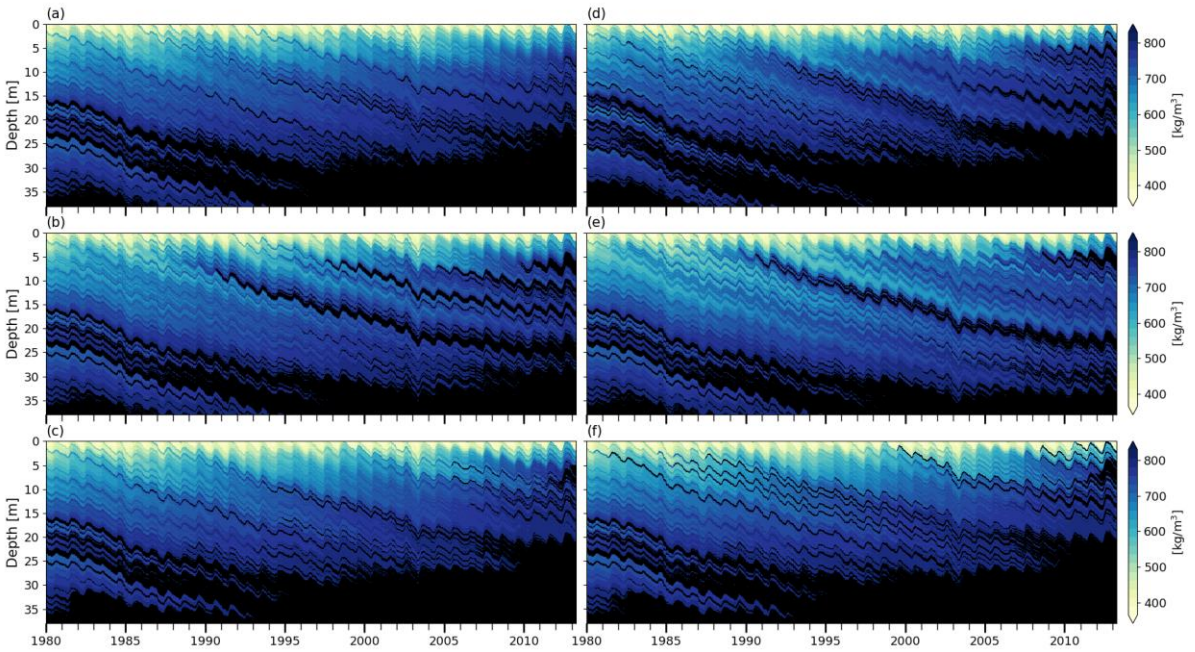


Figure 1.9. Firn density at FA13 (a) BK wh02 ip810, (b) R1M grLK ip810, (c) DPM grLK ip810, (d) BK whCL ip810, (e) R1M grA ip810, (f) DPM grA ip810, black indicates solid ice.

In the absence of any shallow ice layer throughout most of the simulation (Fig. 1.9a, b and c), meltwater is free to percolate through the winter accumulation layers and to deplete their cold content. The flow schemes have different abilities to store liquid water, which leads to small variations in runoff and refreezing rates. In BK, water is retained according to the water-holding capacity (Fig. 1.11a) and refreezes during subsequent winters. In contrast, DPM allows percolation down to the firn – ice sheet transition where it ponds to form an aquifer (Fig. 1.11d and e). This leads to a significant reduction of runoff amounts during the aquifer build-up (-6 % of the water input over the entirety of the transient model run compared to BK) and the water remaining in the firn column is essentially constrained by the maximal amount of water we allow in the aquifer (1.65 m). In theory, the same mechanism could be simulated by R1M but the percolating water is depleted before it reaches the bottom of the firn column (Fig. 1.11b). This is due to refreezing, to the lateral runoff parameterisation and to the presence of ice layers in the upper 10 m. No water persists through the winter seasons, which illustrates the model artefact that the effective saturation must be strictly positive for the stability of the RE (Sect. 3.2). Thus, the refreezing rates are slightly lower than in BK since no residual water is stored and later exposed to winter refreezing (Table 1.5).

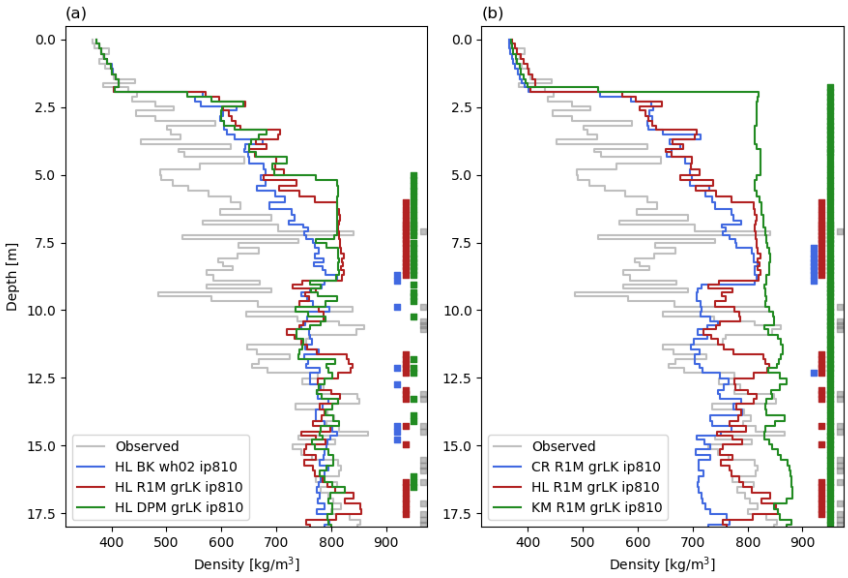


Figure 1.10. Measured and modelled depth-density profiles at FA13 on 10/04/2013. Thick vertical lines show ice layers. The modelled densities are averaged at the vertical resolution of the drilled core. CR: CROCUS, HL: Herron and Langway, KM: Kuipers Munneke.

The build-up of the aquifer starts very early (in the summer of 1981) when DPM is turned on in the transient run due to the low refreezing capacity of the deep firn. The depth of this aquifer is constrained by the impermeability threshold applied, which determines where the model places the firn ice transition. This depth is at 33 m in 1981 and 21 m in 2013, the decrease being caused by

enhanced densification. The aquifer is fed only by preferential flow (Fig. 1.11e) since matrix flow cannot reach the water table due to runoff, refreezing and the presence of ice layers in the firn column.

From 1994 and onwards the total simulated water content in summer is only regulated by the maximum allowed in the model (1.65 m). Since the water table is at a shallow depth towards the end of the simulation (7.5 m), the propagation from the surface of the cold winter temperatures can refreeze part of the saturated layers. This leads to the formation and progressive thickening of the shallow, thick ice layer. Also, the shallowness of the aquifer causes 23% of the porosity in the top 15 m to be filled with liquid water and the 10 m temperature to be at the melting point.

The higher impermeability threshold in DPM grLK ip830 increases the depth of the calculated firn – ice transition, producing a deeper aquifer that extends between 12 and 29 m depth at the end of the simulation, similar to the 12-37 m depth range observed by Koenig et al. (2014). Compared to DPM grLK ip810, the increased depth leads to less refreezing in the shallowest layers of the aquifer and thus a higher FAC value (+3 %) and a 10 m temperature below the melting point.

The grain-size formulation following Arthern et al. (2010) (DPM grA ip810) reduces the ability of preferential flow to transport water down to the firn - ice transition but instead favours formation of discrete ice layers in the firn column (Fig. 1.9f). In this case the aquifer does not start to form until summer 1988, but the final aquifer structure (also between 7.5 and 21 m), the FAC value (-3 % for grA), and the partitioning between refreezing and runoff are similar to those simulated using grLK (Table 1.5). In R1M, the sensitivity to grain-size is noticeable in the firn-structure evolution with differences in ice-layer formation between R1M grLK ip810 and R1M grA ip810 (Fig. 1.9b and e). The final FAC value (+4 % for grA) and the meltwater partitioning remain similar (Table 1.5) between R1M grLK and R1M grA, as for the case of DPM grLK and DPM grA. This can be explained by the total refreezing's stronger dependence on the firn thermal structure than on the percolation pattern at this site.

Increasing the water-holding capacity in BK (BK whCL ip810) leads to a significantly lower FAC value (-12 %): more water refreezes in the near-surface layers, which reduces the runoff and enhances densification in the entire underlying firn column. Also, more water remains stored at depth throughout the different winter seasons (Fig. 1.11c), and some is still present at the end of the simulation (0.09 m) between 16 and 23 m depth. However, this small amount retained by the water-holding capacity is much less than is stored in the saturated layers of the aquifer simulated in DPM.

We compare the three different densification models (CROCUS, HL, KM) using the R1M grLK ip810 flow scheme and these show important differences in the final modelled depth-density profiles

(Fig. 1.10b). CROCUS agrees reasonably well with HL in the top 6 m but, as mentioned above, it has a strong low density bias at greater depths. Since CROCUS simulates lower densification rates, its underestimation of the FAC value in the 15 upper meters (-21 %) is smaller than in HL (-30 %), but it is clearly not representative of the density conditions below 15 m. KM predicts a firn column below the last winter's accumulation entirely at the ice density. The model thus identifies a firn - ice sheet transition at shallow depth (~2 m), which the water can reach before being depleted by the lateral runoff parameterisation and saturated layers can thus build up. This further amplifies the densification since the saturated layers at the transition depth are exposed to refreezing. Hence, in 2012, runoff combined with refreezing exceeds the liquid water input (Table 1.5). This occurs because some layers wherein water had been stored in previous years reach the 810 kg m^{-3} density, causing the stored water to be considered as runoff by the model. Whereas the FAC values are generally close for the different flow schemes and their parameterisations (maximal difference of 12 %), CROCUS and KM reach values 14 % higher and 34 % lower than HL respectively.

1.6 Discussion

The three liquid water schemes show consistent behaviour between sites. R1M generally predicts slower downward percolation of water than the other schemes, which leads to more near-surface refreezing, the formation of near-surface ice layers, more lateral runoff and thus lower densities and lower temperatures in deeper firn. As a result, when compared to observations R1M tends to reach higher FAC values and to underestimate 10 m temperatures. The BK formulation with the Coléou and Lesaffre (1998) parameterisation for the water-holding capacity leads to the same effects, but they are amplified. The underestimation of the 10 m temperature is stronger, suggesting that BK whCL does not allow for deep enough percolation. BK with the lower water-holding capacity (BK wh02) leads to a partitioning of the water input between refreezing and runoff similar to the more complex R1M at the four sites. As a result, the FAC values predicted by BK wh02 and R1M generally agree (maximum difference less than 10 %), as do the temperatures at 10 m depth (maximum difference less than 1 K). The FAC values and 10 m temperatures of R1M at the end of the model runs always lie in the range of the ones obtained with different parameterisations of BK. This suggests that BK can produce results similar to R1M, provided it is parameterised appropriately.

DPM exhibits a different behaviour: it effectively brings water to greater depths, depleting the deep-firn pore space and cold content. Even in the presence of shallow ice layers hindering matrix flow, the preferential flow implementation still ensures efficient vertical water transport, and runoff amounts remain low. This suggests that transfer mechanisms to the preferential flow domain implemented in DPM are more effective in draining ponding water than the lateral runoff parameterisation. Due to large FAC underestimation and 10 m temperature overestimation, the data-model mismatch of DPM

with respect to these variables is significantly greater than that of R1M and BK. DPM is better at producing density variability in depth, which is underestimated in all schemes at all sites. Also, in contrast to the two other schemes, DPM can form ice layers even in summers of average melt, and it is able to simulate the persistence of deep saturated firn layers at the FA13 site. In this respect our findings support those of Wever et al. (2016) who highlight the tendency of DPM to produce ice layers at various depths in alpine snowpack simulations, and thus to reproduce depth-density variability. It is important to bear in mind that we only use the dual permeability water scheme of SNOWPACK in DPM and not the other physics of this model; the results produced by the full SNOWPACK model would be different because it has its own formulations for snow mechanical and thermal properties. In particular, DPM relies heavily on the grain-size, and it would thus benefit from better representations of the firn's structural properties. Moreover, the primary purpose of the DPM implementation in SNOWPACK is to reproduce the occurrence of ice layers in a seasonal alpine snow pack (Wever et al., 2016), whereas in this study we evaluate its ability to simulate representative firn depth-density profiles over the course of numerous decades.

BK with low water-holding capacity is usually used to mimic preferential flow (Reijmer et al., 2012). However, our findings suggest that, in fact, this more closely represents matrix flow as modelled using the Richards Equation. We suggest that in order to use BK in this way, percolation of some meltwater in the presence of ice layers should be considered.

The lack of variability in density in the modelled profiles cannot only be attributed to inaccuracies in the percolation-refreezing process. This is demonstrated in the example of NASA-SE: the layers of the density peak observed around 1 m depth (Fig. 1.4a and b) were deposited during the final winter of the simulation (2015-2016). As such, these have only been influenced by the percolation and refreezing of negligible amounts of liquid water. The consistent underestimation of density variability across all schemes indicates that one or several other factors that are not or poorly represented by firn models likely play a crucial role in firn evolution. These factors may include horizontal water flow, prolonged ponding of water in soaked firn close to the surface, variability in fresh snow density, the effects of firn microstructure and impurity content on densification, wind packing and short-term weather fluctuations. Moreover, the validity of the firn model relies on the accuracy of the climatic forcing.

At all sites, interchanging the HL, KM and CROCUS formulations for firn densification generally leads to more variability in the results than using different water flow schemes. This highlights the need to understand the causes of disparity between densification models under various climatic conditions and thus to proceed to further model intercomparison experiments (Lundin et al., 2017). The existing firn-densification formulations are likely not suited for representing densification in

conditions of high water contents and high refreezing rates. Our study indicates that firn-densification models could be improved by accounting for the latent heat source as well as the effects of liquid water and of refreezing cycles on firn viscosity and densification rates. For example, the KM and HL densification equations were established for dry firn (Herron and Langway, 1980; Kuipers Munneke et al., 2015a). In the CROCUS scheme, firn viscosity is adjusted according to the water content, but our results show that the modification in the parameterisation is insufficient to reproduce the observed densities at KAN-U. A simple example of the densification schemes in HL and KM not representing reality becomes apparent when applying the percolation-refreezing schemes: in reality, densification is dependent on the overburden stress, but these models use accumulation rate as a proxy for stress. Consequently, in these models the redistribution of mass due to runoff and percolation does not affect the densification rates, despite the effect it has on the firn column mass. The absence of a preferential flow scheme is often presented as a possible explanation for firn-density overestimation close to the surface (e.g. Gascon et al., 2014; Kuipers Munneke et al., 2014a; Steger et al., 2017a). However, our results suggest that simply adding a one-dimensional preferential flow scheme, although physically detailed, to firn-densification models does not solve this issue. The water that is transported quickly from the shallow layers must flow back into the matrix domain at some point. If this occurs in the shallow layers, the density-overestimation issue remains; if this occurs in deep layers it can lead to unrealistic temperature signatures. There are several other possible factors for densification errors at high melt sites, including exaggerated sensitivity of the model to temperature because the densification model is not suited for conditions with substantial latent heat release.

We compare the results reached at KAN-U with previous firn modelling studies at this site. Langen et al. (2017) also used the CROCUS densification scheme and a water flow scheme conceptually comparable to R1M but with a simplified solving process of the RE. Their results show a density profile entirely at ice density from 4 m depth and an overestimation of the temperature at 10 m depth (+3 to +5 K). In our study, the R1M results show lower deep densities and a 10 m temperature underestimation. These discrepancies can be attributed to differences in 1) model implementation, 2) climatic forcing and 3) details of the water flow scheme. Model resolution is likely of importance here: Langen et al. (2017) used a coarser vertical resolution, disfavouring the formation of thin impermeable layers, allowing liquid water to flow more readily to greater depths and refreeze, thus causing greater density and higher temperature values. Steger et al. (2017b) used the SNOWPACK densification model (recall: we use different densification physics to SNOWPACK in this study) with a bucket scheme, similar to BK, and the same climatic forcing as in this study. Their model output shows a firn column fully compacted to ice at KAN-U. They attribute this to the overestimation of densification rates by the densification model used and also argued that some other effects could be at play such as the surface density applied. The 10 m temperature is underestimated in their result, likely due to the absence of percolation of water through ice and thus no latent heat release at depth.

Both DPM and R1M exhibit significant sensitivity to the choice of the grain-size formulation (grLK or grA). Modifying this formulation in R1M affects the model results more than changing to the use of BK at all sites apart from FA13 where the magnitudes of change are comparable. For example, the change of grain size at FA13 in R1M leads to a difference of 4 % in FAC and 1.2 K in 10 m temperature, and in DPM, the patterns of ice layer formation are significantly different at all sites where such ice layers form. This highlights another significant difficulty for percolation schemes: the dependence of water flow on the firn's structural properties. Field evidence demonstrates the crucial role of structural transitions, even at the scale of centimeters, on the behaviour of water flow in firn (Marsh and Woo, 1984; Pfeffer and Humphrey, 1996, 1998; Williams et al., 2010). With respect to this, the advanced flow schemes applied in this study have some limitations. Firstly, the structural properties of grains in the firn layer are poorly constrained by observations. Secondly, the parameterisations linking the structural and hydraulic properties on which R1M and DPM rely were derived from a limited number of laboratory experiments. These are typically performed at a very small scale (e.g. shallow snow columns with diameter of 5 cm in the experiments of Yamaguchi et al. (2012)), and are mostly based on homogeneous snow in terms of grain-size and temperature (Yamaguchi et al. 2012; Katsushima et al., 2013). The much larger scale of the GrIS firn layer, the spatial and temporal heterogeneity of its structural properties, and its climatic and glaciological settings render the validity of these idealised parameterisations questionable. Finally, the density dependence of the parameters makes them sensitive to errors in the densification process. Thus, a better knowledge of firn structural properties would only be profitable to water flow schemes if we have a clear understanding of the link between snow structure and its hydraulic properties and vice versa.

Another major limitation of the implementation of physically detailed liquid water flow schemes in one-dimensional firn models is the fact that water flow is in reality three dimensional. Water can flow horizontally on top of buried ice lenses or on thin, near-surface ice crusts caused by daily refreezing (Marsh and Woo, 1984; Pfeffer and Humphrey, 1996). Even at depths greater than 10 m, large masses of liquid water can persist through the winter and move laterally over considerable distances (Humphrey et al., 2012). In one-dimensional models, the key to solving this issue is to accurately partition between vertical percolation and lateral flow; this likely requires a better approach than the lateral runoff parameterisation we implement here. As an example, at FA13 all three water-transport schemes overestimate the density in the 10 upper meters except in the last winter's accumulated layers (0 – 2 m depth), where there is a good agreement with observations (Fig. 1.10a). This suggests a consistent overestimation of the summer meltwater refreezing and underestimation of lateral runoff. Also, the need to use a limit for the PFA water content demonstrates that some processes not represented in the model must regulate its water volume; these are likely lateral movement driven by

hydraulic pressure gradients and connections with englacial and subglacial hydrological systems. A one-dimensional preferential flow scheme aims to correctly partition the water input between matrix flow and fast preferential flow; there are several other difficulties with this approach. These include accurately determining how deep the water can be transported by preferential flow, how much water refreezes and how much is stored as liquid water, and the amount of lateral flow at different depths in the firn column. Similarly, the same considerations apply to liquid water flowing in from upstream grid cells. The representation of preferential flow physics thus requires improvements to address some deficiencies. We highlight (1) that too much water is transported through preferential flow or at exaggerated depths as demonstrated by the consistent overestimation of 10 m temperature, (2) the need to consider lateral flow as the tendency to underestimate FAC shows and (3) the large sensitivity of the ice layer formation process to grain size, which should be overcome or should be addressed by further observational studies of grain metamorphism in firn.

Our results also suggest that more observations of firn-temperature variability in time and depth would likely be useful for the evaluation of existing flow schemes and the development of new ones. Modelled temperature profiles show both the depth and volume of refreezing due to the release of latent heat. Moreover, deep meltwater refreezing causes marked and long-lasting temperature increases due to insulation from the overlying firn, and temperature measurements in depth can be powerful indicators of the occurrence of deep percolation and refreezing events. On the other hand, comparisons between modelled and observed density profiles are strongly affected by the choice and accuracy of the densification formulation, the variability of surface density, several other factors influencing model outputs mentioned above and possible uncertainties in field measurements. Such uncertainties are related to the strong spatial variability of firn structure (Marchenko et al., 2017), which can be observed by comparing density profiles of cores drilled at nearby locations. However, the modelled temperature profile also depends on the accuracy of the climatic forcing, of the heat-transport scheme and of the thermal-conductivity parameterisation. The latter is a function of density, thus erroneous depth-density profiles inevitably lead to an inaccurate heat-transport process. As an example of the influence of various sources of errors, the warm 10 m temperature bias of all the schemes at NASA-SE (Table 1.3) is unlikely to be only due to the percolation-refreezing process. RACMO2.3p2 has a slight negative precipitation bias at this site (-0.21 m w.e. yr-1, see Appendix A1) which likely leads to underestimated advection of cold snow down into the firn column and could partly explain the firn models' temperature bias. Similarly, air temperature is overestimated here by RACMO (+ 1.5 K), which could enhance the problem. It is possible that the consistent over- and underestimation of FAC at FA13 and at KAN-U could be due to snowfall over- and underestimation respectively by RACMO, however no observations of precipitation at these sites is available for comparison.

1.7 Conclusion

We implemented three liquid water schemes of different levels of physical complexity in a firn model using a fine vertical resolution: a bucket scheme, Richards Equation in a matrix flow scheme, and Richards Equation in a preferential flow scheme. To our knowledge, this is the first study to apply the Richards Equation as well as a preferential flow scheme in firn-densification simulations on the GrIS. Our three liquid water flow schemes predict significantly different vertical patterns of refreezing and consequently modelled densities, firn air content values and 10 m temperatures. The preferential flow scheme effectively evacuates meltwater from the surface layers and leads to underestimation of firn air content and overestimation of 10 m temperatures. Compared to the preferential flow scheme, the single-domain Richards Equation scheme generally showed biases of the opposite signs and of much lower magnitudes in both FAC and 10 m depth temperature, suggesting it slightly underestimates percolation depths. The simpler bucket scheme predicted refreezing rates, firn air contents and 10 m temperatures similar to those obtained by the single-domain Richards Equation; by adjusting its water-holding capacity and impermeable density parameters, it could produce the same results. Using the Coléou and Lesaffre (1998) parameterisation for the water-holding capacity in the bucket scheme led to underestimation of percolation depths. The bucket scheme with lower water-holding capacity and the single-domain Richards Equation scheme predicted firn air contents and 10 m temperatures in closest agreement with observations. However, the preferential flow scheme was found to perform better than the simpler flow schemes in reproducing the density variability with depth and the water-saturated conditions at the bottom of the firn column at a site of a perennial firn aquifer.

We identified the multidimensionality of liquid water flow as the prominent challenge for water percolation schemes. Because firn models are currently one dimensional, an accurate partitioning between horizontal and vertical flow is likely to be at least as difficult and as important as the separation between slow matrix and rapid preferential flow. Other difficulties related to water-flow representation include the uncertainties in firn hydraulic properties and in firn micro- and macro-structure on the GrIS. This is further demonstrated by our results showing the sensitivity of the Richards Equation-based schemes on the grain-size formulation. However, the absence of any large-scale field observations of water flow in firn makes it difficult to constrain its implementation and to validate model behaviour. Here, we aimed to assess the applicability of flow schemes developed for seasonal snow models to the Greenland firn. Whilst we did apply some modifications to account for the differences between snow and firn, we suggest that more modifications are likely required since the spatial scales and the structural characteristics of seasonal snowpacks and firn are different.

There is no large scale detailed observation available of liquid water content and percolation pattern during melting events in firn. This renders the validation of a particular flow scheme difficult and

validation relies on temperature and density profiles. However, there are a number of effects that influence firn density and temperature, all potentially contributing to mismatch between modelled and observed values. As an example, the density variability in depth was largely underestimated regardless of the flow scheme. This suggests that there are uncaptured complexities in the percolation and refreezing mechanisms that need to be better taken into account such as water ponding and the effect of grain shape on water flow. Furthermore, this shows that firn-model development must focus on including complex processes currently poorly or not represented, such as surface-density variability, wet firn densification and firn structural effects on densification. We note also that our assessment considers only a single profile from each site, captured at one time point, and further work is required to assess the impact of liquid water percolation on transient firn evolution. A comprehensive exploration of the various firn models and their parameter spaces could help identify priorities for further model developments based on minimising data-model mismatch and overall uncertainty. In line with this, we showed that output from three common firn-densification models shows greater variability than the output from single densification model using the different flow schemes. In order to capture the multiple impacts of liquid water on firn densification, future models require an improved liquid water flow scheme, accurate boundary conditions, and formulations developed explicitly to simulate densification of wet firn.

2. Bayesian calibration of firn densification models

2.1 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.

2.2 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, 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 (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 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).

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., 2015b). 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 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 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., 2015b), 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 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.

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, 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. (2.7)), but because of the high-dimensional parameter space and the non-linearity of firn 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 firn models and improve our understanding of their associated uncertainty.

2.3 Data and Methods

2.3.1 Firn densification data

In order to calibrate three firn densification models, we use observations of firn depth-density profiles from 91 firn cores (see Appendix B) located in different climatic conditions on both the GrIS (27 cores) and the AIS (64 cores) (Fig. 2.1). 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 firn models and wet-firn 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 firn air content. We calculate DIP until 15 m depth ($DIP15$, Eq. (2.1)). For sufficiently deep measurements, we also calculate $DIPpc$, Eq. (2.2), 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:

$$DIP15 = \int_0^{15} \frac{\rho_i - \rho}{\rho_i} dz \quad (2.1)$$

$$DIPpc = \int_{15}^{z_{pc}} \frac{\rho_i - \rho}{\rho_i} dz \quad (2.2)$$

where z (m) increases downwards, ρ is the density of firn (kg m^{-3}) and ρ_i is the density of ice (917 kg m^{-3}). In Eq. (2.2), we consider porosity only below 15 m to avoid dependency between $DIP15$ and $DIPpc$. We choose to use both $DIP15$ and $DIPpc$ 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 $DIP15$ 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 firn 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 Appendix B). The resulting evaluation data has 8 GrIS and 14 AIS cores; 11 of the 22 cores extend to z_{pc}.

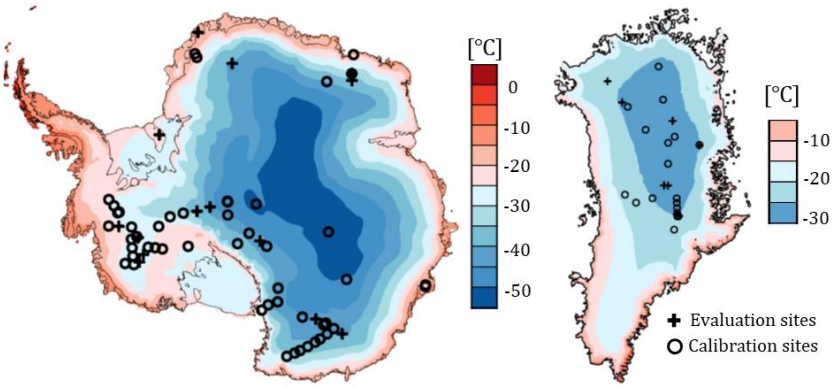


Figure 2.1. Maps of Antarctic (left) and Greenland (right) ice sheets. Background is mean annual air temperature as modelled by RACMO2. Note the different colour scales.

2.3.2 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 Appendix B). 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. 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 Appendix B). We prefer this approach to the use of available parameterisations of ρ_0 (Helsen et al., 2008; Kuipers Munneke et al., 2015b) to avoid any error in the fresh snow parameterisation affecting the calibration process.

2.3.3 Firn densification models

We use the Community Firn 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 firn 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:

$$\begin{cases} \frac{d\rho}{dt} = c_0\,(\rho_i - \rho), & \rho \leq 550\ kg\ m^{-3} \\ \frac{d\rho}{dt} = c_1\,(\rho_i - \rho), & \rho > 550\ kg\ m^{-3} \end{cases} \tag{2.3}$$

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

HL

$$\begin{cases} c_0 = \dot{b}^a k_0^* \exp\left(\frac{-E_0}{RT}\right) \\ c_1 = \dot{b}^b k_1^* \exp\left(\frac{-E_1}{RT}\right) \end{cases} \quad (2.4)$$

Ar

$$\begin{cases} c_0 = \rho_w \dot{b}^\alpha k_0^{Ar} g \exp\left(\frac{-E_c}{RT} + \frac{E_g}{RT_{av}}\right) \\ c_1 = \rho_w \dot{b}^\beta k_1^{Ar} g \exp\left(\frac{-E_c}{RT} + \frac{E_g}{RT_{av}}\right) \end{cases} \quad (2.5)$$

LZ

$$\begin{cases} c_0 = \beta_0 lz_a (273.15 - T)^{lz_b} \dot{b} \\ c_1 = \beta_1 lz_a (273.15 - T)^{lz_b} \dot{b} \end{cases} \quad (2.6)$$

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

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 firn layer because it better approximates the overburden stress than 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 firn cores, from which they inferred the values for the six free parameters (Table 2.1) 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 2.1) 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 (2.6) shows that LZ has eight free parameters (Table 2.1), 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. (2.6) 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. (2.6) against firn cores only from the GrIS. 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. (2.6)). Since one of the goals of this study is to find a densification formulation applicable to firn in both the GrIS and AIS, we choose to apply our calibration method only to the formulations of β_0 and β_1 specified in Li and Zwally (2011) (Eq. (2.6)). However, in our results' analysis (Sect. 2.4), we also consider the performance of the Li and Zwally (2015) model on the AIS cores of our dataset.

Parameter	Value in original model	Prior distribution	MAP	95 % Credible interval
k_0^* [m w.e.$^{-a}$]	11	$N(11, 100)$	17.4	7.58; 28.4
k_1^* [m w.e.$^{-b}$]	575	$N(575, 9 \times 10^4)$	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 2.1. 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 Appendix B). Maximum a posteriori (MAP) estimates and credible intervals are results from the calibration process.

2.3.4 Bayesian calibration

In our approach, the free parameters of the firn 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. (2.7)) which allows to update a prior probability distribution $P(\theta)$ for θ based on observed data Y.

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)} \quad (2.7)$$

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 2.1). 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 Appendix B). 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 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:

$$P(Y|\theta) \propto \exp\left[\frac{-1}{2}(X_{15} - Y_{15})^T \Sigma_{15}^{-1}(X_{15} - Y_{15}) - \frac{1}{2}\left(X_{pc} - Y_{pc}\right)^T \Sigma_{pc}^{-1}\left(X_{pc} - Y_{pc}\right)\right] \quad (2.8)$$

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 firn 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. (2.2)). 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 Appendix B. 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 2.1), the distribution $P(\theta|Y)$ is essentially governed by the likelihood function (Eq. (2.8)). 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. 2.2, on which we base the brief following description. A given model (HL, Ar or LZ) starts with the original model parameter values and simulates firn profiles at all the calibration sites. Its $DIP15$ and $DIPpc$ results are compared with observations and

the general performance of the model is quantified by the likelihood. From there and with the prior distributions assumed, the posterior probability is computed following Eq. (2.7). At this point, the RWM algorithm starts and the state of the chain, θ_i (Fig. 2.2a), 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. 2.2b). 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. 2.2c) and $P(\theta_i^*|Y)$ is computed (Fig. 2.2d). 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. 2.2e). The set saved in the ensemble (Fig. 2.2g) 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. 2.2a) with its associated posterior probability, $P(\theta_{i+1}|Y)$. 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 Appendix B. The proposal 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):

$$\Sigma_{prop} = \frac{2.38^2}{p}\,\Sigma_{cov} \tag{2.9}$$

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.

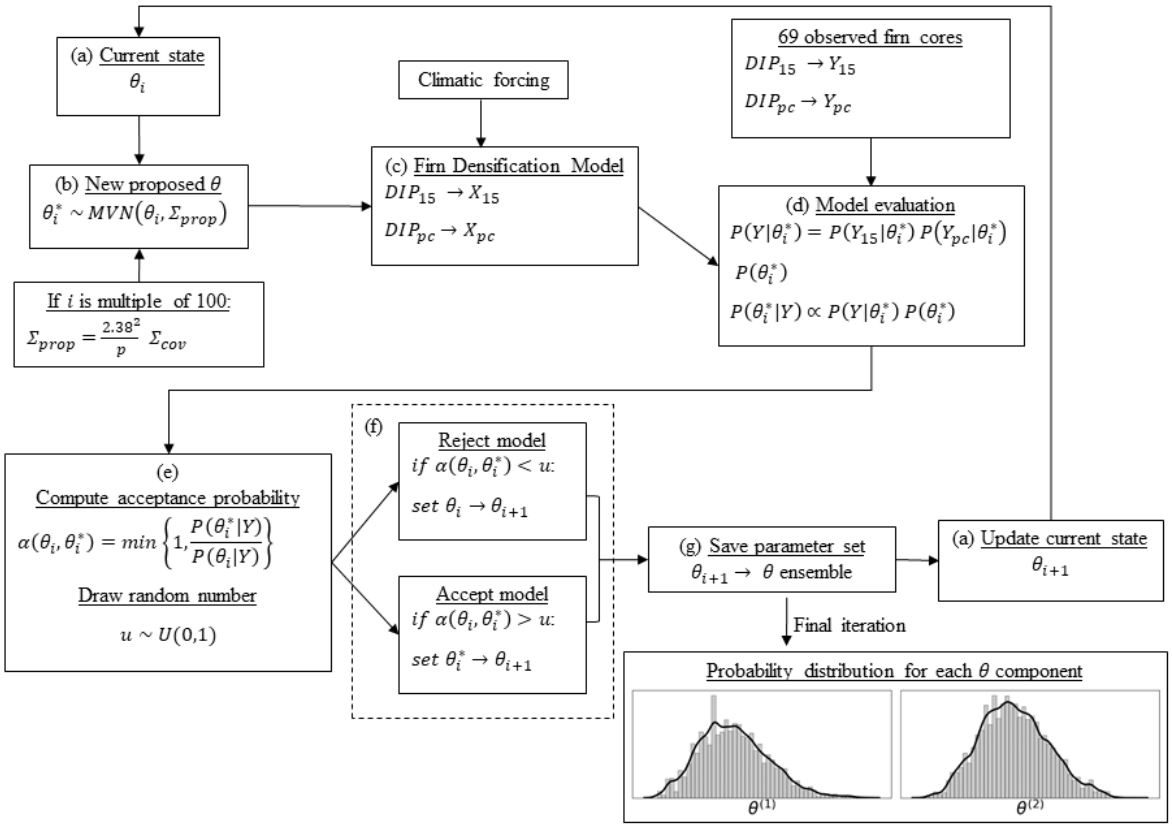


Figure 2.2. Implementation of the Random Walk Metropolis algorithm. θ represents a parameter combination of any given firn densification model investigated.

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 accounting for uncertainty. Intuitively, a large spread in results from 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 firn 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 instead of relying on posterior samples of the MCMC. We provide information about the normal approximations and assess

their validity in the Appendix B. Such normal approximations are asymptotically exact and are commonly applied to analytically intractable Bayesian posterior distributions (Gelman et al., 2013).

2.4 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.

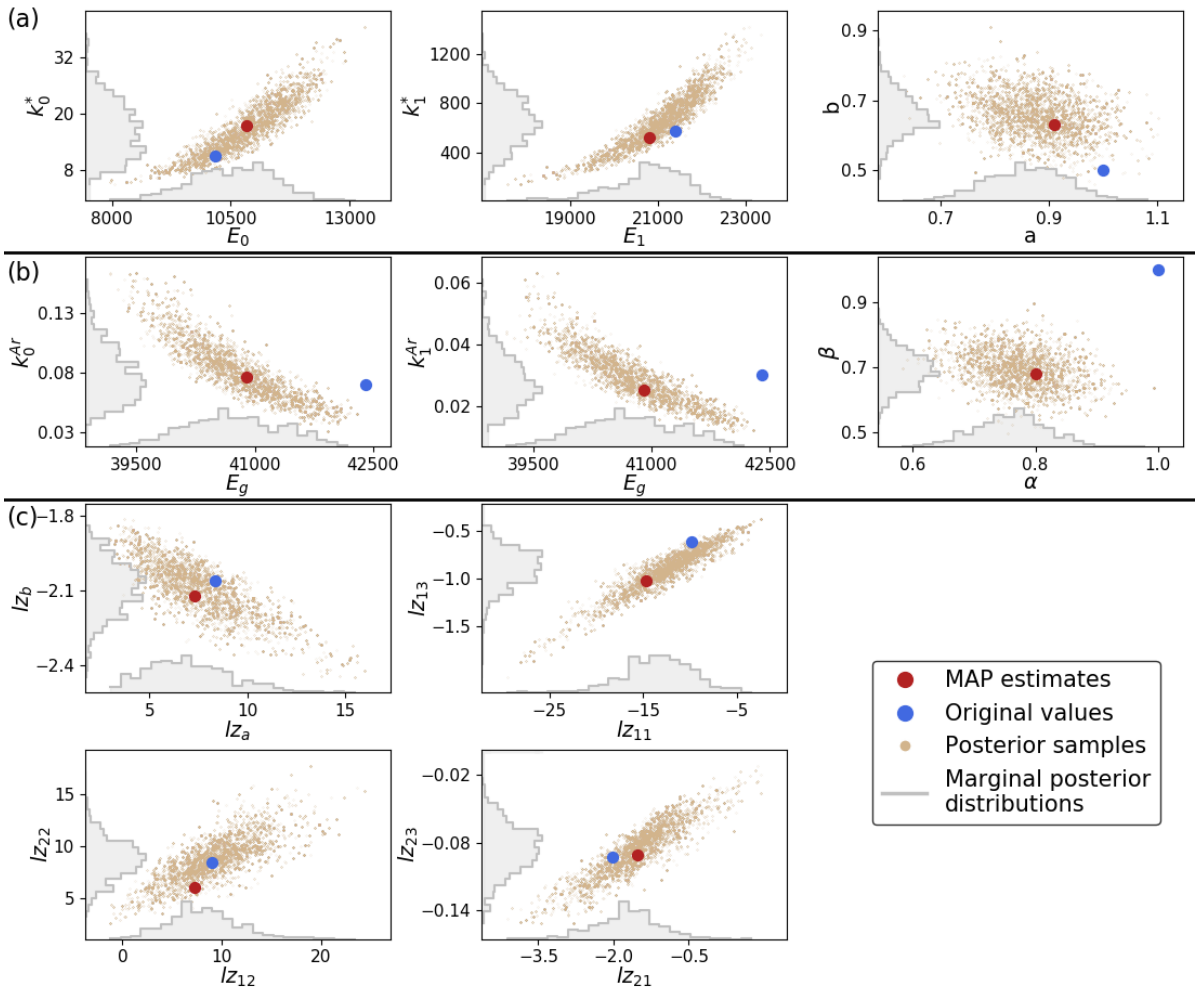


Figure 2.3. Posterior probability distributions, shown for pairs of parameters, for (a) HL, (b) Ar, (c) LZ. Where possible, correlated parameters share the same graph (see Appendix B for full correlation matrices). The posterior samples are 500 randomly selected parameter combinations from the posterior ensembles of each model (HL, Ar, LZ).

For HL and even more so for Ar, the posterior distributions for the parameters demonstrate some strong disagreements with the original values (Figs. 2.3a, 2.3b). The 95% credible intervals for each

parameter (Table 2.1) 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. 2.3a, 2.3b) 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 2.1 and Fig. 2.3c). The posterior distributions show some strong correlation between certain pairs of parameters (Fig. 2.3). 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 Appendix B.

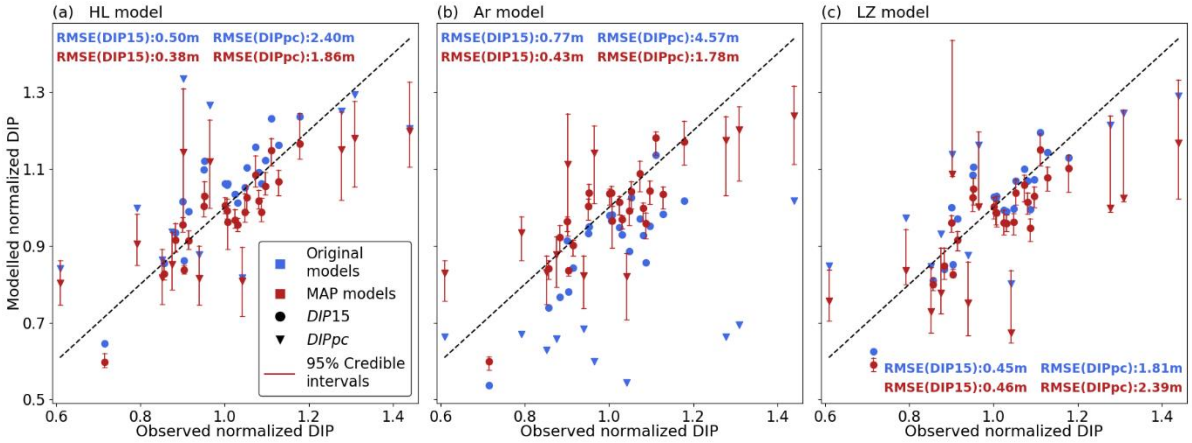


Figure 2.4. Comparison of evaluation data DIP with model results. The 95% credible intervals are computed from results of 500 randomly selected parameter combinations from the posterior ensembles of each model (HL, Ar, LZ). Similar scatter plots for the LZ dual and IMAU results are shown in the Appendix B (Fig. B6).

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. 2.4a) and even more for MAP_{Ar} (Fig. 2.4b), with the original Ar strongly underestimating DIP values. These improvements translate into significantly reduced root mean squared errors (RMSE) in modelled values of both $DIP15$ (-24% for HL and -45% for Ar) and $DIPpc$ (-22% and -61%) (Table 2.2).

Model	RMSE ($DIP15$) [m]	RMSE ($DIPpc$) [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 2.2. Model results on the evaluation data. 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.

For LZ, the relative performance of the MAP_{LZ} model for both $DIP15$ and $DIPpc$ is worse (+2% and +24% in RMSE) but differences are of smaller magnitude (Table 2.2 and Fig. 2.4c). 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. 2.5). 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. 2.5), improvements are reached for the three MAP models (Figs. 2.5g, 2.5h, 2.5i). 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. 2.4). 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.

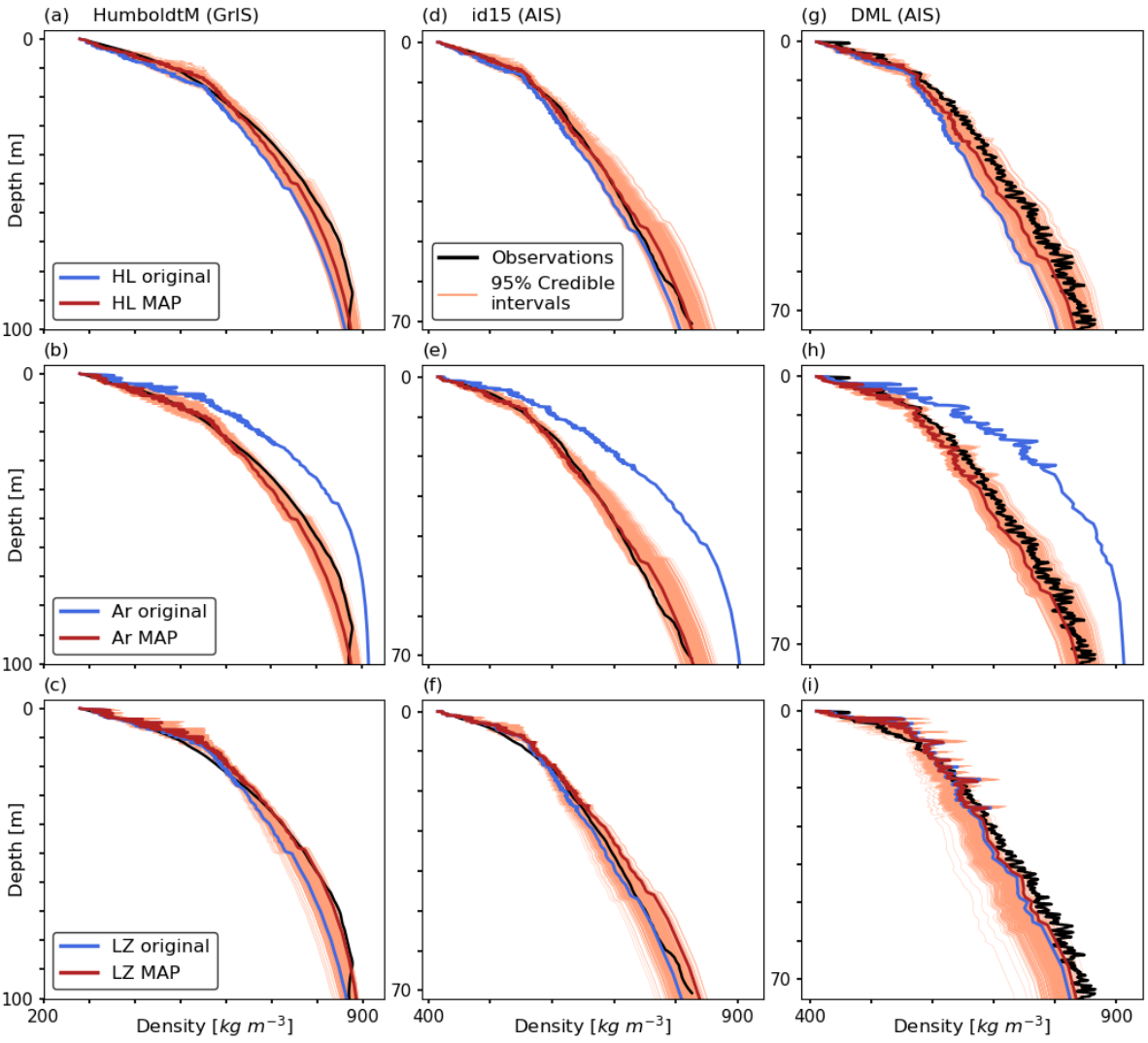


Figure 2.5. Depth-density profiles at three evaluation sites. DML is a climatic outlier of our dataset with particularly high temperatures and accumulation rates. The 95% credible intervals are computed from results of 500 randomly selected parameter combinations from the posterior ensembles of each model (HL, Ar, LZ).

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. 2.6a). 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. 2.6a) 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 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. 2.6b). 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. 2.6c), 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.

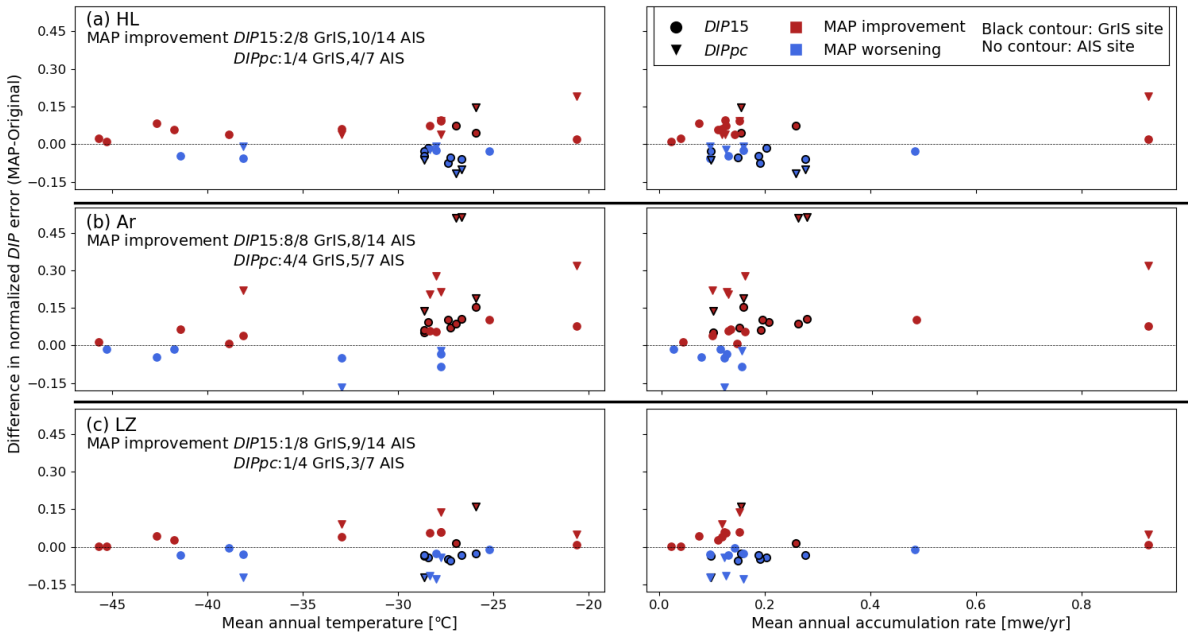


Figure 2.6. Improvements of the MAP models with respect to the original models for the evaluation data. The ratios indicate the ratios of cores for which an improvement is achieved by the corresponding MAP. Graphs in the left column display the mean annual temperature on the x-axis and those in the right column display the mean annual accumulation rate.

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. The RMSE for $DIP15$ of LZ dual is slightly larger (+8 %) than that of MAP_{LZ} and significantly larger (+38 %) for $DIPpc$ (Table 2.2). 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., 2015b), 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 2.2).

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 2.2). This demonstrates a low uncertainty in the optimal parameter combinations identified by calibration. Furthermore, the best performing 95^{th} 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 2.2). MAP_{LZ} performs worse than the other MAP models even when accounting for uncertainty by using the 500-samples random selections (Table 2.2).

2.5 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 firn 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. Under the assumption of linear viscous deformation, firn viscosity changes caused by increasing densities can lead to a sublinear relation between densification and load increase. 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. 2.3a, b). 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 firn, 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, firn 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 firn 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 firn with larger contact areas between grains. The difference in sensitivities of stage-1 and stage-2 densification to accumulation also holds in the LZ model, as illustrated by the posterior correlation between its free parameters. The correlation coefficient 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. (2.6)). However, β_0 appears in the numerator of the β_1 calculation (Eq. (2.6)) 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 Appendix B.

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., 2015b). 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:

$$\frac{\partial c_0}{\partial \dot{b}} = \rho_w \, k_0^{Ar} g \exp\left(\frac{-E_c}{RT} + \frac{E_g}{RT_{av}}\right) \alpha \, \dot{b}^{\alpha-1} \tag{2.10}$$

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. 2.3b). 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 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), firn grain size (Arthern et al., 2010), impurity content (Freitag et al., 2013) and a transition region between stage-1 and 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 firn data remain scarce. As long as more firn 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 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, firn 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 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 firn models are of common use: calculating firn compaction rates and predicting the age of firn 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:

$$cmp_{an,i} = cmp_{tot,i}^{00-17} - 17cmp_{ref,i}^{yr} \quad (2.11)$$

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 2.3). 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 2.3). 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. 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 uncertainties are aggregated when mass balance trends are evaluated. 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.

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

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. 2.7). 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 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-timescale sensitivities (insets in Fig. 2.7), 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.

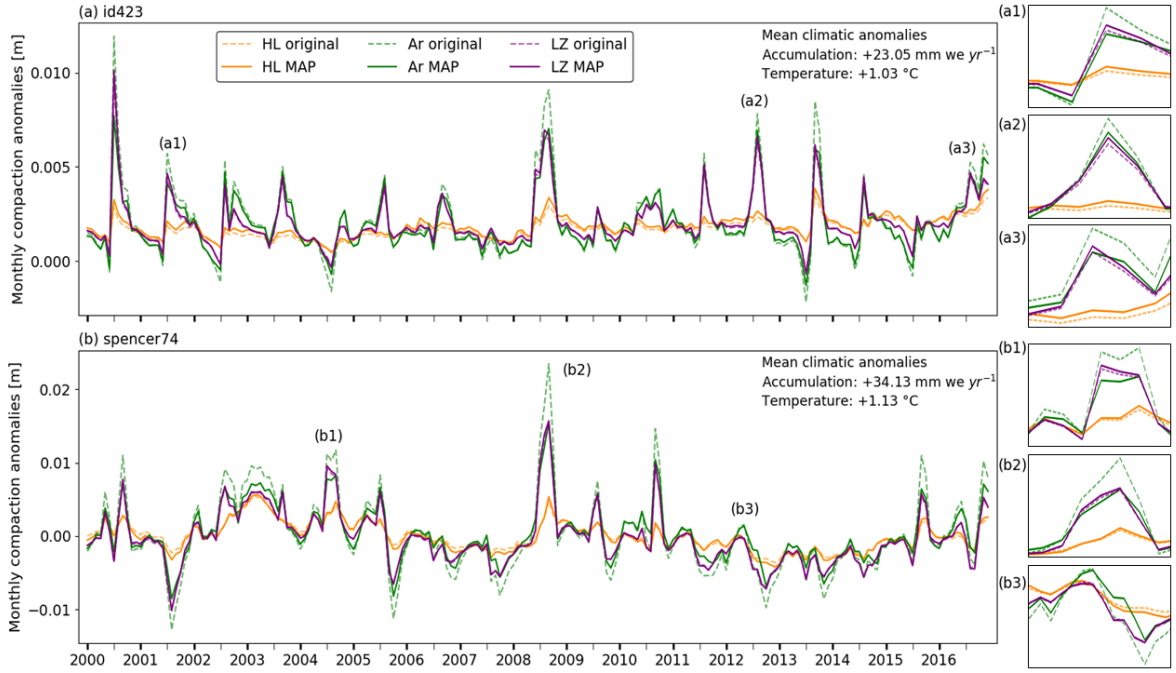


Figure 2.7. Monthly time-series of compaction anomalies at two sites on the GrIS. Insets show details for particular intervals of the time-series. Mean climatic anomalies are calculated as a difference between mean climatic values over the period 2000-2017 with respect to the reference period 1960-1979, and based on RACMO2 values.

2.6 Conclusion

We have implemented a Bayesian calibration method to estimate optimal parameter combinations applicable to GrIS and AIS firn for three benchmark firn densification models (HL, Ar, LZ). An extensive dataset of 91 firn 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 firn densification models. The formulation of models' governing equations impacts the remaining errors with respect to observations, which highlights deficiencies in our understanding of dry firn 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 firn, which are difficult to observe. Our study

demonstrates that, despite these observational limitations, thorough calibration methods relying only on climatic variables can substantially improve firn model accuracy, and constrain uncertainties.

3. Uncertainty in East Antarctic firn thickness constrained using a model ensemble approach

3.1 Abstract

Mass balance assessments of the East Antarctic ice sheet (EAIS) are highly sensitive to changes in firn thickness, causing substantial disagreement in estimates of its contribution to sea-level. To better constrain the uncertainty in recent firn thickness changes, we develop an ensemble of 54 model scenarios of firn evolution between 1992-2017. Using statistical emulation of firn-densification models, we quantify the impact of firn compaction formulation, differing climatic forcing, and surface snow density on firn thickness evolution. At basin scales, the ensemble uncertainty in firn thickness change ranges between 0.2–1.0 cm yr^{-1} (15–300% relative uncertainty), with the choice of climate forcing having the largest influence on the spread. Our results show the regions of the ice sheet where unexplained discrepancies exist between observed elevation changes and an extensive set of modelled firn thickness changes estimates, marking an important step towards more accurately constraining ice sheet mass balance.

3.2 Introduction

The Antarctic ice sheet (AIS) is the largest ice body on Earth, holding a total potential contribution to sea-level rise of ~57.2 m (Rignot et al., 2019). The AIS is divided into three entities: the Antarctic Peninsula (AP) and the West and East Antarctic ice sheets (WAIS and EAIS, respectively). The EAIS has shown less dynamic instabilities than the WAIS and AP over the past four decades, but it holds ~90% of the total AIS ice mass and is the area with highest uncertainty concerning recent mass trends (Shepherd et al., 2018; Rignot et al., 2019). A layer of firn, the intermediary stage between snow and ice, covers ~99% of the AIS (Winther et al., 2001). The firn layer thickness, defined here as the depth from the surface until the firn-ice transition, varies from 0 to more than 100 m (van den Broeke, 2008). Firn thickness also fluctuates in time due to changes in firn compaction rates and climatic conditions, primarily net snow accumulation. These fluctuations affect ice sheet mass balance assessments derived from satellite-based altimetry. Measured surface elevation changes are converted into mass changes, but the conversion requires precise knowledge of variability in firn thickness and mass. Atmospheric reanalysis products, Regional Climate Models (RCMs) and Firn Densification Models (FDMs) are therefore used to simulate changes of firn properties and, ultimately, to evaluate ice sheet mass changes with precision (Li and Zwally, 2011; Kuipers Munneke et al., 2015b; McMillan et al., 2016; Shepherd et al., 2019; Smith et al., 2020).

By simulating mass fluxes (snowfall, sublimation, melt), RCMs estimate the surface mass balance (SMB) of ice sheets, which partly determines the evolution of the firn layer. These fluxes and

modelled surface temperatures also serve as input forcing for FDMs that explicitly simulate firn compaction rates. Such coupling is required to reproduce seasonal and multi-annual fluctuations in compaction rates (Arthern et al., 2010). Uncertainty in SMB estimates across Antarctica are typically assessed by comparing outputs from different RCMs. While SMB is key to firn thickness evolution because it determines the amount of snow removed and added at the surface, it does not capture the effects of fluctuating firn compaction that must be estimated with FDMs. Differences between FDMs can lead to substantial spread in modelled firn thickness and air content (Lundin et al., 2017), especially if scaled up to ice sheet extent.

Compared to the AP and the WAIS, observed elevation changes across the EAIS over the past 25 years have been generally smaller, and largely driven by snowfall and compaction variability (Davis et al., 2005; Shepherd et al., 2018; Shepherd et al., 2019). Altimetry-derived mass balance assessments of the EAIS are very sensitive to estimates of firn thickness fluctuations because these are of the same order of magnitude as measured elevation changes. This sensitivity complicates the interpretation of altimetry measurements in this area, and it is unclear whether to attribute elevation changes to ice dynamical imbalance or firn thickness change (Zwally et al., 2015; Scambos and Shuman, 2016; Shepherd et al., 2018). These conflicting assessments motivate precise uncertainty analyses of coupled RCM-FDM systems over the EAIS. Previously, such analyses have been computationally limited; running multiple FDMs for many years at the spatial resolution of RCM grids over the EAIS requires many thousands of simulations. The extent to which estimates of firn thickness change vary by combining different FDMs with different RCMs remains an open question.

To overcome computational limitations and thus improve evaluation of uncertainty in the evolution of firn thickness, we build statistical emulators of nine FDMs. An emulator is a fast and statistically-driven approximation of a more complex physical model (Sacks et al., 1989; O'Hagan 2006). By combining the FDM emulators with climatic output from three state-of-the-art polar RCMs, we develop an ensemble of 54 scenarios of EAIS firn thickness change. We exploit these scenarios to constrain uncertainty analyses of firn thickness fluctuations on the EAIS and to quantify the contributions of various sources of uncertainty to the spread of modelled results.

3.3 Methods

3.3.1 Ensemble configuration

To generate our ensemble, we first calibrate each of the nine FDM emulators to its corresponding FDM (Table 3.1) in a representative range of EAIS climate conditions, and then we use it to emulate compaction rates across the entire EAIS. Changes in SMB as well as climatic forcing for the emulators are computed from three RCMs: RACMO2, MAR and HIRHAM (Table 3.1). Our

modelled scenarios of firn thickness change span the 1992-2017 period. This period is chosen to match the long-term altimetry record of Shepherd et al. (2019), hence facilitating intercomparison of observed elevation changes and modelled firn thickness change experiments of this study. We limit our analysis to the EAIS because surface melt there is minor compared to the AP and WAIS, and FDM fidelity remains questionable for simulating wet firn compaction, water percolation and refreezing (Steger et al., 2017a; Verjans et al., 2019; Vandecrux et al., 2020).

	References
FDM	
Armap	Arthern et al. (2010); Verjans et al. (2020)
GSFC-FDMv0	Smith et al. (2020)
Cr	Vionnet et al. (2012); van Kampenhout et al. (2017)
HL	Herron and Langway (1980)
HLmap	Herron and Langway (1980); Verjans et al. (2020)
Lig	Ligtenberg et al. (2011)
LZ15	Li and Zwally (2015)
LZmap	Li and Zwally (2011); Verjans et al. (2020)
Morris	Morris and Wingham (2014)
RCM	
RACMO2.3p2 (27 km)	van Wessem et al. (2018)
MARv3.11 (35 km)	Agosta et al. (2019); Kittel et al. (2020)
HIRHAM5 (12.5 km)	Christensen et al. (2007)
ρ_0	
L11	Ligtenberg et al. (2011)
fixed-350	Smith et al. (2020)

Table 3.1. The nine firn densification models (FDM), three regional climatic models (RCM) and two surface density parameterisations (ρ_0) used in this study. The horizontal resolutions of the RCM grids are given in brackets. All RCMs were forced by the ERA-Interim reanalysis at their boundaries (Dee et al., 2011). See Appendix C for details on the FDMs.

3.3.2 Firn thickness change calculations

Observed ice sheet elevation changes, once corrected for glacial isostatic adjustment, are composed of two different signals: one related to ice dynamical imbalance and one to firn thickness change. In this study, we focus on the latter. The change in firn thickness at time step t, $dh_f(t)$, is given by:

$$dh_f(t) = dh_{acc}(t) - dh_M(t) - dh_c(t) - dh_{ice}(t) \quad (3.1)$$

with all components expressed in metres and considered positive, and t set to a daily time step in this study. The subscript M refers to surface firn removal by melting and acc refers to net snow accumulation. Both dh_{acc} and dh_M depend on the RCM-computed mass fluxes and on the value assumed for surface snow density, but they are independent of FDM calculations. The component dh_c is the emulated firn compaction term (Section 3.3.3). The last component, dh_{ice}, quantifies changes in the flux through the lower boundary of the firn column and thus captures changes in the rate of conversion from firn to ice. Equation (3.1) neglects changes in firn thickness caused by temporal variability of horizontal divergence (e.g. Horlings et al., 2020). Changes in dh_{ice} act on much longer

timescales than the other components (Zwally and Li, 2002; Kuipers Munneke et al., 2015b). As such, dh_{ice} can be set constant and equal to the average rate of conversion from firn to ice over a reference period (we use 1979-2009, following Ligtenberg et al., 2011), dh_{ice}^{r}. By assuming firn thickness in steady state, thus without trend, over the reference period, dh_{ice}^{r} balances the reference period averages of the other components:

$$dh_{ice}^{r} = dh_{acc}^{r} - dh_{M}^{r} - dh_{c}^{r} \quad (3.2)$$

Substituting dh_{ice}^{r} for dh_{ice} in Eq. (3.1) yields Eq. (3.3). This is equivalent to calculating firn thickness change by computing anomalies in each of the acc, M and c components with respect to their average value in the reference period (Li and Zwally, 2015).

$$dh_{f}(t) = dh_{acc}(t) - dh_{M}(t) - dh_{c}(t) - dh_{ice}^{r} \quad (3.3)$$

In this study, we are interested in the cumulative 1992-2017 firn thickness changes. We thus integrate Eq. (3.3) over this time period to compute a total firn thickness change dh_{f}^{tot}.

3.3.3 Emulation of firn compaction

The nine FDM emulators are first calibrated at 50 sites on the EAIS and over the entire time span (1979-2017) covered by the output of RCMs (Appendix C for details). The goal of the emulation is to capture both long- and short-term sensitivity of dh_c to climatic forcing. The long-term (1979-2017) mean and trend in dh_c are estimated by linear regressions on the long-term means and trends of temperature, accumulation and melt. These linear regressions are specific to each FDM and show good performance in capturing the FDM-computed means and trends at the calibration sites (R^2>0.99 and R^2>0.97 respectively). Gaussian Process regression complements the linear regression by capturing short-term fluctuations from the long-term trends as a function of detrended values of temperature and accumulation. We evaluate the emulation capabilities in a leave-one-out cross-validation framework; the nine FDM emulators reproduce the FDM output well, both for the total 1979-2017 dh_c (R^2>0.99, RMSE=0.49 m, corresponding to 3.5% of the mean total dh_c) and for daily values (R^2>0.99, RMSE=0.15×10^{-3} m) (Appendix C for details).

3.3.4 Uncertainty contributions

In order to evaluate uncertainty on the time series of cumulative $dh_f(t)$ and on dh_f^{tot}, we construct a model ensemble; the spread arising from a large number of simulations provides an estimate of uncertainty (e.g. Déqué et al, 2007). Our ensemble includes all combinations of the nine FDM emulators and the three RCMs (Table 3.1). Furthermore, surface snow density, ρ_0, contributes to uncertainty in all components of dh_f (e.g. Ligtenberg et al., 2011; Agosta et al., 2019). As such, we use two different possibilities for the value of ρ_0: the climate-dependent parameterisation of Ligtenberg et al. (2011) and the approach of Smith et al. (2020), which takes a constant value of 350 kg m^{-3} (B. Medley, personal communication, 2020) (Table 3.1). The different combinations of RCM,

FDM and ρ_0 provide 54 different firn thickness change scenarios across the EAIS. We refer to the spread in the model ensemble results as the total ensemble uncertainty to distinguish it from the true uncertainty, which may not be captured by the ensemble. We then use the analysis of variance (ANOVA) theory to partition the total ensemble uncertainty among the three factors RCM, FDM and ρ_0 (von Storch and Zwiers, 1999; Déqué et al., 2007; Yip et al., 2011). This approach allows us to decompose the variance in model results into the contribution of each factor and of each interaction between these factors (Eq. (3.4)).

$$\sigma^2 = \eta^2_{RCM} + \eta^2_{FDM} + \eta^2_{\rho_0} + \eta^2_{RCM-FDM} + \eta^2_{RCM-\rho_0} + \eta^2_{FDM-\rho_0} + \eta^2_{RCM-FDM-\rho_0} \quad (3.4)$$

where σ^2 is the variance in the ensemble results (m^2) and the η^2 terms are the contributions from each factor and interaction between factors to σ^2. Interaction effects stem from a non-linear behaviour of the three uncertainty sources. Contributions are calculated by computing the sum of squares associated with each η^2 term.

$$\begin{cases} \eta^2_i = \frac{1}{N_i}\sum_{i=1}^{N_i}(x_{i..} - x_{...})^2 \\ \eta^2_{i-j} = \frac{1}{N_i N_j}\sum_{i=1}^{N_i}\sum_{j=1}^{N_j}\left(x_{ij.} - x_{i..} - x_{.j.} + x_{...}\right)^2 \\ \eta^2_{i-j-k} = \frac{1}{N_i N_j N_k}\sum_{i=1}^{N_i}\sum_{j=1}^{N_j}\sum_{k=1}^{N_k}\left(x_{ijk} - x_{ij.} - x_{i.k} - x_{.jk} + x_{i..} + x_{.j.} + x_{..k} - x_{...}\right)^2 \end{cases} \quad (3.5)$$

where N denotes the number of possible levels for a factor (3 for RCMs, 9 for FDMs, 2 for ρ_0), x denotes the value of the variable of interest (dh_f^{tot}) and a dot represents the arithmetic mean with respect to the index it is substituted for. Because the sums of squares in Eq. (3.5) are averaged departures from a mean, these terms are biased estimates of the variance (Déqué et al., 2007). An unbiased estimate should be divided by $N - 1$, but dividing by N results in η^2 terms fulfilling Eq. (3.4). As such, any ratio η^2/σ^2 is only interpreted as a percentage of contribution to the total ensemble uncertainty. We group together all η^2 terms capturing an interaction effect to quantify the non-linear behaviour of the model experiments with respect to the three factors (η^2_{intr}).

3.4 Results

3.4.1 Ensemble scenarios

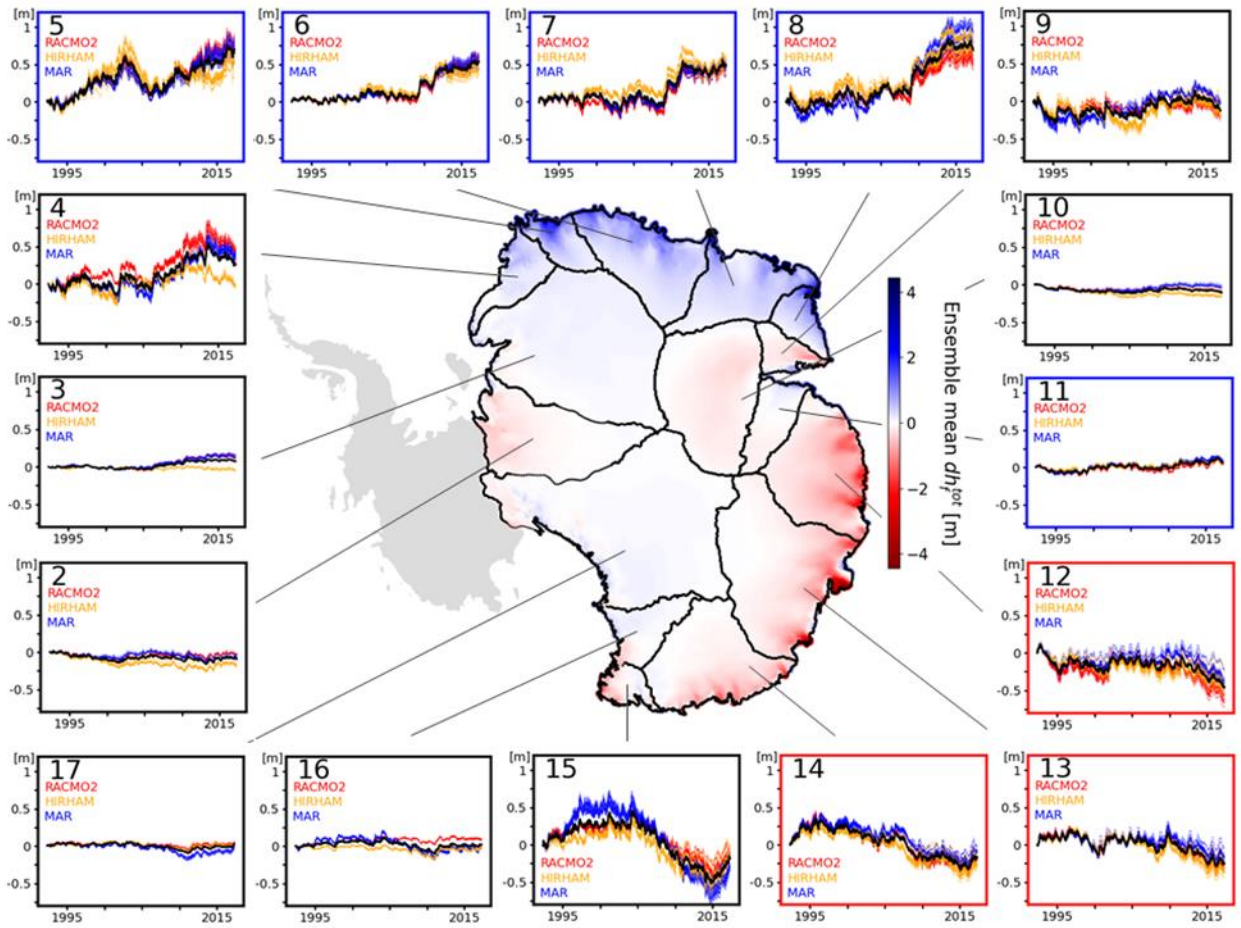


Figure 3.1. Ensemble mean 1992-2017 firn thickness change (dh_f^{tot}) in each EAIS basin. Simulation results are interpolated by nearest-neighbour to a common 12.5 km grid. The map uses a 3×3 median filter. Each inset shows the basin-averaged modelled time series of all the 54 model scenarios. Red, yellow and blue curves represent scenarios forced with RACMO2, HIRHAM and MAR respectively. Each curve represents a particular RCM-FDM-ρ_0 combination. The thick black curve represents the ensemble mean. Basin numbers are displayed within the insets. Frame colours show whether dh_f^{tot} is significantly positive (blue), negative (red) or not significantly different from zero (black) (at $\pm 2\sigma$). Basin limits follow Zwally et al. (2015).

The model ensemble shows a stable firn thickness over 1992-2017 for most of the EAIS, but strong regional changes are evident in several of the 16 basins (Fig. 3.1). The large interior basins (2, 3, 10, 17) show no significant thickness change; the 2σ uncertainty ranges of the ensemble results encompass zero. In contrast, the ensemble shows a significant and pronounced (>0.49 m) firn thickening in Dronning Maud Land (basins 5-8), driven by high snowfall rates since 2009 (Boening et al., 2012; Medley et al., 2018). Conversely, decreases in snowfall rates cause firn thinning (>0.25 m) in the areas of Shackleton ice shelf and Totten glacier (basins 12-13), which coincide with localised zones of high ice flow velocities (Rignot et al., 2019). Low accumulation since 2005 also induced thinning in Victoria Land (basin 14) (Velicogna et al., 2014). In such cases of accumulation anomalies, the firn compaction signal must be accounted for as it partially mitigates the overall change in firn thickness; increased accumulation provides more pore space and thus higher compaction rates, while decreased accumulation has the opposite effect. In other basins, the ensemble

suggests thickening (e.g. basins 3-4) or thinning (e.g. basins 9 and 15) but high variability among model scenarios precludes any firm conclusion.

Basin	dh_f^{tot} [cm]	σ_f^{tot} [cm]	$trend\ dh_f$ [cm yr^{-1}]	$\sigma^{trend}\ dh_f$ [cm yr^{-1}]	$\frac{\sigma^{trend}\ dh_f}{\lvert trend\ dh_f \rvert}$ [%]	Altimetry trend [cm yr^{-1}]	η^2_{RCM}/σ^2 [%]	η^2_{FDM}/σ^2 [%]	$\eta^2_{\rho_0}/\sigma^2$ [%]	η^2_{intr}/σ^2 [%]
2	-9.1	6.5	-0.3	± 0.3	100	0.3 ± 2.3	92.1	6.2	0.2	1.5
3	7.3	8.8	0.5	± 0.4	80	0.7 ± 0.1	97.6	1.4	0.0	1.0
4	26.7	21.3	2	± 0.8	40	3.3 ± 0.4 *	97.3	1.0	0.4	1.3
5	70.9	12.6	2.2	± 0.6	27	4.4 ± 0.6 *	49.4	32.5	15.3	2.8
6	54.1	6.9	2	± 0.3	15	1.8 ± 0.3	28.5	42.6	26.1	2.8
7	49.0	6.8	1.9	± 0.3	16	1.6 ± 0.3	18.9	46.0	28.2	6.9
8	68.9	16.0	3.6	± 1.0	28	4.1 ± 0.4	71.9	11.8	12.6	3.7
9	-13.1	9.1	0.8	± 0.7	88	3.0 ± 0.7 *	69.5	23.9	0.3	6.3
10	-10.2	5.5	-0.1	± 0.3	300	0.0 ± 0.2	92.5	6.2	0.1	1.3
11	6.3	2.2	0.5	± 0.2	40	0.4 ± 0.4	84.9	6.0	5.6	3.5
12	-44.8	12.0	-0.8	± 0.4	50	1.8 ± 0.4 *	51.0	40.8	3.9	4.3
13	-25.5	9.0	-1.2	± 0.5	42	-0.7 ± 0.4	60.8	32.8	3.4	3.1
14	-17.3	7.9	-2.3	± 0.4	17	-1.5 ± 0.2 *	74.6	21.1	1.3	3.1
15	-18.4	9.4	-2.7	± 0.9	33	-6.2 ± 1.5 *	72.1	21.6	1.2	5.2
16	2.2	5.0	-0.2	± 0.4	200	-0.1 ± 0.3	95.5	1.4	0.0	3.1
17	3.3	2.5	-0.3	± 0.2	67	0.3 ± 0.1 *	70.5	21.6	0.1	7.9

Table 3.2. For each EAIS basin, ensemble mean (dh_f^{tot}) and standard deviation (σ_f^{tot}) of firn thickness change. Mean ($trend\ dh_f$) and standard deviation ($\sigma^{trend} dh_f$) of the linear trends fitted to the ensemble scenarios, and their ratio ($\frac{\sigma^{trend}\ dh_f}{\lvert trend\ dh_f \rvert}$). Altimetry trends are from Shepherd et al. (2019). Superscript * denotes non-overlapping uncertainty ranges from altimetry and from the model ensemble. η^2/σ^2 ratios show contributions of the sources of uncertainty to the ensemble spread.

Model uncertainties in basin-averaged rates of firn thickness change range between 0.2–1.0 cm yr^{-1}, translating into relative uncertainties between 15–300% (Table 3.2). Despite low absolute uncertainties, the interior basins (2, 3, 10, 16, 17) show the largest relative values because their trends in dh_f are close to zero (<0.4 cm yr^{-1}). Basins with trends exceeding 1 cm yr^{-1} have lower relative uncertainties. Yet, some of these still exhibit relative uncertainties higher than 25% (4, 5, 8, 13, 15). The relative importance of the RCM, FDM and ρ_0 factors on the model spread varies between basins. An area-weighted averaging demonstrates the general predominance of the RCM factor (72%) followed by the FDM (20%), ρ_0 (4%) and interaction (4%) factors. The high influence of RCM choice is mostly due to the large and direct impact of SMB on firn thickness. In addition, there is an indirect impact of RCM output as forcing for FDMs and for the climate-dependent L11 parameterisation of ρ_0.

Cold basins with low snowfall rates (e.g. 2-3, 10-11) are characterised by particularly high contributions of η^2_{RCM} (>90% of σ^2). In such dry conditions, small discrepancies between RCM-modelled snowfall anomalies translate into large relative differences in firn thickness change. FDM contribution to the total ensemble uncertainty increases in basins with higher temperature and

accumulation (e.g. 5-7, 12-13), with η^2_{FDM} explaining approximately 30 to 45% of the spread. These climatic settings drastically increase both the sensitivity of FDMs to temperature fluctuations and the absolute compaction rates. Consequently, small relative differences in compaction rates between the FDMs result in large absolute differences in firn thickness change. Moreover, high total snowfall amounts mitigate the impact of small differences between RCM estimates of accumulation, thus reducing η^2_{RCM}. Another aspect that favours high η^2_{FDM}/σ^2 values is spatial variability of climatic conditions within basins; within basins spanning many climatic zones, there is more likely to be a region in which the FDMs disagree on compaction rates. Contribution of $\eta^2_{\rho_0}$ is highest in basins with large and positive snowfall anomalies (basins 5-8). In such basins, it accounts for up to 28% of the model spread because the thickening caused by the anomaly is sensitive to the snow density parameterisation. Basins 16 and 17 illustrate the role of interaction effects. In these basins, MAR simulates substantially higher temperatures and accumulation rates, causing larger disagreements between FDMs forced by MAR than between FDMs forced by RACMO2 or HIRHAM. This nonconstancy of variance across FDMs for different RCMs leads to a significant interaction term η^2_{intr}. Because interaction effects account for a non-negligible part of the model spread in all basins (1 to 8% of σ^2), our results demonstrate the importance of combining RCMs, FDMs and ρ_0 within different model experiments to assess firn thickness change uncertainty.

3.4.2 Comparison with altimetry

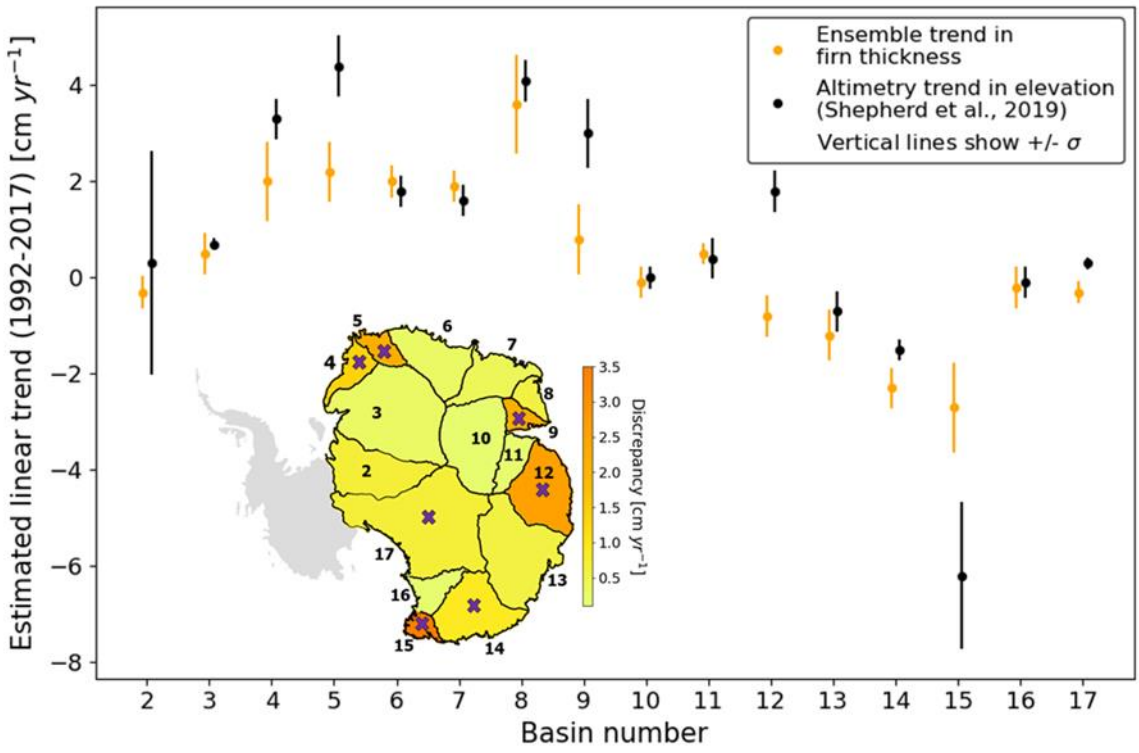


Figure 3.2. Comparison of 1992-2017 altimetry-based elevation trends and firn thickness trends of the ensemble, with their respective 1σ uncertainty ranges. Map shows the absolute ensemble-altimetry differences, crosses highlight basins with non-overlapping uncertainty ranges.

We compare our estimates of basin-wide trends in firn thickness with elevation trends reported by Shepherd et al. (2019) (Table 3.2, Fig. 3.2). Firn thickness change is only a single component of the ice sheet elevation change signal, which also captures ice dynamical imbalance. Thus, accurate estimations of firn thickness change can be compared to measured elevation changes to identify areas of dynamical imbalance (Li and Zwally, 2011; Kuipers Munneke et al., 2015b; Hawley et al., 2020). For 9 of the 16 basins, the model ensemble uncertainty ranges and the altimetry uncertainty ranges overlap. In these cases, our results provide no evidence to support the existence of net ice flow imbalance. However, basin-wide averaging may conceal localised dynamic changes. On Totten glacier (within basin 13) for example, ice dynamical imbalance close to the grounding line makes a substantial contribution to recent mass loss and thus to local elevation decrease (Li et al., 2016). On the other hand, the uncertainty ranges do not overlap for several basins, highlighting the need to better understand the source of the discrepancies in these regions. In such cases, three possibilities, or combinations thereof, should be considered: (1) the model ensemble may fail to represent the true firn thickness change over the 1992-2017 period, (2) the 1σ uncertainty range associated with the altimetry measurements may not adequately capture the true signal or (3) a component of the elevation changes may be related to ice dynamical imbalance.

At this stage, identifying the exact cause of the discrepancy remains speculative. The long response time of ice flow makes any dynamical imbalance challenging to evaluate because long-term trends may still outweigh recent changes (Zwally et al., 2015). Moreover, disagreements persist between simulated SMB and field observations in certain regions (Wang et al., 2016), which can lead to substantial differences in mass balance partitioning (Martin-Español et al., 2017; Mohajerani et al., 2019). Similarly, different sources of altimetry data, inter-satellite bias correction, and other processing steps induce uncertainty in altimetry signals (Shepherd et al., 2019). We use several basins where ensemble- and altimetry-based trends disagree to illustrate these factors. In basins 4 and 5, Medley et al. (2018) demonstrated that global climate models underestimate recent increases in snowfall. A similar underestimation from the RCMs used here would explain the lower ensemble trend compared to the observed elevation trend. In basin 12, significant changes in ice discharge may hint at a dynamic imbalance causing the disagreement (Rignot et al., 2019). However, this area also shows major discrepancies in SMB anomalies from different model estimates (Wang et al., 2016) and from probabilistic inversion techniques (Martin-Español et al., 2017), suggesting that modelling SMB in this region is challenging. Basin 15 is characterised by sparse satellite sampling but also shows a large spread in our model ensemble and is thus poorly constrained. The relatively high model- and altimetry-uncertainties may both be related to the complex topography of this basin. Finally, a robust evaluation of FDM-reliability in all possible EAIS areas and climatic conditions does not exist and models may fail to predict true compaction rates. The objective of comparing ensemble firn thickness

trends and altimetry trends is not to draw hasty conclusions about dynamical imbalance, but rather to highlight areas which deserve greater attention because recent measurements and current state-of-the-art model scenarios do not match.

In critically evaluating our work, it is important to highlight sources of uncertainty that are not accounted for in the ensemble. We use three RCMs forced by ERA-Interim at their boundaries. Different atmospheric reanalyses could, in theory, be used to force the RCMs or could be directly taken over the EAIS domain itself. Both the L11 and fixed-350 parameterisations of ρ_0 assume a time-invariant surface density because possible seasonal and interannual variabilities are unknown. The ensemble is limited by the deterministic RCM, FDM and ρ_0 combinations considered here. In the future, the work could be extended to consider stochastic perturbations and parametric uncertainties in climate input, FDMs and ρ_0, thereby providing a larger range of results. In principle, the process of emulation might lead to localised discrepancies between the emulator and a corresponding FDM, although evaluation (Appendix C) shows that this is unlikely when averaged over large spatial areas, as is done here. One critical assumption is the reference climatic period of 1979-2009. Different ice core analyses and model-based studies disagree on the existence of a trend in Antarctic SMB over the last decades and centuries, but several agree on existing regional trends (Monaghan et al., 2006; Frezzotti et al., 2013; Previdi and Polvani, 2016; Medley and Thomas, 2019). The year 1979 coincides with the start of satellite data assimilation into atmospheric products, and thus with the earliest RCM output, motivating this choice of reference period for practical reasons (e.g. Ligtenberg et al., 2011; Rignot et al., 2019). However, we cannot discount that substantially lower/higher past accumulation rates would result in under/over-estimating recent firn thickness change, thus providing a possible cause of disagreement with elevation change measurements. Nevertheless, because all model scenarios use the same reference period, it has a minor impact on both the total ensemble uncertainty and the uncertainty partitioning; using another reference period could shift the estimates of each scenario but would affect differences between the estimates only marginally.

3.5 Conclusions

Our model ensemble experiment provides a range of modelled scenarios of 1992-2017 firn thickness change on the EAIS that encompass current state-of-the-art modelling capabilities. Using statistical emulation of firn model output, we compute a total of 54 scenarios to assess variability associated with different RCMs, FDMs and surface snow density parameterisations. The ensemble agrees that firn thickness changes in the interior are minor, but there are pronounced thickening and thinning patterns in coastal areas. At basin-level, the uncertainty on the model estimates ranges between 0.2–1.0 cm yr^{-1} and is generally dominated by differences between RCMs due to the strong and direct effect of SMB on firn thickness. However, in basins with high snowfall and with large spatial

variability of climatic conditions, FDM-related variability increases up to 46% of the total ensemble uncertainty. The surface snow density factor has a large impact on uncertainty in basins with recent increases in snowfall rates, reaching a maximum contribution of 28%. Finally, non-linear interactions between the three sources of uncertainty are substantial across the EAIS. Our results demonstrate that refining SMB estimates in RCMs is the priority for constraining future assessments of firn thickness change. However, as snowfall and temperatures are expected to increase in Antarctica (Ligtenberg et al., 2013; Lenaerts et al., 2019), FDMs and snow density will increasingly contribute to model uncertainty and should not be neglected. By comparing the ensemble scenarios with satellite measurements of elevation changes over the same 1992-2017 period, we find that these estimates are consistent over a majority of basins. Nonetheless, we identify several basins where model estimates do not match altimetry measurements. While ice dynamical imbalance could be the source of the discrepancies in these regions, so too could be inadequacies in the respective uncertainty characterisations. As such, our analysis serves to highlight specific areas where further focus on potential sources of errors in model and altimetry results is needed in order to better constrain mass balance assessments in East Antarctica.

4. Synthesis

The work presented in this thesis has focused on the development and the use of firn densification models, on both the Greenland and Antarctic ice sheets (GrIS and AIS). Firn models are crucial to different applications in glaciology: mass balance assessments via satellite altimetry, surface mass balance modelling and paleoclimate analyses from ice cores. Both accuracy and quantification of uncertainty are crucial aspects for these three applications. Therefore, this thesis aimed at improving firn model accuracy and at better constraining uncertainty in model output. This encompassed three main objectives: better representing uncertain processes, constraining parametric uncertainty, and quantifying direct implications of firn model output uncertainty. Uncertainty constraint has often been neglected or poorly accounted for in firn modelling, because no robust framework for uncertainty quantification of firn model output existed until now.

4.1 Summary

4.1.1 Model developments in meltwater treatment

Chapter 1 investigated the process of meltwater percolation and refreezing in firn models. Improving firn model capabilities in the treatment of meltwater is perceived as a priority for firn model developments (van As et al., 2016a), which motivated this work. Such improvements are important for present and future model estimates of the Surface Mass Balance (SMB) of the GrIS for several reasons. First, ~65% of the GrIS is currently affected by surface melt, and this extent is rapidly increasing (Tedesco and Fettweis, 2020). Second, meltwater refreezing is a major component of the GrIS SMB, accommodating ~40% of the total surface meltwater production (van den Broeke et al., 2016). Third, the future capacity of the firn layer to refreeze and retain meltwater is vulnerable to increases in surface melting, which could lead to firn saturation (van Angelen et al., 2013). Finally, perennial firn aquifers are known to store substantial amounts of meltwater in its liquid form, but the timescale and total amount of this storage remains unknown (Foster et al., 2014; Miège et al., 2016). Additionally, there is increasing concern that firn hydrology should be better accounted for in Antarctica to improve SMB estimates and assessments of ice shelf stability (Bell et al., 2018).

The model study of meltwater treatment focused on four different sites of the percolation area of the GrIS that span different climatological and glaciological settings. Three meltwater schemes of different levels of complexity were developed and implemented in the Community Firn Model (Stevens et al., 2020), including advanced water schemes based on hydrological theory that use the Richards Equation and an explicit representation of heterogeneous preferential flow. The study shows that the choice of liquid water scheme does result in significant differences in modelled density and temperature patterns, but that no one scheme exhibits significantly better performance than the others

when compared with observations. The results also showed that firn density is more sensitive to the choice of compaction scheme than to the choice of liquid water scheme, even at high-melt sites.

The finding that simple empirical schemes of water percolation perform similarly to advanced schemes highlighted several elements that hinder the development of physically representative liquid water schemes in firn models. Firstly, results of the advanced schemes are sensitive to the hydraulic parameters in the model formulations, which are constrained by only few experimental studies. Such schemes would thus benefit from a better knowledge of firn structural properties, of their variability in space and of their precise relationship with hydraulic parameters, such as irreducible water content and hydraulic permeability. Secondly, disagreements between existing firn densification formulations are amplified in conditions of higher meltwater percolation. These disagreements are larger than those stemming from the use of different liquid water schemes. Such inter-model disagreements on firn compaction result in differences in modelled density patterns, in turn affecting the treatment of meltwater percolation. Due to their sensitivity to firn structural properties, advanced water schemes are strongly affected by differences in modelled densities. This was illustrated by significant differences in runoff and refreezing amounts simulated from different firn compaction models that use the same meltwater scheme. Nevertheless, the implementation of the Richards Equation and of the preferential flow scheme showed that some features, not captured by the empirical scheme, could be reproduced with better fidelity. This includes a more representative density variability in depth and the reproduction of perennial firn aquifers by fast and heterogeneous vertical water transport in firn.

4.1.2 Parametric uncertainty in dry firn densification

Chapter 1 highlighted large discrepancies attributable to the firn densification model itself, independent of the choice of water scheme. This is despite firn models being structurally similar; they all rely on Arrhenius-type equations as formulation for dry firn densification. The findings of major discrepancies between models motivated the investigation of model parameterisation, which is described in Chapter 2. This work aimed to optimise parameter estimates and better constrain parametric uncertainty in three benchmark firn models. Compared to the original publication of the firn models investigated, the calibration exploited a more extensive set of firn cores and used a Bayesian statistical framework using Markov Chain Monte Carlo methods. The firn core dataset spanned a large range of climatic conditions from both the GrIS and AIS, resulting in a unified calibration representative of all dry firn conditions, and not reliant on location-specific expressions, which have been common until now.

The *maximum a posteriori* models – the optimal estimates reached for the parameter combinations – showed significant improvements for two models and comparable results for the third, when evaluated against an observational dataset independent of the calibration process. The *maximum a*

posteriori models also demonstrated comparable or even improved performances compared to other models of higher degree of complexity. These results show that thorough calibration can yield significant improvements in model accuracy, even without reconsidering the governing equations. Furthermore, the Bayesian approach allows to compute credible intervals for parameter estimates, hence constraining parametric uncertainty, and to evaluate correlation between different parameters. The credible intervals accounted for stochastic perturbations in climatic forcing and in surface snow densities, which are sources of uncertainty propagation in firn densification models. The final parameter estimates also provide better understanding of the adequate mathematical relationships for densification equations. For example, the results showed strong evidence for a sublinear relationship between densification rates and accumulation rates, a proxy for overburden pressure.

Despite improved agreement with respect to observations, the uncertainty intervals produced with the three models still failed to reproduce several of the firn depth-density profiles used as evaluation data. Thus, the formulation of the governing equations affects the remaining errors, which suggests deficiencies in our representation of dry firn densification in models. Alternatively, uncertainty intervals could be underestimated, implying that larger datasets are required to generate wider intervals, representative of true model uncertainty. Together, future refinement of model formulations, and more data against which to evaluate, should result in improved accuracy of firn model estimates. By exploiting the three different firn models in both their original and optimised forms, this study demonstrated important inter-model disagreements, even with improved calibration, that should not be neglected in firn model applications such as altimetry corrections and dating of air bubbles in ice cores.

4.1.3 Firn model output uncertainty at ice sheet scale

The strong inter-model disagreements highlighted in Chapter 2 motivated the objective to perform large-scale intercomparisons of firn model results. Such an approach would allow to evaluate how disagreements translate into firn model output uncertainty for ice sheet wide applications. In this context, Chapter 3 describes work focused on quantifying uncertainty in firn model output, and its implications for altimetry-based mass balance assessments of the East Antarctic ice sheet (EAIS). Such assessments on the EAIS are particularly sensitive to model-estimates of firn thickness change (Shepherd et al., 2018). In this work, a model ensemble of 54 different firn thickness change scenarios was developed. It included climatic forcing from three different Regional Climate Models (RCM), compaction formulations of nine firn densification models (FDM) and two different assumptions on surface snow density. Daily simulations were run for 1992 – 2017 over the entire EAIS (>10 million km^2) and the results were compared to recently published assessments of EAIS surface elevation change from altimetry (Shepherd et al., 2019). The range of results from the ensemble provided a measure of model uncertainty, which was decomposed into its different components using an

ANOVA method. Uncertainty was then partitioned between choice of RCM forcing, of FDM, and of surface density, as well as the nonlinear effects stemming from interactions between these three sources of uncertainty. The computational limitations associated with running the ~89 000 simulations required to cover the EAIS were overcome by developing statistical emulators of the FDMs, which were thoroughly calibrated and evaluated.

The ANOVA demonstrated that RCM forcing contributes most to model spread in general, but that FDM and surface density choices can have large, and even predominant contributions in some areas. Also, nonlinear interaction effects have non-negligible contributions to model uncertainty across the EAIS. Considered on a drainage-basin scale, modelled firn thickness change and published satellite measurements of surface elevation change are in agreement in the majority of basins. This suggests that elevation change in these basins is dominated by firn thickness change and well-captured by model estimates. In several basins however, modelled firn thickness change and measured surface elevation change do not match, even when accounting for their respective uncertainties. Ice dynamical imbalance, not accounted for in firn modelling, could contribute to the discrepancy in those areas. However, further work is needed to improve uncertainty characterisation in both firn models and altimetry, especially in the basins showing disagreements, before this can be confirmed or refuted. Chapter 3 thus underscored the importance of uncertainty quantification in the interpretation of satellite-based altimetry measurements. Ultimately, partitioning observed elevation change between firn thickness change and ice dynamical imbalance is important to reach precise estimates of ice sheet mass balance and to detect potential onsets of ice flow imbalance in the EAIS. As such, the ensemble results highlight specific areas of the EAIS where further focus on potential sources of errors in model and altimetry results is needed.

Through this model ensemble framework, different sources of epistemic uncertainty (uncertainty as a result of lack of information or knowledge) were considered. Firstly, by using different climatic forcings and surface density assumptions, input uncertainty was extensively covered. Secondly, the use of different firn model formulations addressed the model structural uncertainty. Finally, parametric uncertainty was also considered because some model formulations differed only by the re-parameterisations developed in Chapter 2. Importantly, this study was the first to evaluate robustly and comprehensively the potential impacts of firn model output uncertainty on mass balance assessments at the ice sheet scale.

4.2 Scientific contributions

My work through my PhD thesis has made a number of contributions, particularly to the field of firn densification modelling. These are summarised here.

All three of my Chapters have made improvements to the Community Firn Model (CFM, Stevens et al., 2020). The CFM is an open source model that facilitates the applicability of firn modelling for various research interests among the glaciology community because it is open access, user-friendly and transparent about its implementation. As such, it can be used by other researchers needing an accessible and verified firn model in their studies (e.g. Hawley at al., 2020; Smith et al., 2020). Improvements I have made to the CFM as a part of my PhD include implementing new processes into the model (Chapter 1), extending its applicability (Chapter 2), improving the model code (all Chapters) and using it in model-based studies (all Chapters), as well as additional developments of the CFM not explicitly described in Chapters 1-3, such as sublimation and regridding routines. My improvements have been assimilated into the master version of the model.

I have implemented the liquid water schemes described in Chapter 1 into the CFM. As such, I am the first person to have modelled meltwater percolation in firn based on the Richards Equation, building upon earlier developments in snow modelling (Colbeck, 1972; Jordan, 1983; Illangasekare et al., 1990; Hirashima et al., 2010; Wever et al., 2014; D'Amboise et al., 2017). I also modelled the process of preferential flow, which has been observed in snow and firn (e.g. Humphrey et al., 2012) but never represented explicitly in firn models before. Both of these aspects had repeatedly been identified as research gaps by the community (van As et al., 2016a; Steger et al., 2017a). Based on this new framework for meltwater modelling, described in Chapter 1, I led the CFM contribution to the RetMIP firn model intercomparison project (Vandecrux et al., 2020). This study is a community-led effort to intercompare results from nine different firn models at different locations of the GrIS. It allows to better understand the sources of discrepancies between models in their treatment of meltwater, and thus contributes to quantifying firn model uncertainty in the percolation area.

I have extended the applicability of firn models through developing and applying a Bayesian calibration framework in Chapter 2. This brings novel statistical methods into the field of firn densification modelling, and is one of the earliest studies to adopt an advanced Bayesian approach in glaciology more generally. Despite the empirical nature of firn models, little effort had been dedicated to the statistical aspects of model calibration until now. By uniting advanced statistics and glaciology, I thus bring an interdisciplinary approach to improving firn model accuracy and quantifying uncertainty in firn modelling, through the use of a mathematically robust framework.

By exploiting different firn models, RCMs and boundary conditions, I have led the first, thorough ice sheet scale evaluation of firn model output uncertainty through Chapter 3. The ANOVA method allowed the contribution of these three uncertainty sources to be evaluated, with respect both to total uncertainty in firn thickness change and the spatial variability of the contributions. These findings can

help to establish future research priorities in model developments. The direct implications for altimetry-based mass balance assessments, were also considered. The results highlighted areas where surface elevation change measurements and model-based estimates of firn thickness change are not consistent with each other, and thus either where ice dynamical imbalance is occurring, or where there is a need to better characterise uncertainty in one or both of models and altimetry. This work thus contributes to more thorough considerations of mass balance assessment given the typical two- to three-fold difference between ice and firn densities, which has major implications in the conversion from measured elevation changes to estimates of mass changes, and thus altimetry-based estimates of the ice sheet contribution to global sea level rise

Beyond my PhD work I have also contributed to the work of others through performing firn model simulations for colleagues at Lancaster and beyond. This work has been used in publications on which I am a co-author, including Stevens et al. (2020), Vandecrux et al. (2020), Barnes et al. (in prep.) and Arthur et al. (in prep.).

4.3 Limitations

The liquid water schemes that I describe in Chapter 1 are heavily reliant on the Richards Equation which presumes that parameter values are constant at the scale of representative elementary volumes (Beven and Germann, 2013). Given firn heterogeneity, it is probably not reasonable to assume that this holds at the horizontal scale at which firn models are typically run (order of km), which itself is dictated by the spatial resolution of the RCM forcing data. A similar consideration holds for the dual porosity approach adopted in Chapter 1 to model preferential flow. Furthermore, a limitation often neglected in hydrological science is that the Richards Equation was derived under the assumption of steady water fluxes, and is not necessarily applicable in transient cases (Beven and Germann, 2013). Thus, daily and sub-daily variabilities in meltwater input disagree with some of the original assumptions built into the equation. Finally, despite progress in computational capabilities, the numerical solution of the Richards Equation is still challenging and not guaranteed to be reliable (Farthing and Ogden, 2017). In this context, the use of the equation remains debated in the hydrology modelling community (Beven and Germann, 2013). Applied to firn hydrology, the computational requirements also raise questions with respect to the applicability of Richards Equation-based firn models for ice sheet scale studies, and simplifications may be necessary. The Richards Equation is also reliant on precise estimates of hydrological parameters, but firn hydraulic properties are not well understood. Currently, parameter values are only inferred from ideal small-scale laboratory experiments (e.g. Yamaguchi et al., 2012). On the other hand, field studies have demonstrated that Greenland firn permeability is strongly variable depending on both depth and location, is sensitive to presence of ice layers and shows no clear relationship with density (Adolph and Albert, 2014;

Sommers et al., 2017). In this context, it is not yet clear whether laboratory-based parameter estimates are reasonable with respect to natural firn conditions on ice sheets.

In Chapter 2, firn cores were used as model calibration data, however firn cores represent only a snapshot in time of the density profiles. This limits the evaluation of model capability in representing intra- and inter-annual variability of densification rates. Instead, the calibration focuses on accurate representation of mean densification rates as a function of long-term climate. Using observations of firn compaction rates could help overcome this limitation. But such measurements are still scarce, have not been collected in a large variety of climatic settings, and do not extend to great depths. Nevertheless, the Bayesian framework developed is directly applicable to new data and to combinations of different types of data. Errors in the input (climatic forcing and surface density values) and in the calibration data can also alter parameter estimates. Such errors have been accounted for by implementing realistic noise in the input values and by assuming uncertainty in the density measurements. This had a direct impact on the uncertainty intervals for the parameter estimates. If better constraints on model input, and more – and more precise – measurements are used in future calibration efforts using this framework, the Bayesian calibration will naturally lead to narrower uncertainty intervals.

The model ensemble described in Chapter 3 aimed to represent our uncertainty in firn thickness change estimates, based on state-of-the-art modelling capabilities. While limitations can be attributed to the physics and parameterisations of the models that constitute the ensemble, here I focus on the limitations of the ensemble methodology itself. The three RCMs of the ensemble were forced with the same atmospheric reanalysis product, thus the ensemble does not capture uncertainty arising from boundary forcing of RCMs. Also, firn model estimates of compaction rates cannot be perfectly reproduced by statistical emulators. However, the evaluation of the emulators shows a high degree of fidelity and unbiased performances. Furthermore, for computational reasons, ensemble modelling with firn models at ice sheet scale is prohibitive and statistical emulators overcome this limitation. A potentially important, yet difficult to quantify limitation is the assumption of a reference climatic period with respect to which changes in firn thickness are calculated. This is necessary because reliable, fine resolution climatic forcing over the EAIS only goes back to 1979, contrasting with the century timescale of firn layer deposition. Ice core studies provide no clear conclusion on possible trends in snowfall amounts over the past centuries. If past snowfall conditions were lower than those during the reference climatic period, the firn thickness changes over our period of interest (1992-2017) would be calculated with respect to overestimated snowfall rates. This would likely result in underestimated estimates of 1992-2017 firn thickness change. Conversely, higher past snowfall conditions would result in overestimated modelled firn thickness change. While this limitation may affect comparisons with elevation change measurements, it does not affect the total ensemble

uncertainty and the uncertainty partitioning because all model scenarios of the ensemble use the same reference climatic period.

4.4 Outlook

The results from each Chapter show important research avenues to improve firn model accuracy and to refine estimates of the associated uncertainties. Such perspectives for further developments are addressed in this section.

As mentioned in Sections 4.1.1 and 4.3, hydraulic parameters play an important role in the patterns of meltwater percolation simulated by schemes relying on the Richards Equation and on a preferential flow implementation. These parameters govern the relationship between meltwater infiltration and firn structural properties such as density and grain size. As such there are strong interactions between the densification, grain growth and meltwater components of firn models, providing the potential for errors from any of these to propagate into the model. This highlights important necessary developments to fully benefit from the implementation of advanced liquid water flow schemes in firn models. First, firn models currently do not account for the impact of water content and refrozen ice lenses on the densification rates. Such features are known to influence firn viscosity (Vionnet et al., 2012), which governs the dependence of densification rates on firn structural properties. Chapter 1 showed the major and predominant impact of the densification formulation on differences in firn model results. For this reason, developing densification equations which are valid in conditions where wet-firn processes dominate is necessary to better simulate densification, meltwater infiltration and interactions between these processes. Second, better constraints on firn hydraulic properties from laboratory and field studies are needed. Such properties have so far been inferred from only few laboratory measurements (e.g. Yamaguchi et al., 2012), but there is field evidence that the relationship between firn permeability and firn structure is complex (Adolph and Albert, 2014; Sommers et al., 2017). Hydraulic parameterisations for firn are challenging to derive because it is a highly anisotropic porous medium and its hydraulic characteristics are also directly impacted by the solidification of liquid water through refreezing (Sommers et al., 2017). More observational studies will help better understand firn hydraulic properties and how to average these at large spatial scales. Finally, Chapter 1 identified the multidimensionality of liquid water flow as an important challenge for firn model water schemes. The one-dimensionality of current firn models requires parameterisations for horizontal flow treatment, which are difficult to constrain (e.g. Lefebre et al., 2003). This is particularly important given the clear evidence of large subsurface horizontal water movement in GrIS firn (Humphrey et al., 2012) and in perennial firn aquifers (Miller et al., 2018). Addressing these model shortcomings will improve firn model accuracy in the percolation areas of the ice sheets. If such improvements are achieved, analyses similar to the model ensemble study performed on the EAIS in Chapter 3 could include areas where melting and wet firn processes have a large influence on

firn thickness changes. Consequently, robust large-scale uncertainty analyses could include the Antarctic Peninsula, the West-Antarctic ice sheet, ice shelves and the GrIS.

Besides refinements of the densification governing equations, statistical calibration methods for firn densification models, such as the one performed in Chapter 2, would benefit from better constraints on climatic forcing and on the surface snow density boundary condition. For climatic input, firn models rely on atmospheric reanalysis products and RCMs. These model products show significant discrepancies among themselves, and also when compared to field-based observations (e.g. Wang et al., 2016; Fettweis et al., 2020; Mottram et al., 2020). Polar RCMs face the challenge of simulating complex physical processes such as the melt-albedo feedback, the redistribution of snow under strong wind currents and snow sublimation in extreme temperature and atmospheric humidity conditions (e.g. Noël et al., 2018; Agosta et al., 2019). Furthermore, they cannot rely on extensive field measurements for calibration and for constraining model settings because the coverage of the GrIS and AIS by weather stations is low and unequally distributed, leaving extended areas uncovered (Lazzara et al., 2012; Wang et al., 2018). Future improvements in ice sheet climate modelling will lead to more effective calibration of firn models and to improvements in firn model output. Similarly, uncertainties in the surface snow density boundary condition propagate in firn models, affecting the model results. In the upper centimetres of firn, densification is rapid due to breaking of the fine crystal structure of fresh snow and wind-driven compaction (Eisen et al., 2008; Groot Zwaaftink et al., 2013). The physics and the temporal resolution (daily or coarser) of firn models make it impractical to represent such short-lived processes, justifying the use of a surface density value typically around 300 kg m^{-3}. However, the surface density value is affected by spatiotemporal variability (Groot Zwaaftink et al., 2013; Proksch et al., 2015; Fausto et al., 2018). By calibration to upper densities of firn cores, parameterisations of surface densities have been developed for both the GrIS and AIS (e.g. Kaspers et al., 2004; Kuipers Munneke et al., 2015b; Fausto et al., 2018; Agosta et al., 2019). However, their calibration shows large variations between model estimates and field data, and temporal variability is generally not accounted for. Better constraining surface densities will require more field data and a better understanding of relationships between climatic factors and very fresh snow compaction and erosion.

The large majority of current firn models are empirical models that do not include formulations for the different micro-mechanisms that affect densification. Instead, they are parameterised to include the relevant mechanisms in a single equation. Alternative and more physically-based microscale approaches are possible and in progress (e.g. Alley, 1987; Arthern and Wingham 1998; Meyer et al., 2020; Fourteau et al., 2020), but they still face the challenge to be applied on firn columns typically 10–100 m thick. The recent developments of empirical, macro-scale firn models show encouraging results. An important aspect of their future development will be to better validate their accuracy in

short timescale variability of firn compaction (i.e. daily to seasonal). To date, this has been limited by model calibration largely relying on firn core datasets. Validation of short timescale variability will require evaluating models against firn strain rates measurements in a variety of climatic settings. By combining calibration to such measurements and knowledge of the microphysical mechanisms at play, it will be possible to improve firn models to ensure adequate representation of shorter-term densification fluctuations. Refining model formulations to explicitly include relevant small scale mechanisms could also help address (1) the shortcomings of current formulations in melt areas identified in Chapter 1, and (2) the remaining deficiencies of models in representing dry firn densification after recalibration, as identified in Chapter 2.

These avenues of investigation will be crucial to further improve firn model accuracy and quantification of uncertainty. This thesis has highlighted their importance, addressed some of the associated limitations, and thus contributed to progress in firn densification modelling.

References

Adolph, A. C. and Albert, M. R.: Gas diffusivity and permeability through the firn column at Summit, Greenland: measurements and comparison to microstructural properties, The Cryosphere, 8, 319–328, doi:10.5194/tc-8-319-2014, 2014.

Agosta, C., Amory, C., Kittel, C., Orsi, A., Favier, V., Gallée, H., van den Broeke, M. R., Lenaerts, J. T. M., van Wessem, J. M., van de Berg, W. J., and Fettweis, X.: Estimation of the Antarctic surface mass balance using the regional climate model MAR (1979–2015) and identification of dominant processes, The Cryosphere, 13, 281–296, https://doi.org/10.5194/tc-13-281-2019, 2019.

Alley, R. B.: Firn Densification By Grain-Boundary Sliding : a First Model, Journal De Physique, 48, 249−256, doi:10.1051/jphyscol:1987135, 1987.

Alley, R. B., Clark, P. U., Huybrechts, P., and Joughin, I.: Ice sheet and sea-level changes, Science, 310, 456–460, doi: 10.1126/science.1114613, 2005.

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.

Andiranakis, I. and Challenor P.G.: The effect of the nugget on Gaussian process emulators of computer models, Computational statistics and data analysis, 56, 4215–4228. doi:10.1016/j.csda.2012.04.020, 2012.

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, Climatic Change, 40(3–4), 605–624, doi:10.1023/A:1005320713306, 1998.

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, Journal of Geophysical Research, 115, 1–12, doi:10.1029/2009JF001306, 2010.

Arthur, J., Stokes, C., Jamieson, S., Carr, R., Leeson, A. and Verjans V.: Inter-annual variability in supraglacial lakes around East Antarctica, in prep.

Aster, R. C., Borchers, B. and Clifford, H. T.: Parameter estimation and inverse problems, Elsevier, Amsterdam., 2005.

Avanzi, F., Hirashima, H., Yamaguchi, S., Katsushima, T. and De Michele, C.: Observations of capillary barriers and preferential flow in layered snow during cold laboratory experiments, The Cryosphere, 10, 2013–2026, doi:10.5194/tc-10-2013-2016, 2016.

Babonis, G. S., Csatho, B. and Schenk, T.: Mass balance changes and ice dynamics of Greenland and Antarctic ice sheets from laser altimetry, ISPRS J. Photogramm. Remote Sens., 41(B8), 481–487,, doi:10.5194/isprsarchives-XLI-B8-481-2016, 2016.

Bader, H.: Sorge's Law of Densification of Snow on High Polar Glaciers, Journal of Glaciology, 2, 319–323, https://doi.org/10.3189/S0022143000025144, 1954.

Barnes, T. J., Leeson, A. A., McMillan, M., Verjans, V., Carter, J., Kittel, C.: High densities of Supraglacial Lakes on George VI Ice Shelf in 2020 and comparison with past events, in prep.

Barnola, J.-M., Pimienta, P., Raynaud, D., and Korotkevich, Y. S.: CO2-climate relationship as deduced from the Vostok ice core: a re-examination based on new measurements and on a reevaluation of the air dating, Tellus, 43B, 83–90, 1991

Bell, R. E., Banwell, A. F., Trusel, L. D., and Kingslake, J.: Antarctic surface hydrology and impacts on icesheet mass balance, Nature Climate Change, 8, 1044–1052, https://doi.org/10.1038/s41558-018-0326-3, 2018.

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, Journal of Glaciology, 54(187), 705–714, 2008.

Beven, K. J.: Environmental Modelling: An Uncertain Future?, Routledge, London, 2009.

Beven, K. and Germann, P.: Macropores and water flow in soils revisited, Water Resour. Res., 49, 3071–3092, doi:10.1002/Wrcr.20156, 2013.

Blackford, J. R.: Sintering and microstructure of ice: a review, Journal of Physics D: Applied Physics, 40, R355–R385, 2007.

Boening, C., Lebsock, M., Landerer, F. and Stephens, G.: Snowfall‐driven mass change on the East Antarctic ice sheet, Geophysical Research Letters, 39, L21501. https://doi.org/10.1029/2012GL053316, 2012.

Brinkerhoff, D. J., Aschwanden, A. and Truffer, M.: Bayesian Inference of Subglacial Topography Using Mass Conservation, Frontiers in Earth Science, 4, 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, Annals of Glaciology, 57(72), 84–95, doi:10.1017/aog.2016.3, 2016b.
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, 345(6201), 1177–1180, doi:10.1126/science.1254961, 2014.
Buizert, C., Cuffey, K. M., Severinghaus, J. P., Baggenstos, D., Fudge, T. J., Steig, E. J., Markle, B. R., Winstrup, M., Rhodes, R. H., Brook, E. J., Sowers, T. A., Clow, G. D., Cheng, H., Edwards, R. L., Sigl, M., McConnell, J. R., and Taylor, K. C.: The WAIS Divide deep ice core WD2014 chronology – Part 1: Methane synchronization (68–31 ka BP) and the gas age-ice age difference, Climate of the Past, 11, 153–173, doi:10.5194/cp-11-153-2015, 2015.
Calonne, N., Geindreau, C., Flin, F., Morin, S., Lesaffre, B., Rolland Du Roscoat, S. and Charrier, P.: 3-D image-based numerical computations of snow permeability: Links to specific surface area, density, and microstructural anisotropy, The Cryosphere, 6, 939–951, doi:10.5194/tc-6-939-2012, 2012.
Capron, E., Landais, A., Buiron, D., Cauquoin, A., Chappellaz, J., Debret, M., Jouzel, J., Leuenberger, M., Martinerie, P., MassonDelmotte, V., Mulvaney, R., Parrenin, F., and Prié, F.: Glacial– interglacial dynamics of Antarctic firn columns: comparison between simulations and ice core air-δ 15N measurements, Climate of the Past, 9, 983–999, doi:10.5194/cp-9-983-2013, 2013.
Celia, M. A., Bouloutas, E. T. and Zarba, R. L.: A General Mass-Conservative Numerical Solution for the Unsaturated Flow Equation, Water Resources Research, 26, 1483–1496, doi:10.1029/90wr00196, 1990.
Christensen, O. B., Drews, M., Christensen, J. H., Dethloff, K., Ketelsen, K., Hebestadt, I. and Rinke, A.: The HIRHAM regional climate model, version 5 (Technical Report No. 06-17), Danish Meteorological Institute, Copenhagen, Denmark, 2006.
Coléou, C. and Lesaffre, B.: Irreducible water saturation in snow: experimental results in a cold laboratory, Annals of Glaciology, 26, 64–68, doi:10.3189/1998AoG26-1-64-68, 1998.
Colbeck, S. C.: A theory of water percolation in snow, Journal of Glaciology, 11, 369–385, 1972.
Colbeck, S.: Water flow through heterogeneous snow, Cold Regions Science and Technology, 1, 37–45, 1979.
Cox, C., Humphrey, N. and Harper, J.: Quantifying meltwater refreezing along a transect of sites on the Greenland Icesheet, The Cryosphere, 8, 5485–5509, doi:10.5194/tcd-8-5485-2014, 2015.
Cuffey, K. and Paterson, W. S. B.: The Physics of Glaciers, Elsevier, Butterworth-Heineman, Burlington, MA, USA, 2010.
D'Amboise, C. J. L., Müller, K., Oxarango, L., Morin, S., and Schuler, T. V.: Implementation of a physically based water percolation routine in the Crocus/SURFEX (V7.3) snowpack model, Geoscientific Model Development, 10, 3547–3566, https://doi.org/10.5194/gmd-10-3547-2017, 2017.
Davis, C. H., Li, Y., McConnell, J. R., Frey, M. M. and Hanna, E.: Snowfall-driven growth in East Antarctic Ice Sheet mitigates recent sea-level rise, Science, 308, 1898‑1901. doi:10.1126/science.1110662, 2005.
Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Koehler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.
Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., Jacob, D., Kjellström, E., de Castro, M., and van den Hurk, B.: An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections, Climatic Change, 81(S1), 53–70, https://doi.org/10007/s10584- 006-9228-x, 2007.
Eisen, O., Frezzotti, M., Genthon, C., Isaksson, E., Magand, O., van den Broeke, M. R., Dixon, D. A., Ekaykin, A., Holmlund, P., Kameda, T., Karlof, L., Kaspari, S., Lipenkov, V. Y., Oerter, H., Takahashi, S., and Vaughan, D. G.: Ground-based measurements of spatial and temporal variability of snow accumulation in East Antarctica, Reviews of Geophysics, 46, RG2001, doi:10.1029/2006RG000218, 2008.
Farthing, M. W. and Ogden, F. L.: Numerical Solution of Richards' Equation: A Review of Advances and Challenges, Soil Science Society of America Journal, 81, 1257, https://doi.org/10.2136/sssaj2017.02.0058, 2017.
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, Frontiers in Earth Science, 6, 51 pp., https://doi.org/10.3389/feart.2018.00051, 2018.
Fettweis, X.: Reconstruction of the 1979–2006 Greenland ice sheet surface mass balance using the regional climate model MAR, The Cryosphere, 1, 21–40, doi:10.5194/tc-1-21-2007, 2007.

Fettweis, X., Hofer, S., Krebs-Kanzow, U., Amory, C., Aoki, T., Berends, C. J., Born, A., Box, J. E., Delhasse, A., Fujita, K., Gierz, P., Goelzer, H., Hanna, E., Hashimoto, A., Huybrechts, P., Kapsch, M.-L., King, M. D., Kittel, C., Lang, C., Langen, P. L., Lenaerts, J. T. M., Liston, G. E., Lohmann, G., Mernild, S. H., Mikolajewicz, U., Modali, K., Mottram, R. H., Niwano, M., Noël, B., Ryan, J. C., Smith, A., Streffing, J., Tedesco, M., van de Berg, W. J., van den Broeke, M., van de Wal, R. S. W., van Kampenhout, L., Wilton, D., Wouters, B., Ziemen, F., and Zolles, T.: GrSMBMIP: intercomparison of the modelled 1980–2012 surface mass balance over the Greenland Ice Sheet, The Cryosphere, 14, 3935–3958, https://doi.org/10.5194/tc-14-3935-2020, 2020.

Forster, R. R., Box, J. E., van den Broeke, M. R., Miège, C., Burgess, E. W., van Angelen, J. H., Lenaerts, J. T. M., Koenig, L. S., Paden, J., Lewis, C., Gogineni, S. P., Leuschen, C. and McConnell, J. R.: Extensive liquid meltwater storage in firn within the Greenland ice sheet, Nature Geoscience, 7, 95–98, doi:10.1038/ngeo2043, 2014.

Forsyth, P. A., Wu, Y. S. and Pruess, K.: Robust numerical methods for saturated-unsaturated flow with dry initial conditions in heterogeneous media, Advances in Water Resources, 18, 25–38, doi:10.1016/0309-1708(95)00020-J, 1995.

Fourteau, K., Gillet-Chaulet, F., Martinerie, P., Faïn, X.: A Micro-Mechanical Model for the Transformation of Dry Polar Firn Into Ice Using the Level-Set Method, Frontiers in Earth Science, 8(101), 1–15, doi: 10.3389/feart.2020.00101, 2020.

Freitag, J., Kipfstuhl, S., Laepple, T. and Wilhelms, F.: Impurity-controlled densification: a new model for stratified polar firn, Journal of Glaciology, 59(218), 1163–1169, doi:10.3189/2013jog13j042, 2013.

Frezzotti, M., Scarchilli, C., Becagli, S., Proposito, M., and Urbini, S.: A synthesis of the Antarctic surface mass balance during the last 800 yr, The Cryosphere, 7, 303–319, https://doi.org/10.5194/tc-7-303-2013, 2013.

Gascon, G., Sharp, M., Burgess, D., Bezeau, P., Bush, A. B. G., Morin, S. and Lafaysse, M.: How well is firn densification represented by a physically based multilayer model? Model evaluation for Devon Ice Cap, Nunavut, Canada, Journal of Glaciology, 60, 694–704, doi:10.3189/2014JoG13J209, 2014.

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.

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, Journal of Geophysical Research, 108, doi:10.1029/2002JD003319, 2003.

Gow, A. J.: Deep core studies of the accumulation and densification of snow at Byrd Station and Little America V, Antarctica. U.S. Cold Regions Research and Engineering Laboratory, Research Report 197, 1968.

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, Annals of Glaciology, 39, 101–109, doi:10.3189/172756404781814690, 2004.

Gregory, S. A., Albert, M. R. and Baker, I.: Impact of physical properties and accumulation rate on pore close-off in layered firn, The Cryosphere, 8, 91–105, doi:10.5194/tc-8-91-2014, 2014.

Groot Zwaaftink, C. D., Cagnati, A., Crepaz, A., Fierz, C., Macelloni, G., Valt, M., and Lehning, M.: Event-driven deposition of snow on the Antarctic Plateau: analyzing field measurements with SNOWPACK, The Cryosphere, 7, 333–347, doi:10.5194/tc7-333-2013, 2013.

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.

Harper, J., Humphrey, N., Pfeffer, W. T., Brown, J. and Fettweis, X.: Greenland ice-sheet contribution to sea-level rise buffered by meltwater storage in firn, Nature, 491, 240–243, doi:10.1038/nature11566, 2012.

Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, Biometrika, 57(1), 97–109, 1970.

Hawley, R. L. and Waddington, E. D.: Instruments and Methods in situ measurements of firn compaction profiles using borehole optical stratigraphy, Journal of Glaciology, 57(202), 289–294, doi:10.3189/002214311796405889, 2011.

Hawley, R. L., Neumann T. A., Stevens, C. M., Brunt. K. M. and Sutterley T. C.: Greenland Ice Sheet Elevation Change: Direct Observation of Process and Attribution at Summit, Geophysical Research Letters, https://doi.org/10.1029/2020GL088864, 2020.

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, 320, 1626–1629, doi:10.1126/science.1153894, 2008.

Herron, M. and Langway, C.: Firn densification: an empirical model, Journal of Glaciology, 25(93), 373–385 https://doi.org/10.3189/S0022143000015239, 1980.

Hirashima, H., Yamaguchi, S., Sato, A. and Lehning, M.: Numerical modeling of liquid water movement through layered snow based on new measurements of the water retention curve, Cold Regions Science

and Technology, 64, 94–103, doi:10.1016/j.coldregions.2010.09.003, 2010.
Hirashima, H., Yamaguchi, S. and Katsushima, T.: A multi-dimensional water transport model to reproduce preferential flow in the snowpack, Cold Regions Science and Technology, 108, 80–90, doi:10.1016/j.coldregions.2014.09.004, 2014.
Horlings, A. N., Christianson, K., Holschuh, N., Stevens, C. M. and Waddington E.: Effect of horizontal divergence on estimates of firn-air content, Journal of Glaciology, 67, 287–296, https://doi.org/10.1017/jog.2020.105, 2020.
Huang, K., Mohanty, B., and van Genuchten, M.: A new convergence criterion for the modified Picard iteration method to solve the variably saturated flow equation, Journal of Hydrology, 178, 69–91, doi:10.1016/0022-1694(95)02799-8, 1996.
Humphrey, N. F., Harper, J. T. and Pfeffer, W. T.: Thermal tracking of meltwater retention in Greenland's accumulation area, Journal of Geophysical Research, 117, 1–11, doi:10.1029/2011JF002083, 2012.
Illangasekare, T. H., Walter Jr., R. J., Meier, M. F., and Pfeffer, W. T.: Modeling of meltwater infiltration in subfreezing snow, Water Resources Research, 26, 1001–1012, doi:10.1029/WR026i005p01001, 1990.
Jordan, P.: Meltwater movement in a deep snowpack: 2. Simulation model, Water Resources Research, 19, 979–985, doi:10.1029/WR019i004p00979, 1983.
Jordan, R.: Effects of capillary discontinuities on water flow and water retention in layered snowcovers, Defence Science Journal, 45, 79–91, 1995.
Kaspers, K. A., van de Wal, R. S. W., van den Broeke, M. R., Schwander, J., van Lipzig, N. P. M., and Brenninkmeijer, C. A. M.: Model calculations of the age of firn air across the Antarctic continent, Atmospheric Chemistry and Physics, 4, 1817– 1853, doi:10.5194/acpd-4-1817-2004, 2004.
Katsushima, T., Kumakura, T. and Takeuchi, Y.: A multiple snow layer model including a parameterization of vertical water channel process in snowpack, Cold Regions Science and Technology, 59, 143–151, doi:10.1016/j.coldregions.2009.09.002, 2009.
Katsushima, T., Yamaguchi, S., Kumakura, T. and Sato, A.: Experimental analysis of preferential flow in dry snowpack, Cold Regions Science and Technology, 85, 206–216, doi:10.1016/j.coldregions.2012.09.012, 2013.
Kittel, C., Amory, C., Agosta, C., Jourdain, N. C., Hofer, S., Delhasse, A., Doutreloup, S., Huot, P.-V., Lang, C., Fichefet, T., and Fettweis, X.: Diverging future surface mass balance between the Antarctic ice shelves and grounded ice sheet, The Cryosphere Discussion, https://doi.org/10.5194/tc-2020-291, 2020.
Koenig, L. S., Miège, C., Forster, R. R. and Brucker, L.: Initial in situ measurements of perennial meltwater storage in the Greenland firn aquifer, Geophysical Research Letters, 41, 81–85, doi:10.1002/2013GL058083, 2014.
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., van den Broeke, M. R., van Angelen, J. H. and Forster, R. R.: Explaining the presence of perennial liquid water bodies in the firn of the Greenland Ice Sheet, Geophysical Research Letters, 41, 476–483, doi:10.1002/2013GL058389, 2014a.
Kuipers Munneke, P., Ligtenberg, S. R. M., Van Den Broeke, M. R., and Vaughan, D. G.: Firn air depletion as a precursor of Antarctic ice-shelf collapse, J. Glaciol., 60, 205–214, https://doi.org/10.3189/2014JoG13J183, 2014b.
Kuipers Munneke, P., Ligtenberg, S. R. M., Suder, E. A. and van den Broeke, M. R.: A model study of the response of dry and wet firn to climate change, Annals of Glaciology, 56, 1–8, doi:10.3189/2015AoG70A994, 2015a.
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, 2015b.
Langen, P. L., Fausto, R. S., Vandecrux, B., Mottram, R. H. and Box, J. E.: Liquid Water Flow and Retention on the Greenland Ice Sheet in the Regional Climate Model HIRHAM5: Local and Large-Scale Impacts, Frontiers in Earth Science, 4, doi:10.3389/feart.2016.00110, 2017.
Lazzara, M. A., Weidner, G. A., Keller, L. M., Thom, J. E., and Cassano, J. J.: Antarctic Automatic Weather Station Program: 30 years of polar observations, Bulletin of the American Meteorological Society, 93, 1519– 1537, doi:10.1175/BAMS-D-11-00015.1, 2012.
Leeson, A. A., Forster, E., Rice, A., Gourmelen, N., and van Wessem, J. M.: Evolution of supraglacial lakes on the Larsen B ice shelf in the decades before it collapsed, Geophysical Research Letters, 47, 1–3, https://doi.org/10.1029/2019GL085591, 2020.
Lefebre, F., Gallee, H., van Ypersele, J. P., and Greuell, W.: Mod-´ eling of snow and ice melt at ETH Camp (West Greenland): a study of surface albedo, Journal of Geophysical Research, 108, 4231, doi:10.1029/2001jd001160, 2003.

Lenaerts, J. T. M., Medley, B., van den Broeke, M. R., and Wouters, B.: Observing and modeling ice sheet surface mass balance, Reviews in Geophysics, 57, 376–420, https://doi.org/10.1029/2018RG000622, 2019.

Li, J. and Zwally, H. J.: Modeling of firn compaction for estimating ice-sheet mass change from observed ice-sheet elevation change, Annals of Glaciology, 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, Journal of Glaciology, 61, 1037–1047, doi:10.3189/2015JoG14J182, 2015.

Li, X., Rignot, E., Mouginot, J. and Scheuchl, B.: Ice flow dynamics and mass loss of Totten glacier, East Antarctica, from 1989 to 2015, Geophysical Research Letters, 43, 6366–6373. https://doi.org/10.1002/2016GL069173, 2016.

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., van de Berg, W. J., van den Broeke, M. R., Rae, J. and van Meijgaard, E.: Future surface mass balance of the Antarctic ice sheet and its influence on sea level change, simulated by a regional atmospheric climate model, Climate Dynamics, 41(3-4), 867–884. https://doi.org/10.1007/s00382-013-1749-1, 2013.

Ligtenberg, S. R. M., Kuipers Munneke, P., Noël, B. P. Y. and Van Den Broeke, M. R.: Brief communication: Improved simulation of the present-day Greenland firn layer (1960-2016), The Cryosphere, 12, 1643–1649, doi:10.5194/tc-12-1643-2018, 2018.

Linow, S., Hörhold, M. W. and Freitag, J.: Grain-size evolution of polar firn: A new empirical grain growth parameterization based on X-ray microcomputer tomography measurements, Journal of Glaciology, 58, 1245–1252, doi:10.3189/2012JoG11J256, 2012.

Liu, H., Ong, Y. S., Shen, X. and Cai, J.: When Gaussian Process Meets Big Data: A Review of Scalable GPs, IEEE Transactions on Neural Networks and Learning Systems. doi: 10.1109/TNNLS.2019.2957109, 2020.

Loulergue, L., Parrenin, F., Blunier, T., Barnola, J.-M., Spahni, R., Schilt, A., Raisbeck, G., and Chappellaz, J.: New constraints on the gas age-ice age difference along the EPICA ice cores, 0–50 kyr, Climate of the Past, 3, 527–540, doi:10.5194/cp-3-527-2007, 2007.

Löwe, H., Riche, F., and Schneebeli, M.: A general treatment of snow microstructure exemplified by an improved relation for thermal conductivity, The Cryosphere, 7, 1473–1480, https://doi.org/10.5194/tc-7-1473-2013, 2013.

Lundin, J. M. D., Stevens, C. M., Arthern, R., Buizert, C., Orsi, A., Ligtenberg, S. R. M., Simonsen, S. B., Cummings, E., Essery, R., Leahy, W., Harris, P., Helsen, M. M. and Waddington, E. D.: Firn Model Intercomparison Experiment (FirnMICE), Journal of Glaciology, 63, 401–422, doi:10.1017/jog.2016.114, 2017.

Lyapustin, A., Tedesco, M., Wang, Y., Aoki, T., Hori, M., and Kokhanovsky, A.: Retrieval of snow grain size over Greenland from MODIS, Remote Sensing of Environment, 113, 2463–2475, doi:10.1016/j.rse.2009.05.008, 2009.

Machguth, H., Macferrin, M., van As, D., Box, J. E., Charalampidis, C., Colgan, W., Fausto, R. S., Meijer, H. A. J., Mosley-Thompson, E. and van de Wal, R. S. W.: Greenland meltwater storage in firn limited by near-surface ice formation, Nature Climate Change, 6, 390–393, doi:10.1038/nclimate2899, 2016.

Maeno, N. and Ebinuma, T.: Pressure sintering of ice and its implication to the densification of snow at polar glaciers and ice sheets, Journal of Physical Chemistry, 87(21), 4103–4110, doi:10.1021/j100244a023, 1983.

Marchenko, S., Pohjola, V., Pettersson, R., van Pelt, W. J. J., Vega, C., Machguth, H., Boggild, C. E. and Isaksson, E.: A plot-scale study of firn stratigraphy at Lomonosovfonna, Svalbard, using ice cores, borehole video and GPR surveys in 2012–14, Journal of Glaciology, 63(237), 67–78, doi:10.1017/jog.2016.118, 2017.

Marsh, P. and Woo, M. -K: Wetting front advance and freezing of meltwater within a snow cover: 1. Observations in the Canadian Arctic, Water Resources Research, 20, 1853–1864, doi:10.1029/WR020i012p01853, 1984.

Martin-Español, A., Bamber, J. L. and Zammit-Mangion, A.: Constraining the mass balance of East Antarctica. Geophysical Research Letters, 44, 4168–4175. https://doi.org/10.1002/2017GL072937, 2017.

McMillan, M., Leeson, A., Shepherd, A., Briggs, K., Armitage, T., Hogg, A., Kuipers Munneke, P., van den Broeke, M., 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, Geophysical Research Letters, 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, Annals of Glaciology, 56(70),

155–166, doi:10.3189/2015AoG70A203, 2015.
Medley, B., McConnell, J., Neumann, T. A., Reijmer, C. H., Chellman, N., Sigl, M. and Kipfstuhl, S.: Temperature and snowfall in Western Queen Maud Land increasing faster than climate model projections, Geophysical Research Letters, 45, 1472–1480. https://doi.org/10.1002/2017GL075992, 2018.
Medley, B. and Thomas, E. R.: Increased snowfall over the Antarctic Ice Sheet mitigated twentieth-century sea-level rise, Nature Climate Change, 9, 34–39. https://doi.org/10.1038/s41558-018-0356-x, 2019.
Meyer, C. R., Keegan, K. M., Baker, I., and Hawley, R. L.: A model for French-press experiments of dry snow compaction, The Cryosphere, 14, 1449–1458, https://doi.org/10.5194/tc-14- 1449-2020, 2020.
Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., Kofinas, G., Mackintosh, A., MelbourneThomas, J., Muelbert, M. M. C., Ottersen, H., Pritchard, H., and Schuur, E. A. G.: Polar Regions, in IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, in press, 2019.
Meyer, C. R. and Hewitt, I. J.: A continuum model for meltwater flow through compacting snow, The Cryosphere, 11, 2799-2813, https://doi.org/10.5194/tc-11-2799-2017, 2017.
Miège, C., Forster, R. R., Brucker, L., Koenig, L. S., Solomon, D. K., Paden, J. D., Box, J. E., Burgess, E. W., Miller, J. Z., McNerney, L., Brautigam, N., Fausto, R. S., and Gogineni, S.: Spatial extent and temporal variability of Greenland firn aquifers detected by ground and airborne radars, Journal of Geophysical Research, 121, 2381–2398, https://doi.org/10.1002/2016JF003869, 2016.
Miller, O., Solomon, D. K., Miège, C., Koenig, L., Forster, R., Schmerr, N., and Montgomery, L.: Direct evidence of meltwater flow within a firn aquifer in southeast Greenland, Geophysical Research Letters, 45, 207–215, https://doi.org/10.1002/2017GL075707, 2018.
Mohajerani, Y., Velicogna, I. and Rignot, E.: Evaluation of regional climate models using regionally-optimized GRACE Mascons in the Amery and Getz ice shelves basins, Antarctica, Geophysical Research Letters, 46, 13,883–13,891. https://doi.org/10.1029/2019GL084665, 2019.
Monaghan, A. J., Bromwich, D. H., Fogt, R. L., Wang, S.- H., Mayewski, P. A., Dixon, D. A., Ekaykin, A., Frezzotti, M., Goodwin, I., Isaksson, E., Kaspari, S. D., Morgan, V. I., Oerter, H., Van Ommen, T. D., Van der Veen, C. J., and Wen, J.: Insignificant change in Antarctic snowfall since the international geophysical year, Science, 313, 827–831, https://doi.org/10.1126/science.1128243, 2006.
Montgomery, L., Koenig, L. and Alexander, P.: The SUMup dataset: Compiled measurements of surface mass balance components over ice sheets and sea ice with analysis over Greenland, Earth System Science Data, 10(4), 1959–1985, doi:10.5194/essd-10-1959-2018, 2018.
Montgomery, L., Miège, C., Miller, J., Scambos, T., Wallin, B., Miller, O., Solomon, D. K. C., Forster, R., and Koenig, L.: Hydrologic properties of a highly permeable firn aquifer in the Wilkins Ice Shelf, Antarctica, Geophysical Research Letters, 47, https://doi.org/10.1029/2020GL089552, 2020.
Morris, E. M. and Wingham, D. J.: Densification of polar snow: Measurements, modeling, and implications for altimetry, Journal of Geophysical Research, 119, 349–365, doi:10.1002/2013JF002898, 2014.
Morris, E. M. and Wingham, D. J.: Uncertainty in mass balance trends derived from altimetry; a case study along the EGIG line, Central Greenland, Journal of Glaciology, 61(226), 345−356, https://doi.org/10.3189/2015JoG14J123, 2015.
Morris, E. M., Mulvaney, R., Arthern, R. J., Davies, D., Gurney, R. J., Lambert, P., De Rydt, J., Smith, A. M., Tuckwell, R. J. and Winstrup, M.: Snow Densification and Recent Accumulation Along the iSTAR Traverse, Pine Island Glacier, Antarctica, Journal of Geophysical Research, 122, 2284–2301, doi:10.1002/2017JF004357, 2017.
Morris, E. M.: Modeling Dry-Snow Densification without Abrupt Transition, Geosciences, 8(12), 464, doi:10.3390/geosciences8120464, 2018.
Mottram, R., Hansen, N., Kittel, C., van Wessem, M., Agosta, C., Amory, C., Boberg, F., van de Berg, W. J., Fettweis, X., Gossart, A., van Lipzig, N. P. M., van Meijgaard, E., Orr, A., Phillips, T., Webster, S., Simonsen, S. B., and Souverijns, N.: What is the Surface Mass Balance of Antarctica? An Intercomparison of Regional Climate Model Estimates, The Cryosphere Discussion, https://doi.org/10.5194/tc-2019-333, 2020.
Mualem, Y.: A new model for predicting the hydraulic conductivity of unsaturated porous media, Water Resources Research, 12, 564–566, 1976.
Nghiem, S. V., Hall, D. K., Mote, T. L., Tedesco, M., Albert, M. R., Keegan, K., Shuman, C. A., DiGirolamo, N. E., and Neuman, G.: The Extreme Melt across the Greenland Ice Sheet in 2012, Geophysical Research Letters, 39, L20502, doi:10.1029/2012GL053611, 2012.
Noël, B., van de Berg, W. J., van Wessem, J. M., van Meijgaard, E. and van As, D.: Modelling the climate and surface mass balance of polar ice sheets using RACMO2 – Part 1 : Greenland (1958 – 2016), The Cryosphere, 811–831, 2018.
Noël, B., van de Berg, W. J., Lhermitte, S. and van den Broeke, M. R.: Rapid ablation zone expansion amplifies north Greenland mass loss, Science Advances, 5(9), 1–9, doi:10.1126/sciadv.aaw0123, 2019.

O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, Reliability Engineering and System Safety, 91(10–11), 1290–1300. https://doi.org/10.1016/j.ress.2005.11.025, 2006.
Parry, V., Nienow, P., Mair, D., Scott, J., Hubbard, B., Steffen, K. and Wingham, D.: Investigations of meltwater refreezing and density variations in the snowpack and firn within the percolation zone of the Greenland ice sheet, Annals of Glaciology, 46, 61–68, doi:10.3189/172756407782871332, 2007.
Pfeffer, W. T. and Humphrey, N. F.: Determination of timing and location of water movement and ice-layer formation by temperature measurements in sub-freezing snow, Journal of Glaciology, 1996.
Pfeffer, W. T. and Humphrey, N. F.: Formation of ice layers by infiltration and refreezing of meltwater, Annals of Glaciology, 83–91, 1998.
Poinar, K., Joughin, I., Lilien, D., Brucker, L., Kehrl, L. and Nowicki, S.: Drainage of Southeast Greenland Firn Aquifer Water through Crevasses to the Bed, Frontiers in Earth Science, 5, 1–15, doi:10.3389/feart.2017.00005, 2017.
Poinar, K., Dow, C. F., and Andrews, L. C.: Long-Term Support of an Active Subglacial Hydrologic System in Southeast Greenland by Firn Aquifers, Geophysical Research Letters, 46, 4772-4781, 2019.
Previdi, M. and Polvani, L. M.: Anthropogenic impact on Antarctic surface mass balance, currently masked by natural variability, to emerge by mid-century, Environmental Research Letters, 11(9), 94001, https://doi.org/10.1088/1748-9326/11/9/094001, 2016.
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.
Proksch, M., Löwe, H., and Schneebeli, M.: Density, specific surface area and correlation length of snow measured by highresolution penetrometry, Journal of Geophysical Research, 120, 346–362, doi:10.1002/2014JF003266, 2015.
Quiñonero-Candela, J., Rasmussen, C. E. and Williams, C. K. I.: Approximation methods for Gaussian process regression, Large Scale Learning Machines, 203–223. MIT Press, 2007.
Rasmussen, C. E., and Williams C. K. I.: Gaussian Processes for Machine Learning, MIT Press, Cambridge, Mass., 2006.
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.
Reeh, N., Fisher, D. A., Koerner, R. M. and Clausen, H. B.: An empirical firndensification model comprising ice lenses, Annals of Glaciology, 42, 101–106, https://doi.org/10.3189/172756405781812871, 2005.
Reeh, N.: A nonsteady-state firn-densification model for the percolation zone of a glacier, Journal of Geophysical Research, 113(3), 1–13, doi:10.1029/2007JF000746, 2008.
Refsgaard, J., Van der Sluijs, J., Højberg, A., and Vanrolleghem, P.: Uncertainty in the environmental modelling process–A framework and guidance, Environmental Modelling and Software, 22(11), 1543– 1556, doi:10.1016/j.envsoft.2007.02.004, 2007.
Reijmer, C. H., van den Broeke, M. R., Fettweis, X., Ettema, J. and Stap, L. B.: Refreezing on the Greenland ice sheet: A comparison of parameterizations, The Cryosphere, 6, 743–762, doi:10.5194/tc-6-743-2012, 2012.
Rignot, E., Mouginot, J., Scheuchl, B., van den Broeke, M., van Wessem, M. J. and Morlighem, M.: Four decades of Antarctic Ice Sheet mass balance from 1979–2017, Proceedings of the National Academy of Sciences, 116(4), 1095–1103. https://doi.org/10.1073/pnas.1812883116, 2019.
Rios, G. and Tobar, F.: Learning non-Gaussian time series using the box-cox Gaussian process, Proceedings of the International Joint Conference on Neural Networks 2018-July, 2018
Robin, G. d. Q.: Glaciology III: Seismic shooting and related investigations, in Norwegian-British-Swedish Antarctic Expedition, 1949–52, Scientific Results, vol. 5, Norsk Polarinstit, Oslo, 1958.
Rosenthal, J.: Optimal Proposal Distributions and Adaptive MCMC, (1), 1–25, doi:10.1201/b10905-5, 2010.
Sacks, J., Welch W. J., Mitchell W. J., & H. P. Wynn H. P.: Design and Analysis of Computer Experiments. Statistical Science, 4(4): 409–435. doi:10.1214/ss/1177012413, 1989.
Scambos, T. and Shumman, C.: Comment on 'Mass gains of the Antarctic ice sheet exceed losses' by H. J. Zwally and others, Journal of Glaciology, 62(233), 599–603, doi:10.1017/jog.2016.59, 2016.
Schleef, S. and Löwe, H.: X-ray microtomography analysis of isothermal densification of new snow under external mechanical stress, Journal of Glaciology, 59(214), 233–243, doi:10.3189/2013JoG12J076, 2013.
Seeger, M., Williams, C. and Lawrence, N.: Fast forward selection to speed up sparse Gaussian process regression, Artificial Intelligence and Statistics, 9, EPFL-CONF-161318, 2003.
Shepherd, A., Ivins, E., Geruo, A., Barletta, V., Bentley, M., Bettadpur, S., Briggs, K., Bromwich, D., Forsberg, R., Galin, N., Horwath, M., Jacobs, S., Joughin, I., King, M., Lenaerts, J., Li, J., Ligtenberg, S., Luckman, A., Luthcke, S., McMillan, M., Meister, R., Milne, G., Mouginot, J., Muir, A., Nicolas, J.,

Paden, J., Payne, A., Pritchard, H., Rignot, E., Rott, H., Sandberg Sørensen, L., Scambos, T., Scheuchl, B., Schrama, E., Smith, B., Sundal, A., van Angelen, J., van de Berg, W., van den Broeke, M., Vaughan, D., Velicogna, I., Wahr, J., Whitehouse, P., Wingham, D., Yi, D., Young, D. and Zwally, H. J.: A Reconciled Estimate of Ice-Sheet Mass Balance, Science, 338(6111), 1183–1189, https://doi.org/10.1126/science.1228102, 2012.

Shepherd, A., Ivins, E., Rignot, E., Smith, B., van den Broeke, M., Velicogna, I., Whitehouse, P., Briggs, K., Joughin, I., Krinner, G., Nowicki, S., Payne, T., Scambos, T., Schlegel, N., Geruo, A., Agosta, C., Ahlstrøm, A., Babonis, G., Barletta, V., Blazquez, A., Bonin, J., Csatho, B., Cullather, R., Felikson, D., Fettweis, X., Forsberg, R., Gallee, H., Gardner, A., Gilbert, L., Groh, A., Gunter, B., Hanna, E., Harig, C., Helm, V., Horvath, A., Horwath, M., Khan, S., Kjeldsen, K. K., Konrad, H., Langen, P., Lecavalier, B., Loomis, B., Luthcke, S., McMillan, M., Melini, D., Mernild, S., Mohajerani, Y., Moore, P., Mouginot, J., Moyano, G., Muir, A., Nagler, T., Nield, G., Nilsson, J., Noel, B., Otosaka, I., Pattle, M. E., Peltier, W. R., Pie, N., Rietbroek, R., Rott, H., Sandberg-Sørensen, L., Sasgen, I., Save, H., Scheuchl, B., Schrama, E., Schröder, L., Seo, K. W., Simonsen, S., Slater, T., Spada, G., Sutterley, T., Talpe, M., Tarasov, L., van de Berg, W. J., van der Wal, W., van Wessem, M., Vishwakarma, B. D., Wiese, D. and Wouters, B.: Mass balance of the Antarctic Ice Sheet from 1992 to 2017, Nature, 558, 219–222, doi:10.1038/s41586-018-0179-y, 2018.

Shiklomanov, I. A. and Rodda, J. C. (Eds.): World water resources at the beginning of the twenty-first century, Cambridge University Press, Cambridge, UK, 2003.

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, Journal of Glaciology, 59, 545–558, doi:10.3189/2013JoG12J158, 2013.

Smith, B., Fricker, H. A., Gardner, A. S., Medley, B., Nilsson, J., Paolo, F. S., Holschuh, N., Adusumilli, S., Brunt, K., Csatho, B., Harbeck, K., Markus, T., Neumann, T., Siegfried, M. R., and Zwally, H. J.: Pervasive ice sheet mass loss reflects competing ocean and atmosphere processes, Science, 368, 1239–1242, https://doi.org/10.1126/science.aaz5845, 2020.

Sommers, A. N., Rajaram, H., Weber, E. P., MacFerrin, M. J., Colgan, W. T. and Stevens, C. M.: Inferring Firn Permeability from Pneumatic Testing: A Case Study on the Greenland Ice Sheet, Frontiers in Earth Science, 5, 1–12, doi:10.3389/feart.2017.00020, 2017.

Spencer, M. K., Alley, R. B. and Creyts, T. T.: Preliminary firn-densification model with 38-site dataset, Journal of Glaciology, 47, 671–676, https://doi.org/10.3189/172756501781831765, 2001.

Steffen, K., Box, J. E. and Abdalati, W.: Greenland Climate Network: GC-Net, in Glaciers, Ice Sheets and Volcanoes: A Tribute to Mark F. Meier, vol. Report 96-27, pp. 98–103, Cold Regions Research and Engineering Laboratory, Hanover, NH., 1996.

Steger, C. R., Reijmer, C. H., van den Broeke, M. R., Wever, N., Forster, R. R., Koenig, L. S., Kuipers Munneke, P., Lehning, M., Lhermitte, S., Ligtenberg, S. R. M., Miège, C. and Noël, B. P. Y.: Firn Meltwater Retention on the Greenland Ice Sheet: A Model Comparison, Frontiers in Earth Science, 5, doi:10.3389/feart.2017.00003, 2017a.

Steger, C. R., Reijmer, C. H., and van den Broeke, M. R.: The modelled liquid water balance of the Greenland Ice Sheet, The Cryosphere, 11, 2507-2526, https://doi.org/10.5194/tc-11-2507-2017, 2017b.

Stevens, C. M.: Investigations of physical processes in polar firn through modeling and field measurements, University of Washington, Seattle., 2018.

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, Geoscientific Model Development, 13, 4355–4377, https://doi.org/10.5194/gmd-13-4355-2020, 2020.

Szymkiewicz, A.: Approximation of internodal conductivities in numerical simulation of one-dimensional infiltration, drainage, and capillary rise in unsaturated soils, Water Resources Research, 45, 1–16, doi:10.1029/2008WR007654, 2009.

Tedesco, M. and Fettweis, X.: Unprecedented atmospheric conditions (1948–2019) drive the 2019 exceptional melting season over the Greenland ice sheet, The Cryosphere, 14, 1209–1223, https://doi.org/10.5194/tc-14-1209-2020, 2020.

van Angelen, J. H., Lenaerts, J. T. M., van den Broeke, M. R., Fettweis, X., and van Meijgaard, E.: Rapid loss of firn pore space accelerates 21st century Greenland mass loss, Geophysical Research Letters, 40, 2109–2113, https://doi.org/10.1002/grl.50490, 2013.

van Angelen, J. H., van den Broeke, M. R., Wouters, B. and Lenaerts, J. T. M.: Contemporary (1960–2012) Evolution of the Climate and Surface Mass Balance of the Greenland Ice Sheet, Surveys in Geophysics, 35, 1155–1174, doi:10.1007/s10712-013-9261-z, 2014.

van As, D., Box, J. E. and Fausto, R. S.: Challenges of Quantifying Meltwater Retention in Snow and Firn: An Expert Elicitation, Frontiers in Earth Science, 4, 1–5, doi:10.3389/feart.2016.00101, 2016a.

van As, D., Fausto, R. S., Cappelen, J., Van De Wal, R. S. W., Braithwaite, R. J., Machguth, H., Charalampidis,

C., Box, J. E., Solgaard, A. M., Ahlstrom, A. P., Haubner, K., Citterio, M. and Andersen, S. B.: Placing Greenland ice sheet ablation measurements in a multi-decadal context, Geological Survey of Denmark and Greenland Bulletin, 35, 71–74, 2016b.

van den Broeke, M. R.: Depth and density of the Antarctic firn layer depth and density of the Antarctic firn layer. Arctic, Antarctic, and Alpine Research, 40(2), 432–438, https://doi.org/10.1657/1523-0430(07-021)[BROEKE]2.0.CO;2, 2008.

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.

van Genuchten, M.: A closed form equation for predicting the hydraulic conductivity of unsaturated soils., Soil Science Society of America Journal, 44, 892–898, 1980.

van Kampenhout, L., Lenaerts, J. T. M., Lipscomb, W. H., Sacks, W. J., Lawrence, D. M., Slater, A. G. and van den Broeke, M. R.: Improving the Representation of Polar Snow and Firn in the Community Earth System Model, Journal of Advances in Modeling Earth Systems, 9, 2583–2600, doi:10.1002/2017MS000988, 2017.

van Pelt, W. J. J., Oerlemans, J., Reijmer, C. H., Pohjola, V. A., Pettersson, R. and van Angelen, J. H.: Simulating melt, runoff and refreezing on Nordenskiöldbreen, Svalbard, using a coupled snow and energy balance model, The Cryosphere, 6, 641–659, doi:10.5194/tc-6-641-2012, 2012.

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 Ulft, 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.

Vandecrux, B., Mottram, R., Langen, P. L., Fausto, R. S., Olesen, M., Stevens, C. M., Verjans, V., Leeson, A., Ligtenberg, S., Kuipers Munneke, P., Marchenko, S., van Pelt, W., Meyer, C. R., Simonsen, S. B., Heilig, A., Samimi, S., Marshall, S., Machguth, H., MacFerrin, M., Niwano, M., Miller, O., Voss, C. I., and Box, J. E.: The firn meltwater Retention Model Intercomparison Project (RetMIP): evaluation of nine firn models at four weather station sites on the Greenland ice sheet, The Cryosphere, 14, 3785–3810, https://doi.org/10.5194/tc-14-3785-2020, 2020.

Velicogna, I., Sutterley, T. and van den Broeke, M.: Regional acceleration in ice mass loss from Greenland and Antarctica using GRACE time-variable gravity data, Geophysical Research Letters, 41(22), 8130–8137. https://doi.org/https://doi.org/10.1002/2014GL061052, 2014.

Verjans, V., Leeson, A. A., Stevens, C. M., 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, 1819–1842, https://doi.org/10.5194/tc-13-1819-2019, 2019.

Verjans, V., Leeson, A. A., Nemeth, C., Stevens, C. M., Kuipers Munneke, P., Noël, B., and van Wessem, J. M.: Bayesian calibration of firn densification models, The Cryosphere, 14, 3017–3032, https://doi.org/10.5194/tc-14-3017-2020, 2020.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E. and Willemet, J. M.: The detailed snowpack scheme CROCUS and its implementation in SURFEX v7.2, Geoscientific Model Development, 5, 773–791, doi:10.5194/gmd-5-773-2012, 2012.

von Storch, H. and Zwiers, F. W.: Statistical analysis in climate research, Cambridge University Press, Cambridge, U. K., 1999.

Waldner, P. A., Schneebeli, M., Schultze-Zimmermann, U. and Flühler, H.: Effect of snow structure on water flow and solute transport, Hydrological Processes, 18, 1271–1290, doi:10.1002/hyp.1401, 2004.

Wang, Y., Ding, M., van Wessem, J., Schlosser, E., Altnau, S., van den Broeke, M., Lenaerts, J., Thomas, E., Isaksson, E., Wang, J., and Sun, W.: A Comparison of Antarctic Ice Sheet Surface Mass Balance from Atmospheric Climate Models and In Situ Observations, Journal of Climate, 29, 5317–5337, https://doi.org/10.1175/JCLID-15-0642.1, 2016.

Wang, W. S., Zender, C. S., and van As, D.: Temporal Characteristics of Cloud Radiative Effects on the Greenland Ice Sheet: Discoveries From Multiyear Automatic Weather Station Measurements, Journal of Geophysical Research, 123, 11348–11361, https://doi.org/10.1029/2018JD028540, 2018.

Wever, N., Fierz, C., Mitterer, C., Hirashima, H. and Lehning, M.: Solving Richards Equation for snow improves snowpack meltwater runoff estimations in detailed multi-layer snowpack model, The Cryosphere, 8, 257–274, doi:10.5194/tc-8-257-2014, 2014.

Wever, N., Würzer, S., Fierz, C. and Lehning, M.: Simulating ice layer formation under the presence of preferential flow in layered snowpacks, The Cryosphere, 10, 2731–2744, doi:10.5194/tc-10-2731-2016, 2016.

Wilkinson, D.: A pressure-sintering model for the densification of polar firn and glacier ice, Journal of Glaciology, 34(116), 40–45, doi:10.3189/S0022143000009047, 1988.

Williams, M. W., Erickson, T. A. and Petrzelka, J. L.: Visualizing meltwater flow through snow at the centimetre-to-metre scale using a snow guillotine, Hydrological Processes, 24, 2098–2110, doi:10.1002/hyp.7630, 2010.
Winther, J. G., Jespersen, M. N. and Liston, G.E.: Blue-ice areas in Antarctica derived from NOAA AVHRR satellite data, Journal of Glaciology, 47(157), 325–334. doi:10.3189/172756501781832386, 2001.
Yamaguchi, S., Katsushima, T., Sato, A. and Kumakura, T.: Water retention curve of snow with different grain sizes, Cold Regions Science and Technology, 64, 87–93, doi:10.1016/j.coldregions.2010.05.008, 2010.
Yamaguchi, S., Watanabe, K., Katsushima, T., Sato, A. and Kumakura, T.: Dependence of the water retention curve of snow on snow characteristics, Annals of Glaciology, 53, 6–12, doi:10.3189/2012AoG61A001, 2012.
Yen, Y.-C.: Review of the thermal properties of snow, ice and sea ice, Tech. Rep. 81-10, Cold Regions Research and Engineering Laboratory, Hanover, NH, 1981.
Yip, S., Ferro, C. A. T., Stephenson, D. B. and Hawkins, E.:. A simple, coherent framework for partitioning uncertainty in climate predictions, Journal of Climate, 24(17), 4634–4643. https://doi.org/10.1175/2011jcli4085.1, 2011.
Zuo, Z. and Oerlemans, J.: Modelling albedo and specific balance of the Greenland ice sheet: Calculations for the Søndre Stromfjord transect, Journal of Glaciology, 42, 305–316, doi:10.3198/1996JoG42-141-305-317, 1996.
Zwally, H. J. and Li J.: Seasonal and interannual variations of firn densification and ice-sheet surface elevation at the Greenland Summit, Journal of Glaciology, 48(161), 199–207. https://doi.org/10.3189/172756502781831403, 2002.
Zwally, H. J., Li, J., Robbins, J. W., Saba, J. L., Yi, D. and Brenner, A. C.: Mass gains of the Antarctic ice sheet exceed losses, Journal of Glaciology, 61(230), 1019–1036. https://doi.org/10.3189/2015JoG15J071, 2015.

Appendix A

This appendix includes the Supplementary Information to Chapter 1.

A.1 Regional Climate Model performance

The climatic forcing (temperature, melt, snowfall, rain and sublimation) is provided by the 5.5 km resolution RACMO2.3p2 regional climate model. Table A1 provides summary statistics of model performance with respect to weather station data from the Institute for Marine and Atmospheric research at Utrecht University (IMAU), from the Programme for Monitoring of the Greenland Ice Sheet (PROMICE) (van As et al., 2016) and from the Greenland Climate Network GC-Net (Steffen et al., 1996). Detailed pointwise comparisons between modelled and observed values of temperature, melt rate and precipitation rate are shown in Figs. A1, A2 and A3. In general, RACMO reaches good agreement the with observed data, and root mean squared errors (RMSE) for daily 2 m height temperatures, annual melt rates and annual precipitation rates are of the order of 5 K, 0.1 m w.e. yr^{-1} and 0.15 m w.e. yr^{-1} respectively.

At both FA13 and KAN-U, modelled temperatures and melt rates show very low bias (smaller than 1 K and 0.06 m w.e. yr^{-1}). The low RMSE in melt rates (0.11 and 0.08 m w.e. yr^{-1} respectively) is noteworthy because the total melt rates are large. We note that no precipitation data is available at these sites and only two years of melt data for FA13, but RACMO2.3p2 has shown great agreement with other surface mass balance measurements, especially in accumulation areas (RMSE of 0.08 m w.e. yr^{-1}), although weather stations remain sparse on the GrIS (Noël et al., 2018). There is no evidence that the climatic forcing is a source for bias in the firn models' simulations at these sites.

At the stations of DYE-2 and NASA-SE, observed GC-Net data is of lower quality, which partly explains their relatively larger RMSE in temperature and melt rates. At DYE-2, RMSE in precipitation is still low (0.09 m w.e. yr^{-1}) and the bias is only slightly positive (0.04 m w.e. yr^{-1}). Melt rates tend to be underestimated but the RMSE is reasonable (0.13 m w.e. yr^{-1}). It is difficult to partition the larger errors in temperature (positive bias of 1.2 K) between model and measurement errors, but no indication of firn temperature overestimation appears in the firn models' simulations. At NASA-SE, we hypothesize that the negative bias in precipitation rates (-0.21 m w.e. yr^{-1}) leads to an underestimation of downward advection of cold snow and partly causes the 10 m temperature overestimation in our simulations. Additionally, the temperature overestimation (bias of 1.5 K) could enhance the problem. The negative bias in melt rates (-0.07 m w.e. yr^{-1}) plays an opposite role by underestimating latent heat release. However, annual melt rates are low and the effects of this underestimation are likely minor.

RACMO2.3p2 vs. AWS			Temperature 2 m [K]			Melt [m w.e. yr^{-1}]			Precipitation [m w.e. yr^{-1}]		
Station	Period	number of observations	R^2	Bias	RMSE	R^2	Bias	RMSE	R^2	Bias	RMSE
NASA-SE	1998-2015	6209	0.61	1.47	6.94	0.77	$-68.34\ 10^{-3}$	$84.48\ 10^{-3}$	0.85	$-206.1\ 10^{-3}$	$228.8\ 10^{-3}$
DYE-2	1998-2015	6209	0.60	1.19	7.32	0.88	$-101.4\ 10^{-3}$	$128.2\ 10^{-3}$	0.78	$43.58\ 10^{-3}$	$90.30\ 10^{-3}$
KAN-U	2009-2016	2718	0.97	0.88	2.09	0.93	$14.00\ 10^{-3}$	$81.92\ 10^{-3}$	-	-	-
FA13	2014-2016	352	0.92	0.87	2.20	-	$-51.56\ 10^{-3}$	$105.8\ 10^{-3}$	-	-	-

Table A1. Statistics of comparison between RACMO2.3p2 climatic output and Automatic Weather Station measurements, hyphens indicate absence or paucity of data

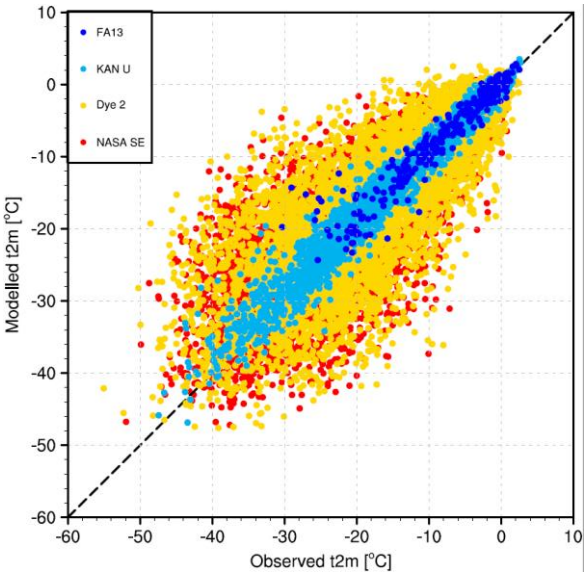


Figure A1. Pointwise comparison between RACMO2.3p2 and Automatic Weather Station data of temperature at 2 m height

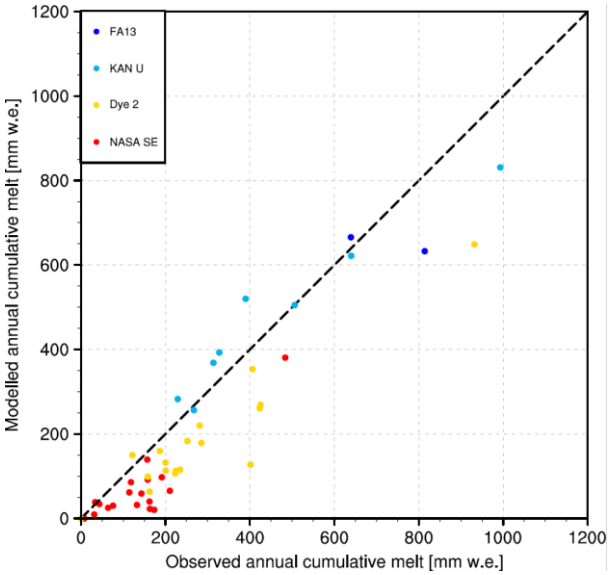


Figure A2. Pointwise comparison between RACMO2.3p2 and Automatic Weather Station data of annual cumulative melt

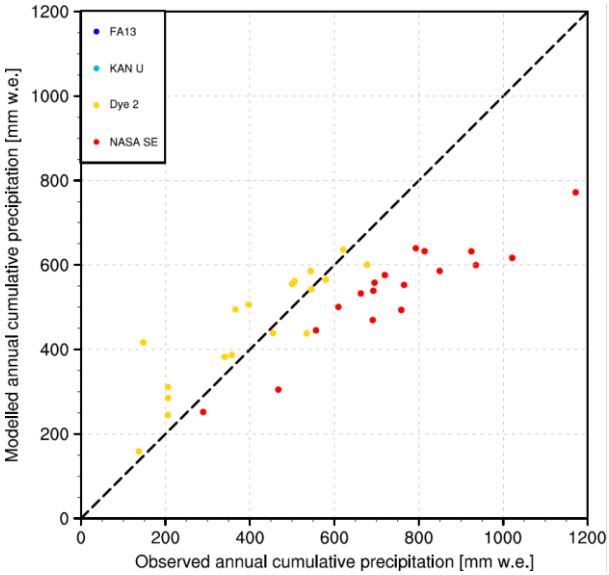


Figure A3. Pointwise comparison between RACMO2.3p2 and Automatic Weather Station data of annual cumulative precipitation

A.2 Model Implementation

The model uses a finite-volume scheme with each layer being an independent volume. We use the general mixed-form Picard iteration scheme to solve the RE, as it has been demonstrated that the mixed form of RE can be efficiently used in finite-difference schemes because of its accuracy and its robustness with respect to mass conservation (Celia et al., 1990). The Picard scheme discretises the model using central finite differences for the space derivative and a backward Euler method for the time derivative. The iterative process calculates the value of the pressure head at each iteration and then adjusts the liquid water content according to the water retention curve, Eq. (1.12). Hydraulic parameters are updated and iterations are repeated until convergence of the solution is achieved. Boundary conditions are the rate of meltwater input at the surface and a no-flow condition at the bottom. Solving the RE in the firn column presents numerical challenges. We adopt an implementation strategy based on the works of Wever et al. (2014) and D'Amboise et al. (2017) who implemented the RE in the snow models SNOWPACK and CROCUS respectively. Here, we give more details about this methodology.

A.2.1 Convergence criteria

For the solution reached by the Picard iteration scheme to be considered convergent, it must fulfil different criteria. The convergence criteria between two successive iterations are defined for the head pressure (ε_h) and liquid water content (ε_θ) values as well as for the mass balance error (ε_{MB}) of individual layers. These three criteria are fixed to 10^{-3} m, 10^{-5} and 10^{-8} m respectively, following Huang et al. (1996) and Wever et al. (2014). For each layer, we select ε_θ or ε_h according to the

effective saturation. Huang et al. (1996) showed that using ε_θ allows faster convergence. However, it cannot be used in very saturated layers and thus we apply ε_h for layers where effective saturation exceeds 0.99, in accordance with Wever et al. (2014). The mass-balance criterion is always applied to every layer, regardless of the saturation.

A.2.2 Hydraulic conductivity calculation

As we use a central finite difference approach to compute RE, the fluxes are assumed to occur on the interface between adjacent layers. Incoming and outgoing fluxes are computed and this requires the hydraulic conductivity value to be calculated at the top and bottom of every layer *i* and not at the centre. We use the upstream-weighting technique (Forsyth et al., 1995):

$$K_{i+\frac{1}{2}} = \begin{cases} K_i, \ if \ \frac{\Delta h}{\Delta z} - 1 \leq 0 \\ K_{i+1}, \ if \ \frac{\Delta h}{\Delta z} - 1 > 0 \end{cases} \quad \text{(A1)}$$

The advantage of this formulation over a simple arithmetic mean is that it does not lead to oscillatory solutions, regardless of the mesh size (Forsyth et al., 1995; Szymkiewicz, 2009).

A.2.3 Dry layers

For numerical stability, a snow layer cannot be completely dry (i.e. θ = 0). Therefore, two cases must be considered: dry layers and refreezing layers. At the start of the flow routine, all layers are initialised with a very low θ value, θ_{dry}. The value must be sufficiently low to avoid influencing the refreezing process but sufficiently high to lead to a convergent solution (D'Amboise et al., 2017). In this study, the θ_{dry} value is fixed at 10^{-6} as this is a tenth of the ε_θ criterion. This corresponds to a 1 m thick snow layer holding 1 μm of liquid water. When the flow routine is called in the firn model, the water content of every dry layer is thus synthetically raised to θ_{dry}, which corresponds to a pre-wetting. The porosity of ice layers that are at high densities (>900 kg m^{-3}) is thus adjusted in order to raise their water content to θ_{dry} in both domains.

Similarly, there is a risk for θ reaching too-low values when refreezing occurs. Therefore, refreezing is allowed only if the θ value is above 0.01% (Wever et al., 2014). This value is above θ_{dry} to avoid refreezing and corresponding latent heat release of the very low amounts of water resulting from the fluxes between layers that are initialised at θ_{dry}. Only at the last time step of the flow routine is the refreezing process allowed to decrease the volumetric water content until θ_{dry}. After that, the pre-wetting amounts of liquid water are subtracted at the end of the flow routine to maintain the mass conservation property of the firn model. At the end of the flow routine, if all the layers have a water content below 10^{-5}, we consider the firn column to be completely dry again so that the flow routine does not have to be called until the next melt event and computational time is largely saved.

A.2.4 Dynamical time step adjustment

The numerical solving of RE uses a dynamically adjusted time step. Certain situations, such as the arrival of the wetting front at a stratigraphic transition, require a very small time step whereas larger time steps can be used in other cases without affecting numerical stability. Thus, the time step is adjusted according to the number of iterations, n_{it}, required to achieve convergence of the solution at the previous time step: decreased for a large number of iterations and increased for few iterations. Also, as in Wever et al. (2014) and D'Amboise et al. (2017), a back step case is used: the calculation is stopped and the time step automatically decreased if the solution fails to converge in 15 iterations or if warning signs of instability appear (positive pressure head values, effective saturation exceeding 1 or differences in successive pressure head values exceeding 10^3 m). The time step is bounded between 10^{-20} s and 900 s. The procedure can be summarised as follows:

$$\Delta t_{RE}^{t} = \begin{cases} 1.25\ \Delta t_{RE}^{t-1}, & if\ n_{it} \leq 5 \\ \Delta t_{RE}^{t-1}, & if\ 5 < n_{it} < 10 \\ 0.5\ \Delta t_{RE}^{t-1}, & if\ 10 \leq n_{it} \leq 15 \\ back\ step, & if\ n_{it} > 15 \end{cases} \quad \text{(A2)}$$

A.2.5 Saturated layers and aquifer treatment

If water reaches an impermeable ice layer, the layer above progressively becomes saturated. This means that its hydraulic conductivity progressively increases. As a consequence, the incoming flow becomes very large whereas the outgoing flow is forced to be zero. To deal with this issue, the layer has to be set impermeable once close to saturation and this process must go on for layers above when these reach saturation in turn. When an aquifer is present at the bottom of the domain, the amount of water is held in memory at the start of the flow routine and the end of the domain is set as the top of the aquifer. All the percolating water reaching the end of the domain is added to the aquifer amount and at the end of the routine, this total amount is redistributed in the bottom layers.

A.2.6 Partial RE solving

In order to save computational time, the RE is not necessarily solved for the entire domain. If a significant part of the lowest layers is dry, we do not proceed to the calculations for this lower dry part. Starting from the surface, we look for the lowest layer where the water content is at least ε_θ + 0.01 % (above the minimum water content after refreezing). Then we take as lower limit for the RE calculation the layer situated 50 cm below this lowest wet layer. This is recalculated at every time step of the RE solving, making the 50 cm addition largely sufficient to capture the wetting of the dry lower part. If the lowest wet layer is less than 50 cm above the end of the domain, then the RE is calculated on the entire domain. This is applied in both the matrix and the preferential flow domains.

A.2.7 Refreezing in the preferential flow domain

Contrarily to the SNOWPACK model, there is a particular circumstance for which we apply refreezing directly in the preferential flow domain: if a cold front (subfreezing temperatures) propagates from the surface into a wet firn column, all the water present in the matrix flow domain will progressively be refrozen, starting from the surface layer. It would be unrealistic to keep liquid water present in the preferential flow domain of layers that are above this cold front. Thus, if starting from the surface, the entire firn column until a particular layer that holds some liquid water in the preferential flow domain is at subfreezing temperatures, this liquid water is refrozen. In such cases, the firn column is dry in both domains until the depth delimited by the cold front. Simulations without this refreezing implementation reached very similar results but required more computational time.

A.2.8 Merging process

The CFM usually considers every accumulation event as a new layer. However, as we use three hourly accumulation forcing, the firn layer could consist of a high number of extremely fine layers. Because the calculation time for the RE is very dependent on the number of distinct layers in the firn column, we chose to merge any layer thinner than 2 cm with the underlying layer. If this was applied to the surface layer, every accumulation event of less than 2 cm snow would be immediately merged with the previous surface layer. In the case where a high number of successive snowfall events would be below the 2 cm threshold, these would all be merged within the same layer, possibly becoming very thick. To avoid this, the newly added snow layer is merged with the previous surface layer only if the latter is below the 2 cm threshold. However, newly-added layers that are less than 0.01 mm thick are always merged with the layer below. It is important to keep a high vertical resolution when simulating the percolation process with the RE, as this flow equation is highly sensitive to structural heterogeneities in the firn. If the merging process is too lenient, this leads to the smoothing of heterogeneities such as sharp grain-size or density transitions. Moreover, using a coarse resolution would lead to only an approximation of the water percolation because water content is always homogeneous in a single layer. Thus, as soon as water percolates at the top of a given layer, it is distributed in the entire layer.

Appendix B

This appendix includes the Supplementary Information to Chapter 2.

B.1 Separation between the calibration and evaluation data

The 91 sites of the dataset span a broad range of temperature and accumulation rate conditions (Table B1 and Fig. B1, Table B1 is provided at the end of Appendix B). 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 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 B1) with high T_{av} and $\dot{b}$ values (Fig. B1). We select randomly one of these 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.

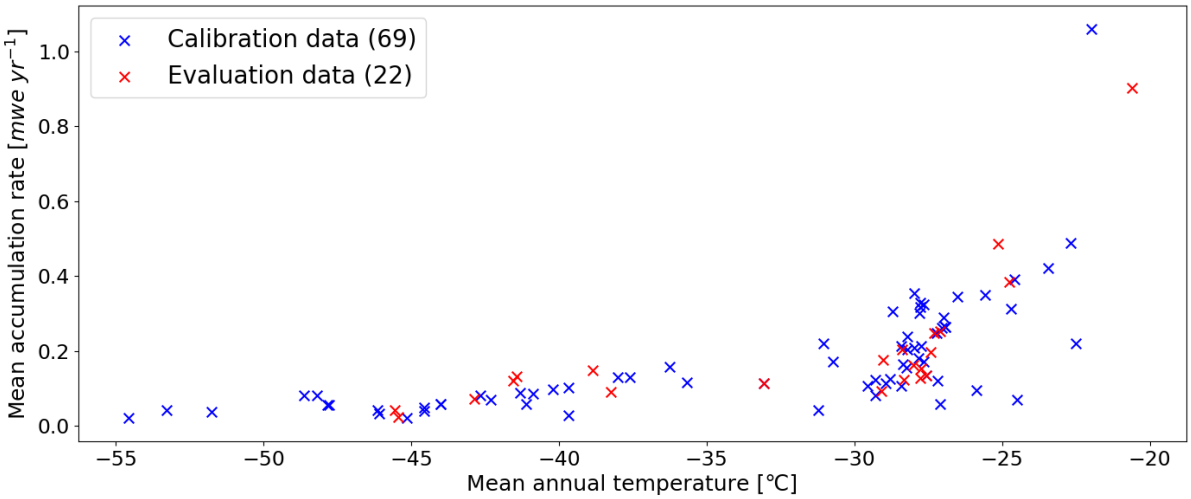


Figure B1. Climatic conditions at the 91 sites of the dataset

B.2 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 firn models. At each iteration (a round of 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. (B1)). 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 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. (B2), (B3), (B4)).

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 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.

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. (B3), (B4)).

Our approach is summarized in Eq. (B1), (B2), (B3) and (B4). These equations are applied at all iterations of the calibration process.

$$c_p \sim N(0,1) \quad \text{at all calibration sites} \tag{B1}$$

$$T_t^* = T_t + c_p \sigma_T \quad \text{at all } t \tag{B2}$$

$$\dot{b}_{tot}^* = n_{yr} c_p \sigma_{SMB} \tag{B3}$$

$$\dot{b}_t^* = \dot{b}_t + \dot{b}_{tot}^* \frac{\dot{b}_t}{\sum_t \dot{b}_t} \tag{B4}$$

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, σ_T = 2.1 K and σ_{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 σ_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. (B3) and (B4) are replaced by Eq. (B5).

$$\dot{b}_t^* = \dot{b}_t + 0.05\, c_p\, \dot{b}_t \quad \text{(B5)}$$

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 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. 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). As such, adding noise to ρ_0 simplifies to Eq. (B6).

$$\rho_{0,t}^* \sim N(\rho_0^{core}, 25) \quad \text{(B6)}$$

We emphasize that the aim of this study is not to conduct a complete sensitivity analysis of firn 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.

B.3 Prior correlations in HL and Ar

The Arrhenius form of HL and Ar (Eq. (2.4) and (2.5)) 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. (2.8)).

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

$$k_{0,mv}^* = k_{0,HL}^* \exp\left(\frac{E_{0,mv} - E_{0,HL}}{R\,T}\right) \quad \text{(B7)}$$

$$k_{1,mv}^* = k_{1,HL}^* \exp\left(\frac{E_{1,mv} - E_{1,HL}}{R\,T}\right) \quad \text{(B8)}$$

$$k_{0,mv}^{Ar} = k_{0,Ar}^{Ar} \exp\left(\frac{E_{g,Ar} - E_{g,mv}}{R\,T}\right) \quad \text{(B9)}$$

$$k_{1,mv}^{Ar} = k_{1,Ar}^{Ar} \exp\left(\frac{E_{g,Ar} - E_{g,mv}}{R\,T}\right) \quad \text{(B10)}$$

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. 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 ±500 J mol^{-1} around the original values. Thirdly, we calculate the corresponding values in the pre-exponential factors from Eq. (B7), (B8), (B9) and (B10). This results in 10000 pairs of $(k^*_{0,mv}, E_{0,mv})$, $(k^*_{1,mv}, E_{1,mv})$, $(k^{Ar}_{0,mv}, E_{g,mv})$ and $(k^{Ar}_{1,mv}, E_{g,mv})$, 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 $k^{Ar}_{0,mv}$ and $k^{Ar}_{1,mv}$. 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. B.7.

B.4 The likelihood function (Eq. (2.8))

The covariance matrices Σ_{15} and Σ_{pc} that appear in Eq. (2.8) are diagonal matrices with the site-specific variances on the diagonal. At each site, we take 10% of the observed $DIP15$ and 20% of the observed $DIPpc$ as the standard deviation, and the variance value is the square of the standard deviation. We take the higher value of 20% for $DIPpc$ because density errors propagate in firn models. Equation (2.3) 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 $DIPpc$ values.

The form of Eq. (2.8) 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 $DIP15$ $(X_{15} - Y_{15})$ and $DIPpc$ $\left(X_{pc} - Y_{pc}\right)$ 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 $DIP15$ and in $DIPpc$. 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 B2. As explained in the main text, the form of Eq. (2.8) also assumes independence between errors in $DIP15$ and $DIPpc$, which is the reason why $DIPpc$ is calculated only from depths below 15 m. As such, observations-

model discrepancies are essentially governed by parameter values of stage-1 densification for $DIP15$ and by parameter values of stage-2 densification for $DIPpc$, with little interaction between both. The same preliminary verification as mentioned above allows us to evaluate the correlation between $DIP15$ and $DIPpc$ 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.

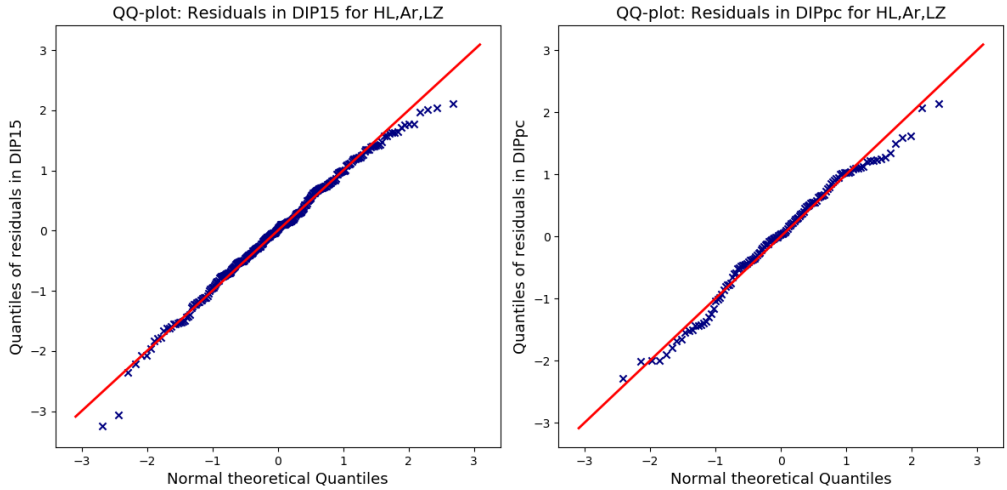


Figure B2. Quantiles-Quantiles plots for the errors of the three original models (HL, Ar, LZ) computed on the entire dataset. The alignment of the points along the red line informs about the fit to a normal distribution.

B.5 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. B3. 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.

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:

$$W = \frac{s_1^2+s_2^2+s_3^2}{3} \quad \text{(B11)}$$

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

$$B = \frac{n}{3-1}\sum_{i=1}^{3}\left(\bar{\theta}_i - \bar{\bar{\theta}}\right)^2 \quad \text{(B12)}$$

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:

$$\hat{\sigma}^2 = \frac{n-1}{n}W + \frac{1}{n}B \quad \text{(B13)}$$

And the Gelman-Rubin statistic is defined as:

$$R = \sqrt{\frac{\hat{\sigma}^2}{W}} \quad \text{(B14)}$$

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 $R < 1.1$.

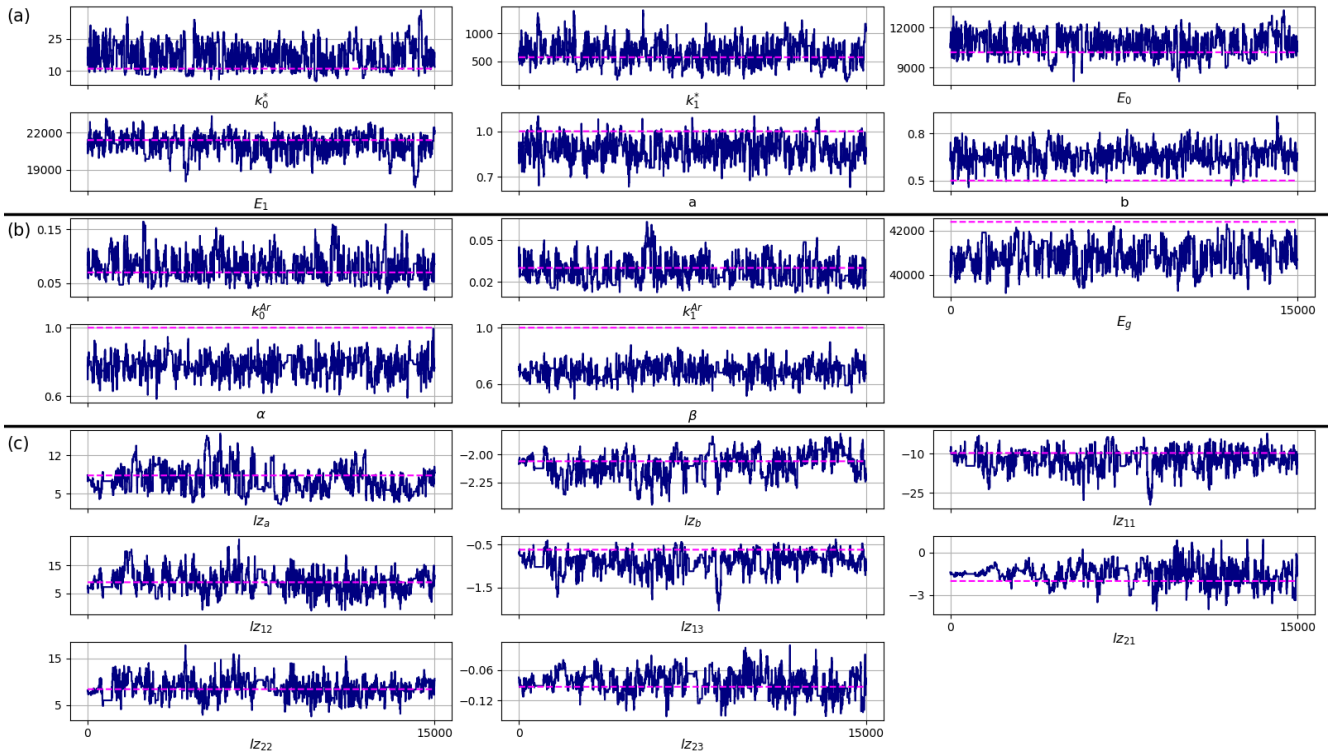


Figure B3. Sampling chains of each parameter for (a) HL, (b) Ar, (c) LZ. The x-axis displays the iteration number, the y-axis displays the parameter value. The dashed pink line shows the value of the original model, which is also the starting point of each chain.

B.6 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 firn 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 B2 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. B4. Posterior samples and the normal approximations are very similar, with correlations only slightly less well captured in the tails of the distributions. 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).

	Parameters	Posterior mean	Posterior covariance matrix
HL	$k_0^*, k_1^*, E_0, E_1, a, b$	$\begin{bmatrix}16.7, 649, 10760,\\ 21000, 0.88, 0.66\end{bmatrix}$	$\begin{bmatrix} 34.4 & 40.2 & 4500 & 324 & -0.0685 & -0.0195 \\ 40.2 & 44000 & 618 & 161000 & 1.087 & -3.670 \\ 4502 & 618 & 710000 & 7080 & -29.95 & 1.94 \\ 324 & 1610000 & 7080 & 694000 & 7.86 & -27.51 \\ -0.0685 & 1.087 & -29.95 & 7.86 & 0.0051 & -0.0012 \\ -0.0195 & -3.670 & 1.94 & -27.51 & -0.0012 & 0.0036 \end{bmatrix}$
Ar	$k_0^{Ar}, k_1^{Ar}, E_g, \alpha, \beta$	$\begin{bmatrix}0.080, 0.028, 40900,\\ 0.78, 0.69\end{bmatrix}$	$\begin{bmatrix} 5.62\times10^{-4} & 1.55\times10^{-4} & -12.66 & 9.65\times10^{-5} & -3.23\times10^{-4} \\ 1.55\times10^{-4} & 7.41\times10^{-5} & -4.64 & -2.04\times10^{-4} & 1.05\times10^{-4} \\ -12.66 & -4.64 & 360000 & 11.0 & 4.67 \\ 9.65\times10^{-5} & -2.04\times10^{-4} & 11.0 & 0.00330 & -0.00101 \\ -3.23\times10^{-4} & 1.05\times10^{-4} & 4.67 & -0.00101 & 0.00312 \end{bmatrix}$
LZ	$lz_a, lz_b, lz_{11}, lz_{12}, lz_{13}, lz_{21}, lz_{22}, lz_{23}$	$\begin{bmatrix}7.56, -2.091, -14.71,\\ 7.269, -1.019, -1.513,\\ 6.0203, -0.09127\end{bmatrix}$	$\begin{bmatrix} 5.27 & -0.198 & -1.20 & -1.68 & -0.0239 & 0.00553 & -0.0606 & 0.00413 \\ -0.198 & 0.0116 & 0.218 & -0.0612 & 0.0134 & -0.0158 & -0.00229 & -7.37\times10^{-4} \\ -1.20 & 0.218 & 14.6 & -3.96 & 0.801 & 0.368 & 0.354 & 0.0129 \\ -1.68 & -0.0612 & -3.96 & 13.3 & -0.309 & -0.0850 & 5.40 & 0.0166 \\ -0.0239 & 0.0134 & 0.801 & -0.309 & -0.0502 & -0.0173 & 0.0252 & -4.42\times10^{-4} \\ 0.00553 & -0.0158 & 0.368 & -0.0850 & -0.0173 & 0.446 & -0.429 & 0.0131 \\ -0.0606 & -0.00229 & 0.354 & 5.40 & 0.0252 & -0.429 & 3.94 & -2.59\times10^{-4} \\ 0.00413 & -7.37\times10^{-4} & 0.0129 & 0.0166 & -4.42\times10^{-4} & 0.0131 & -2.59\times10^{-4} & 4.80\times10^{-4} \end{bmatrix}$

Table B2. 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.

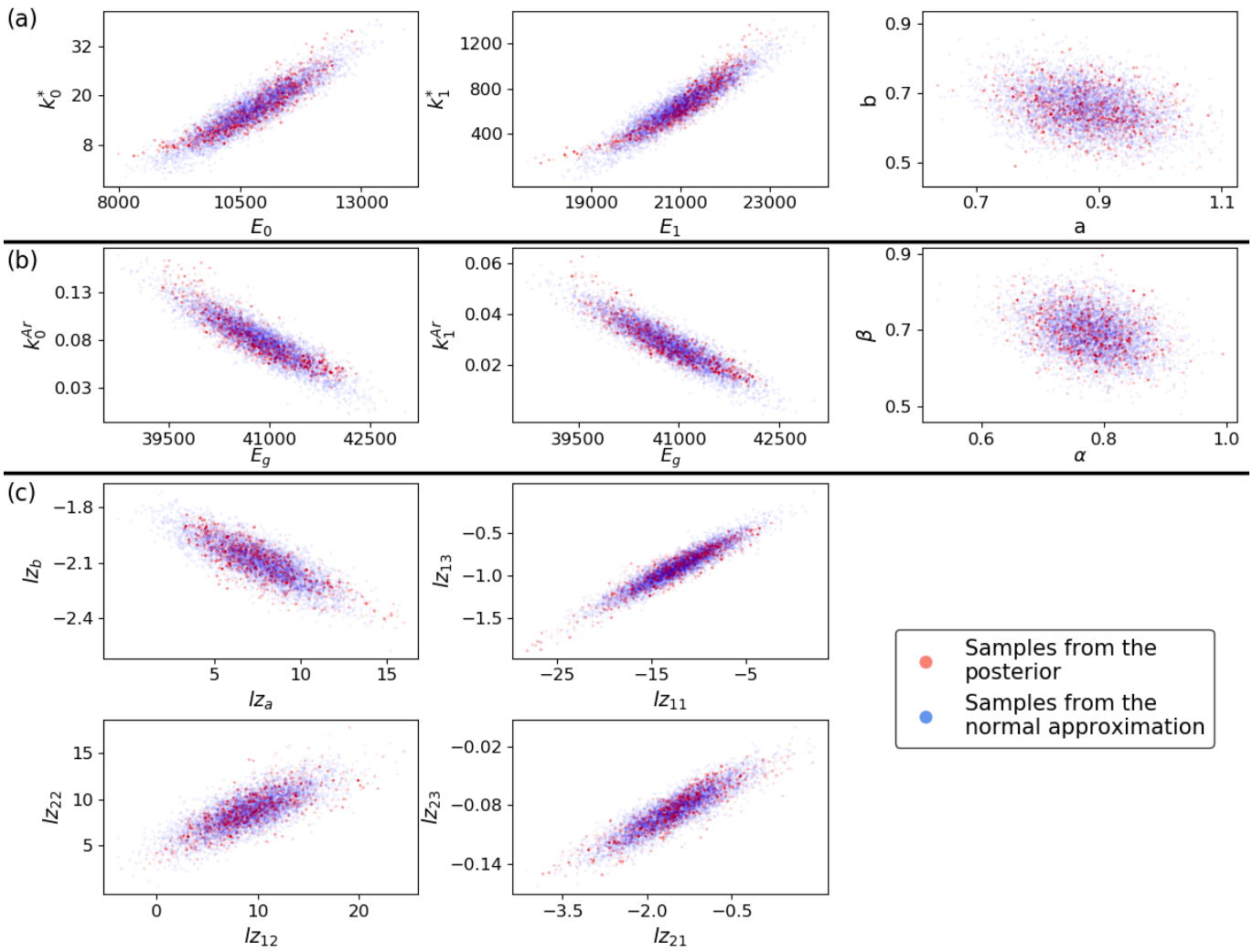


Figure B4. Evaluation of the normal approximations to the posterior distributions for (a) HL, (b) Ar, (c) LZ. Where possible, correlated parameters share a same graph.

B.7 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. B5. 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 $DIP15$ and $DIPpc$ respectively. Higher values

of a tend to cause lighter firn at 15 m depth. Lower E_0 values compensate for this effect on $DIP15$ because the shallow firn densifies faster due to its greater sensitivity to temperature variations. The lighter 15 m depth density also affects $DIPpc$, 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 firn to be buried rapidly, which may cause lower density firn governed by the fast stage-1 densification to extend below 15 m. To avoid underestimation of $DIPpc$ 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 $DIPpc$ 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. (2.6), 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. (2.6), 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.

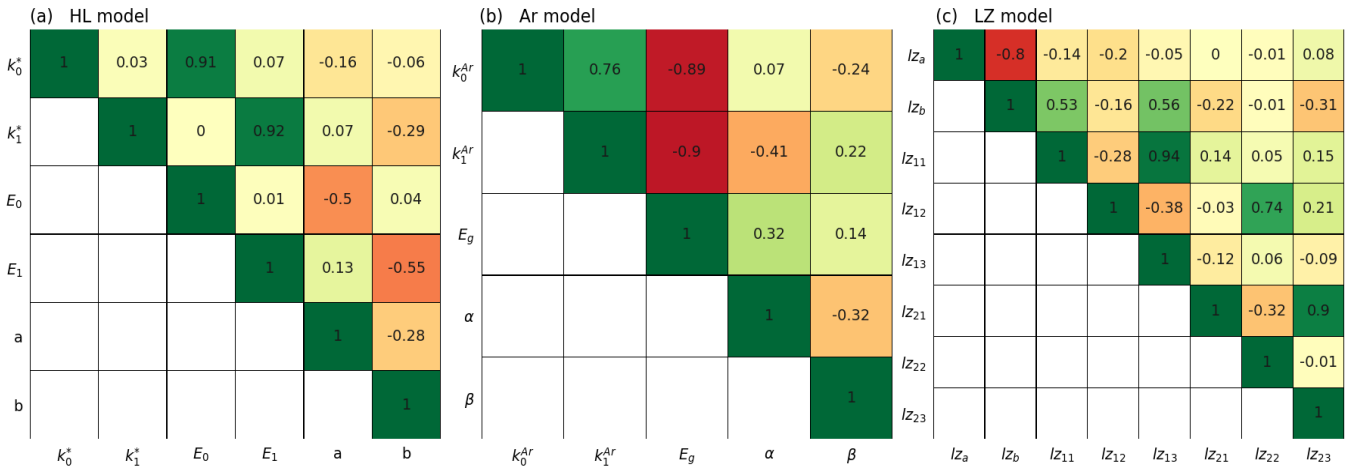


Figure B5. Posterior correlation matrices.

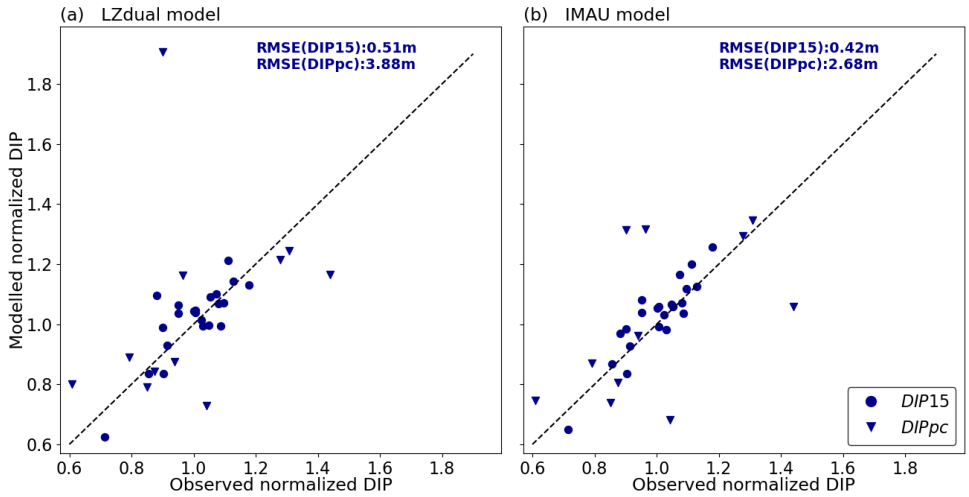


Figure B6. Comparison of evaluation data *DIP* with model results for the LZ dual and IMAU models.

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]	$DIPpc$ [m]	Var $DIPpc$ [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 B1. 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.

Appendix C

This appendix includes the Supplementary Information to Chapter 3. It includes additional details about the construction of the firn densification models emulators (C1), the validation of the emulators (C2) and the original formulations of the firn densification models (C3). It also includes a map that compares the ensemble mean and the spread in the ensemble results (Figure C4) and a map that compares the mean accumulation anomaly and the spread in accumulation anomaly from the three RCMs used in this study (RACMO2, MAR, HIRHAM) (Figure C5).

C.1 Emulator Construction

We first describe the construction of a single FDM emulator since all nine emulators for the nine FDMs are constructed in the same way. A FDM simulates firn compaction values, dh_c (m), and the emulator is a statistical approximation of the FDM computing estimations $\widehat{dh_c}$. Calibration of the emulator requires a calibration set, which is given by computations of the FDM at specific calibration sites of the EAIS. In order to select the calibration sites, we perform k-means clustering on the mean annual temperature and accumulation rate values from the RACMO2 model on the EAIS domain. By selecting the cluster centroids, our calibration sites span the entire range of climatic conditions on EAIS. We test the clustering with various number of clusters, which shows that the decrease in differences in mean climatic values between centroids becomes marginal for numbers of centroids greater than 50 (Fig. C1). As such, we keep the 50 centroids of the 50-means clustering as our calibration sites.

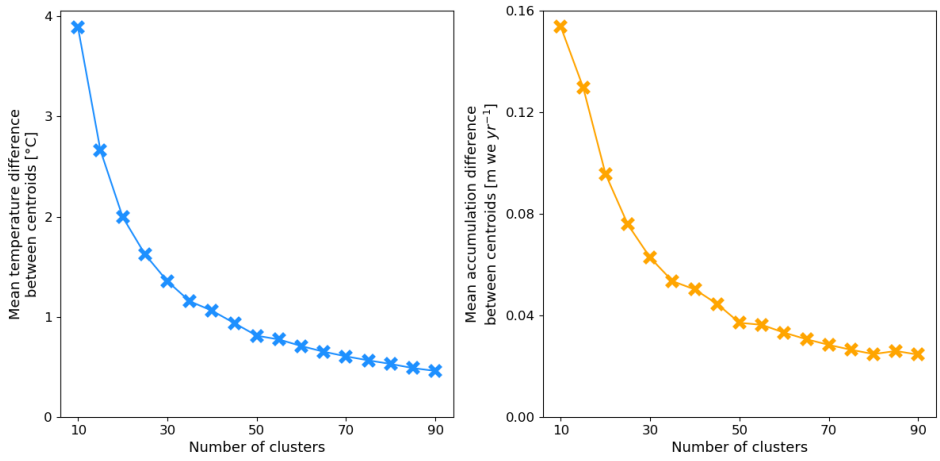


Figure C1. Mean differences in average surface temperature (left) and accumulation rate (right) between cluster centroids as a function of the number of clusters taken in EAIS.

At each calibration site, we compute entire time series of daily FDM computations over the 1979-2017 period. As such, the FDM requires 1979-2017 time-series of accumulation, surface temperature, melt rates and rainfall computed by RCMs as forcing inputs. We use the RACMO2 RCM for these

computations but this has no effect on the calibration and climatic sensitivity of the emulator. Indeed, climatic values depend much more on the calibration site than on the particular RCM, and the range of climatic values of the three RCMs at the locations of the 50 calibrations sites is small. Each FDM simulation proceeds to a spin-up until refreshment of 80 m w.e. of firn under the reference period (1979-2009) climate and we use the L11 parameterisation for ρ_0 (Ligtenberg et al., 2011). A simple bucket scheme is used to treat the small amounts of melt and rain at calibration sites at the coastal margins of EAIS. We use the Community Firn Model to perform the FDM simulations (Stevens et al., 2020).

The first step in the statistical emulation is to capture long term signals in the modelled dh_c. At each calibration site, the FDM 1979-2017 time series is fitted with a linear trend:

$$dh_c(t^*) \approx \beta_0 + \beta_1 t^* \quad \text{(C1)}$$

where t^* is the time centred about the mid-date of 1979-2017 (i.e. 1998). As such, the coefficients β_0 and β_1 capture the mean dh_c and the trend in dh_c respectively. For the purpose of estimating the coefficients at non-calibration sites, we perform linear regression of β_0 and β_1 on climatic values of the calibration sites:

$$\begin{cases} b_0 = a_{01} + a_{02}T_{av} + a_{03}\varphi_{av} + a_{04}M_{av} \\ b_1 = a_{11} + a_{12}T_{td} + a_{13}\varphi_{td} + a_{14}M_{td} \end{cases} \quad \text{(C2)}$$

where b_0 and b_1 are the estimates of β_0 and β_1 by linear regression. The coefficients a are the regression coefficients, the subscripts av and td stand for the average and the trend in the 1979-2017 values, T is surface temperature (°C), φ is rate of porosity gain (m yr^{-1}) (Eq. C3), and M is melt rate (m ice equivalent (i.e.) yr^{-1}). The use of φ instead of the accumulation rate $\overline{acc}$ (m i.e. yr^{-1}) is motivated by the differences in ρ_0 across EAIS, which leads to different inputs of pore space in the firn for same accumulation rates. Firn compaction is strongly sensitive to the total pore space in the firn column, making dh_c more sensitive to φ than to $\overline{acc}$.

$$\varphi = \frac{\rho_i - \rho_0}{\rho_0}\,\overline{acc} \quad \text{(C3)}$$

The regression coefficients a are common across sites but are specific to each of the 9 FDMs, since these have different climatic sensitivities. The respective fits of b_0 and b_1 to β_0 and β_1 across the calibration sites and for all FDMs explain more than 97% of their variances and show negligible biases (Fig. C2). This demonstrates that the long-term compaction signal modelled by FDMs can be estimated by linear regression on the long-term means and trends of the climatic values that force the FDMs.

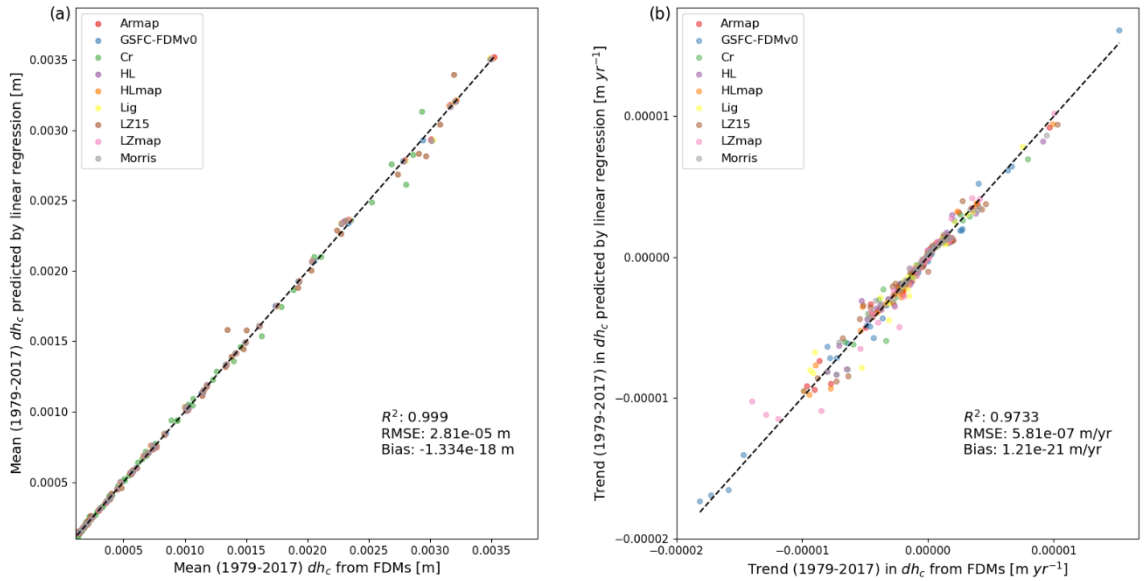


Figure C2. Fit (a) of the mean compaction values estimated by linear regression (b_0) to the mean compaction values computed by the FDMs (β_0) and (b) of the trend in compaction values estimated by linear regression (b_1) to the trend in compaction values computed by the FDMs (β_1).

We calculate FDM-specific uncertainties on the estimated linear signals in dh_c. First, we compute the sample error standard deviations, $\hat{\sigma}(\beta_0)$ and $\hat{\sigma}(\beta_1)$. In addition to these, we must account for the uncertainty associated with the estimation of b_0 and b_1. For this purpose, we compute the root mean square error (RMSE) between the 50 daily 1979-2017 dh_c time series reconstructed using the b coefficients and those reconstructed using the β coefficients. As expected, for each FDM, the RMSE values are larger at sites with strong compaction. By using the centred t^*, b_0 is an estimate of the mean dh_c and we compute this component of the uncertainty as the across sites mean ratio

$$\sigma_b^{ratio} = \frac{1}{50}\sum_{i=0}^{50}\frac{\mathrm{RMSE}_i}{b_{0,i}} \quad \text{(C4)}$$

As such, each FDM has its own value of σ_b^{ratio}. Multiplying the b_0 value estimated at a new site by σ_b^{ratio} provides an uncertainty estimate due to the estimations of b_0 and b_1. We sum this term linearly with the contributions from $\hat{\sigma}(\beta_0)$ and $\hat{\sigma}(\beta_1)$ to reach an uncertainty related to the linear fit for each daily $\widehat{dh_c}$ value at this new site.

FDM computed dh_c time series are non-linear, and a statistical emulator must go beyond the linear regression (Eq. C1) to simulate shorter term features such as seasonality and daily spikes in compaction. In order to capture short term patterns in dh_c, we complement the linear regression with Gaussian Process (GP) regression. The GP contribution to the emulator is to capture deviations away from the estimated linear trend. The statistical emulator is thus expressed as:

$$\widehat{dh_c(t^*)} = b_0 + b_1 t^* + \varepsilon(t^*) \quad \text{(C5)}$$

where the terms $\varepsilon(t^*)$ are the stochastic terms simulated by the GP. As input, the GP uses a vector of predictor variables $\boldsymbol{x}(t^*)$, which we construct to account for different time spans of the short-term

features in dh_c. As for the linear regression, the predictor variables are entirely determined from the RCM climatic values. We consider as predictors the daily values as well as the weekly, monthly and yearly rolling averages of the T and φ variables. We discard the melt rate input for the GP calibration because melt rates are mostly negligible across the EAIS, and because their association with surface temperatures close to the melting point still allows the emulator to capture melt-related patterns in dh_c. Since the GP must capture deviations from the trend in dh_c, the T and φ values used as its inputs are detrended. The GP regression is most efficient when the target variable, dh_c in this case, approximately follows a Gaussian distribution. However, dh_c values are always positive and are mostly small during colder periods with some peak values in summer. This motivates a log transformation of dh_c such that $\log(dh_c)$ can be modelled as a variable with Gaussian errors (Rios and Tobar, 2018) and then be converted back for the results analysis. The stochastic term ε in Eq. (C5) is modelled as a GP with a 0 prior mean because it only captures deviations from the trend in dh_c.

$$\varepsilon \sim GP(0, K(\boldsymbol{x}, \boldsymbol{x}')) \quad \text{(C6)}$$

For the covariance function $K(x, x')$, we choose the commonly used squared exponential covariance function:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp[(\boldsymbol{x} - \boldsymbol{x}')^T D^{-1} (\boldsymbol{x} - \boldsymbol{x}')] + \nu\, \delta(\boldsymbol{x}, \boldsymbol{x}') \quad \text{(C7)}$$

where σ_f^2, D and ν are the hyperparameters of the covariance function and $\delta(\boldsymbol{x}, \boldsymbol{x}')$ is the Dirac delta function. The choice of this covariance function form assumes a certain degree of smoothness in the response with respect to the predictor variables, which is motivated by the log transformation. The diagonal matrix D is formed of the length scales for every predictor, σ_f^2 is the variance and ν is a noise term, also called the nugget. While ν is often omitted when emulating deterministic models such as FDMs, it brings significant advantages such as numerical stability and prevention of overfitting. Moreover, the predictor variables used by the GP are only a subset of the full RCM time series used by the FDM and ν can account for the variability related to inactive inputs (Andrianakis and Challenor, 2012). The hyperparameters are fixed by maximum likelihood estimation with respect to the calibration data (Rasmussen and Williams, 2006). We selected this covariance function by assessing the emulator performance with different possible function choices in a leave-one-out cross validation framework (see Section S2). We use the software package GPy for the GP regression.

Every daily time step of the 50 calibration FDM simulations over the 1979-2017 period provides a potential training point for the GP regression. However, for n training points, the GP calibration time has a scaling of $O(n^3)$ and the memory requirements have a scaling of $O(n^2)$ (Quiñonero-Candela et al., 2007). Using the entirety of the possible training points for the GP calibration is thus computationally impractical. It would also be highly redundant because many daily steps share similar predictor $\boldsymbol{x}(t^*)$ and response $dh_c(t^*)$ values, both within a time series and across time series of

different sites. Furthermore, climatic conditions vary strongly across EAIS, and dh_c emulation at a given new site does not require information from all the 50 calibration sites. As such, for any new site, we select relevant calibration sites based on their climatic similarities. Since b_0 is calculated at any site and represents its mean dh_c estimated from the long term climate, we select the two calibration sites with the b_0 values closest to the new site b_0. By doing so, the number of calibration sites used for the GP calibration is reduced from 50 to the 2 sites deemed most relevant to emulation of daily dh_c values at a given new site. In order to address the redundancy of information in the two selected FDM time series, we proceed to sparse GP regression by following the Deterministic Training Conditional approximation (Seeger et al., 2003; Quiñonero-Candela et al., 2007; Liu et al., 2018). Sparse GP regression essentially approximates the full GP regression by selecting a smaller set of inducing inputs and cutting the computational burden of a high number of daily values in the two calibration 1979-2017 time series. The emulator allows a computational speedup of two orders of magnitude compared to the original FDMs.

C.2 Emulator validation

To evaluate the predictive accuracy of the emulator (Eq. C5), we use a leave-one-out cross-validation (LOOCV) technique. In total, each FDM performs simulations at the 50 calibration sites. The LOOCV framework consists in leaving out each site in turn from the emulator calibration and predict the daily dh_c time series at the left-out site. As such, every emulator prediction can only use the results of the FDM at the 49 other sites for calibration. From the emulated dh_c time series, we compute the total firn compaction over the 1979-2017 period and compare these values with the ones computed by the FDM at the corresponding site. This process is repeated for each of the nine FDMs and executed with the RACMO2 forcing. Again, the choice of the RCM for this evaluation has no impact, since the emulator does not retrieve any information from the RCM used to compute the forcing climatic time series, and climatic differences across sites largely cover climatic differences across RCMs. We compute several statistics from the LOOCV experiment to assess the performance of the emulator (Fig. C3a). The coefficient of determination (R^2) shows that the emulator captures more than 99% of the variance in the FDM simulated total dh_c values. The RMSE is of 0.49 m, which corresponds to only 3.5% of the mean total compaction values of the LOOCV experiment and the small bias (+0.04 m) demonstrates that the emulator does not over- or underestimate the densification rates of the FDMs. Finally, the emulator reproduces low and high total compaction equally well, with confidence intervals that are understandably larger for larger magnitudes of compaction. The emulator is conservative in its prediction uncertainties because 99% of the FDM computed values fall within their corresponding 95% confidence intervals computed by the emulator. The performance of the emulator is slightly worse for the LZ15 (RMSE = 0.92 m) and Cr (RMSE = 0.74 m) models with respect to the other FDMs. We attribute this to the very strong sensitivity of

LZ15 to daily spikes in temperature, and to the strongly different formulation of Cr. The latter is an alpine snow model with significantly different governing equations for firn densification than the eight other FDMs (see Section S3). It is nevertheless regularly used to simulate firn compaction over ice sheets (e.g. van Kampenhout et al., 2017; Agosta et al., 2019). The emulator performance on total 1979-2017 dh_c values is essentially governed by the accuracy of the linear trend estimations. In order to evaluate the ability of the emulator on a shorter timescale and thus the GP contribution, we also show the match between all daily emulated dh_c values from the 1979-2017 LOOCV experiment and the daily FDMs-computed values (Fig. C3b). Furthermore, we perform the same LOOCV experiment on only the last year of the period (i.e. 2016-2017). The results shown in Figure C3c confirm the predictive accuracy of the emulator over this single year. We tested this LOOCV technique with three other possibilities for the covariance function of the GP (Eq. C7): the exponential, the Matérn(3/2) and the Matérn(5/2) functions. None of these showed an improvement in RMSE, bias and confidence interval coverage.

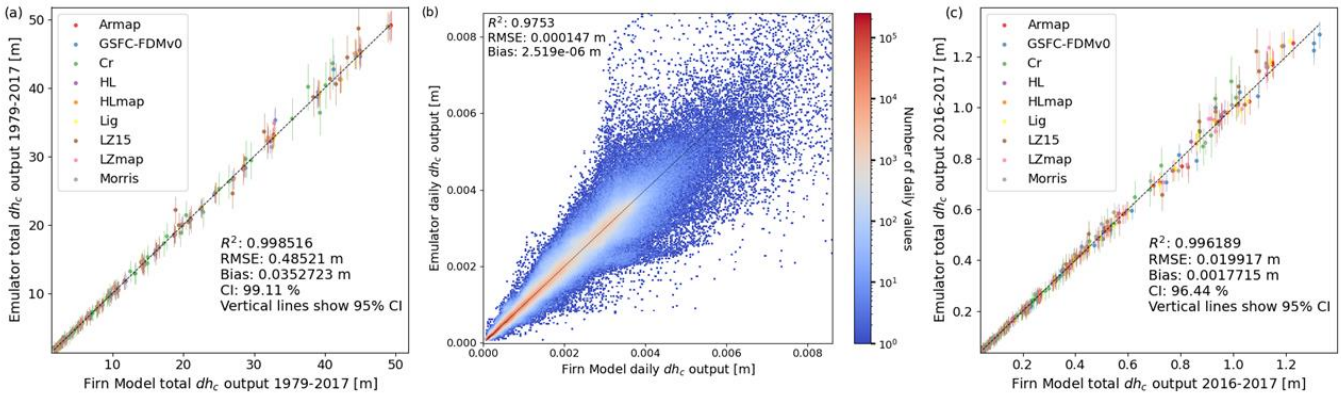


Figure C3. Results of the leave-one-out cross validation experiment. Fit (a) of the total (1979-2017) compaction values emulated to the values computed by the FDMs. Fit (b) of each daily compaction value of 1979-2017 emulated to the corresponding value computed by the FDMs. Note the logarithmic colour scale in (b). And (c) is same as (a) but computed only over the last 2016-2017 year.

C.3 Densification equations of the firn densification models

Eight FDMs (Armap, GSFC-FDMv0, HL, HLmap, Lig, LZ15, LZmap, Morris) use Eq. (C8) as governing formulation for densification.

$$\begin{cases} \frac{d\rho}{dt} = c_0\,(\rho_i - \rho), & \rho \leq 550\ kg\ m^{-3} \\ \frac{d\rho}{dt} = c_1\,(\rho_i - \rho), & \rho > 550\ kg\ m^{-3} \end{cases} \quad \text{(C8)}$$

with ρ the firn density (kg m^{-3}), ρ_i ice density (917 kg m^{-3}) and t time (s). They vary by the parameterisation of the coefficients c_0 and c_1.

Armap and GSFC-FDMv0

$$\begin{cases} c_0 = \rho_w \, \dot{b}^{\alpha} \, k_0^{Ar} \, g \exp\left(\frac{-E_{c0}}{RT} + \frac{E_g}{RT_{av}}\right) \\ c_1 = \rho_w \, \dot{b}^{\beta} \, k_1^{Ar} \, g \exp\left(\frac{-E_{c1}}{RT} + \frac{E_g}{RT_{av}}\right) \end{cases} \quad \text{(C9)}$$

where $\dot{b}$ is the mean accumulation rate over the lifetime of each specific firn layer (m w.e. yr^{-1}), T the firn temperature (K), T_{av} the annual mean temperature (K), R the gas constant (8.314 J K^{-1} mol^{-1}), g gravity (9.81 m s^{-2}) and ρ_w water density (1000 kg m^{-3}). The remaining terms are calibration parameters. For Armap, α=0.80, β=0.68, k_0^{Ar}= 0.077 m w.e.$^{-\alpha}$, k_1^{Ar}= 0.025 m w.e.$^{-\beta}$, E_{c0}=E_{c1}= 60 000 J mol^{-1}, E_g= 40 900 J mol^{-1} (Arthern et al., 2010; Verjans et al., 2020). For GSFC-FDMv0, α=0.8517, β=0.6490, k_0^{Ar}= 0.07 m w.e.$^{-\alpha}$, k_1^{Ar}= 0.03 m w.e.$^{-\beta}$, E_{c0}= 59 269 J mol^{-1}, E_{c1}= 57 139 J mol^{-1}, E_g= 42 400 J mol^{-1} (Arthern et al., 2010; Smith et al., 2020).

HL and HLmap

$$\begin{cases} c_0 = \dot{b}^{a} \, k_0^{*} \exp\left(\frac{-E_0}{RT}\right) \\ c_1 = \dot{b}^{b} \, k_1^{*} \exp\left(\frac{-E_1}{RT}\right) \end{cases} \quad \text{(C10)}$$

For HL, a= 1, b= 0.5, k_0^*= 11 m w.e.$^{-a}$, k_1^*= 575 m w.e.$^{-b}$, E_0= 10 160 J mol^{-1}, E_1= 21 400 J mol^{-1} (Herron and Langway, 1980). For HLmap, a= 0.90, b= 0.64, k_0^*= 16.3 m w.e.$^{-a}$, k_1^*= 627 m w.e.$^{-b}$, E_0= 10 790 J mol^{-1}, E_1= 21 100 J mol^{-1} (Herron and Langway, 1980; Verjans et al., 2020).

LZ15 and LZmap

$$\begin{cases} c_0 = \beta_0 \, lz_a (273.15 - T)^{lz_b} \, \dot{b} \\ c_1 = \beta_1 \, lz_a (273.15 - T)^{lz_b} \, \dot{b} \end{cases} \quad \text{(C11)}$$

For LZ15 (Li and Zwally, 2015), lz_a= 8.36, lz_b= -2.061 and

$$\begin{cases} \beta_0 = -1.218 - 0.403 \, T_{av} \\ \beta_1 = \beta_0 \left(0.792 - 1.080 \, \dot{b} + 0.00465 \, T_{av}\right)^{-1} \end{cases} \quad \text{(C12)}$$

For LZmap (Li and Zwally, 2011; Verjans et al., 2020), lz_a= 7.31, lz_b= -2.124 and

$$\begin{cases} \beta_0 = -14.710 + 7.269 \, \dot{b} - 1.019 \, T_{av} \\ \beta_1 = \beta_0 \left(-1.513 + 6.0203 \, \dot{b} - 0.09127 \, T_{av}\right)^{-1} \end{cases} \quad \text{(C13)}$$

For both LZ15 and LZmap, the term $(273.15 - T)$ is set to a maximum value of -10 K to avoid numerical instabilities and singularities in Eq. (C11).

Lig

$$\begin{cases} c_0 = \left[1.435 - 0.151 \ln(\dot{b})\right] \rho_w \, \dot{b} \, k_0 \, g \exp\left(\frac{-E_c}{RT} + \frac{E_g}{RT_{av}}\right) \\ c_1 = \left[2.366 - 0.293 \ln(\dot{b})\right] \rho_w \, \dot{b} \, k_1 \, g \exp\left(\frac{-E_c}{RT} + \frac{E_g}{RT_{av}}\right) \end{cases} \quad \text{(C14)}$$

with k_0= 0.07 m w.e., k_1= 0.03 m w.e., E_c= 60 000 J mol^{-1}, E_g= 42 400 J mol^{-1} (Arthern et al., 2010; Ligtenberg et al., 2011).

Morris

$$\begin{cases} c_0 = \frac{k_0^M}{\rho_w\, g\, \rho}(1 - 3.3\, m)\frac{1}{H(\tau)}\, \exp\left(\frac{-E_\alpha}{RT}\right)\sigma \\ c_1 = \dot{b}^b\, k_1^*\, \exp\left(\frac{-E_1}{RT}\right) \end{cases} \quad \text{(C15)}$$

with b, k_1^*, E_1 taken from the HL model, with τ the age of a specific firn layer (s), σ being the weight (N) exerted by the firn overlying a specific firn layer and m is the normalized deviation of ρ to a quadratic curve fit to the vertical profile of ρ. The parameter E_α is assigned a value of 60 000 J mol^{-1}, which is one of the values suggested by Morris and Wingham (2014). The remaining terms, k_0^M and $H(\tau)$ are functions of T_{av} and of the firn temperature history respectively, and as defined in Eqs. (C16) and (C17).

$$k_0^M = k_0^*\, \exp\left(\frac{-\left(E_0 + (0.05606\, T_{av} - 15.0205)\times 10^3\right)}{RT_{av}}\right) \quad \text{(C16)}$$

$$H(\tau) = \int_0^\tau \exp\left(-\frac{E_\alpha}{R\, T(\tau')}\right) d\tau' \quad \text{(C17)}$$

with k_0^*, E_0 taken from the HL model (Morris and Wingham, 2014).

Finally, the Cr model (Vionnet et al., 2012; van Kampenhout et al., 2017) does not use Eq. (C8). Densification rates are given by Eq. (C18).

$$\frac{d\rho}{dt} = \rho\, \frac{\sigma}{\eta} \quad \text{(C18)}$$

$$\eta = f_1 f_2 \eta_0 \frac{\rho}{c_\eta} \exp\left[a_\eta(T_0 - T) + b_\eta \rho\right] \quad \text{(C19)}$$

with η_0= 7.62237 kg s^{-1} m^{-1}, a_η= 0.1 K^{-1}, b_η= 0.023 m^3 kg^{-1}, c_η= 358 kg m^{-3}, f_2= 4 and

$$f_1 = \frac{1}{1 + 60\,\theta} \quad \text{(C20)}$$

with θ being the volumetric liquid water content of a specific firn layer.

Appendix D. Official statements of authorship

Chapter 1
Verjans, V., Leeson, A., Stevens, C. M., 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, 1819-1842, https://doi.org/10.5194/tc-13-1819-2019, 2019.

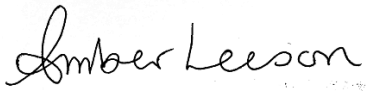
Vincent Verjans is the lead author of this publication. Vincent Verjans and Amber Leeson conceived the study. Vincent Verjans developed the new model schemes, performed model experiments, analysed the results, led writing the manuscript, submitted the paper and did the revisions required by the peer-review process.

All authors have signed below to confirm this.

Vincent Verjans

List of co-authors:

Amber Leeson:



Max Stevens:

Michael MacFerrin:

Brice Noël:

Michiel van den Broeke:

Chapter 2
Verjans, V., Leeson, A. A., Nemeth, C., Stevens, C. M., Kuipers Munneke, P., Noël, B., and van Wessem, J. M.: Bayesian calibration of firn densification models, The Cryosphere, 14, 3017–3032, https://doi.org/10.5194/tc-14-3017-2020, 2020.

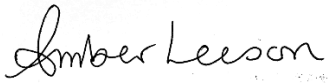
Vincent Verjans is the lead author of this publication. Vincent Verjans, Amber Leeson and Christopher Nemeth conceived the study. Vincent Verjans developed the statistical calibration method, performed model experiments, analysed the results, led writing the manuscript, submitted the paper and did the revisions required by the peer-review process.

All authors have signed below to confirm this.

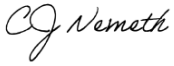
Vincent Verjans

List of co-authors:

Amber Leeson:

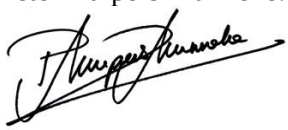


Christopher Nemeth:



Max Stevens:

Peter Kuipers Munneke:



Brice Noël:

Melchior van Wessem:

Publication III
Verjans, V., Leeson, A. A., McMillan, M., Stevens, C. M., van Wessem, J. M., van de Berg, W. J., van den Broeke, M. R., Kittel, C., Amory, C., Fettweis, X., Hansen, N., Boberg, F., and Mottram, R.: Uncertainty in East Antarctic firn thickness constrained using a model ensemble approach, Geophysical Research Letters, doi: 10.1029/2020GL092060, 2021.

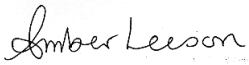
Vincent Verjans is the lead author of this publication. Vincent Verjans, Amber Leeson and Malcolm McMillan conceived the study. Vincent Verjans developed the statistical emulator, developed the model ensemble framework, performed model experiments, analysed the results, led writing the manuscript, submitted the paper and did the revisions required by the peer-review process.

All authors have signed below to confirm this.

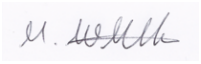
Vincent Verjans

List of co-authors:

Amber Leeson:



Malcolm McMillan:



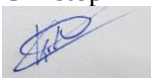
Max Stevens:

Melchior van Wessem:

Willem Jan van de Berg:

Michiel van den Broeke:

Christoph Kittel:



Charles Amory:

Xavier Fettweis:

Nicolaj Hansen:

Fredrik Boberg:

Ruth Mottram: