https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

Predicting meltwater ponding on Antarctic ice shelves using a firn model emulator

Sanne Veldhuijsen[1], Peter Kuipers Munneke[1], Willem Jan van de Berg[1], Michiel van den Broeke[1], and Luke Trusel[2]

[1]Institute for Marine and Atmospheric research, Utrecht University, PO Box 80125, 3508 TC Utrecht, The Netherlands
[2]Department of Geography, Penn State University, University Park, PA, United States

Correspondence: Peter Kuipers Munneke (p.kuipersmunneke@uu.nl)

Abstract. Depletion of firn air can lead to meltwater ponding. If conditions are unfavourable, it can cause hydrofracturing and disintegration of Antarctic ice shelves. Projections of firn air content can therefore be used in ice-sheet model simulations to prescribe which ice-shelf sections become vulnerable to collapse, and when. However, using a full firn model for simulating firn air content for a wide range of climate forcings is computationally expensive. We therefore developed an XGBoost emulator for firn air content, a fast machine-learning model trained to reproduce the firn air content from a full firn densification model. The emulator needs annual means of snowfall, surface melt, rain, air temperature, and wind speed as input. Using scenario and spatial blocking cross-evaluation, we find that the emulator predicts > 94% of the variance in the firn air content of the full firn densification model. When a simulated firn air content of less than 2 m is assumed to represent the onset of melt ponding, the emulator shows 6-7% false positives and 14-17% false negatives. We demonstrate that surface melt is an important input variable, by constructing an alternative emulator that does not require surface melt. This emulator explains > 82% of the variance in firn air content, and has poorer agreement with firn-model simulated melt pond locations. Including surface melt results in a better emulator, because the non-linear behaviour of snow surface albedo is hard to predict from atmospheric variables only.

1 Surface melt and ice-shelf instability

Ice shelves are important for the mass balance of the Antarctic ice sheet, as they buttress the discharge of grounded ice through drag forces at their contact with walls and pinning points (Goldberg et al., 2009). The Antarctic Ice Sheet is losing mass primarily due to increased ice discharge into ice shelves, especially in the Amundsen Sea Embayment (Smith et al., 2020). This enhanced ice discharge is most likely caused by stronger basal melt beneath ice shelves, which thins and damages the ice shelves and reduces ice-shelf buttressing (Thoma et al., 2008; Paolo et al., 2015). Ice shelves can also be vulnerable to atmospheric warming. They are covered by a layer of firn, a matrix of air and snow that is more than one year old. The pockets of air provide space for surface meltwater to percolate into, be retained, and to refreeze. When a warming atmosphere enhances snowmelt, increased refreezing may deplete the firn of air pockets. It reduces the firn's capacity to retain and refreeze surface meltwater. Instead, water becomes mobile, and can move both into meltponds and into crevasses. When water collects in

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

crevasses, the mechanical force exerted at the crevasse tip by a column of standing meltwater enables the crevasse to penetrate the full ice shelf (Van der Veen, 1998). In the Antarctic Peninsula, the warmest region of Antarctica, firn air depletion by atmospheric warming was likely the driver of the collapse of the Larsen B ice shelf in 2002 (Scambos et al., 2003; Banwell et al., 2013). It was followed by accelerated mass loss from tributary glaciers (Rignot et al., 2004). Assuming that firn air depletion prompts flow of water into melt ponds and crevasses, it can be considered as a proxy for vulnerability to ice-shelf collapse by hydrofracturing.

By area, 60% of the ice shelves both buttress upstream ice and would currently be vulnerable to hydrofracturing, if they were experiencing firn air depletion and meltwater ponding (Lai et al., 2020). In future climate scenarios, the firn on many ice shelves indeed becomes depleted (Van Wessem et al., 2023). For a global warming of 4 $^{\circ}$C above the pre-industrial era, about 34 % of the Antarctic ice-shelf area could become vulnerable to meltwater ponding (Gilbert and Kittel, 2021). To improve projections of future global mean sea level, it is therefore essential to include firn air depletion as a proxy for ice-shelf instability into ice-sheet model simulations.

It is important to note that meltwater ponding is merely a prerequisite for ice-shelf collapse (Kuipers Munneke et al., 2014a), and not by itself sufficient. Collapse is only triggered when tensile stresses are sufficiently large (Banwell et al., 2013). For example, the George VI ice shelf has been experiencing meltwater ponding for decades (Wager, 1972), but has nonetheless remained relatively stable (Cook and Vaughan, 2010) because it is confined between land masses and therefore lacks the appropriate tensile stresses for crevasses to appear (Smith et al., 2007). Accelerated ice flow in remaining grounded ice following collapse is expected only when the ice shelf provides buttressing (Lai et al., 2020). Ice-sheet models therefore need to combine information about firn air depletion with their stress state to determine actual collapse and acceleration of ice flow.

Conceptually, the processes that lead from atmospheric warming to ice-shelf collapse are reasonably well understood. But they are poorly quantified, mainly because firn modelling is a challenging subject. The time scales required for a detailed simulation of firn evolution are too small to be included directly in an ice-sheet model. Similar to iceberg calving, explicit parameterizations for hydrofracturing are rarely included in ice sheet models and challenging to implement. The Ice Sheet Model Intercomparison Project for the 6th Climate Model Inter-comparison Project (CMIP6, Eyring et al. (2016), ISMIP6, Nowicki et al. (2020)) was designed to estimate the contribution of the Antarctic Ice Sheet to future sea-level change. In ISMIP6, hydrofracturing-induced ice-shelf collapse was included in some of the experiments (Seroussi et al., 2024). Ice-shelf removal was prescribed when the average melt rate for a decade exceeded 725 $\mathrm{mm\,w.e.\,yr^{-1}}$, the average estimated melt over the Larsen A and B ice shelves in the decade before their collapse (Trusel et al., 2015). However, the potential for meltwater ponding is not solely determined by melt rates. It depends on the balance of processes that add and remove air to and from the firn. Most importantly, firn air content is also influenced by accumulation rates (Van Wessem et al., 2023). A more accurate assessment of firn air depletion, and thus the potential for meltwater ponding during melt events, can be achieved using a dedicated firn model that considers all such processes.

Full firn models that include meltwater hydrology can provide a detailed assessment of firn air depletion but are computationally expensive, making them impractical for generating numerous projections of future Antarctic firn structure for centuries. To address this, statistical emulators, which mimic the outputs of firn models, have recently been developed (Dunmire et al., 2024;

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

Verjans et al., 2021; Veldhuijsen et al., 2025). When properly trained, computationally efficient emulators are useful especially in ice-sheet model inter-comparison projects, where time and computational resources are often constrained. Emulators readily ingest multiple climate forcing datasets and provide results quickly.

The emulator in this study was designed with ISMIP7 in mind. It is the Ice Sheet Model Intercomparison Project 7, and it aims to provide update sea-level projections for the next IPCC reporting cycle. ISMIP7 seeks to implement an ice-shelf collapse protocol for ice-sheet models that better reflects the underlying physical mechanisms that influence ice-shelf stability and susceptibility to hydrofracture. Therefore, the ISMIP7 collapse protocol puts forward firn air depletion as a precursor for meltwater pond formation and ice-shelf stress states as a governing mechanism for hydrofracture potential. The emulator presented here forms the backbone of this protocol, identifying regions of potential meltwater ponding due to future firn air depletion.

In this contribution, we present an XGBoost-based emulator to predict firn air content across Antarctic ice shelves in a future warming climate. We train and evaluate the firn air content emulator using output from the IMAU Firn Densification model v1.2AD (IMAU-FDM v1.2AD) (Veldhuijsen et al., 2024). Inspired by Dunmire et al. (2024), we use an XGBoost model (Breiman, 2001) rather than a Random Forest model (Chen and Guestrin, 2016), because XGBoost outperforms Random Forest in capturing complex nonlinear relationships, particularly in high-dimensional datasets with feature interactions (Anilkumar et al., 2023; Iban and Bilgilioglu, 2023; Sun et al., 2024). We avoid the model evaluation to be impacted by model training, by employing a strategic data-splitting approach, namely by warming scenario and spatially by region. We use the emulator to predict firn air content and meltwater ponding potential on Antarctic ice shelves under various historical and future climate forcing scenarios.

2 Methods

In this study, we aim to emulate the evolution of firn air content as it is simulated in the IMAU-FDM firn model (section 2.1). Here, we motivate our choice for XGBoost (section 2.2), and explain what input variables are selected to emulate firn air content (section 2.3). Finally, we present our training strategy (section 2.4), the training datasets, and the climate scenario runs to which the emulator is applied (sections 2.5 and 2.6).

2.1 IMAU-FDM firn model and forcing

IMAU-FDM version v1.2AD is a semi-empirical one-dimensional firn densification model that simulates the evolution of firn properties such as firn layer depth, density, temperature, grain size, and liquid water content (Veldhuijsen et al., 2024). Firn compaction is based on the semi-empirical dry-snow densification equations from Arthern et al. (2010). These equations assume a steady state, whereas IMAU-FDM is used here to simulate firn under changing climate conditions. Therefore, version v1.2AD features an updated densification parameterization that copes with significantly changing climate forcing (Veldhuijsen et al., 2024). This revised densification treatment depends on firn temperature, grain size, and overburden pressure, rather than on climatological means of temperature and accumulation. The model simulates meltwater percolation using the bucket

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

method, where each firn layer has a maximum irreducible water content that decreases as density increases (Coléou et al., 1999). Meltwater can percolate through all layers within a single time step, is not blocked by ice layers, and will (partially) refreeze when it encounters a layer below the freezing point. If the meltwater concentration in the bottom firn layer exceeds the maximum irreducible water content, it is assumed to exit the firn column as runoff immediately. While computationally efficient, the bucket method does not account for saturated pore spaces, preferential vertical flow, standing water above ice layers, or horizontal flow.

In the simulations used here, IMAU-FDM was forced at the snow surface with 3-hourly values of snowfall, sublimation, snowdrift erosion, 10 m wind speed, surface temperature, surface melt and rainfall from the regional climate model RACMO2 version 2.3p2, at a horizontal resolution of $27{\times}27$ km^2(Van Wessem et al., 2023). RACMO2, in turn, was forced by simulations from the Community Earth System Model version 2 (CESM2) (Danabasoglu et al., 2020). These simulations consist of a historical climate realisation (1950–2014), and future projections for low-, middle- and high-emission scenarios, SSP1-2.6, SSP2-4.5 and SSP5-8.5, respectively. The historical run compares well with an ERA5 reanalysis-driven RACMO simulation (Esch et al., 2025). The projected Antarctic end-of-century (2090-2100) warming in CESM2 under SSP5-8.5 compared to 2005-2015 (+6.7 $^{\circ}\mathrm{C}$) exceeds the mean Antarctic warming in CMIP6 models (+4.8 $^{\circ}\mathrm{C}$), which allowed us to train the firn model on a wide range of temperatures (Dunmire et al., 2022; Kittel et al., 2021). The coupling between RACMO2 and IMAU-FDM is unidirectional, precluding feedback from the snow surface to the atmosphere. However, running IMAU-FDM offline allows for a more realistic initialisation and higher vertical resolution than RACMO2's built-in firn model, which uses similar physical parameterizations. An initial firn layer is obtained by looping over the forcing of the 1950-2000 reference period until the firn layer is in equilibrium with the surface climate of that period. The model has up to 3000 layers of 3 to 15 cm thickness that are advected along with the firn in a Lagrangian way (Veldhuijsen et al., 2024).

2.2 Firn air content emulator

We developed an emulator that mimics how IMAU-FDM simulates the evolution of firn air content across Antarctic ice shelves. We set up extreme gradient boosting (XGBoost) machine learning models (Chen and Guestrin, 2016) in Python using the open-source scikit-learn and XGBoost packages. XGBoost sequentially builds an ensemble of weak decision trees, fitting each new tree to the residuals of the previous ones and minimizing prediction error through an objective function. XGBoost was selected for this study for its high predictive accuracy and its proven ability to outperform other machine learning algorithms, such as neural networks, random forest regression and linear regression methods, for various applications as predicting snow water equivalent, snow avalanche susceptibility and glacier mass balance (Anilkumar et al., 2023; Iban and Bilgilioglu, 2023; Sun et al., 2024). Additionally, XGBoost is well-suited for medium to large tabular datasets (> 10,000) (Grinsztajn et al., 2022). XGBoost also includes regularization techniques that prevent overfitting, and it can estimate the feature importance.

XGBoost hyperparameters control the construction and complexity of the decision trees, as well as the overall behaviour of the model and the learning process. The hyperparameters were tuned using the BayesSearchCV function from scikit-learn during the spatial blocking cross-evaluation (Turner et al., 2021). BayesSearchCV is an automated technique to find optimal hyperparameters for the architecture of decision trees, using bayesian optimization. The technique begins with a prior set of

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

hyperparameters, and by learning from the data, the hyperparameters are optimized in subsequent steps. Here, the optimization focused on eight key hyperparameters in the XGBoost algorithm. Firstly, the number of estimators (`n_estimators` = 120) specifies the number of trees in the model. The learning rate (`learning_rate` = 0.139) determines the step size at which the optimizer updates the weights. Regularization hyperparameters that were tuned include: the maximum tree depth (`max_depth` = 6), minimum weight required in a leaf node (`min_child_weight` = 88), the subsample ratio of training instances (`subsample` = 0.775), subsample ratio of columns when building each tree (`colsample_bytree` = 0.775), the minimum loss reduction (`gamma` = 1.2288), and the regularisation degree (`reg_lambda` = 2.784). The number of estimators, the learning rate, and the six regularization hyperparameters together are optimized such that the model captures relevant patterns without having overly complex trees that are overfitted to the dataset.

2.3 Target variable and input features

We used the annual average firn air content as the target variable. With the purpose of predicting firn air depletion in mind, non-depleted firn needs to be contrasted strongly to depleted firn for the emulator to recognize the difference. However, in the IMAU-FDM training data set, there is only a small difference between the annual average firn air content of a firn layer that is depleted for a brief period during the year, and that of a firn layer that is depleted for most of the year. In addition, the target variable, firn air content, cannot take negative values. This prevents negative values from counterbalancing positive biases in the predictions (e.g. an asymmetric distribution of the errors). To address this, we artificially assigned negative values if the firn air content in a given year falls below 0.1 m at any point in time. Specifically, a value of -1 was assigned for each 5-day period with less than 0.1 m of firn air content. For instance, if there is less than 0.1 m of firn air for 100 days in a given year, the target value becomes -20. This value replaces the true average firn air content. Allowing the target variable to have negative values enables individual decision trees in XGBoost to offset positive biases of other decision trees. Although the units of the negative and positive quantities are different, the same order of magnitude enables a relatively smooth transition around zero, representing the transition from non-depleted to depleted firn (Veldhuijsen et al., 2025).

The climate variables used as input features for the emulator are (1) total annual snow accumulation (snowfall minus evaporation/sublimation), (2) total annual surface melt, (3) total annual rainfall, (4) annual Melt-over-Accumulation (MoA) ratio, (5) mean annual 2 m temperature and (6) mean annual 10 m wind speed. These are the most important mass fluxes and boundary conditions governing the firn density (e.g., The Firn Symposium Team, 2024). Note that we use 2 m air temperature for the emulator, rather than surface temperature that is used to force the IMAU-FDM model. We choose 2 m air temperature, because it is less dependent on the treatment of e.g. albedo or thermal properties of snow, allowing for compatibility with a broader range of forcing datasets. Annual MoA is the ratio between total annual surface melt plus rain and total annual snow accumulation. When MoA exceeds a certain value (often taken as 0.7, but there is some uncertainty), firn pore space can no longer be maintained, and firn air depletion is initiated (Pfeffer et al., 1991). Although MoA is derived from two input features already included in the emulator, this is a common practice in machine learning. By combining these two variables into a single ratio, which is a strong predictor of the target variable, the complexity of the trained XGBoost model is reduced. To account for the time required for a firn column to adjust to changing climatic conditions, we use trailing multi-year averages

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

of the input features. Specifically, we calculate 10-, 30-, and 50-year averages for snow accumulation, surface melt, rainfall, MoA, and 2 m air temperature. For 10 m wind speed, only the 10- and 30-year averages are retained, as excluding the 50-year wind-speed feature slightly improved overall emulator performance. These trailing averages include the year being predicted and the preceding years in the averaging window. For some of the firn air content predictions with the emulator, not all 50 preceding years of data are available. In that case, the average of the maximum available preceding years are used instead.

For discussing the importance of surface melt for emulator performance, we trained an alternative emulator, which we will refer to as the no-melt emulator. Details on this emulator are given in Section 3.3.

2.4 Emulator training

The emulator was trained to predict annual average firn air content, individually for all Antarctic ice shelves. We also include grid cells on the grounded ice below 1,000 m a.s.l. in the training data to cover a wider range of climate regimes. In total, we have 4,613 IMAU-FDM grid cells with 85 yearly values each from the period 2015-2100, amounting to 1.2 million data points across three climate scenarios. For evaluation, we only consider ice shelf locations, which amounts to 0.5 million data points for the testing data. We evaluate the emulator performance across the full range of firn air content values. In addition, since we are mainly interested in predicting meltwater ponding, we also evaluate the performance of the emulator to simulate firn air depletion specifically. Here, we used a threshold of 2 m of firn air to define firn air depletion, discussed in Section 3.2.

The training and testing data have a spatial and temporal structure, meaning there are similarities between grid cells within regions of Antarctica and for the different time segments of the simulation. To ensure independence between training and testing data, we split the training and testing data strategically rather than randomly (Roberts et al., 2017). This approach prevents overfitting and avoids overestimating the emulator performance during evaluation (Veldhuijsen et al., 2025). First, we evaluate the emulator performance using spatial blocking cross-evaluation. For this, we divide the selected dataset into eight regions (Fig. 1) and leave out one region at a time. The cross-evaluation ensures that the model is evaluated on the full dataset. Additionally, we also evaluate the performance through scenario blocking cross-evaluation. Thereby, we test if the emulator can predict FAC under SSP5-8.5 scenarios while only using SSP1-2.6 and SSP2-4.5 for training, while for testing SSP1-2.6 and SSP2-4.5, we only use SSP5-8.5 for training. To minimize dependency between scenarios, the first 30 years of future simulations (2015–2045) are excluded from validation scores as the input features do not sufficiently diverge during this period. For the training of the final version of the emulator, used for the scientific analysis, we used all scenarios and all regions, to ensure that the widest possible range of warming trends and climatic conditions are included.

2.5 Data for emulator training and forcing

To train the firn air emulator, we used IMAU-FDM simulations forced by RACMO2, which was in turn forced by CESM2 for the historical period and three future scenarios (SSP1-2.6, SSP2-4.5 and SSP5-8.5), thereby covering the period 1950-2100 (Veldhuijsen et al., 2024). No other similar simulations were available for training. The simulations show that ice shelves in the Antarctic Peninsula and the Roi Baudouin ice shelf in Dronning Maud Land are particularly vulnerable to firn air depletion (> 50 % decrease by 2100), even for the low-emission (SSP1-2.6) and intermediate-emission (SSP2-4.5) scenarios (Fig. 2a,b).

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

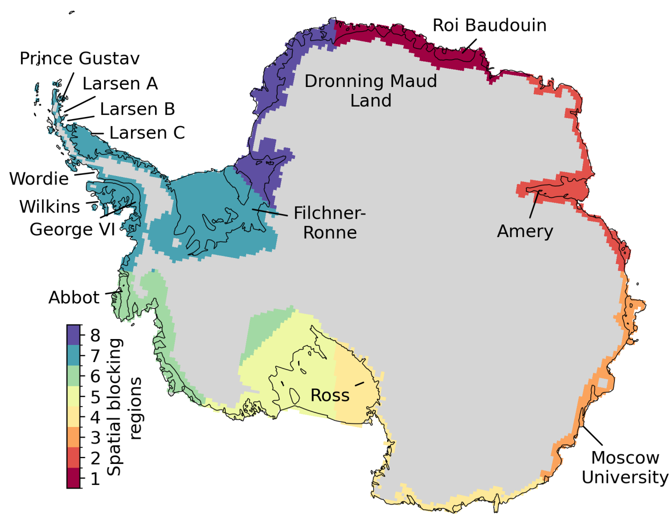

Figure 1. Eight regions of the Antarctic Ice Sheet used for spatial blocking cross-evaluation. The ice shelves and regions referred to in this study are labelled, and black lines connect the labels and the exact locations of the time series in Fig. 11.

In the high-emission (SSP5-8.5) scenario, all ice shelves experience considerable firn air depletion by 2100 (> 19 % decrease, Fig. 2c).

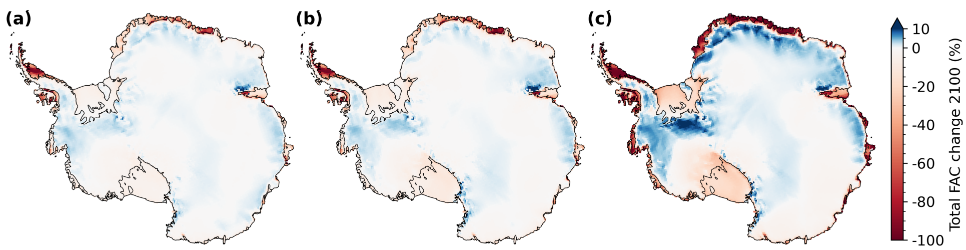

Figure 2. Relative change in total firn air content by 2100 compared to 2014 for **(a)** SSP1-2.6, **(b)** SSP2-4.5 and **(c)** SSP5-8.5. Note the different scales for decreasing (red) and increasing (blue) firn air. Reproduced from Veldhuijsen et al. (2024).

Once the emulator is trained, we explore its skill in simulating the recent past. We assess historical firn air content over previously collapsed ice shelves, by forcing it with historical ERA5 reanalysis data (1940-2020). This is done for locations on the Larsen A, Larsen B, Prince Gustav, Wordie and Wilkins ice shelves (Cook and Vaughan, 2010). Ice-shelf collapse on these ice shelves started in the 1980s. Therefore, we directly use ERA5 as forcing data for the emulator as it extends back to 1940, compared to 1979 for RACMO2 forced with ERA5 (Van Wessem et al., 2018). However, ERA5 is unreliable before 1979 due to lack of satellite observations (Bromwich et al., 2024). Since ERA5 does not include reliable melt estimates, we used melt estimates obtained from a machine learning model trained on ERA5 reanalysis data, explained in Section 2.6.

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

Table 1. List of available regional climate models (RCMs) driven by earth system models (ESMs) along with the ESM era and ensemble member number, the RCM horizontal resolution (HR), the available scenarios, and the time period.

RCM	ESM	ESM era	RCM HR	Scenarios	Period
RACMO2	CESM2	CMIP6	27 km	SSP1-2.6	1950-2100
				SSP2-4.5	
				SSP5-8.5	
MAR	ACCESS1.3	CMIP5	35 km	SSP5-8.5	1980-2100
HIRHAM	CESM2	CMIP6	12 km	SSP5-8.5	1980-2100

As a final investigation, we explore how the emulator predicts future firn air content when forced with other regional climate models than RACMO2 that they were trained with. We forced the emulator with SSP5-8.5 projections from the Modèle Atmosphérique Régional (MAR) (1980-2100) driven by ACCESS1.3 (Kittel et al., 2021), and from HIRHAM (1980-2100) forced by CESM2 (Hansen et al., 2021) (see Table 1 for details). The horizontal resolution of each climate forcing is preserved in the emulator simulations. From MAR, only 8 m wind speed is available. To convert this to 10 m wind speed, we assumed a logarithmic wind speed profile.

2.6 Surface melt data

Although ERA5 provides a long and internally consistent record of near-surface meteorological and radiative conditions over Antarctica (Hersbach et al., 2020; Gossart et al., 2019), native ERA5 surface melt is not considered reliable over Antarctic ice shelves, largely because melt depends strongly on snow and firn processes such as albedo evolution that are not well represented in the reanalysis (Carter et al., 2022). We therefore used ERA5 atmospheric and radiative fields to force an empirical melt-estimation model.

Daily surface meltwater production was estimated using a Random Forest regression model trained against surface energy balance (SEB) melt rates from the automated weather station (AWS) dataset of Jakobs et al. (2020). The model was trained using daily data from AWS14, AWS15, AWS17, and AWS18 on Larsen C and Scar Inlet ice shelves. The target variable was daily total surface meltwater production from the AWS-derived SEB calculations, and predictors were limited to the following variables available from both the AWS observations and ERA5: downward shortwave radiation, downward longwave radiation, 2 m air temperature, sensible heat flux, and latent heat flux.

The Random Forest model was calibrated using data from January 2009 through June 2017, with a random 75/25 train-test split within this period, and July 2017 through December 2018 was retained for independent evaluation at the AWS sites. In the random train-test split, the model explained approximately 81% of the variance in daily AWS-derived melt rates. To generate gridded melt estimates, daily ERA5 predictor fields were prepared over the Antarctic Peninsula domain and passed through the trained model at each grid cell to produce daily meltwater-production estimates for 1940-2023. The resulting daily melt fields

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

were then aggregated to annual melt inputs for the firn air content emulator, together with the other ERA5-derived forcing variables.

3 Emulator training and evaluation

3.1 Design and range of application

To assess whether the regions used for the spatial blocks (Fig. 1) give a representative result of the emulator's performance when used with different climate forcing datasets, we examine their climate regimes. We focus on snow accumulation and liquid water input (melt and rain), as they are good predictors of simulated firn facies types (Brils et al., 2024). The climate regimes of the eight regions used for the spatial blocking cross-evaluation, based on average snow accumulation and liquid water for 1980-2015, are shown in Figs. 3a,b. To show both low and high accumulation and melt locations in sufficient detail, we present the data twice: once using a logarithmic scale in Fig. 3a, and again with a linear scale in Fig. 3b. Two clusters can be identified: a "high melt" cluster, characterized by high melt rates (> 20 to 300 mm w.e. yr^{-1}), and a "low melt" cluster with melt rates up to 10 mm w.e. yr^{-1}. Region 2, which includes Amery ice shelf, is the only region with low accumulation (20-90 mm w.e. yr^{-1}) and some melt (20-50 mm w.e. yr^{-1}). Region 7, which includes the Antarctic Peninsula, is notable for having the highest melt and accumulation rates.

We visually compare the differences between the regions to the differences between the entire RACMO2 dataset and the MAR and HIRHAM simulations (Figs. 3c,d). HIRHAM has a larger spread in climate regimes compared to MAR and RACMO, which could be related to the finer horizontal resolution. Although this comparison is limited to the historical period and is merely visual, it suggests that the variability among the regions is at least as large as that among the forcing datasets, and that the emulator is not applied outside the training range.

The optimal hyperparameters obtained during the spatial blocking cross-evaluation are listed in Section 2.4. The permutation importance of the emulator is shown in Figure 4a. This metric measures the increase in the prediction error of the model when the feature's values are randomly shuffled, thereby breaking its relationship with the target variable. Figure 4a shows that MoA is the most important input feature.

3.2 Emulator performance

Among the most important input features for the emulator are accumulation, surface melt, and its ratio MoA. These inputs are among the most discerning variables in IMAU-FDM itself (Brils et al., 2024). Therefore, it is expected that this emulator is able to predict firn air content, and firn air depletion, as long as the provided surface melt and accumulation estimates are reliable. The spatial blocking cross-evaluation shows that the emulator explains at least 94 % of the variance in IMAU-FDM firn air content (R^2 = 0.94 to 0.96; Figs. 5a-c). The scenario-blocking evaluation yields a slightly better performance (R^2 = 0.97 to 0.98, Figs. 5d-f). In general, the highest biases are found for low values of firn air content (< 8 m). The results also show that the emulator performs well when predicting a colder scenario if trained on a warmer scenario, also for low firn air

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

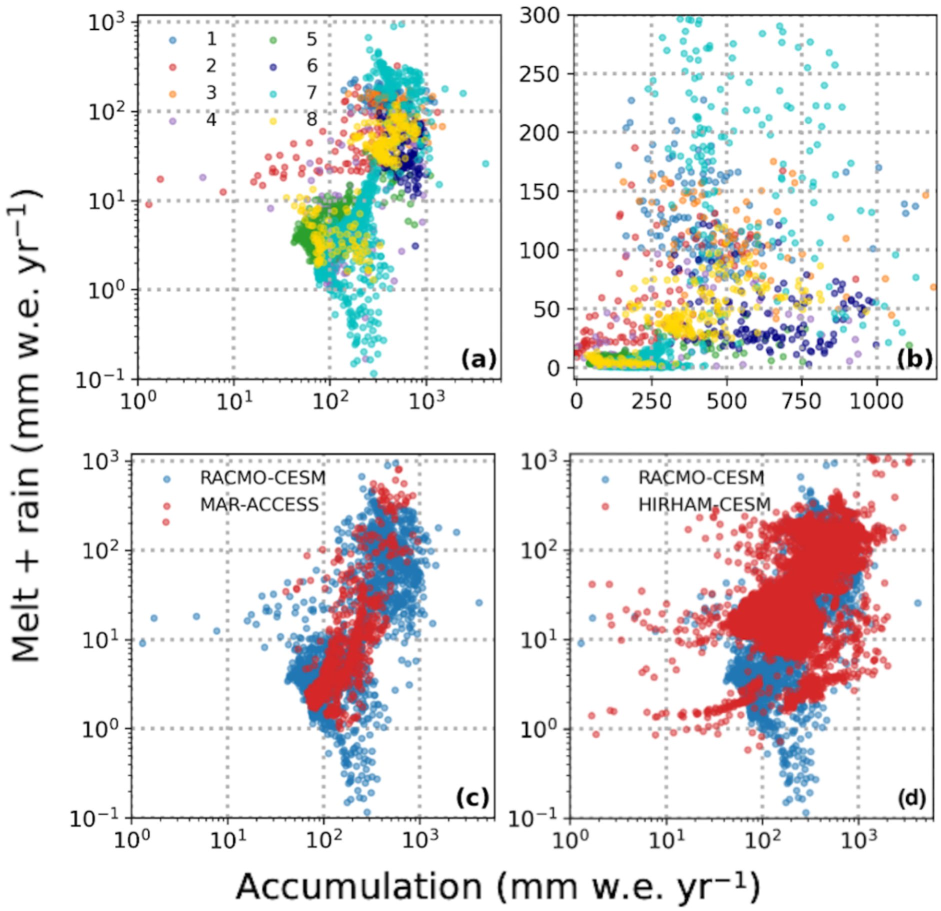

Figure 3. Average 1980-2015 combined surface melt and rain rate (liquid water input) against snow accumulation (snowfall minus sublimation/evaporation minus snowdrift erosion) rate for all ice shelf locations for **(a,b)** the eight RACMO2 regions used for spatial blocking cross-evaluation, **(c)** RACMO2 and MAR, and **(d)** RACMO2 and HIRHAM.

content values (Figs. 5e,f). When extrapolating to a warmer scenario (Fig. 5g), the performance is somewhat worse, with a small negative bias (-0.2 m), notably visible in the low firn air content range.

Since we are eventually interested in predicting the potential for meltwater ponding, Table 2 (leftmost columns) also shows the agreement between the emulator and IMAU-FDM on the presence of depleted firn (< 2 m FAC). The evaluation for the spatial blocking shows that the emulator predicts 7 % false positives (cases where the emulator indicates depleted firn, but IMAU-FDM does not) and 17 % false negatives (cases where the emulator fails to identify depleted firn, despite IMAU-FDM doing so) for all scenarios combined. The highest fraction of false negatives is found for SSP1-2.6 and SSP2-4.5 (24 % and 25 %, respectively). Excluding region 2, which includes Amery ice shelf, reduces the false negatives for SSP1-2.6 and SSP2-4.5 rates to 13 % and 11 %, respectively. This region has a unique climate with low accumulation and intermediate melt (Fig. 2a),

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

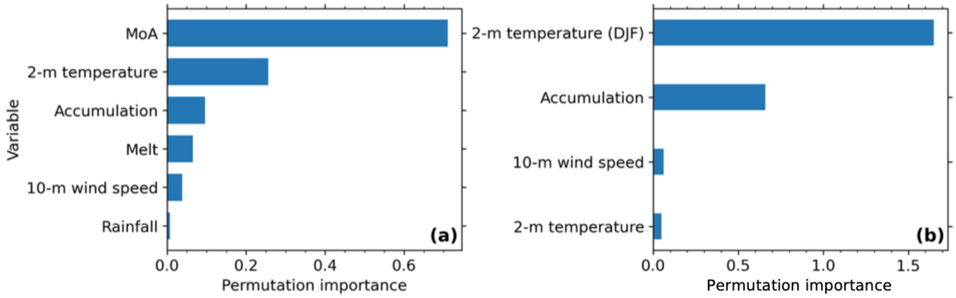

Figure 4. Permutation importance of features for (a) the emulator, and (b) the no-melt emulator, calculated based on the R^2 score. Importance of features over different time periods are summed, e.g. 10-year, 30-year and 50-year averaged values.

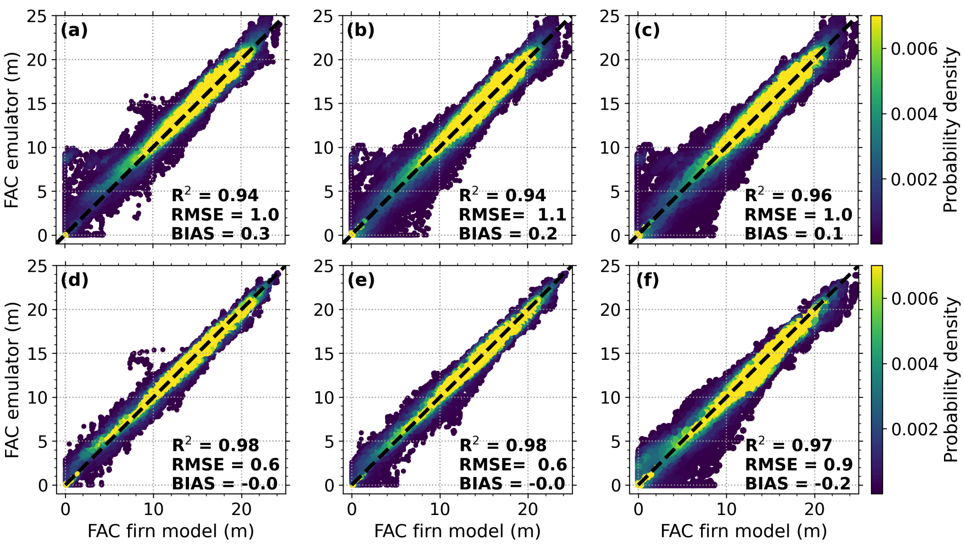

Figure 5. Firn air content simulated by the emulator from the spatial blocking **(a-c)** and scenario blocking **(d-f)** cross-evaluation for **(a,d)** SSP1-2.6, **(b,e)** SSP2-4.5 and **(c,f)** SSP5-8.5.

and is therefore difficult to approximate with the emulator if similar training data from other regions are missing. Again, the scenario blocking evaluation yields a slightly better performance than spatial blocking evaluation (left half of Table 2). In both evaluation approaches, the emulator tends to underestimate rather than overestimate depletion, given that there are more false negatives than false positives.

Figure 6 shows the spatial pattern of the mean firn air content differences between the emulator and IMAU-FDM over the period 2090-2100, for each scenario separately, for both the spatial blocking (Figs. 6a-c) and scenario blocking (Figs. 6d-f) evaluation. The spatial blocking results are aggregated for all blocked regions per scenario (Fig, 6a-c). In the scenario blocking evaluation, SSP1-2.6 and SSP2-4.5 (Figs. 6a,b) were trained on SSP5-8.5 only, while SSP5-8.5 (Fig. 6f) was trained on both

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

Table 2. Spatial blocking and scenario blocking cross-evaluation results for predicting depleted firn air (< 2 m firn air content). Results are given as false negatives and false positives. The no-melt emulator is introduced in Section 3.3.

Scenario	Emulator		No-melt emulator	
	False negatives	False positives	False negatives	False positives
Spatial blocking				
All	17%	7%	32%	14%
SSP1-2.6	24%	3%	37%	14%
SSP2-4.5	25%	7%	39%	19%
SSP5-8.5	10%	8%	26%	12%
Scenario blocking				
All	14%	6%	19%	16%
SSP1-2.6	11%	5%	10%	21%
SSP2-4.5	13%	2%	10%	22%
SSP5-8.5	16%	8%	26%	10%

SSP1-2.6 and SSP2-4.5. The spatial and scenario blocking results for SSP1-2.6 and SSP2-4.5 are similar and show a generally good performance (RMSD = 0.63 to 0.95 m). The scenario blocking evaluation for SSP5-8.5 exhibits the poorest performance (RMSD = 1.3 m), with the emulator underestimating FAC on Ross ice shelves, and overestimating FAC in West Antarctica and Dronning Maud Land. This contrasts with the results in Fig. 5f, which shows a better performance (RMSD = 0.9 m). This discrepancy suggests that the emulator's accuracy decreases further into the future when extrapolating to a warmer scenario. It highlights the importance of including a strong warm scenario in the training data.

Figure 7 presents the results for depleted firn air, showing the difference in number of years with depleted firn (< 2 m) between the emulator and IMAU-FDM over the period 2015-2100. In the spatial blocking evaluation (Figs. 7a-c), the emulator performs poorest on Amery and George VI ice shelves, where it underestimates firn air depletion. On George VI ice shelf in 2015, the melt and accumulation in RACMO2 are both around 250 $\mathrm{mm\,w.e.\,yr^{-1}}$, i.e. MoA approximately equals 1, which is not found in other regions, explaining the discrepancy (Fig. 2a). Similarly, the low accumulation conditions of Amery ice shelf are absent in the other spatial blocks. In contrast, with scenario blocking, the firn air content at these locations is well approximated by the emulator (Figs. 7d-f). In the scenario blocking evaluation for SSP1-2.6 and SSP2-4.5 (Figs. 7d,e), the emulator in general underestimates firn air depletion, but with relatively low biases. In contrast, under SSP5-8.5 scenario blocking (Fig. 7f), the emulator underestimates firn air depletion for several ice shelves in West Antarctica, such as Wilkins and Abbot, and ice shelves in Dronning Maud Land, as discussed above.

Figure 8 shows six example time series of firn air content predicted by the emulator and IMAU-FDM for individual grid points, from both the spatial blocking and scenario blocking evaluations. Predictions from the no-melt emulator are also included and will be discussed in Section 5.2. In a number of examples, the emulator predicts a firn air content decrease similar to IMAU-FDM for the spatial blocking and scenario blocking evaluation, namely at Moscow University ice shelf for SSP5-8.5 (panel 8a), at Wilkins ice shelf for SSP1-2.6 (panel 8c), and at Ross ice shelf for SSP1-2.6 (panel 8f). However, at Amery ice

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

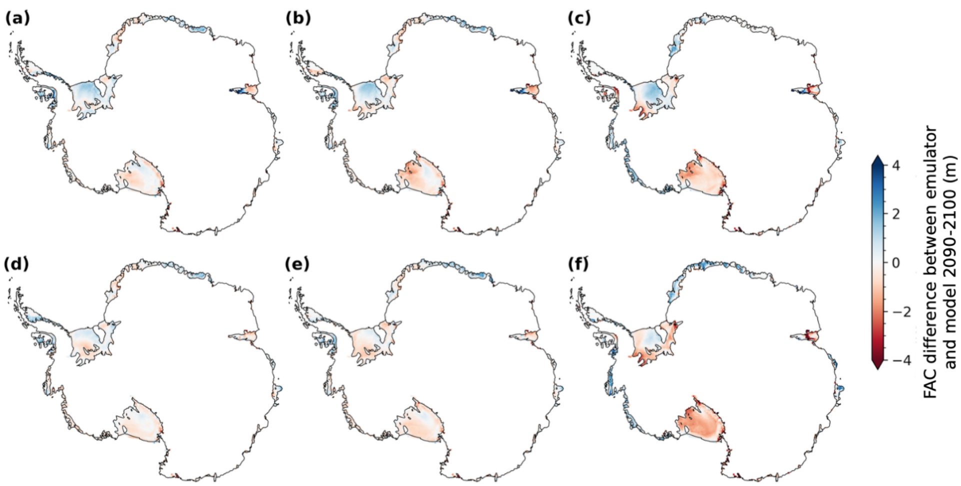

Figure 6. Difference in firn air content between the output of the firn model (IMAU-FDM) and the emulator (emulator minus firn model) for spatial blocking cross-evaluation **(a-c)** and scenario blocking cross-evaluation **(d-f)** for **(a,d)** SSP1-2.6, **(b,e)** SSP2-4.5 and **(c,f)** SSP5-8.5 over the period 2090-2100. The spatial blocking cross-evaluation results are aggregated for all blocked regions per scenario. Red indicates that the emulator predicts lower firn air content compared to IMAU-FDM.

shelf for the spatial blocking evaluation for SSP2-4.5 (panel 8b), the emulator predicts just under 10 m FAC until 2060, with decreases occurring only after substantial melt begins, whereas IMAU-FDM simulates a persistent complete depletion (0 m FAC). The discrepancy is caused by the unique climate of this region, as previously discussed. In addition, the emulator overestimates FAC in the final decades at Abbot ice shelf for SSP5-8.5 for the scenario blocking (panel 8d), and at Roi Baudouin for SSP2-4.5 (panel 8e) for the spatial blocking evaluation.

3.3 No-melt emulator

Many climate models do not provide robust snow melt rates to force a firn emulator. To demonstrate how important it is to use surface melt as an input feature, we also train an alternative XGBoost emulator, that does not use surface melt. This *no-melt emulator* instead requires the following input features: (1) total annual snow accumulation (snowfall minus evaporation/sublimation), (2) mean summer (DJF) 2 m temperature, (3) mean annual 2 m temperature and (4) mean annual 10 m wind speed. We expect the average summer 2 m temperature to be loosely related to surface melt. It is apparent that this input feature is the single most important feature for the no-melt emulator (Figure 4b).

The evaluation of the no-melt emulator is shown in Figure 9. The performance of this emulator is poorer compared to the standard one (R^2 = 0.82 – 0.92, instead of 0.94 - 0.98). The no-melt emulator displays considerable spread around the 1:1 line for the entire range of firn air content values. Also, it shows a substantial increase in the false positives and false negatives for

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

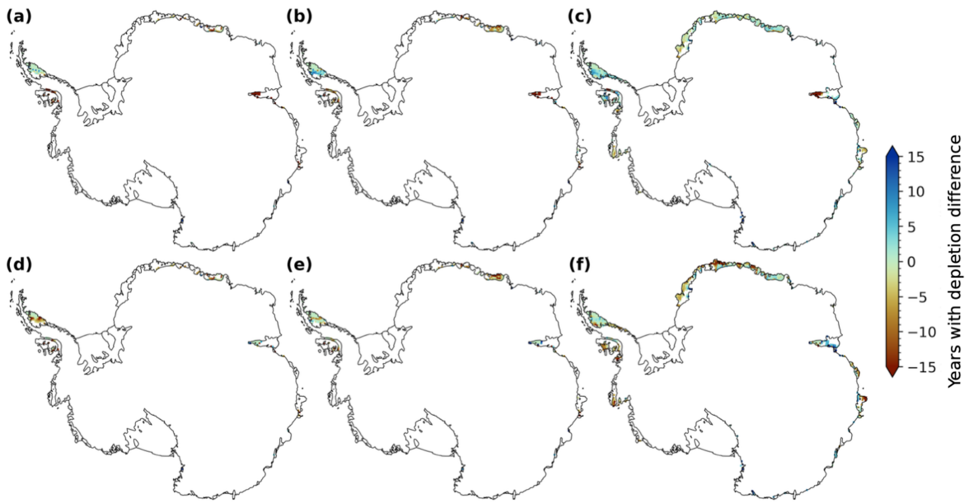

Figure 7. Difference in number of years with firn air depletion (< 2 m FAC) between the output of the firn model (IMAU-FDM) and the emulator (emulator minus firn model) over the period 2015-2100 for spatial blocking **(a-c)** and scenario blocking **(d-f)** evaluation for scenarios **(a,d)** SSP1-2.6, **(b,e)** SSP2-4.5 and **(c,f)** SSP5-8.5. Brown indicates that the emulator predicts shorter periods of firn air depletion compared to IMAU-FDM.

depleted firn air (< 2 m) (Table 2). For example, for scenario blocking the false negatives increase from 14 % to 19 % and the false positives from 6 % to 16 %.

Figure 10 shows the difference between the no-melt emulator and IMAU-FDM in years of firn air depletion (< 2 m) for spatial blocking of SSP5-8.5 and for the scenario blocking of SSP1-2.6. Comparison to the standard emulator (Figs. 7c,d) reveals that the overall performance is poorer. The RMSD increased from 12.0 to 17.4 years for spatial blocking of SSP5-8.5, and from 12.0 to 19.3 years for scenario blocking of SSP1-2.6. Figure 8 also includes example time series of the no-melt emulator. Compared to its training data, it strongly underestimates firn air depletion for spatial blocking of Amery (SSP5-8.5) and Roi Baudouin ice shelves (SSP2-4.5) and for scenario blocking of Abbot ice shelf (SSP5-8.5). For the other examples, the differences between the two emulators are less pronounced.

4 Emulator application

Having trained and evaluated the firn air content emulator, we now demonstrate two applications. In the first, we supply the emulator with ERA5 input for the period 1940-2020. In the second, the emulator is supplied with input from MAR and HIRHAM, two regional climate models that were not used for model training.

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

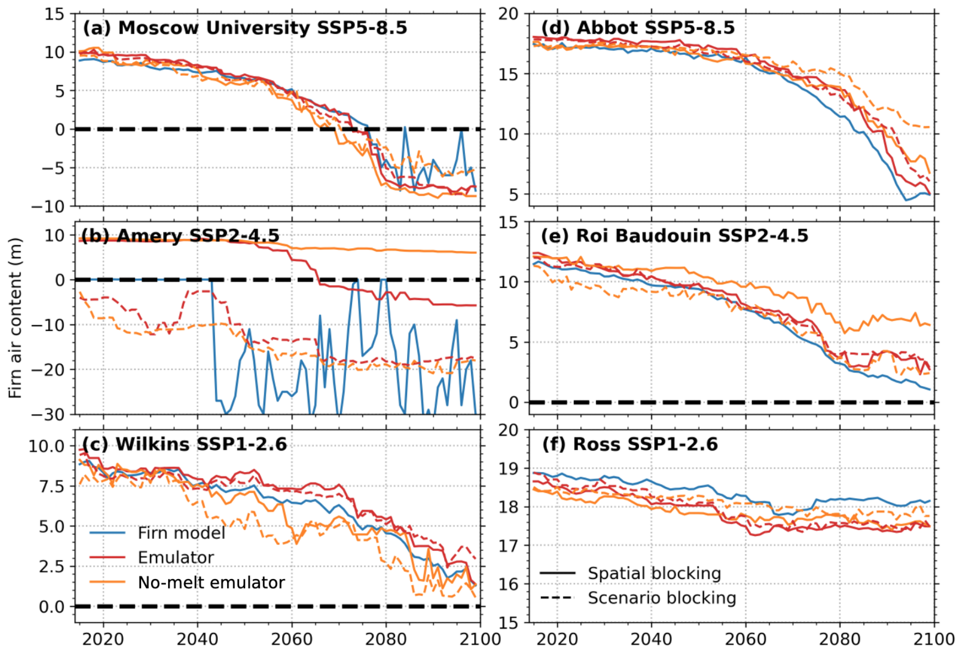

Figure 8. Example timeseries of firn air content (FAC) output of the IMAU-FDM firn model (blue), the emulator (red) and the no-melt emulator (orange) for individual grid points. Negative values are assigned when the FAC in a given year falls below 0.1 m at any point in time. Specifically, a value of -1 was assigned for each 5-day period with less than 0.1 m of FAC. The solid lines indicate the spatial blocking and the dashed lines the scenario blocking results. The locations of the grid points are indicated in Fig. 1.

4.1 Historical ERA5

Figure 11 shows firn air content evolution simulated with the emulator, forced by ERA5 with machine learning melt (see Sect. 2.6), over six ice shelf locations in the Antarctic Peninsula that have collapsed in recent decades (Cook and Vaughan, 2010). For Larsen A and Larsen B ice shelves, the emulator predicts firn air content to be depleted prior to the observed collapse. According to the emulator, Prince Gustav, Wilkins and Wordie ice shelf still had > 5 m of firn air at the timing of their collapse. We are not certain that firn was actually depleted when they collapsed. These locations have climate conditions that are favourable for the formation of firn aquifers (Van Wessem et al., 2023; Veldhuijsen et al., 2025), with more than 500 $\mathrm{mm\,w.e.\,yr^{-1}}$ of melt and more than 800 $\mathrm{mm\,w.e.\,yr^{-1}}$ of accumulation. It is possible that firn aquifers were present at the timing of their collapse, and that they contributed to it.

4.2 MAR and HIRHAM data

In this section, we apply the emulator to climate time series of a MAR projection forced by ACCESS1.3, and a HIRHAM projection forced by CESM2, both under SSP5-8.5 for 1980-2100. These results give an impression of the consistency of the

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

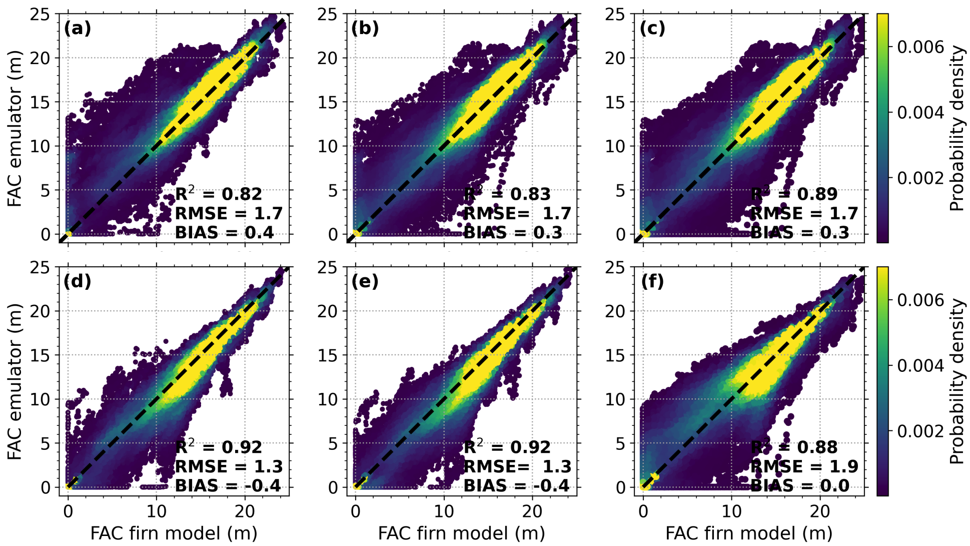

Figure 9. Firn air content (FAC) simulated by the no-melt emulator from the spatial blocking (upper row) and scenario blocking (lower row) cross-evaluation for **(a,d)** SSP1-2.6, **(b,e)** SSP2-4.5 and **(c,f)** SSP5-8.5.

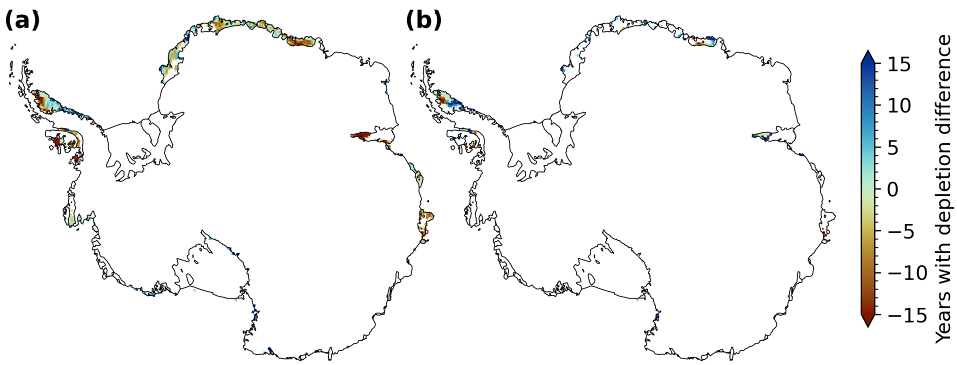

Figure 10. Difference in number of years with firn air depletion (< 2 m FAC) between the output of the firn model (IMAU-FDM) and the no-melt emulator (emulator minus firn model) for **(a)** spatial blocking for SSP5-8.5 and **(b)** for scenario blocking SSP1-2.6 evaluation. Brown indicates that the no-melt emulator predicts shorter periods of firn air depletion compared to IMAU-FDM.

emulator, but are not intended to thoroughly discuss future firn air content evolution. The results are shown in Figures 12, 13, and 14. These figures also show results from the non-melt emulator that we develop and discuss in section 3.3 and 5.1.

Figure 12 shows firn air content in 2100 for the emulator forced by MAR (panel a) and HIRHAM (panel d). For both projections, firn air depletion is predicted over most of Larsen C, George VI, Roi Baudouin, Wilkins ice shelves and parts of Abbot and Amery ice shelves. In the MAR-forced emulator simulations, depletion is predicted for large parts of the ice

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

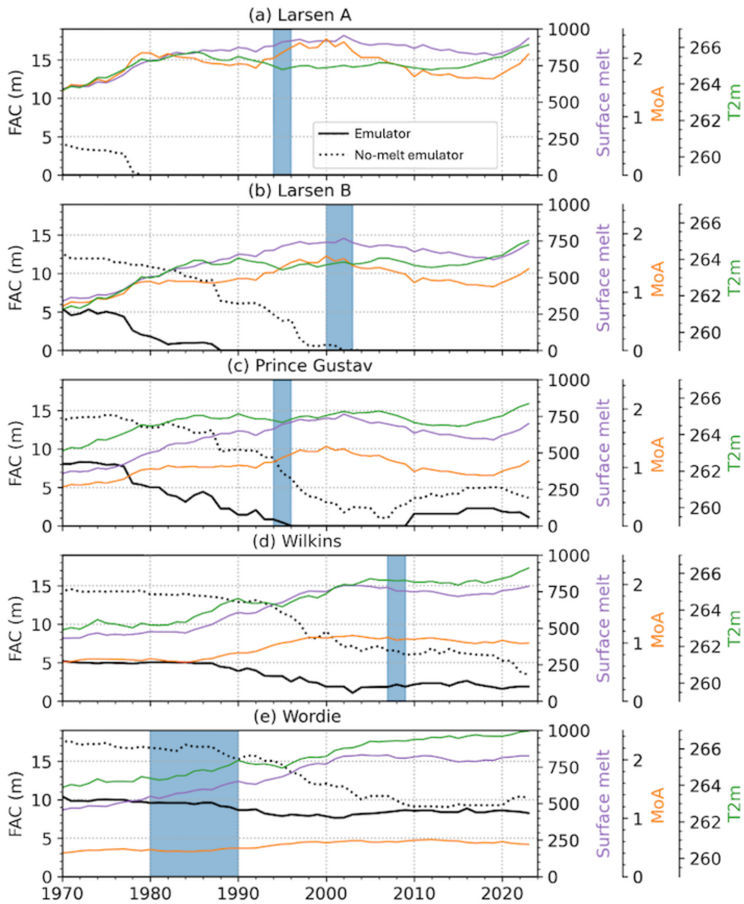

Figure 11. Time series of firn air content (FAC) simulated with the emulator and the no-melt emulator for six ice shelf locations that have collapsed in recent decades. The time series also include surface melt, Melt-over-Accumulation ratio (MoA) and 2 m temperature (T2m). The emulator is forced by ERA5 data and the surface melt estimates are obtained from a machine learning model trained on ERA5 data. The timing of the collapse events is indicated with the blue band (Cook and Vaughan, 2010). In panel **(a)**, the emulator persistently predicts 0 m of firn air. The locations of the grid points are indicated in Fig. 1.

shelves in Dronning Maud Land and other ice shelves in East Antarctica (such as Amery, Shackleton and Moscow University ice shelves), while in the HIRHAM simulations depletion is less widespread.

Figure 13 shows how simulated firn air content from the emulator (upper row) and the no-melt emulator (lower row) over the period 2015-2100 for the three RCMs are related to rates of accumulation and melt (plus rain), shown here as averages for the preceding 10 years. For given surface melt and accumulation, the emulator predicts comparable firn air content for both MAR and RACMO2. For HIRHAM however, the firn air content is higher (Figs. 13a-c). The differences between the emulator and its no-melt counterpart, revealed by Figures 13 and 12, do align. For HIRHAM, the no-melt emulator predicts higher firn air content than the normal emulator for the same conditions (Figs. 13d,f), and this is reflected in Figures 12d-f too. For MAR,

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

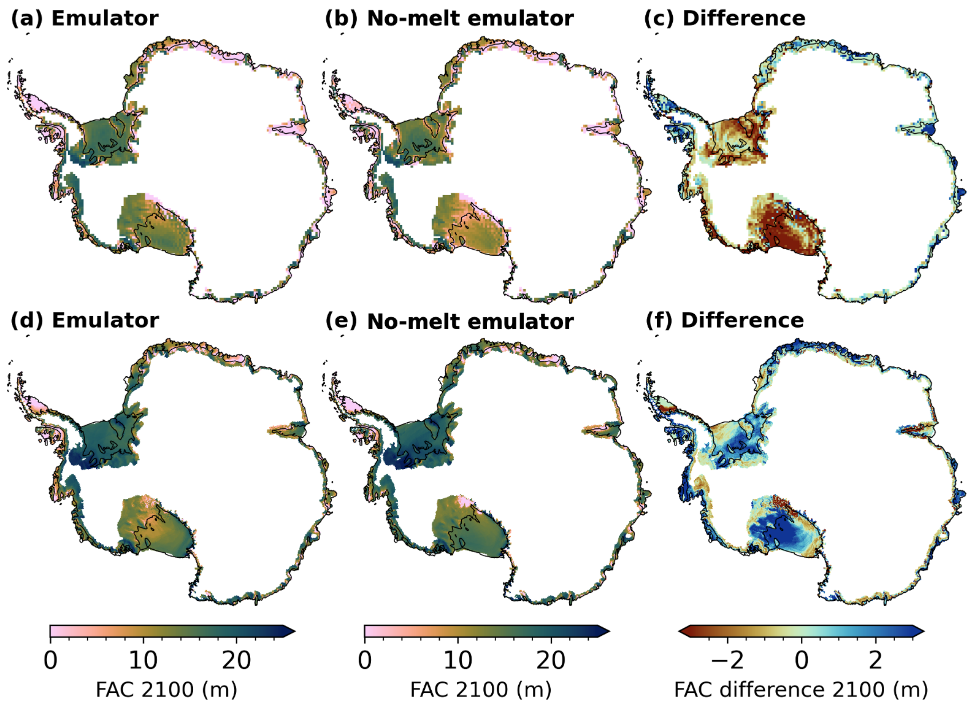

Figure 12. Firn air content (FAC) at 2100 for SSP5-8.5 from MAR **(a-c)** and HIRHAM **(d-f)** simulated with **(a,d)** the emulator, **(b,e)** the no-melt emulator and **(c,f)** the corresponding difference (non-melt emulator minus emulator).

the results are less intuitive, and it appears that for MoA < 0.7, the firn air content is lower for the no-melt emulator, and for MoA > 0.7, it is higher (Figs. 13d,e).

Figure 14 is meant to make us understand the differences between models and emulators. For RACMO2, there is not a very consistent relation between annual average temperature and melt (Fig. 14a), but for summer (DJF) temperature there is (Fig. 14d). The figure shows that for a constant accumulation rate, the surface melt increases with increasing temperature. Also, it shows the impact of snowfall on surface melt through albedo: warmer conditions are needed at snowy locations than at dry locations to end up with a similar melt amount. Surprisingly, this effect of accumulation on albedo and surface melt seems absent for both MAR and HIRHAM (Figs. 14e,f). In those models, the relation between summer temperature and surface melt is not quite uniform. MAR does not show a clear impact of accumulation on the temperature dependency of surface melt, while for HIRHAM this relation even seems to be opposite to that for RACMO2, such that, for a given summer temperature, surface melt increases with higher accumulation.

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

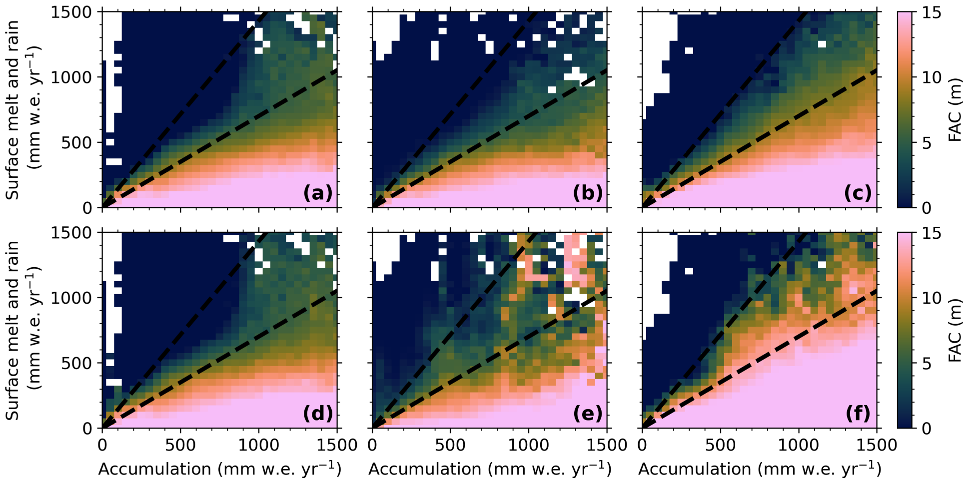

Figure 13. Annual average firn air content (FAC) for each year in the period 2015-2100 simulated with the emulator **(a-c)** and the no-melt emulator **(d-f)** as a function of preceding 10-year average annual surface melt and rainfall (y-axis) and snow accumulation (x-axis). Panels depict **(a,d)** RACMO2, **(b,e)** MAR, and **(c,f)** HIRHAM, each for SSP5-8.5. The dashed lines indicate the 0.7 and 1.7 Melt-over-Accumulation (MoA) thresholds.

5 Discussion

5.1 Performance of a no-melt emulator

The no-melt emulator that was designed to emphasize the importance of surface melt, was also applied to historical ERA5 data as shown in Section 4.1 and Figure 11. Like the standard emulator, it predicts firn air content to be depleted prior to the observed collapse of Larsen A and B ice shelves (Figure 11). In the application with MAR and HIRHAM forcings, more differences with the regular emulator become apparent. For HIRHAM, the no-melt emulator predicts higher firn air content for the same conditions (Figs. 13d,f), and this is reflected in Figures 12d-f too. For MAR, the results are less intuitive, and it appears that for MoA < 0.7, the firn air content is lower for the no-melt emulator, and for MoA > 0.7, it is higher (Figs. 13d,e).

Given that the no-melt emulator is trained with RACMO2 data only, the differences between Figs. 14d-f can explain the differences between Figs. 13d-f. For MAR, firn air content is somewhat lower below the 0.7 MoA line, and clearly higher above (Figs. 13d,e). This is in line with the higher and lower summer temperatures, respectively, modelled by MAR compared to RACMO2 for these accumulation and melt conditions. For HIRHAM, firn air content is higher overall (Fig. 13f), caused by a lower summer temperature compared to RACMO2 for modelled melt rates (Fig. 14f). Thus, summertime 2 m temperature is a relatively vulnerable approximation to estimate surface melt, because the relation between summer 2 m temperature and melt differs from model to model. As shown here, this is true even for climate models that have relatively sophisticated surface parameterizations for e.g. albedo.

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

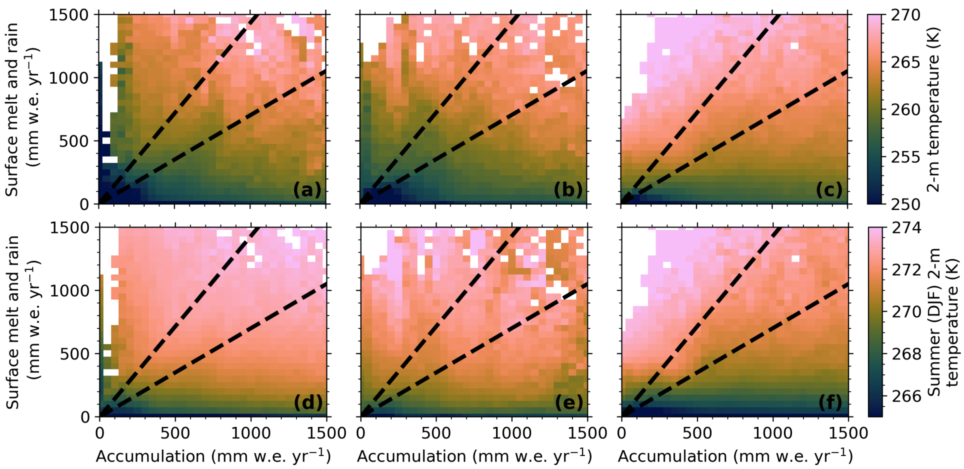

Figure 14. 10-year annual average 2 m temperature **(a-c)** and summer (DJF) 2 m temperature **(d-f)** for each year in the period 2015-2100 as a function of the preceding 10-year averages of annual surface melt and rainfall (y-axis) and snow accumulation (x-axis). Panels depict SSP5-8.5 results for **(a,d)** RACMO2, **(b,e)** MAR, and **(c,f)** HIRHAM. The dashed lines indicate the 0.7 and 1.7 Melt-over-Accumulation (MoA) thresholds.

5.2 The importance of surface melt as a feature

The performance of the emulator is better than its no-melt counterpart 3.3. The major difference is that the standard emulator is explicitly trained with surface melt data. Accumulation is the dominant source for new firn air. For locations with low or no surface melt, firn densification is the most important sink for firn air. But in locations with even a modest amount of surface melt, refreezing of percolating meltwater is the prime mechanism for firn air loss. Surface melt, in turn, is determined by the sum of all energy fluxes at the snow or ice surface. Those energy fluxes include (i) absorbed solar radiation, which strongly depends on albedo; (ii) net longwave radiation, which strongly depends on cloud cover and snow surface temperature; (iii) fluxes of latent and sensible heat that depend on wind speed and surface roughness; and (iv) the ground heat flux, which is determined by the thermal properties of the firn. For surface melt to be accurately simulated by a model, a somewhat realistic description of at least albedo, clouds, wind, surface roughness and snow thermodynamics is therefore needed. Also, an accurate digital elevation model at sufficient spatial resolution is needed for the simulated surface melt to be meaningful. The regional climate models in this study all have such descriptions at high spatial resolution, and thus their surface melt estimates can in principle be used as an input feature for the XGBoost model. However, in many global climate models, one or more of these ice sheet-specific processes typically is poorly represented. The resulting surface melt estimates cannot be relied on, and providing them as an input to an emulator will give biased, or simply meaningless results.

In the absence of a sophisticated snow scheme, near-surface air temperature has some predictive skill for surface melt over glaciated surfaces (e.g., Morris and Vaughan, 2003; Trusel et al., 2015). It is the foundation for temperature index-based methods to estimate surface melt, like positive degree-day models (e.g., Braithwaite, 1985). However, a known drawback to

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

such methods is that absorbed solar radiation, which strongly modulates surface melt over glaciated surfaces, is a function of albedo and not directly of temperature. Therefore, the accuracy of temperature-based surface melt estimates is always compromised. Nonetheless, it is the next best method to represent the effect of surface melt on firn air content in a framework like the emulators developed in this study.

By prescribing surface melt, the emulator doesn't have to "guess" a relation between air temperature, accumulation, and surface melt. The underlying processes and feedbacks like changes in snow albedo (Jakobs et al., 2019), roughness length, or snow thermal properties, are in a way included in the surface melt input. For the no-melt emulator, such relationships need to be implicitly "invented" by the emulator. It was therefore expected that the no-melt emulator would be less accurate.

It is doubtful whether the no-melt emulator can become as accurate as the normal emulator: Antarctic surface melt is intermittent in many places and therefore strongly modulated by the snow melt-albedo feedback, a process poorly captured by temperature index methods. It is expected to be hard to estimate the state of the firn by atmospheric parameters alone.

5.3 Emulator performance and limitations

Dunmire et al. (2024) emulated firn air content from a firn model using 10-year averages of annual and summer 2 m temperature, total annual precipitation, and mean annual wind speed achieving an R^2 of 0.96. This is a better performance than the no-melt emulator presented here, which uses similar input. The poorer performance in our study arises from our approach of strategically splitting the training and testing data rather than randomly. While this approach may result in a poorer performance, we think it also provides a more realistic estimate of emulator performance and reduces the risk of overfitting.

When forced by RACMO2 climate data, the differences between the emulator and the no-melt emulator are solely caused by the differences in performance. However, the comparison between the emulators of predicted firn air content using MAR and HIRHAM model output (Fig. 12) shows regional differences that are not only due to the lesser performance of the no-melt emulator, but also due to the driving regional climate models, which have different modelled relations between air temperature, accumulation and surface melt. It highlights a drawback of the emulator, namely that it requires melt estimates which are inherently uncertain also for models that have a relatively sophisticated snow scheme (and unavailable for models without such schemes). Nevertheless, using the emulator forced with RACMO2, MAR and HIRHAM shows reasonable agreement on the snowmelt and accumulation conditions required for firn air depletion (defined here as firn air content < 2 m), highlighting the emulator's robustness when applied to these regional climate models (Figs. 13a-c). While the surface melt estimates are uncertain, using the emulator accounts for these uncertainties by incorporating the respective melt estimates from each regional climate model, thereby implicitly reflecting the differences in e.g. their albedo schemes. Similar results would also be observed if IMAU-FDM were directly forced with the outputs of these regional climate models. In contrast, the no-melt emulator implicitly includes the temperature-surface melt relationships of RACMO2, which are then applied to the other regional climate models.

It is important to note that we did not train the emulator using IMAU-FDM simulations forced by overshoot scenarios, which are scenarios where global temperatures temporarily exceed a target threshold before eventually decreasing to pre-overshoot values. When projections involving these scenarios are needed, we recommend performing such IMAU-FDM simulations and

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

incorporating them into the training data and adjusting the predictive variables to accommodate for the asymmetry in the response. Firn response time scales differ and can be long, especially upon reversal of climate conditions (Van Wessem et al., 2023). Additionally, extending IMAU-FDM simulations and emulator training beyond 2100 would be valuable when applying forcing data beyond 2100 to the emulator.

5.4 Meltwater ponding

Although the focus of this study was to design, test, and implement an emulator of firn air content, the emulator runs also yielded predictions of firn air depletion, a proxy for meltwater ponding and vulnerability to hydrofracturing on Antarctic ice shelves. For this a threshold of 2 m of firn air content was used. However, this threshold likely varies across climatic regions. For example, in areas with low melt rates (e.g. 50 $\mathrm{mm\,w.e.\,yr^{-1}}$), a firn air content of 2 m may suffice to retain all meltwater, given a typical irreducible water content value of 5-15 % (Brils et al., 2024). In contrast, under high melt rates (e.g. 500 $\mathrm{mm\,w.e.\,yr^{-1}}$) a firn air content of 2 m may result in meltwater runoff. Determining an appropriate threshold requires further investigation, which could involve satellite observations of meltwater ponds in combination with FAC simulations, or the development of theoretical thresholds.

Historical firn air content simulations suggest that, at the time of collapse of Wilkins, Prince Gustav and Wordie ice shelves, firn air content was still considerable (> 5 m) (Fig. 11), although admittedly the climate prior to the satellite era (< 1979) is poorly known and therefore the model runs uncertain. Typically, these ice shelves experienced melt rates exceeding 500 $\mathrm{mm\,w.e.\,yr^{-1}}$ and accumulation rates more than 800 $\mathrm{mm\,w.e.\,yr^{-1}}$, which are favorable conditions for perennial firn aquifers to form (Kuipers Munneke et al., 2014b; Brils et al., 2024). In situ observations have confirmed the presence of a firn aquifer on the Wilkins ice shelf (Montgomery et al., 2020). Perennial firn aquifers are year-round subsurface bodies of liquid water within the firn, which can potentially contribute to ice-shelf collapse by year-round water drainage into fractures, thereby preventing refreezing and eventually triggering hydrofracturing. The high remaining modelled firn air content upon collapse suggests that aquifers may also play a role in the collapse of ice shelves, and that meltwater ponds may not be the only prerequisite for this process. Thus, solely using firn air content to assess ice-shelf vulnerability to hydrofracturing may underestimate the vulnerability in such regions.

In this study, we use firn air content, which is the vertically integrated amount of pore space, to determine the potential for meltwater ponding on Antarctic ice shelves. However, Veldhuijsen et al. (2024) show that near-surface multi-meter thick impermeable ice slabs can form in future IMAU-FDM simulations (2015-2100), especially over dry ice shelves. These ice slabs can partly or fully impede vertical meltwater percolation to deeper firn layers, which limits the fraction of the firn that is accessible to meltwater. Thus, our approach of using total firn air content could potentially underestimate meltwater ponding potential for dry locations where ice slabs form in the future. However, ice slab formation and its interaction with meltwater, is a poorly understood process for which almost no observations exist (Tedstone et al., 2025).

Since our emulator is trained on IMAU-FDM data, uncertainties of the firn model are conveyed to the emulator. For instance, IMAU-FDM does not include any lateral transport of energy and mass. Figures 12a,b show that firn air content on some grounded parts of the ice sheet is expected to reach depletion levels under high warming scenarios. Meltwater generated here

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

will likely drain to the lower-lying ice shelves, a process not included in the training data. In addition, uncertainties in the climate forcing of RACMO2 will persist in the emulator.

6 Conclusions

An XGBoost machine learning emulator was developed to predict future firn air content across Antarctic ice shelves. The emulator, which requires surface melt to be prescribed, explains > 94 % of the firn air content variance. An alternative, no-melt emulator, which requires no surface melt as input, explains only > 82 %. The no-melt emulator had higher rates of false negatives and false positives compared to the original emulator. Using 1980-2100 climate projections from the regional climate models MAR and HIRHAM as a forcing, notable differences in firn air content predictions are revealed between the emulator and its no-melt counterpart (RMSD = 2.5-2.6 m). The lower robustness of the latter can be linked to, for example, the importance of the snowmelt-albedo feedback.

Using the emulator forced with RACMO2, MAR and HIRHAM shows reasonable agreement on the snowmelt and accumulation conditions required for firn air depletion (firn air content < 2 m). While the surface melt estimates are uncertain, also for climate models with a relatively sophisticated snow surface scheme, using the emulator accounts for these uncertainties by incorporating the surface melt estimates from each regional climate model, thereby implicitly reflecting the differences in e.g. their albedo schemes. Similar results would also be observed if IMAU-FDM were directly forced with the outputs of these models. In contrast, the no-melt emulator implicitly includes the temperature-surface melt relations of RACMO2, which are applied to the other regional climate models. It could be improved by including more temporal details of the 2 m temperature, such as positive degree days. This would be useful, as the no-melt emulator would enable estimating firn air content for climate forcings from a wider range of climate models. All in all, it appears that emulating firn air content, a prerequisite to rapidly model meltwater ponding potential, is viable. This opens new avenues to communicate to ice-sheet models the locations and timing of ice shelves becoming susceptible to hydrofracturing following firn air depletion.

Code and data availability. The code and data for training and evaluation of the emulator will be made publicly available through Zenodo upon acceptance of this paper. Also, the emulator itself will be published on GitHub.

Author contributions. S.V. conceived this study, performed the analysis and the figures, and wrote a chapter in her PhD thesis that formed the basis for this manuscript. W.J.B., S.V., M.v.d.B. and P.K.M. discussed the results. P.K.M. rewrote the thesis chapter into this manuscript, expanded the discussion, and was responsible for the review process. L.T. provided the surface melt emulator input for the no-melt emulator.

Competing interests. At least one of the (co-)authors is a member of the editorial board of The Cryosphere.

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

Acknowledgements. MvdB is supported by the Horizon Europe OCEAN:ICE project (grant No. 101059388), the EMBRACER project financed by the Netherlands Organisation for Scientific Research (NWO, Summit grant SUMMIT.1.034) and the European Research Council Synergy Grant project FirnMelt (ERC grant No. 101224055). This work was funded by the Netherlands Organization for Scientific Research (grant no. OCENW.GROOT.2019.091)

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

References

Anilkumar, R., Bharti, R., Chutia, D., and Aggarwal, S. P.: Modelling point mass balance for the glaciers of the Central European Alps using machine learning techniques, The Cryosphere, 17, 2811–2828, https://doi.org/10.5194/tc-17-2811-2023, 2023.

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: Earth Surface, 115, https://doi.org/10.1029/2009JF001306, 2010.

Banwell, A. F., MacAyeal, D. R., and Sergienko, O. V.: Breakup of the Larsen B Ice Shelf triggered by chain reaction drainage of supraglacial lakes, Geophysical Research Letters, 40, 5872–5876, https://doi.org/10.1002/2013GL057694, 2013.

Braithwaite, R. J.: Calculation of degree-days for glacier-climate research, Z. Gletscherkd. Glazialgeol., 20, 1–8, 1985.

Breiman, L.: Random forests, Machine learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Brils, M., Munneke, P. K., Jullien, N., Tedstone, A. J., Machguth, H., van de Berg, W., and van den Broeke, M.: Climatic drivers of ice slabs and firn aquifers in Greenland, Geophysical Research Letters, 51, e2023GL106 613, https://doi.org/10.1029/2023GL106613, 2024.

Bromwich, D. H., Ensign, A., Wang, S.-H., and Zou, X.: Major artifacts in ERA5 2-m air temperature trends over Antarctica prior to and during the modern satellite era, Geophysical Research Letters, 51, e2024GL111 907, https://doi.org/10.1029/2024GL111907, 2024.

Carter, J., Leeson, A., Orr, A., Kittel, C., and van Wessem, J. M.: Variability in Antarctic surface climatology across regional climate models and reanalysis datasets, The Cryosphere, 16, 3815–3841, https://doi.org/10.5194/tc-16-3815-2022, 2022.

Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining,13–17 August 2016, San Francisco, CA, USA, pp. 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Coléou, C., Xu, K., Lesaffre, B., and Brzoska, J.-B.: Capillary rise in snow, Hydrological processes, 13, 1721–1732, 1999.

Cook, A. J. and Vaughan, D. G.: Overview of areal changes of the ice shelves on the Antarctic Peninsula over the past 50 years, The cryosphere, 4, 77–98, https://doi.org/10.5194/tc-4-77-2010, 2010.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., Emmons, L., Fasullo, J., Garcia, R., Gettelman, A., et al.: The community earth system model version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, e2019MS001 916, https://doi.org/10.1029/2019MS001916, 2020.

Dunmire, D., Lenaerts, J., Datta, R. T., and Gorte, T.: Antarctic surface climate and surface mass balance in the Community Earth System Model version 2 during the satellite era and into the future (1979–2100), The Cryosphere, 16, 4163–4184, https://doi.org/10.5194/tc-17-3847-2023, 2022.

Dunmire, D., Wever, N., Banwell, A. F., and Lenaerts, J. T.: Antarctic-wide ice-shelf firn emulation reveals robust future firn air depletion signal for the Antarctic Peninsula, Communications Earth & Environment, 5, 100, https://doi.org/10.1038/s43247-024-01255-4, 2024.

Esch, J., van den Broeke, M., and van Tiggelen, M.: Making the case for a two-step evaluation of regional climate models: application to the melt-over-accumulation ratio in Antarctica in RACMO2.3p2, Annals of Glaciology, 66, e24, https://doi.org/10.1017/aog.2025.10021, 2025.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Gilbert, E. and Kittel, C.: Surface melt and runoff on Antarctic ice shelves at 1.5 C, 2 C, and 4 C of future warming, Geophysical Research Letters, 48, e2020GL091 733, https://doi.org/10.1029/2020GL091733, 2021.

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

Goldberg, D., Holland, D., and Schoof, C.: Grounding line movement and ice shelf buttressing in marine ice sheets, Journal of Geophysical Research: Earth Surface, 114, https://doi.org/10.1029/2008JF001227, 2009.

Gossart, A., Helsen, S., Lenaerts, J., Broucke, S. V., Van Lipzig, N., and Souverijns, N.: An evaluation of surface climatology in state-of-the-art reanalyses over the Antarctic Ice Sheet, Journal of Climate, 32, 6899–6915, https://doi.org/10.1175/JCLI-D-19-0030.1, 2019.

Grinsztajn, L., Oyallon, E., and Varoquaux, G.: Why do tree-based models still outperform deep learning on typical tabular data?, Advances in neural information processing systems, 35, 507–520, https://proceedings.neurips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf, 2022.

Hansen, N., Langen, P. L., Boberg, F., Forsberg, R., Simonsen, S. B., Thejll, P., Vandecrux, B., and Mottram, R.: Downscaled surface mass balance in Antarctica: impacts of subsurface processes and large-scale atmospheric circulation, The Cryosphere, 15, 4315–4333, https://doi.org/10.5194/tc-15-4315-2021, 2021.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Iban, M. C. and Bilgilioglu, S. S.: Snow avalanche susceptibility mapping using novel tree-based machine learning algorithms (XGBoost, NGBoost, and LightGBM) with eXplainable Artificial Intelligence (XAI) approach, Stochastic Environmental Research and Risk Assessment, 37, 2243–2270, https://doi.org/10.1007/s00477-023-02392-6, 2023.

Jakobs, C. L., Reijmer, C. H., Kuipers Munneke, P., König-Langlo, G., and Van Den Broeke, M. R.: Quantifying the snowmelt–albedo feedback at Neumayer Station, East Antarctica, The Cryosphere, 13, 1473–1485, https://doi.org/10.5194/tc-13-1473-2019, 2019.

Jakobs, C. L., Reijmer, C. H., Smeets, C. J. P. P., Trusel, L. D., van de Berg, W. J., van den Broeke, M. R., and van Wessem, J. M.: A benchmark dataset of in situ Antarctic surface melt rates and energy balance, Journal of Glaciology, 66, 291–302, https://doi.org/10.1017/jog.2020.6, 2020.

Kittel, C., Amory, C., Agosta, C., Jourdain, N. C., Hofer, S., Delhasse, A., Doutreloup, S., Huot, P.-V., Lang, C., Fichefet, T., et al.: Diverging future surface mass balance between the Antarctic ice shelves and grounded ice sheet, The Cryosphere, 15, 1215–1236, https://doi.org/10.5194/tc-15-1215-2021, 2021.

Kuipers Munneke, P., Ligtenberg, S. R., Van Den Broeke, M. R., and Vaughan, D. G.: Firn air depletion as a precursor of Antarctic ice-shelf collapse, Journal of Glaciology, 60, 205–214, https://doi.org/10.3189/2014JoG13J183, 2014a.

Kuipers Munneke, P., M. Ligtenberg, S., Van Den Broeke, M., Van Angelen, J., and Forster, R.: Explaining the presence of perennial liquid water bodies in the firn of the Greenland Ice Sheet, Geophysical Research Letters, 41, 476–483, 2014b.

Lai, C.-Y., Kingslake, J., Wearing, M. G., Chen, P.-H. C., Gentine, P., Li, H., Spergel, J. J., and van Wessem, J. M.: Vulnerability of Antarctica's ice shelves to meltwater-driven fracture, Nature, 584, 574–578, https://doi.org/10.1038/s41586-020-2627-8, 2020.

Montgomery, L., Miège, C., Miller, J., Scambos, T. A., Wallin, B., Miller, O., Solomon, D. K., Forster, R., and Koenig, L.: Hydrologic properties of a highly permeable firn aquifer in the Wilkins Ice Shelf, Antarctica, Geophysical Research Letters, 47, e2020GL089 552, https://doi.org/10.1029/2020GL089552, 2020.

Morris, E. M. and Vaughan, D. G.: Spatial and temporal variation of surface temperature on the Antarctic Peninsula and the limit of viability of ice shelves, vol. 79, pp. 61–68, AGU, Washington, DC, ISBN 0-87590-973-6, https://doi.org/10.1029/AR079p0061, 2003.

https://doi.org/10.5194/egusphere-2026-3453
Preprint. Discussion started: 17 August 2026

Nowicki, S., Payne, A. J., Goelzer, H., Seroussi, H., Lipscomb, W. H., Abe-Ouchi, A., Agosta, C., Alexander, P., Asay-Davis, X. S., Barthel, A., et al.: Experimental protocol for sealevel projections from ISMIP6 standalone ice sheet models, The Cryosphere Discussions, 2020, 1–40, https://doi.org/10.5194/tc-14-2331-2020, 2020.

Paolo, F. S., Fricker, H. A., and Padman, L.: Volume loss from Antarctic ice shelves is accelerating, Science, 348, 327–331, https://doi.org/10.1126/science.aaa0940, 2015.

Pfeffer, W. T., Meier, M. F., and Illangasekare, T. H.: Retention of Greenland runoff by refreezing: implications for projected future sea level change, Journal of Geophysical Research: Oceans, 96, 22 117–22 124, https://doi.org/10.1002/qj.4138, 1991.

Rignot, E., Casassa, G., Gogineni, P., Krabill, W., Rivera, A., and Thomas, R.: Accelerated ice discharge from the Antarctic Peninsula following the collapse of Larsen B ice shelf, Geophysical research letters, 31, https://doi.org/10.1029/2004GL020697, 2004.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography, 40, 913–929, https://doi.org/10.1111/ecog.02881, 2017.

Scambos, T., Hulbe, C., and Fahnestock, M.: Climate-induced ice shelf disintegration in the Antarctic Peninsula, Antarctic Peninsula climate variability: Historical and paleoenvironmental perspectives, pp. 79–92, https://doi.org/10.1029/ar079p0079, 2003.

Seroussi, H., Pelle, T., Lipscomb, W. H., Abe-Ouchi, A., Albrecht, T., Alvarez-Solas, J., Asay-Davis, X., Barre, J.-B., Berends, C. J., Bernales, J., et al.: Evolution of the Antarctic Ice Sheet over the next three centuries from an ISMIP6 model ensemble, Earth's Future, 12, e2024EF004 561, https://doi.org/10.1029/2024EF004561, 2024.

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/10.5194/tc-5-809-2011, 2020.

Smith, J. A., Bentley, M. J., Hodgson, D. A., and Cook, A. J.: George VI Ice Shelf: past history, present behaviour and potential mechanisms for future collapse, Antarctic Science, 19, 131–142, https://doi.org/10.1017/S0954102007000193, 2007.

Sun, L., Zhang, X., Xiao, P., Wang, H., Wang, Y., and Zheng, Z.: Fusing daily snow water equivalent from 1980 to 2020 in China using a spatiotemporal XGBoost model, Journal of Hydrology, 632, 130 876, https://doi.org/10.1016/j.jhydrol.2024.130876, 2024.

Tedstone, A., Machguth, H., Clerx, N., Jullien, N., Picton, H., Ducrey, J., van As, D., Colosio, P., Tedesco, M., and Lhermitte, S.: Concurrent superimposed ice formation and meltwater runoff on Greenland's ice slabs, Nature Communications, 16, 4494, https://doi.org/10.1038/s41467-025-59237-9, 2025.

The Firn Symposium Team: Firn on ice sheets, Nature Reviews Earth & Environment, 5, 79–99, https://doi.org/10.1038/s43017-023-00507-9, 2024.

Thoma, M., Jenkins, A., Holland, D., and Jacobs, S.: Modelling circumpolar deep water intrusions on the Amundsen Sea continental shelf, Antarctica, Geophysical Research Letters, 35, https://doi.org/10.1029/2008GL034939, 2008.

Trusel, L. D., Frey, K. E., Das, S. B., Karnauskas, K. B., Kuipers Munneke, P., Van Meijgaard, E., and Van Den Broeke, M. R.: Divergent trajectories of Antarctic surface melt under two twenty-first-century climate scenarios, Nature Geoscience, 8, 927–932, https://doi.org/10.1038/ngeo2563, 2015.

Turner, R., Eriksson, D., McCourt, M., Kiili, J., Laaksonen, E., Xu, Z., and Guyon, I.: Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020, in: NeurIPS 2020 Competition and Demonstration Track, pp. 3–26, PMLR, 2021.

Van der Veen, C. J.: Fracture mechanics approach to penetration of surface crevasses on glaciers, Cold Reg. Sci. Technol., 27, 31–47, 1998.

Van Wessem, J. M., Van De Berg, W. J., Noël, B. P., Van Meijgaard, E., Amory, C., Birnbaum, G., Jakobs, C. L., Krüger, K., Lenaerts, J., Lhermitte, S., et al.: Modelling the climate and surface mass balance of polar ice sheets using RACMO2–Part 2: Antarctica (1979–2016), The Cryosphere, 12, 1479–1498, https://doi.org/10.5194/tc-12-1479-2018, 2018.

Van Wessem, J. M., van den Broeke, M. R., Wouters, B., and Lhermitte, S.: Variable temperature thresholds of melt pond formation on Antarctic ice shelves, Nature Climate Change, 13, 161–166, https://doi.org/10.1038/s41558-022-01577-1, 2023.

Veldhuijsen, S. B., Van De Berg, W. J., Kuipers Munneke, P., and Van Den Broeke, M. R.: Firn air content changes on Antarctic ice shelves under three future warming scenarios, The Cryosphere, 18, 1983–1999, https://doi.org/10.5194/tc-18-1983-2024, 2024.

Veldhuijsen, S. B. M., van de Berg, W. J., Kuipers Munneke, P., Hansen, N., Boberg, F., Kittel, C., Amory, C., and van den Broeke, M. R.: Emulating the expansion of Antarctic perennial firn aquifers in the 21st century, The Cryosphere, 19, 5157–5173, https://doi.org/10.5194/tc-19-5157-2025, 2025.

Verjans, V., Leeson, A., McMillan, M., Stevens, C., van Wessem, J. M., van de Berg, W. J., van den Broeke, M., Kittel, C., Amory, C., Fettweis, X., et al.: Uncertainty in East Antarctic firn thickness constrained using a model ensemble approach, Geophysical Research Letters, 48, e2020GL092 060, https://doi.org/10.1029/2020GL092060, 2021.

Wager, A.: Flooding of the ice shelf in George VI Sound, British Antarctic Survey Bulletin, 28, 71–74, 1972.