

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

V. Verjans¹, A. A. Leeson¹, M. McMillan¹, C. M. Stevens^{2,3,4}, J. M. van Wessem⁵, W. J. van de Berg⁵, M. R. van den Broeke⁵, C. Kittel⁶, C. Amory^{6,7}, X. Fettweis⁶, N. Hansen^{8,9}, F. Boberg⁸ and R. Mottram⁸

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

²Department of Earth and Space Sciences, University of Washington, Seattle, WA, USA

³Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

⁴Department NASA Goddard Space Flight Center, Greenbelt, MD, USA

⁵Institute for Marine and Atmospheric research Utrecht, Utrecht University, Utrecht, the Netherlands

⁶Laboratory of Climatology, Department of Geography, SPHERES research unit, University of Liège, Liège, Belgium

⁷Univ. Grenoble Alpes, CNRS, Institut des Géosciences de l'Environnement, Grenoble, France

⁸Danish Meteorological Institute, Lyngbyvej 100, Copenhagen, 2100, Denmark

⁹DTU Space, Department of Geodynamics, Technical University of Denmark, Kongens Lyngby, Denmark

Contents of this file

Text S1 to S3

Figures S1 to S5

Introduction

The Supplementary information includes additional details about the construction of the firn densification models emulators (S1), the validation of the emulators (S2) and the original formulations of the firn densification models (S3). It also includes a map that compares the ensemble mean and the spread in the ensemble results (Figure S4) 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 S5).

Text S1. 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 (Figure S1). As such, we keep the 50 centroids of the 50-means clustering as our calibration sites.

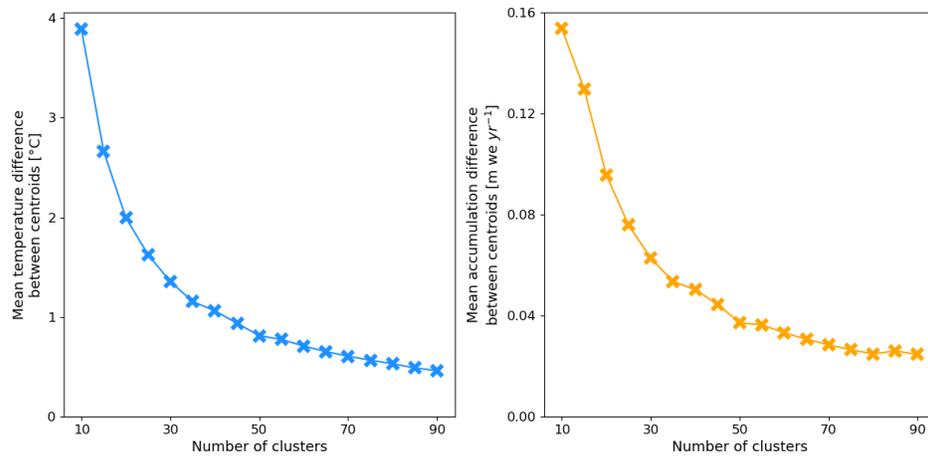


Figure S1: Mean differences in average surface temperature (left) and accumulation rate (right) between cluster centroids as a function of the number of clusters taken in EAIS.

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 firm 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:

$$dh_c(t^*) \approx \beta_0 + \beta_1 t^* \quad (1)$$

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:

$$\begin{cases} b_0 = a_{01} + a_{02}T_{av} + a_{03}\varphi_{av} + a_{04}M_{av} \\ b_1 = a_{11} + a_{12}T_{td} + a_{13}\varphi_{td} + a_{14}M_{td} \end{cases} \quad (2)$$

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 ($^{\circ}\text{C}$), φ is rate of porosity gain (m yr^{-1}) (Eq. 3), and M is melt rate ($\text{m ice equivalent (i.e.) yr}^{-1}$). 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} .

$$\varphi = \frac{\rho_i - \rho_0}{\rho_0} \overline{acc} \quad (3)$$

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 (Figure S2). 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.

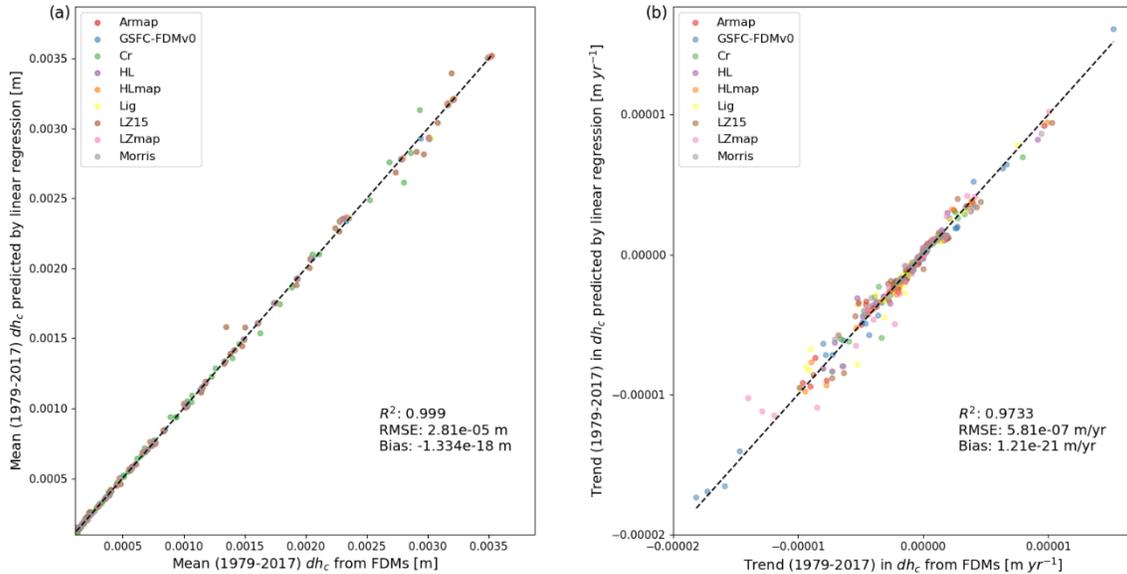


Figure S2: Fit (a) of the mean compaction values estimated by linear regression (b_0) to the mean compaction values computed by the FDMs (β_0) and (b) of the trend in compaction values estimated by linear regression (b_1) to the trend in compaction values computed by the FDMs (β_1).

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

$$\sigma_b^{ratio} = \frac{1}{50} \sum_{i=0}^{50} \frac{RMSE_i}{b_{0,i}} \quad (4)$$

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. 1) 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:

$$\widehat{dh}_c(t^*) = b_0 + b_1 t^* + \varepsilon(t^*) \quad (5)$$

where the terms $\varepsilon(t^*)$ are the stochastic terms simulated by the GP. As input, the GP uses a vector of predictor variables $\mathbf{x}(t^*)$, 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. (5) is modelled as a GP with a 0 prior mean because it only captures deviations from the trend in dh_c .

$$\varepsilon \sim GP(0, K(\mathbf{x}, \mathbf{x}')) \quad (6)$$

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

$$K(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp[(\mathbf{x} - \mathbf{x}')^T D^{-1} (\mathbf{x} - \mathbf{x}')] + \nu \delta(\mathbf{x}, \mathbf{x}') \quad (7)$$

where σ_f^2 , D and ν are the hyperparameters of the covariance function and $\delta(\mathbf{x}, \mathbf{x}')$ 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 $x(t^*)$ 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.

Text S2. Emulator validation

To evaluate the predictive accuracy of the emulator (Eq. 5), 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 (Figure S3a).

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 low 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 (Figure S3b). Furthermore, we perform the same LOOCV experiment on only the last year of the period (i.e. 2016-2017). The results shown in Figure S3c 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. 7): 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.

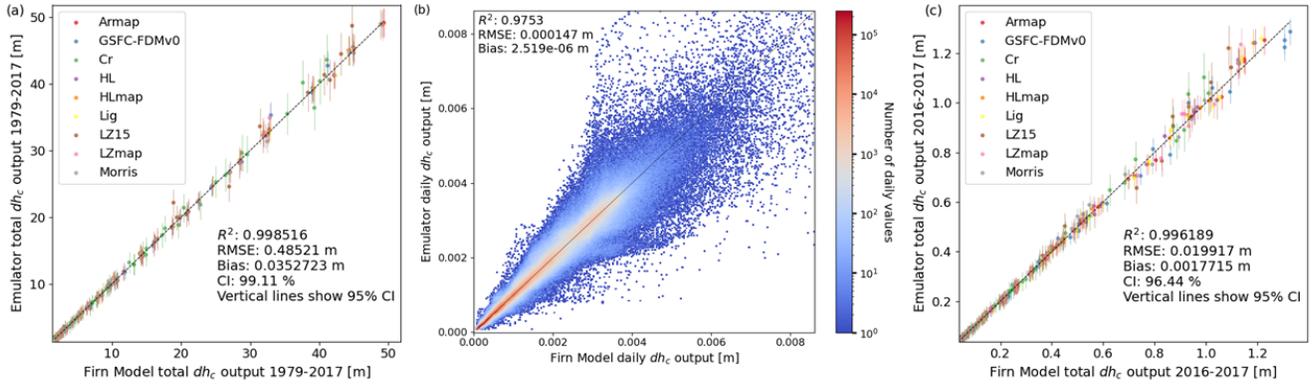


Figure S3: Results of the leave-one-out cross validation experiment. Fit (a) of the total (1979-2017) compaction values emulated to the values computed by the FDMs. Fit (b) of each daily compaction value of 1979-2017 emulated to the corresponding value computed by the FDMs. Note the logarithmic colour scale in (b). And (c) is same as (a) but computed only over the last 2016-2017 year.

Text S3. Densification equations of the firn densification models

Eight FDMs (Armap, GSFC-FDMv0, HL, HLmap, Lig, LZ15, LZmap, Morris) use Eq. (8) as governing formulation for densification.

$$\begin{cases} \frac{d\rho}{dt} = c_0 (\rho_i - \rho), & \rho \leq 550 \text{ kg m}^{-3} \\ \frac{d\rho}{dt} = c_1 (\rho_i - \rho), & \rho > 550 \text{ kg m}^{-3} \end{cases} \quad (8)$$

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

$$\begin{cases} c_0 = \rho_w \dot{b}^\alpha k_0^{Ar} g \exp\left(\frac{-E_{c0}}{RT} + \frac{E_g}{RT_{av}}\right) \\ c_1 = \rho_w \dot{b}^\beta k_1^{Ar} g \exp\left(\frac{-E_{c1}}{RT} + \frac{E_g}{RT_{av}}\right) \end{cases} \quad (9)$$

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, $\alpha=0.80$, $\beta=0.68$, $k_0^{Ar}=0.077$ m w.e.^{- α} , $k_1^{Ar}=0.025$ m w.e.^{- β} , $E_{c0}=E_{c1}=60\,000$ J mol⁻¹, $E_g=40\,900$ J mol⁻¹ (Arthern et al., 2010; Verjans et al., 2020). For GSFC-FDMv0, $\alpha=0.8517$, $\beta=0.6490$, $k_0^{Ar}=0.07$ m w.e.^{- α} , $k_1^{Ar}=0.03$ m w.e.^{- β} , $E_{c0}=59\,269$ 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

$$\begin{cases} c_0 = \dot{b}^a k_0^* \exp\left(\frac{-E_0}{RT}\right) \\ c_1 = \dot{b}^b k_1^* \exp\left(\frac{-E_1}{RT}\right) \end{cases} \quad (10)$$

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

$$\begin{cases} c_0 = \beta_0 l_{z_a} (273.15 - T)^{l_{z_b}} \dot{b} \\ c_1 = \beta_1 l_{z_a} (273.15 - T)^{l_{z_b}} \dot{b} \end{cases} \quad (11)$$

For LZ15 (Li and Zwally, 2015), $l_{z_a}=8.36$, $l_{z_b}=-2.061$ and

$$\begin{cases} \beta_0 = -1.218 - 0.403 T_{av} \\ \beta_1 = \beta_0 (0.792 - 1.080 \dot{b} + 0.00465 T_{av})^{-1} \end{cases} \quad (12)$$

For LZmap (Li and Zwally, 2011; Verjans et al., 2020), $l_{z_a}=7.31$, $l_{z_b}=-2.124$ and

$$\begin{cases} \beta_0 = -14.710 + 7.269 \dot{b} - 1.019 T_{av} \\ \beta_1 = \beta_0 (-1.513 + 6.0203 \dot{b} - 0.09127 T_{av})^{-1} \end{cases} \quad (13)$$

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. (11).

Lig

$$\begin{cases} c_0 = [1.435 - 0.151 \ln(\dot{b})] \rho_w \dot{b} k_0 g \exp\left(\frac{-E_c}{RT} + \frac{E_g}{RT_{av}}\right) \\ c_1 = [2.366 - 0.293 \ln(\dot{b})] \rho_w \dot{b} k_1 g \exp\left(\frac{-E_c}{RT} + \frac{E_g}{RT_{av}}\right) \end{cases} \quad (14)$$

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

$$\begin{cases} c_0 = \frac{k_0^M}{\rho_w g \rho} (1 - 3.3 m) \frac{1}{H(\tau)} \exp\left(\frac{-E_\alpha}{RT}\right) \sigma \\ c_1 = \dot{b}^b k_1^* \exp\left(\frac{-E_1}{RT}\right) \end{cases} \quad (15)$$

with b , k_1^* , E_1 taken from the HL model, with τ the age of a specific firm layer (s), σ being the weight (N) exerted by the firm overlying a specific firm 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 firm temperature history respectively, and as defined in Eqs. (16) and (17).

$$k_0^M = k_0^* \exp\left(\frac{-(E_0 + (0.05606 T_{av} - 15.0205) \times 10^3)}{RT_{av}}\right) \quad (16)$$

$$H(\tau) = \int_0^\tau \exp\left(-\frac{E_\alpha}{RT(\tau')}\right) d\tau' \quad (17)$$

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. (8). Densification rates are given by Eq. (18).

$$\frac{d\rho}{dt} = \rho \frac{\sigma}{\eta} \quad (18)$$

$$\eta = f_1 f_2 \eta_0 \frac{\rho}{c_\eta} \exp[a_\eta(T_0 - T) + b_\eta \rho] \quad (19)$$

with $\eta_0 = 7.62237 \text{ kg s}^{-1} \text{ m}^{-1}$, $a_\eta = 0.1 \text{ K}^{-1}$, $b_\eta = 0.023 \text{ m}^3 \text{ kg}^{-1}$, $c_\eta = 358 \text{ kg m}^{-3}$, $f_2 = 4$ and

$$f_1 = \frac{1}{1 + 60 \theta} \quad (20)$$

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

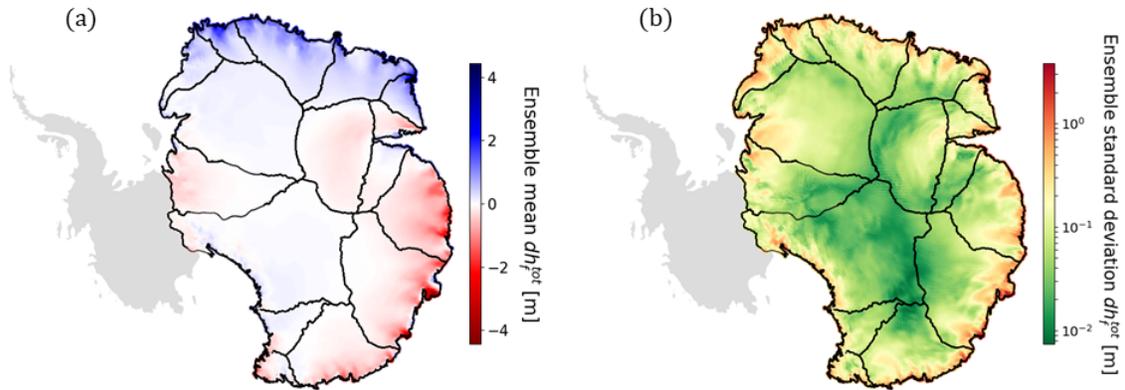


Figure S4: (a) Ensemble mean 1992-2017 firn thickness change (dh_f^{tot}) (same as Figure 1 of main text) and (b) ensemble standard deviation in 1992-2017 firn thickness change.

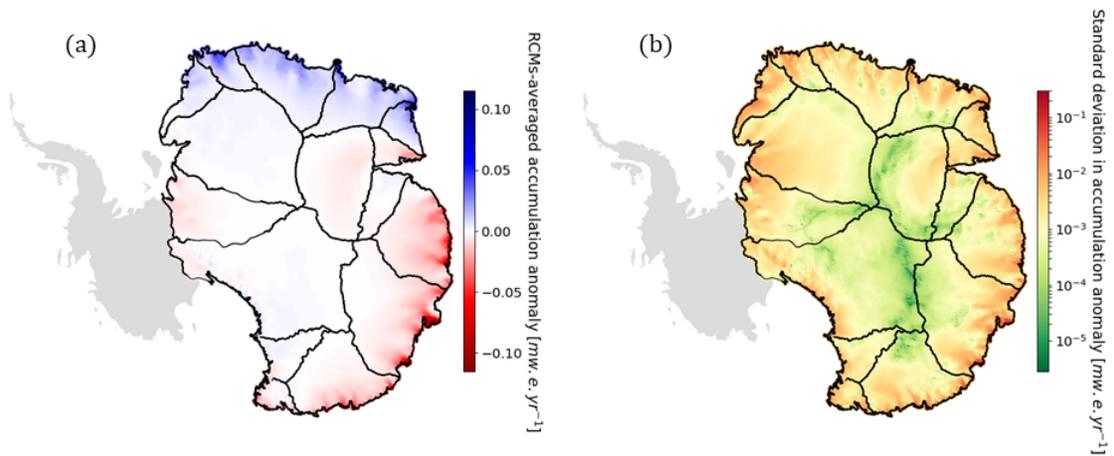


Figure S5: (a) Average of the 1992-2017 annual mean accumulation anomalies of the three RCMs (RACMO2, MAR, HIRHAM) and (b) standard deviation in the 1992-2017 annual mean accumulation anomalies.