https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

Quantifying the Role of Parametric Uncertainty in Projections of Large-Scale Glacier Change

Megan C. James[1], Tamsin L. Edwards[1], Tom Matthews[1], Alexander T. Bradley[1], James O'Neill[2] and Harry Zekollari[3]

[1] Department of Geography, King's College London, London, United Kingdom
[2] University of Exeter, Exeter, United Kingdom
[3] Department of Water and Climate, Vrije Universiteit Brussel, Brussels, Belgium

Correspondence to: Megan C. James (megan.james@kcl.ac.uk)

Abstract. Large-scale glacier evolution models are widely used to generate projections of glacier mass change at regional- to global-scales. In model intercomparison projects, these projections come from multiple different models, allowing for the uncertainties associated with different model structures to be assessed. However, these intercomparisons tend to ignore the uncertainties associated with poorly constrained parameters. Therefore, these projections may miss important contributions to uncertainty, but we lack estimates of the size of these uncertainties. To bridge this gap, we quantify parametric uncertainty in projections of glacier volume change in Iceland under experiments from the glacier model intercomparison exercise, GlacierMIP3. To do so, we perform experiments with a large-scale glacier evolution model, 'GO-VA', using an ensemble of calibrated parameter sets, rather than with a single set of model parameters as was the case in GlacierMIP3. Our results show that parametric uncertainty can be a major, and in some cases dominant, source of uncertainty in projections of glacier volume change. We find that failing to account for parametric uncertainty reduces overall projection uncertainty by 7–91 % across scenarios of global mean temperature change, with the largest reductions occurring for scenarios where climate forcing uncertainty is highly constrained. Comparison with the GlacierMIP3 ensemble suggests that parametric uncertainty is comparable to structural model uncertainty and, depending on the strength of the forcing, can even be larger. Taken together, these findings highlight the importance of accounting for parametric uncertainty, alongside structural model uncertainty, in model intercomparison projects to more comprehensively characterise projection uncertainty.

1 Introduction

Glaciers are losing mass globally in response to anthropogenic climate change (IPCC, 2021), with profound implications for regional water availability (Immerzeel et al., 2020) and global mean sea level rise (Zemp et al., 2025). To understand how these impacts might unfold, numerical models are used to make projections of future glacier mass changes. However, these projections are subject to considerable uncertainty, arising in large part from the differences in model structure (structural model uncertainty), poorly constrained model parameters (parametric uncertainty), and uncertainty in future climate forcing (forcing uncertainty). To account for model structural and forcing uncertainty in projections, multi-model ensemble studies

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

by the Glacier Modelling Intercomparison Project (GlacierMIP: Hock et al., 2019; Marzeion et al., 2020; Zekollari et al., 2025) have simulated future glacier changes under different scenarios of emissions or climate change, using a range of global climate models and glacier evolution models.

However, parametric uncertainty has not been explicitly accounted for in existing model intercomparison projects. In the GlacierMIP ensemble, simulations from each model are derived from a fixed set of model parameter values (though parameters may vary from glacier-to-glacier or region-to-region). The value of these parameters is obtained through calibration, which is typically framed as an optimisation problem: this involves searching for the set of model parameters which minimise the difference between simulations and observations. This classical optimisation approach to calibration is widely used in glacier modelling (e.g. Huss and Hock, 2015; Maussion et al., 2019) and other fields, but has several critical limitations: (i) multiple parameter sets may be equally consistent with observed data, within the observational error bounds; (ii) structural model deficiencies may be compensated for by varying parameters, giving improved agreement with observations for the wrong reasons; and (iii) there are more model parameters than can be constrained by available observations. These limitations lead to a substantial possibility of over-tuning, and thus an under-estimation of projection uncertainties.

Calibration methods that incorporate parametric uncertainty have, to our knowledge, only been applied to one large-scale glacier evolution model. Rounce et al. (2020b) used a Bayesian calibration method – where simulations are weighted based on their agreement with glacier-specific mass balance observations (Shean et al., 2020) – to constrain the mass balance component of the Python Glacier Evolution Model (PyGEM: Rounce et al., 2020a). The projection uncertainties resulting from this calibration procedure are reported to be comparable to the structural model uncertainties in GlacierMIP2 (Felikson et al., 2025). To improve the characterisation of model uncertainty in projections, there is a need to quantify parametric uncertainty across additional large-scale glacier evolution models.

In this work, we use an alternative calibration method known as history matching (Craig et al., 1997). History matching is simpler than Bayesian calibration, and particularly suitable for the initial exploration and assessment of parametric uncertainty (Williamson et al., 2013). In this approach, regions of parameter space that produce simulations that are too far from observations, given model and observational uncertainties, are ruled out. The remaining region, termed the "not-ruled-out-yet" (NROY) space, contains parameter sets that are consistent with observations within a specified uncertainty threshold. History matching is an established statistical methodology that has been used to calibrate models and quantify parametric uncertainty across various fields (e.g. Williamson et al., 2017; Courvreux et al., 2021; Hourdin et al., 2023; Raoult et al., 2024; Lecavalier and Tarasov, 2024).

In this study, we quantify parametric uncertainty in projections of glacier volume change in Iceland in the framework of GlacierMIP3 (Zekollari et al., 2025). For this, we use the 'GO-VA' glacier model, which combines a modified version of the Giesen and Oerlemans (2012) mass balance model ('GO') with volume-area scaling methods ('VA'). The GO-VA model is included in the GlacierMIP3 ensemble for Iceland, where it has seven free model parameters that were calibrated to match glacier-specific mass balance observations. Here, we extend this work by performing an additional suite of simulations in

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

which these seven model parameters are systematically perturbed. Using a history matching calibration approach, we identify parameter sets that are deemed not-ruled-out-yet (NROY) based on observations of regional glacier mass change. We continue to focus on the single region of Iceland as a tractable proof of concept.

Through our modelling framework, we aim to (1) demonstrate the utility of history matching to calibrate glacier mass balance models for large-scale applications, and (2) quantify parametric uncertainty in an individual model from the GlacierMIP3 ensemble and compare this with model structural and forcing uncertainties. The second aim is particularly relevant, since the outputs from model intercomparison projects are used to inform climate impact assessments such as the IPCC (2021). A clearer understanding of model uncertainties, and the factors that drive them, is therefore essential for improving the credibility and robustness of these projections.

2 Methods

2.1 Glacier Evolution Model

We simulate the evolution of all glaciers in Iceland using a modified version of the mass balance model presented in Giesen and Oerlemans (2012), which we couple to a volume-area scaling method based on the approach of Radic et al. (2008); collectively referred to as the 'GO-VA' model. Several key model components relevant to this study are described below (see original publications for more detailed descriptions).

2.1.1 Mass Balance

The annual climatic mass balance (B) of a glacier is computed as a function of mass gain from solid precipitation (P_{snow}) and mass loss from melt runoff:

$$B = \int_{year} \left\{ P_{snow} + (1 - r) \min\left(0; -\frac{Q_m}{\rho_w L_f}\right) \right\} dt \quad (1)$$

In Eq. (1), melt occurs whenever the surface energy flux Q_m is positive; division by the density division by the density of water (ρ_w; 1000 kg m^{-3}) and latent heat of fusion (L_f; 3.34 x 10^5 J kg^{-1}) yields the melt rate. If the glacier surface is ice covered, all meltwater will immediately runoff from the glacier. However, in the presence of a snowpack, a fraction r of meltwater is allowed to refreeze, depending on the temperature in the subsurface layer of a glacier.

Each component of the mass balance (1) is updated in each glacier elevation band (Sect. 2.2) at a daily time-step (dt = 24 h) using precipitation (P) and air temperature (T_a) input data, which are distributed across each glacier using a linear precipitation (γ) and temperature (T_{lapse}) lapse rate (see Sect. 2.2). Precipitation is assumed to fall as snow whenever T_a is below a threshold temperature, T_{snow}. The surface energy flux Q_m is expressed as the sum of contributions from incoming solar radiation (first term below) and all other energy fluxes (second term):

$$Q_m = (1 - \alpha)\tau S_{in,toa} + \psi \quad (2)$$

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

Here, $S_{in,toa}$ is the magnitude of incoming solar radiation at the top of the atmosphere, which is determined from the latitude of a glacier using standard geometrical methods (e.g. Mölg et al., 2003). The fraction of $S_{in,toa}$ that reaches the surface is calculated by multiplying by the atmospheric transmissivity (τ) parameter. The surface albedo (α) of a glacier controls the fraction of solar radiation reflected at the glacier surface. The value of α depends on the surface cover type: after a snowfall event, α is given the value for snow albedo (α_{snow}), which exponentially decreases to the value of (α_{ice}) over time; for a thin snowpack, is a function of firn albedo (α_{firn}) and α_{ice} (Oerlemans and Knap, 1998; Giesen and Oerlemans, 2012). The melt contributions from all other energy fluxes (ψ) are combined into a bulk function of air temperature:

$$\psi = \begin{cases} \psi_{min} + cT_a & T_a \geq T_{tip} \\ \psi_{min} & T_a < T_{tip} \end{cases} \quad (3)$$

For air temperatures below the threshold temperature T_{tip}, the surface energy flux ψ takes the constant value ψ_{min}; otherwise, ψ increases linearly with T_a at a rate given by c. Note that in Eq. (3), ψ scales with the absolute air temperature T_a (in units of °C) rather than T_a relative to T_{tip} i.e., the first line of Eq. (3) is $\psi_{min} + cT_a$ rather than $\psi_{min} + c(T_a - T_{tip})$. This is consistent with the formulation in Giesen and Oerlemans (2012), but means that ψ is discontinuous at the threshold temperature T_{tip}. If both T_{tip} and T_a are negative, and $T_a \geq T_{tip}$, ψ will decrease relative to ψ_{min}. If T_{tip} and T_a are positive, an abrupt increase in ψ will occur when T_a exceeds T_{tip}.

2.1.2 Volume-Area Scaling

Geometry change is accounted for annually using volume-area scaling methods (Bahr et al., 1997). At the end of each mass balance year, the annual volume change $\Delta V\ (t)$ is calculated as

$$\Delta V(t) = \frac{\rho_{ice}}{\rho_w} \cdot A(t) \cdot B(t) \quad (4)$$

where $A(t)$ is the surface area at the start of the mass-balance year, $B(t)$ is the annual mass-balance and ρ_w and ρ_{ice} are the densities of water and ice, respectively (Marzeion et al., 2012).

At a given year t, the glacier volume $V(t)$ is updated to the next year as $V(t+1) = V(t) + \Delta V(t)$, and the area is updated using volume-area scaling (Bahr et al., 1997):

$$A(t+1) = \left[\frac{V(t+1)}{\theta}\right]^{\frac{1}{\gamma}} \quad (5)$$

where θ and γ are empirical parameters which determine the relationship between the volume and area of a glacier. Values of the scaling exponent γ are typically fixed for different geometric classes of glaciers: 1.375 for valley glaciers and 1.25 for ice caps (see Bahr et al., 2015). We estimate the value of the denominator θ for each glacier individually using observations of glacier volume and area (Sect. 2.2).

Following the approach from Radic et al. (2008), glacier geometry is updated by adding or removing area from the glacier front, depending on the calculated annual area change. In years with negative mass balance, elevation bands are removed if

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

annual area change exceeds the area in the lowest elevation band, to account for glacier retreat. For positive mass balance years, area is first added to the lowest elevation band. If the updated area in this elevation band is more than 1.5 times larger than the area of the elevation band immediately above, the area is capped at this threshold and any excess area is moved into new elevation bands below the glacier front, accounting for glacier advance. Each newly added elevation band is progressively larger than the one above to prevent rapid advance during sustained periods of positive mass balance (as in Huss and Hock, 2015).

2.2 Experiment Design

Experiments were set up following the GlacierMIP3 protocol (Schuster et al., 2024; Zekollari et al., 2025), whereby glacier equilibration is simulated under a range of constant climate change scenarios. In this section, we briefly describe this experimental setup.

The Randolph Glacier Inventory (RGI) version 6.0, which is a global dataset of individual glacier outlines, provides the starting point for the GlacierMIP3 experiments (RGI Consortium, 2017). The inventory is separated into 19 regions, and it is a requirement for participating modelling groups to submit simulations for at least one complete RGI region. In this study, the GlacierMIP3 experiments were carried out for the 568 individual glaciers in Iceland (RGI region 06, Figure 1). For each glacier, the ice geometry was initialised using RGI hypsometry data and consensus volume estimates from Farinotti et al. (2019), which were also used to obtain the volume-area scaling parameter (θ) for each glacier.

The central coordinates for each glacier were extracted from the RGI. Climate model data used as model inputs are from the nearest climate model grid cell to these central locations. To force the model for the calibration period (see Sect. 2.4), monthly (2m) air temperature and precipitation data were obtained for the 1990-2020 period from the fifth global reanalysis of the European Centre for Medium-Range Weather Forecasts, ERA5 (Hersbach et al., 2020). Monthly air temperatures at eight pressure levels between 500 and 1000 hPa were used to compute lapse rates following Huss and Hock (2015), resulting in a spatially variable lapse rate ranging from – 6.2 to – 5.0 K km^{-1} to transfer the temperatures from the ERA5 grid to the glacier level.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

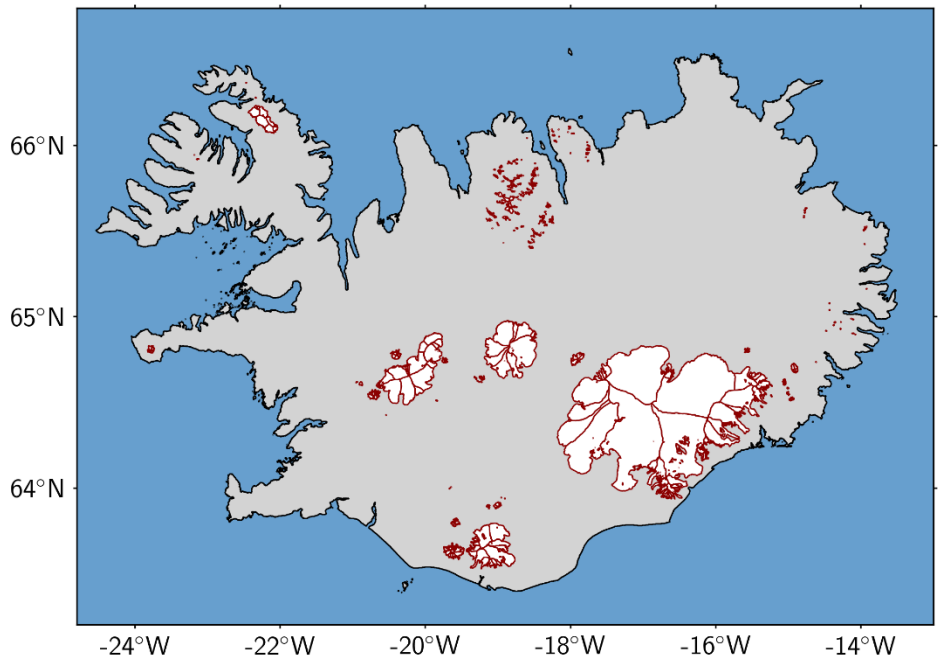

Figure 1. Distribution of glaciers (white polygons) in Iceland. Outlines (red) correspond to glacier geometries in RGI v6.0 (RGI Consortium, 2017).

Following the GlacierMIP3 protocol, we construct 5000-year climate forcings by randomly repeating years of climate model data from selected 20-year periods (Schuster et al., 2024; Zekollari et al., 2025). Sixteen different 20-year periods were taken from monthly temperature and precipitation simulations from five global climate models (GCMs) from the Coupled Model Intercomparison Project Phase 6 (CMIP6: Eyring et al., 2016), provided by the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP3b). This includes four historical periods (1851-1870, 1901-1920, 1951-1970, 1995-2014), and four future periods (2021-2040, 2041-2060, 2061-2080, 2081-2100) for each of three different Shared Socioeconomic Pathways corresponding to low, medium and very high emissions (SSP1-2.6, SSP3-7.0 and SSP5-8.5, respectively). The data were used to generate 80 series of 5000-year forcings, from the four historical periods and twelve future periods for each of the five GCMs. Each of these climate forcings was individually bias-corrected over the full time-series (1850-2100) with fixed additive (temperature) and multiplicative (precipitation) monthly biases, derived from a comparison with monthly ERA5 data for an overlapping reference period from 2000-2019. An example of the forcing experiments extracted from one GCM (MRI-ESM2-0) is given in Fig. 2.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

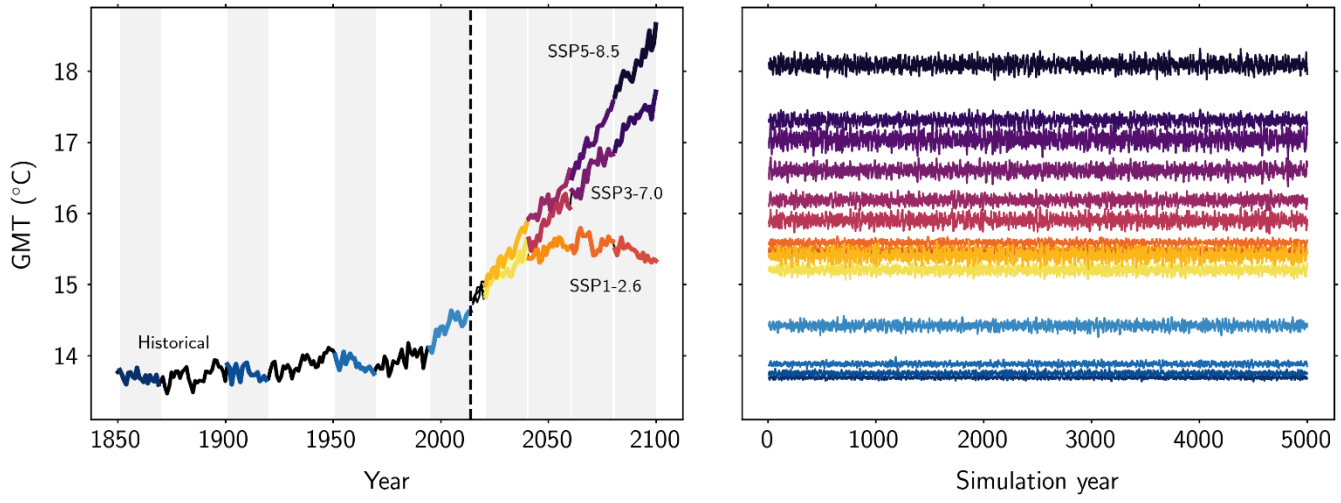

Figure 2. (a) Example of GCM (MRI-ESM2-0) time series, which were used to construct (b) 5000-year shuffled time-series of the climate in four historical periods (1851-1870, 1901-1920, 1951-1970, 1995-2014), and four future periods (2021-2040, 2041-2060, 2061-2080, 2081-2100) from the SSP1-2.6, SSP3-7.0 and SSP5-8.5 scenarios (see main text for a full description of these time-series).

2.3 Ensemble Design

All experiments were performed using a perturbed-parameter ensemble containing seven model parameters, which are summarised in Table 1. These include the five parameters described as 'tuneable' in Giesen and Oerlemans (2012), alongside the precipitation correction factor (p) and the phase partitioning temperature (T_{snow}). In Giesen and Oerlemans (2012), p is derived from winter balance profiles, and T_{snow} is fixed; here, all seven parameters are varied within the ensemble. These parameters can be broadly categorised based on their influence on the different mass balance components: temperature-dependent heat fluxes (ψ_{min}, T_{tip}, c), net shortwave radiative fluxes (α_{ice}, τ) and accumulation (p, T_{snow}).

For the parameters described as tuneable by Giesen and Oerlemans (2012), i.e., $\psi_{min}, T_{tip}, c, \alpha_{ice}, \tau$, parameter ranges were determined from an iterative procedure: we used the 'global' values in Giesen and Oerlemans (2012) as the starting point, and expanded the ranges in successive iterations until we observed a decline in agreement between the model and observations during the calibration period at the edges of the parameter ranges. This approach ensures that the uncalibrated ensemble comprehensively samples parameter space, without omitting regions that the calibration might identify as plausible (Sect. 2.4).

For the other two parameters, precipitation correction factor (p) and phase partitioning threshold temperature (T_{snow}), ranges were derived from the literature. For the precipitation correction factor parameter, we chose a lower bound of 0.5 (corresponding to a halving of the precipitation) because of the suspected positive ERA5 bias in Iceland (Compagno et al., 2022), while the upper limit of 2.5 is in line with values used in previous studies (e.g. Huss and Hock, 2015; Zekollari et al., 2024). We use a range of – 1 °C to + 3 °C for the precipitation phase partitioning threshold temperature, which is in line with previous large-scale modelling studies ranging from -1 °C (Huss and Hock, 2015) to + 3 °C (Marzeion et al., 2012).

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

Table 1. Descriptions, ranges and units for the seven model parameters included in the ensemble design.

Description	Parameter	Min	Max	Units
Ice albedo	α_{ice}	0.15	0.55	-
Atmospheric transmissivity	τ	0.3	0.7	-
Temperature-dependent flux sensitivity to temperature change	c	5	55	$Wm^{-2}\,K^{-1}$
Minimum value of temperature-dependent fluxes	ψ_{min}	- 55	- 15	Wm^{-2}
Threshold temperature for the temperature-dependent fluxes	T_{tip}	- 1.5	4	°C
Precipitation correction factor	p	0.5	2.5	-
Precipitation phase partitioning threshold temperature	T_{snow}	-1	3	°C

The ensemble of parameter sets was generated using Latin hypercube sampling of the underlying parameter space. Latin hypercube sampling is an efficient sampling method that partitions parameter ranges into equal probability bins, then randomly draws a sample from each of these bins (see for instance McKay et al., 1979). Here, a sample size of 250 is used, which provides good coverage of parameter space whilst remaining computationally feasible. This sample size is larger than the 'rule of thumb' of 10 times the input dimensions (Loeppky et al., 2009) because we found this number resulted in inadequate sampling of certain areas of parameter space. A total of 20,000 projections (250 ensemble members x 80 climate forcings) were therefore generated. Owing to the simple volume-area scaling approach used in this study, and our focus on a single region, computational time remained manageable even with the long simulations, which amounted to 10^8 model years for each individual glacier.

A small fraction of ensemble members produced glacier volume changes that were unphysical: either growing unrealistically large or failing to reach steady-state conditions by the end of the simulation period. These simulations were therefore excluded from the ensemble. We applied a data quality filter requiring at least 98% of regional glacier volume to be successfully simulated, which reduced the ensemble size by 0.8%. Remaining ensemble members were upscaled so that the initial regional glacier volume matched consensus estimates from Farinotti et al. (2019), assuming that any missing glaciers follow the same volume change trajectory as those successfully simulated.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

2.4 Ensemble Calibration

For the calibration, the 250-member perturbed-parameter ensemble is used to simulate all glaciers from 1990 to 2019, including a 10-year spin up period (1990-1999). Glacier geometry was fixed during this period to limit the uncertainty introduced from the volume-area scaling methods (as in Huss and Hock, 2015; Rounce et al., 2020a).

We use two different methods to calibrate the ensemble. In the first, referred to as the 'tuning' procedure and used for the GlacierMIP3 GO projections, a single 'optimal' parameter set (x^*) is selected from the ensemble for each individual glacier which minimises the absolute error between simulated and observed glacier mass balance:

$$x^* = \underset{x}{\operatorname{argmin}} \left| z_j - f_j(x) \right| \tag{5}$$

where $f_j(x)$ and z_j are the simulated and observed mass balance, respectively, for each individual glacier j. The glacier-specific mass balance observations are from January 2000 to December 2019 (Hugonnet et al., 2021); we compare the mean mass balance over this 20-year period to reduce observational uncertainties.

The second calibration procedure is history matching (Craig et al., 1997), where regions of the parameter space are ruled out if they are inconsistent with observations, given observational and model uncertainties. To implement this approach, each parameter set is uniformly applied across all glaciers and annual glacier mass changes are calculated. For each ensemble member, mass changes are then summed across all glaciers, and averaged over four five-year periods between 2000 and 2019. The modelled regional glacier mass changes are then compared with regional mass change observations at the same temporal resolution (also from Huggonet et al., 2021). Comparisons were made at the regional scale, rather than for individual glaciers, to avoid over-tuning of parameters and to minimise the need for judgements about model error (see below, and Discussion). Five-year means were used because the uncertainties in regional mass change observations are smaller than the glacier-specific mass balance observations.

In more detail, for each ensemble member, we quantify the difference between simulated and observed regional mass balance using the implausibility metric, which is given by

$$I_i(x) = \frac{f_i(x) - z_i}{\sqrt{\sigma_z^2 + \sigma_f^2}} \tag{6}$$

where $f_i(x)$ and z_i are the modelled and observed mean regional mass changes for each 5-year period denoted by the subscript i. The denominator is the total model error, comprised of observational uncertainty (σ_z) and model error (σ_f). Observational uncertainties are provided by Hugonnet et al. (2021). The model error term, often described in terms of model discrepancy variance (σ_f^2), represents our tolerance to error (e.g. Williamson et al., 2013). Not including σ_f would represent

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

an implicit assumption that our model could, at its best parameter values, reproduce observations perfectly. Following Edwards et al., (2019), we define our model error as a multiple of observational uncertainty, and set σ_f equal to five times σ_z. The implications of this choice are discussed in Appendix A.

The implausibility metric is calculated for each of the 5-year periods over the calibration period using Eq. (6). Here, we consider the maximum implausibility value, and define a threshold so that if $\max(I_i(x)) > a$, x is deemed to be implausible and removed from the ensemble; otherwise, it is deemed 'not-ruled-out-yet' (NROY) and retained. A value of 3 is typically used for a when calibrating with one observation (Williamson et al., 2015; Edwards et al., 2019), based on Pukelsheim's (1994) 3-sigma rule that at least 95 % of the probability density for any unimodal distribution will be contained within 3 standard deviations of the mean. Since multiple observations were used in this study, applying the single-observation threshold would underestimate the overall error rate: with four tests, the probability of falsely rejecting at least one plausible input rises to ~19 %, compared with 5 % for a single observation. Following methods described in Gladstone et al. (2012), we therefore derived a new value of a = 5.91 to account for the multiple observations.

2.5 Sensitivity Analysis

To assess the relationship between the model parameters and model outputs, we fit a LOcally WEighted Scatterplot Smoothing (LOWESS) model to each parameter-output combination, returning the median fit and 17-83 % uncertainty range. Each local regression is based on 80 % of the data, with multiple resampling fits performed to ensure the stability of the median trends and two robust iterations applied to down-weight the influence of outliers (Bourn, 2021). To assess these relationships over the calibration period, we compute the mean annual mass balance, as well as the total daily accumulation, temperature-dependent flux, and shortwave radiative flux for each individual glacier. These values are averaged over the calibration period for each glacier, and then averaged across all glaciers (area-weighted) for each ensemble member. This aggregation allows us to relate regionally-averaged glacier behaviour and model parameter values.

3. Results

3.1. Ensemble Behaviour over the Calibration Period

In this section, we describe the output of the mass balance model over the calibration period (2000 to 2019) using the full (Sect. 3.1.1) and calibrated (Sect. 3.1.2) ensembles.

3.1.1. Full Ensemble

Relationships between the seven perturbed mass balance model parameters and outputs over the calibration period are shown in Fig. 3 for the full 250-member ensemble. Over the calibration period, the mean regional mass balance B (Figure 3a-g) values range from –7.8 to 3.2 m w.e, with a median value of –1.6 m w.e. The parameter controlling the sensitivity of the

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

temperature-dependent fluxes to temperature change (temperature-dependent flux gradient, c) is the primary control on the annual mass balance over the calibration period (Figure 3b), with increasing values of c generally resulting in more negative mass balance values. This is to be expected because higher c increases the temperature-dependent flux gain (once the air temperature is above the threshold T_{tip}). Simulations with lower values of c (5 $Wm^{-2}K^{-1}$) produce mass balance values with a median (17-83% range) of -0.3 (-1.4, 0.2) m w.e, whereas higher values of c (45 $Wm^{-2}K^{-1}$) reduce this to -3.8 (-4.8, -3.2) m w.e. All intervals presented are 17-83% ranges, also described as 'likely' (IPCC, 2021), unless otherwise indicated.

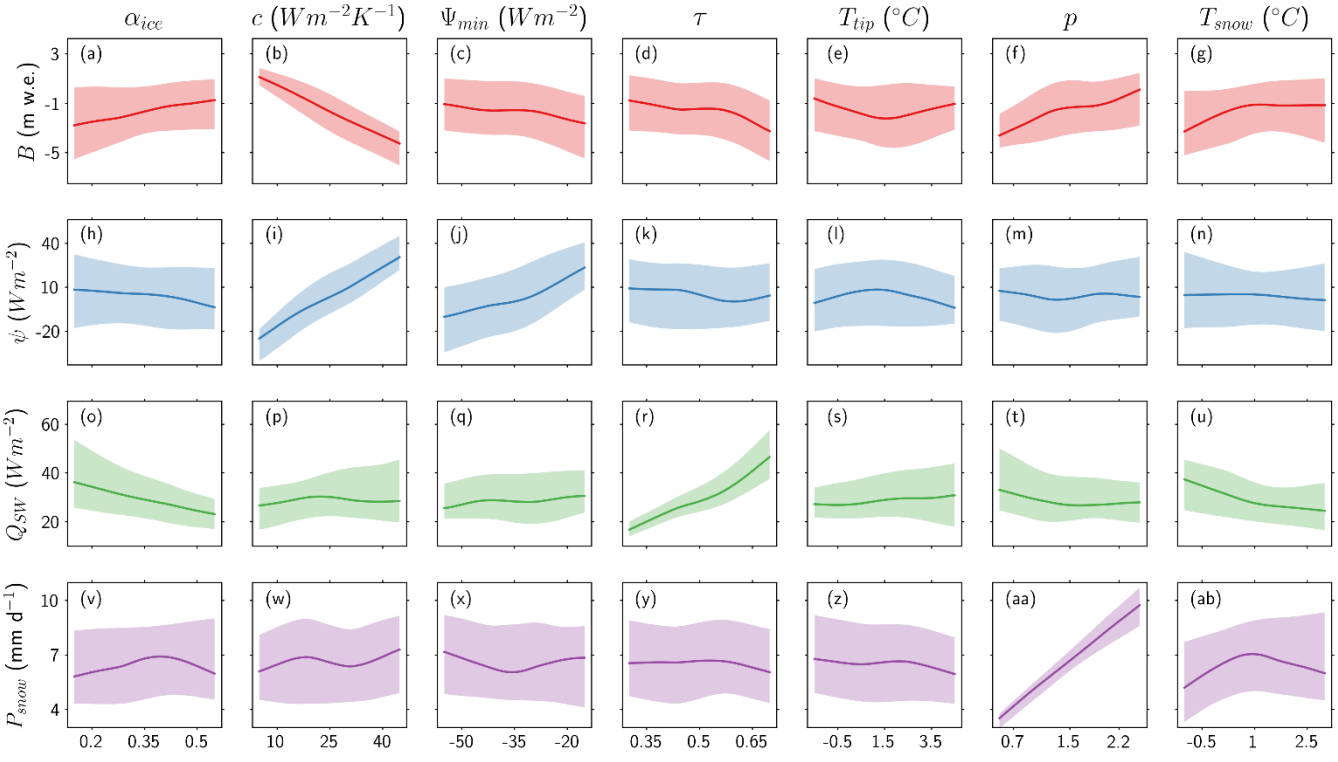

Figure 3. Sensitivity of model outputs to input parameters over the calibration period (2000-2019) from simulations using the uncalibrated ensemble (i.e., all 250 parameter sets). Different rows correspond to different mass balance components as follows: annual mass balance ($B, m\,w.e$) (first row), temperature-dependent heat fluxes (ψ, $Wm^{-2}K^{-1}$) (second row), short-wave radiative fluxes (Q_m, $Wm^{-2}K^{-1}$) (third row) and accumulation (P_{snow}, mm d^{-1}) (fourth row). Columns correspond to model parameters as labelled. For each parameter-output relationship, the solid curve indicates the LOWESS median fit and shaded area indicates the 17-83% uncertainty range (defined as the 'likely' range).

The other model parameters have a less dominant effect on B, though do show discernible impacts on individual mass balance components (Figure 3h-ab). Consistent with physical expectations, the temperature-dependent flux (ψ) are most sensitive to c, though this is also sensitive to the minimum value of the temperature-dependent flux (ψ_{min}) (Figure 3i). The short-wave radiative flux (Q_{SW}) is most sensitive to the atmospheric transmissivity (τ) parameter, with an increasing influence at higher values (Figure 3r). Higher values of the ice albedo (α_{ice}) lead to lower values of Q_{SW}, but the effect is

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

less significant than the atmospheric transmissivity (τ) parameter (Figure 3o). Accumulation (P_{snow}) has a strong linear dependence on the precipitation correction factor (p) parameter, with the median increasing by 5.3 mm d^{-1} over the range of p values. The dominance of this parameter is further reflected in the near-uniform uncertainty across its range (Figure 3aa). However, some expected relationships are not seen in the ensemble. For instance, the threshold temperature below which ψ is equal to ψ_{min} is defined by the parameter T_{tip}, yet this shows no clear relationship with ψ (Figure 3l). Similarly, the threshold temperature for precipitation phase partitioning (T_{snow}) has minimal impact on accumulation (P_{snow}) (Figure 3ab). The limited influence of these threshold temperatures, as well as the significant impact of c and p, reflects the dominance of winter accumulation and summer ablation as modelled mass balance processes over the calibration period. Given the distinct seasonality of modelled snowfall, the influence of the ice albedo (α_{ice}) parameter is restricted to the summer months, when snowpack depth reduces or disappears entirely (Sect. 2.1.1). The value of α_{ice} therefore has a less significant impact than the atmospheric transmissivity (τ) parameter, which is unaffected by model processes (Figure 3o).

3.1.2 Calibrated Ensembles

The ensemble is constrained using the two calibration approaches described in Sect. 2.4. To investigate the distribution of the calibrated parameter sets within the 7-dimensional parameter space, we examine two-dimensional density plots (Figure 4). In more detail, for each pairwise combination of the seven parameters, we visualise how the calibrated values are distributed across their pre-defined ranges.

For the tuning calibration procedure, in which an optimal parameter set is selected for each individual glacier, 93 unique parameter sets (of the 250) were selected as providing the best match with glacier-specific mass balance observations, across all 568 glaciers. Of these parameter sets, low values of the temperature-dependent flux gradient c generally produced the best matches with glacier-specific mass balance observations, regardless of the values of the other parameters (Figure 4a, second row of the right triangle, where highest densities, i.e. more yellow, are seen at low c values). We also find that the tuning calibration method predominantly selects high values of the ice albedo (α_{ice}), although a wider spread remains across the ensemble compared to c. This pattern suggests that α_{ice} plays an important role for some individual glaciers, despite only having a limited influence on regional mean mass balance (Sect 3.1.1). For the other model parameters, calibrated values span widely across the pre-defined ranges, but there are some localised high-density clusters, indicating that groups of glaciers favour similar parameter combinations.

After performing our history matching calibration, which selects an ensemble of NROY regional-scale parameter sets, a broader region of parameter space is preserved (Figure 4b, lower left triangle). Of the full ensemble, 58% are ruled out through calibration, leaving 105 NROY parameter sets. The intersection of parameter sets retained by both calibration procedures— defined as the union space—is only 16 ensemble members (~6 %), indicating that relatively few parameter sets overlap between glacier-specific and regional-scale constraints. Despite this limited overlap, a similar dependence on c can be observed in NROY parameter space. In particular, varying c with other model parameters rules out regions of parameter

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

space (second column of lower triangle of figure 4); low values of c are ruled out when combined with high values of the precipitation correction factor (p) and the precipitation phase partitioning temperature (T_{snow}), and high values of c are ruled out with high values of the atmospheric transmissivity (τ) parameter.

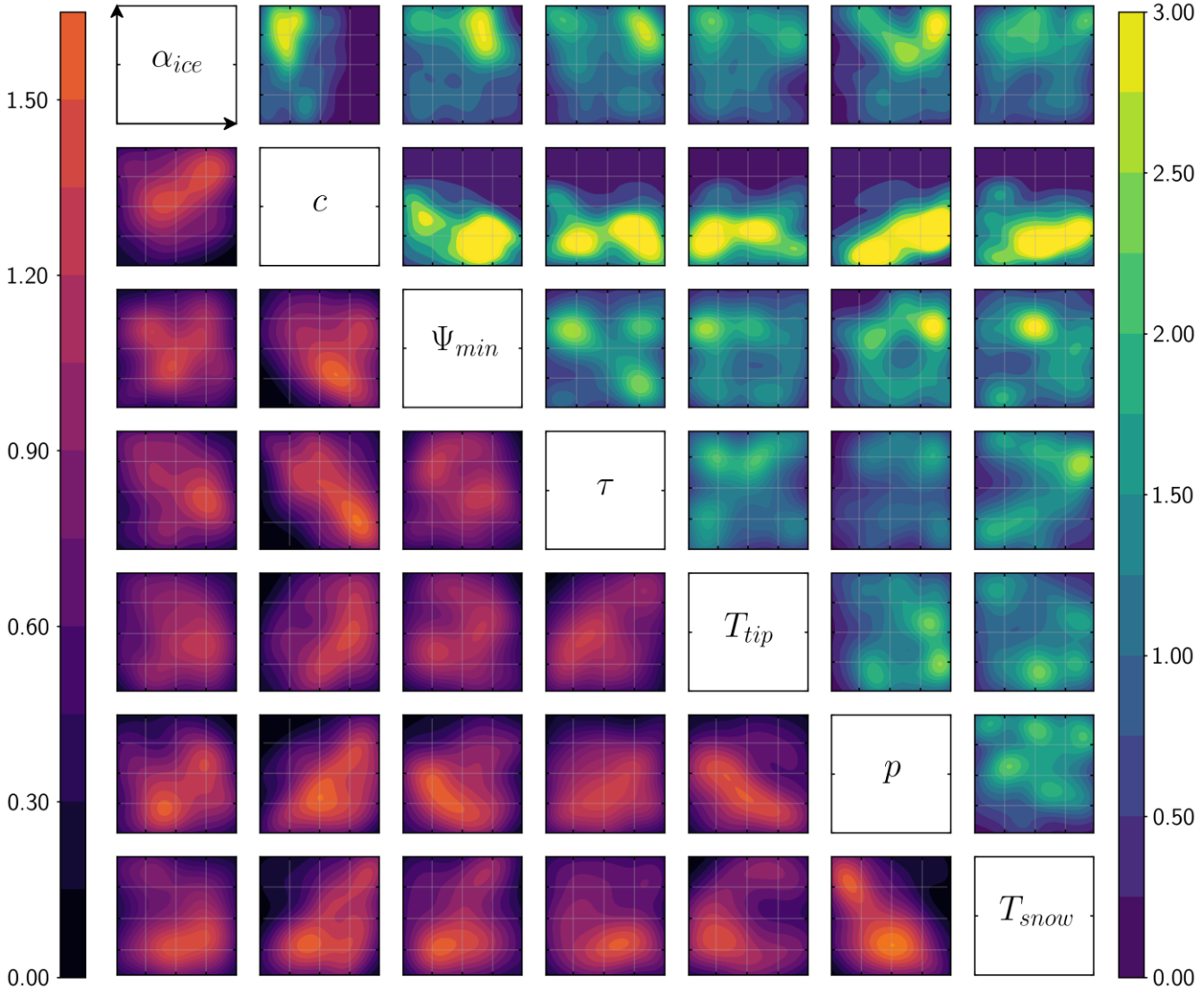

Figure 4. Density of calibrated parameter values (a) tuned for each glacier individually (n=93) (*upper right plots in the green-blue colourmap*) and (b) not-ruled-out-yet from history matching (n=105) (*lower left plots in the black-red colourmap*). Scales on the right and left axis correspond to the upper right and lower left triangle, respectively.

This interdependence among parameters leads to competing controls on the mass balance. To illustrate this, we inspect the relationship between NROY parameter sets and the mass balance model outputs by fitting a LOWESS model to each parameter-output combination (Sect. 2.5; Figure 4). Many patterns from the full ensemble are preserved, such as the dominant control of c on mass balance, but some notable differences emerge due to compensating effects between model

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

parameters. In particular, in order to satisfy the mass balance constraint, high values of parameters that control ablation must be compensated by high values of parameters that control accumulation. This produces weakly counter-intuitive responses in the regional mean mass balance, with higher values of p and T_{snow} resulting in decreasing mass balance values (Figure 5f, 5g). These compensatory relationships produce strongly counter-intuitive responses in the individual mass balance components. Lower values of c generally lead to less accumulation (P_{snow}) (Figure 5w), when we would expect the opposite to be true, and increasing values of τ result in lower values of the temperature-dependent fluxes (ψ) (Figure 5k), even though τ does not directly influence ψ within the model.

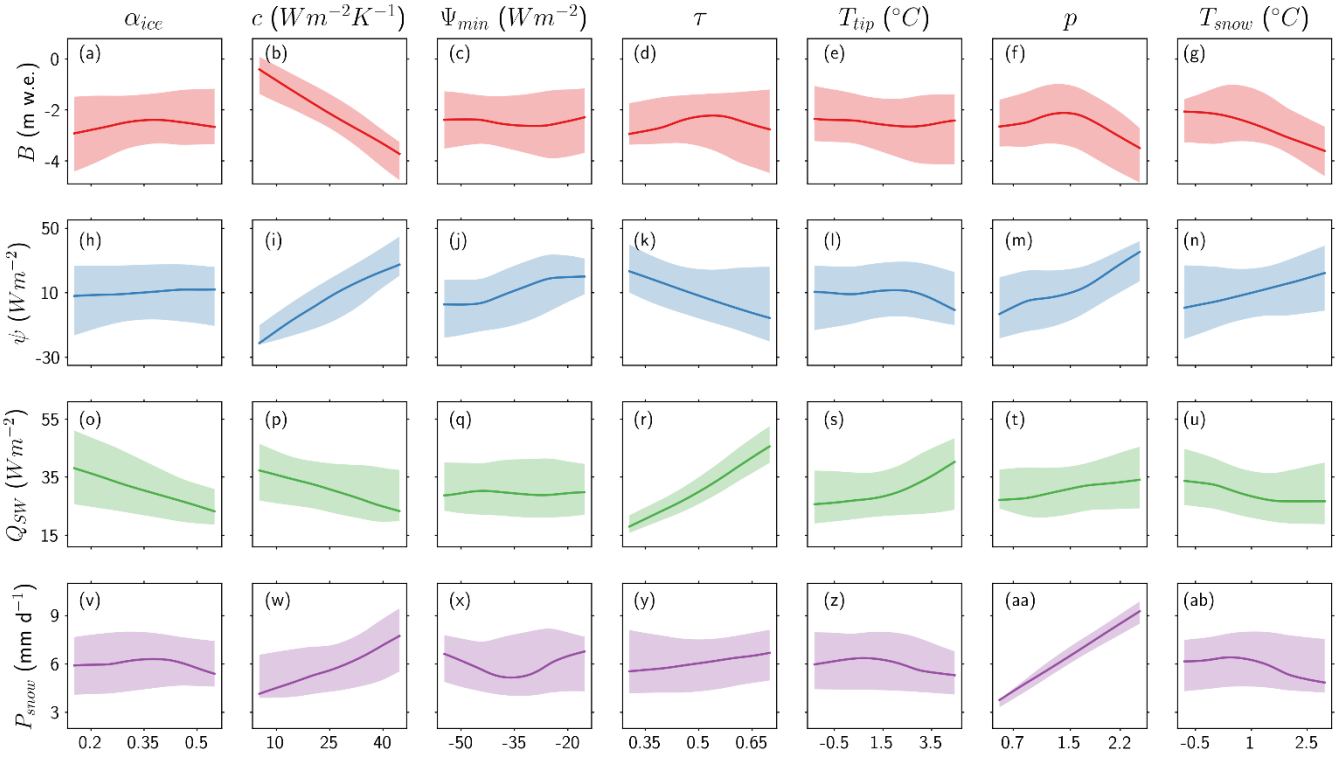

Figure 5. Same as Figure 3, but only considering not-ruled-out-yet (NROY) ensemble members from history matching. Note the different y-axis scales compared to Figure 3.

3.2. Projections of Steady-State Volume

In this section, we show projections of glacier volume change in Iceland under the experiments from GlacierMIP3. Sect. 3.2.1 presents projections generated using parameter values tuned for each glacier individually, which constitute the GO model contribution to GlacierMIP3. In Sect. 3.2.2, we present projections generated using the full (uncalibrated) and NROY (not-ruled-out-yet) ensemble. Note that relative volume changes are directly comparable to relative mass changes (as presented in GlacierMIP3; Zekollari et al., 2025), assuming a constant ice density.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

3.2.1. Tuned Projections

Figure 6 shows the projected glacier volume evolution over the 5000-year simulation period for all GlacierMIP3 forcings. All simulations reach an approximately constant volume, or steady state (as defined by Zekollari et al., 2025), within the first 300 years. Glacier response timescales are typically faster in experiments with higher warming, due to the evolution towards the same ice-free state for small glaciers, which occurs faster at higher temperatures. After reaching steady-state conditions, the projected volume change (as a percentage of initial volume) ranges from an increase of 15 % to a total loss of ice volume (-100 %). Of the 80 simulations (4 periods × 5 climate models x 4 scenarios), only two produce a volume increase; both of which were forced by IPSL-CM6A-LR under the 1851-1870 (+ 8 %) and 1901-1920 (+ 15 %) historical periods (Figure 6a, 6e). For the later historical periods, all simulations produce volume losses, with projected steady-state volumes of 86 (83-96) % and 68 (58-75) % for the 1951-1970 and 1995-2014 periods, respectively (Figure 6i, 6m).

All simulations forced with climate data for future periods project median volume losses greater than 40 % (i.e. steady-state volumes below 60%). For the SSP1-2.6 scenario, projected volume changes show little variation between the different time periods, with the multi-GCM median varying by less than 4 % (Figure 6, second column), because this low emissions scenario does not vary much through time (Figure 2). However, there is a substantial spread among projections from different GCMs, with maximum differences in relative volume losses exceeding 70 % (Figure 6n). This is largely attributable to UKESM1-0-LL, which has a high equilibrium climate sensitivity (Meehl et al., 2020) that produces substantially higher volume losses than the other GCM members and dominates the ensemble spread. Simulations forced with climate from the near-term 2021-2040 period show minimal dependence on emissions scenario, with projected steady-state ice volumes of 57 (22-62) %, 59 (25-66) % and 53 (18-60) % under the SSP1-2.6, SSP3-7.0 and SSP5-8.5 scenarios, respectively (Figure 6f, 6g, 6h). Differences between scenarios increase over time: for the 2081-2100 period, projected steady-state volumes range from 56 (31-71) % for the SSP1.2-6 scenario (Figure 6h) to 3 (0-18) % for the SSP5.8-5 scenario (Figure 6p).

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

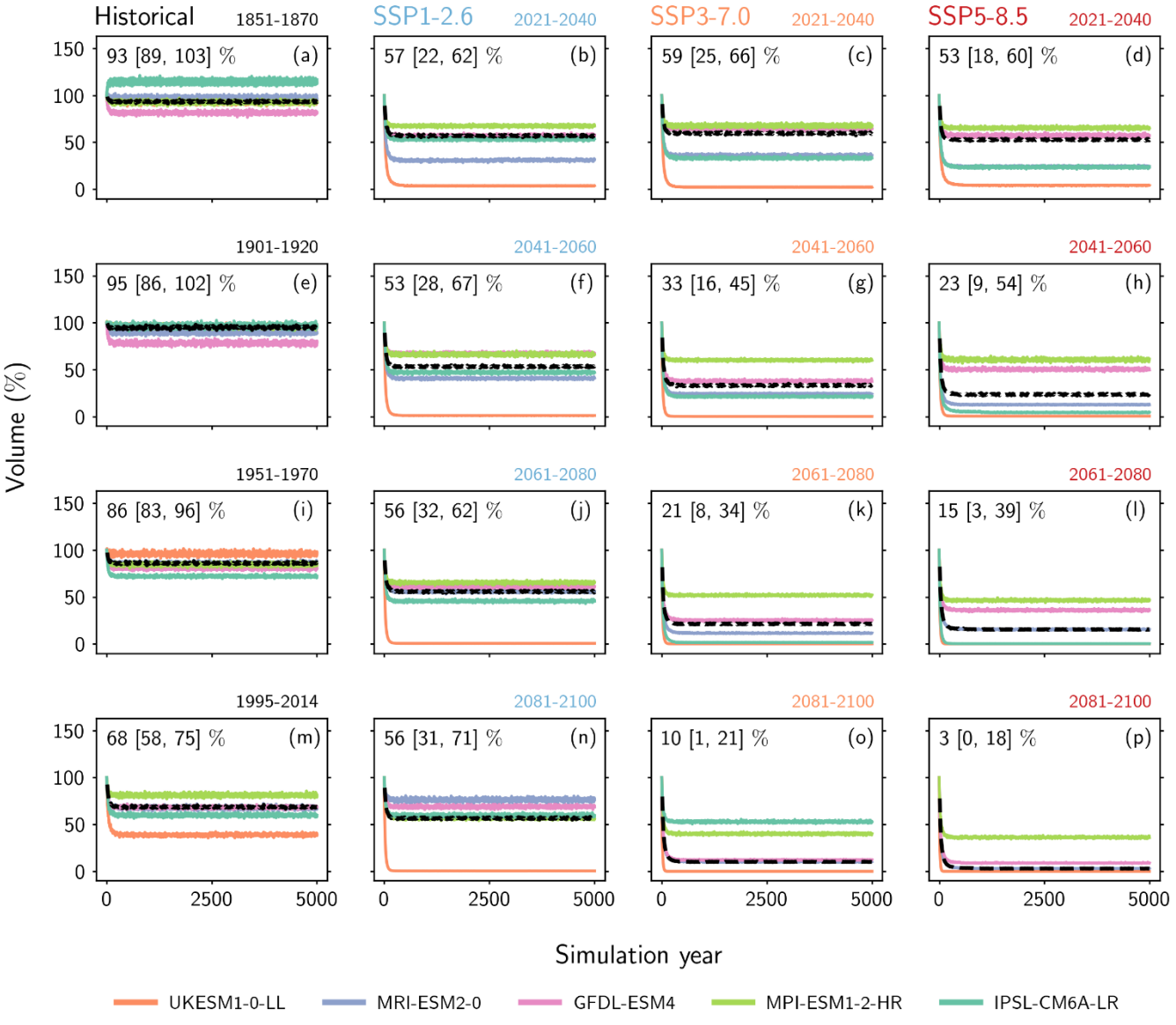

Figure 6. Evolution of glacier volume (as percentage of initial volume) over the 5000-year simulation period for all forcing experiments, generated using the tuned ensemble. Each panel presents simulations forced with climatic conditions from five GCMs (as indicated in the legend), extracted from four historical periods (first column) (1851-1870, 1901-1920, 1951-1970, 1995-2014) and four future periods (2021-2040, 2041-2060, 2061-2080, 2081-2100) under the SSP1-2.6 (second column), SSP3-7.0 (third column) and SSP5-8.5 (fourth column) scenarios. The multi-GCM median and likely range for each experiment are presented on panels as dashed lines, as well as text in upper left corner.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

3.2.2. Ensemble Projections

Figure 7 shows the regional glacier volume projections from the full (grey shading) and NROY (coloured shading) ensembles for each of the GlacierMIP3 experiments (scenario-period combination). For comparison, we also present the tuned projections (pink dashed line and shading).

The full ensemble forced with early historical climates, 1851-1870 and 1901-20, produces steady-state ice volumes of 111 (83-153) % and 109 (82-150) % respectively. These are reduced to 107 (88-132) % and 103 (84-131) % in the NROY ensemble (Figure 7a, 7e): i.e. history matching primarily removes high steady-state volumes, though many remaining NROY ensemble members project an increase in ice volume. Compared with the tuned projections, which generally show a slight decrease in ice volume for these early historical periods, the NROY ensemble produces median steady-state ice volumes 14 % and 8 % higher, and 83rd percentiles 29 % and 28 % higher, for the 1851–1870 and 1901–1920 experiments, respectively.

For the two later historical periods, the probability density functions of the NROY ensemble projections are noticeably narrower than that of the full ensemble (Figure 7i, 7m), reflecting the increased influence of calibration for more recent climate forcings. For the 1951-1970 and 1995-2014 periods, the NROY projections show median relative ice losses of 3 % and 22 %, and a 17-83 % uncertainty range of 27 % and 29 %, respectively.

Under future climate conditions, the NROY ensemble projections indicate that, irrespective of the climate forcing, a reduction in ice volume of at least 15 % is likely. For the 2081-2100 period, simulations forced with climate derived from SSP1-2.6 (Figure 7n) generally produce the smallest volume losses of all scenarios, with a median steady-state volume of 70 %; this is 14 % higher than the median of the tuned projections. However, many simulations in the NROY ensemble project very large volume losses, resulting in a bimodal distribution. As a result, the 17th percentile volume is much smaller for the NROY ensemble (2%) than for the tuned projections (32%). For the forcings using SSP3-7.0 (Figure 7o) and SSP5-8.5 (Figure 7p), most simulations produce large volume losses. The distributions are skewed because only large glaciers can support non-zero steady state, and even then, only under a limited set of forcing and parameter combinations.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

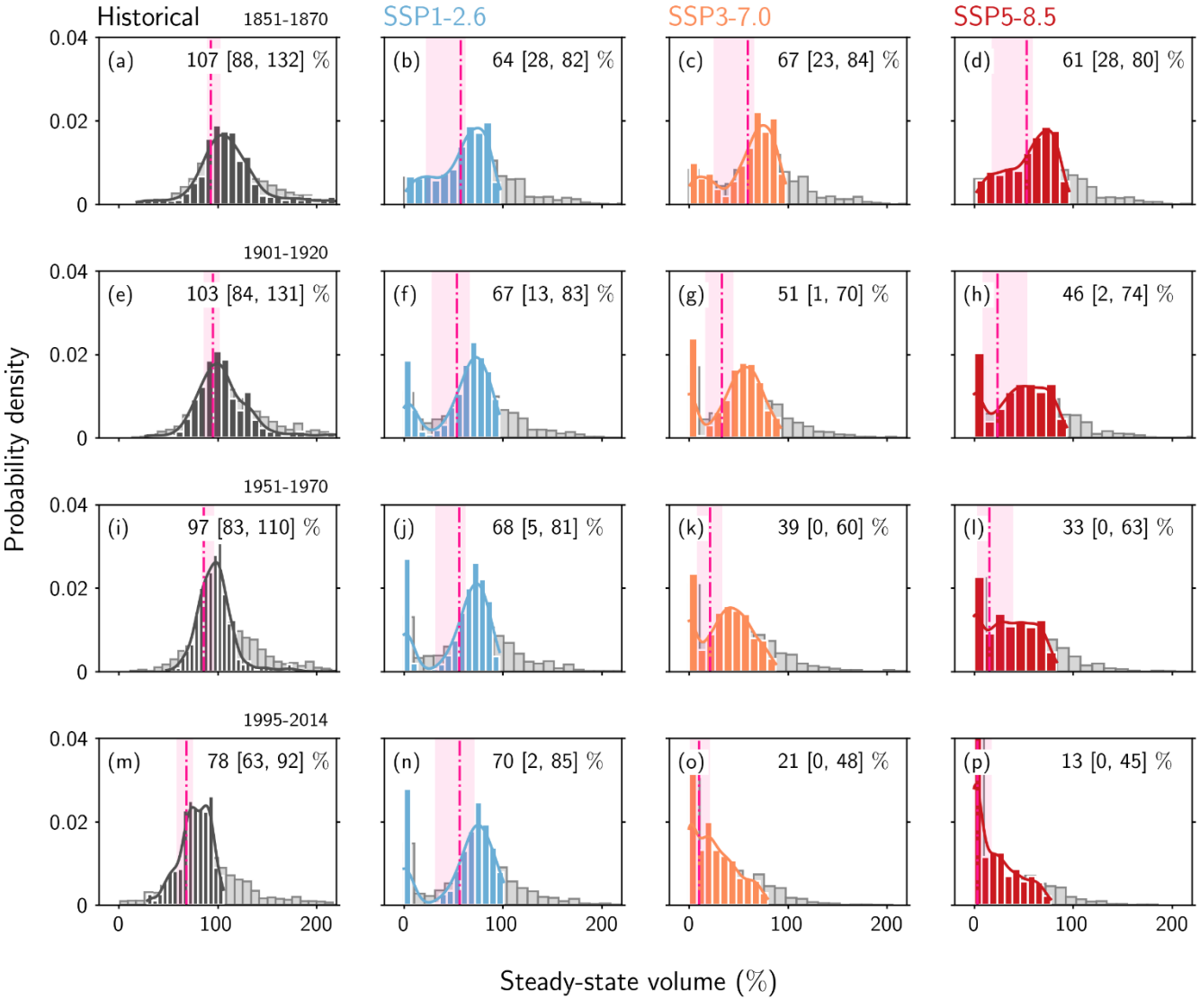

Figure 7. Projected steady-state ice volume for all GlacierMIP3 experiments generated using the full (grey shading) and NROY (coloured shading) ensembles. The pink dashed line and shading represents the median and likely range for the tuned projections (Figure 6). Each panel presents projections from one experiment, forced with climatic conditions from four historical periods (first column: 1851-1870, 1901-1920, 1951-1970, 1995-2014) and four future periods (2021-2040, 2041-2060, 2061-2080, 2081-2100) under the SSP1-2.6 (second column), SSP3-7.0 (third column) and SSP5-8.5 (fourth column) scenarios. The median and likely range for the NROY ensemble projections are indicated as text in the upper-right corner of each panel.

3.3. Sensitivity to Temperature Change

Each of the GlacierMIP3 forcings corresponds to a change in the global mean temperature relative to the pre-industrial period (Sect. 2.2). Figure 8 shows glacier volume projections from the tuned and NROY ensembles aggregated into discrete global mean temperature change scenarios, ranging from 0 °C to 7 °C at 1 °C intervals.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

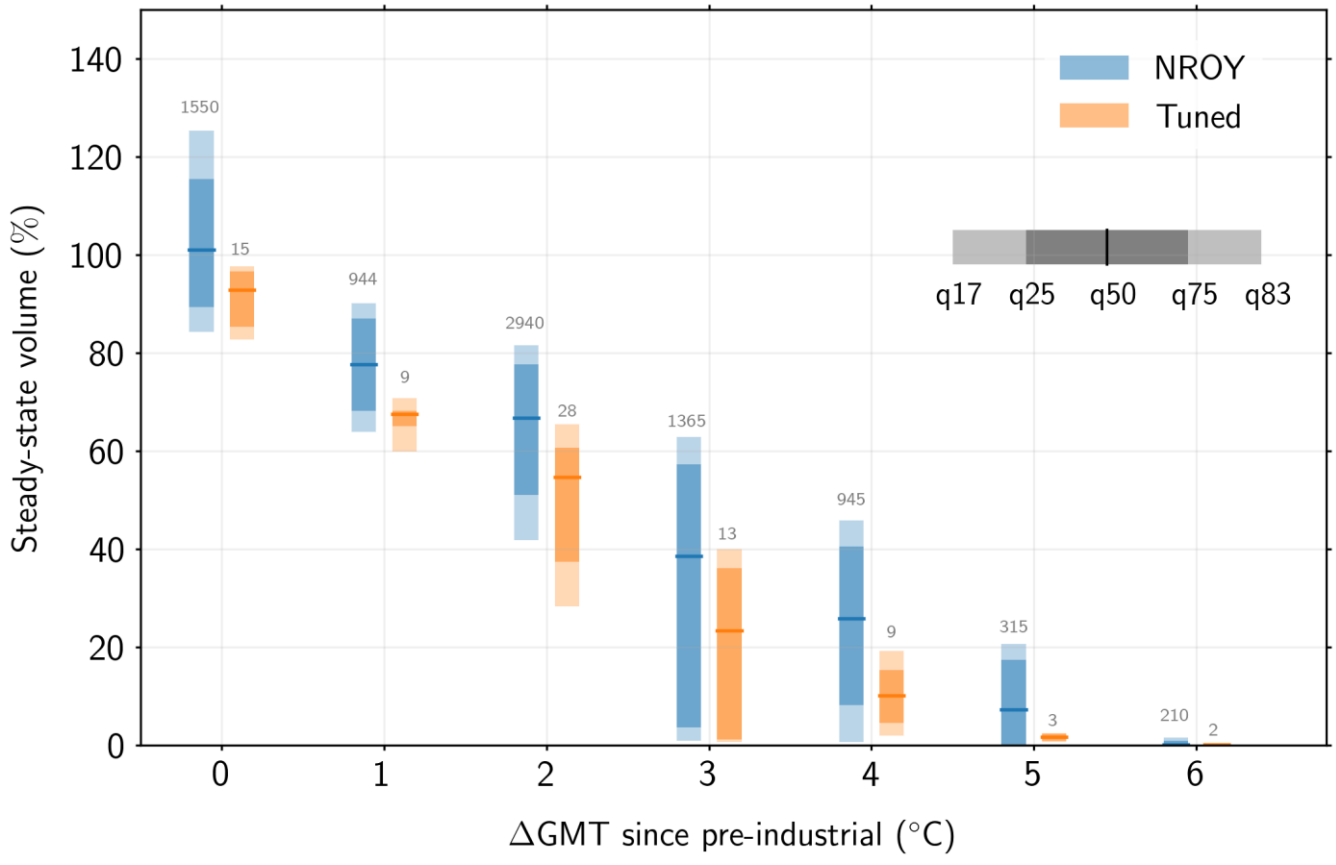

Figure 8. Projections of steady-state volume generated using our not-ruled-out-yet (NROY, blue) and tuned (orange) ensemble, aggregated into different global mean temperature change (ΔGMT) scenarios relative to the pre-industrial period. For each bar, the bold line represents the median, the darker shading indicates the interquartile range, and the lighter shading shows the 1likely range. The number above each bar denotes the total number of regional glacier volume projections aggregated within a given temperature change scenario. The 7 °C scenario is not shown (n=1).

As expected, the projected steady-state volume depends strongly on the applied global mean temperature change scenario, but the central estimates differ between the two ensembles. For temperature changes up to 5 °C, the median is systematically lower from the tuned ensemble than the NROY ensemble projections. Under historical climate conditions (predominantly binned to 0 °C), the tuned ensemble produces median volume loss relative to present-day (–8 %), whereas the median of the NROY ensemble shows a slight increase in volume under this scenario (+ 2 %). For all other temperature change scenarios, both ensembles project volume losses, but the magnitude differs substantially. The largest discrepancies occur under the 3 °C and 4 °C scenarios, where the tuned ensemble projects median volume losses that are 15 % and 16 % higher, respectively, than those of the NROY ensemble. Only under scenarios exceeding 5 °C of warming are absolute differences between the two ensembles negligible, since median relative volume losses exceed 99 %.

For the tuned ensemble, differences between GCMs (hereafter forcing uncertainty) are the sole contributor to projection uncertainty (compared with the NROY ensemble, which includes both forcing and parametric uncertainties). The largest

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

likely range occurs under the 3 °C scenario, where the projected steady-state volume ranges from 1 to 43 % (outer limits of the 'likely' range). Part of this uncertainty reflects differences in the global mean temperature change simulated by different GCMs: for instance, in the 3 °C binned scenario, the simulated global mean temperature changes range from 2.6 °C to 3.4 °C. This variability is further amplified at the regional-scale, as local temperatures deviate from the global mean and simulated precipitation differs across GCMs.

The spread of the NROY ensemble projections reflects both forcing and parametric uncertainty, and – given the common climate forcings – the difference between the tuned and NROY ensembles can be interpreted as the influence of parametric uncertainty. As with the tuned ensemble, we find that the likely range of our NROY ensemble projections is largest under the 3 °C scenario. Here, the lower limit of the likely range is equal to that of the tuned ensemble (<1 % of present-day volume), but the upper limit produces steady-state volumes which are 18 % higher, reaching 61 % of present-day volume. For all scenarios exceeding 3 °C, the NROY ensemble produces a much larger range of steady state volumes: the lower limit of the likely range is approximately equal to the tuned projections, but the upper limit is markedly higher. The largest absolute difference in projection uncertainty occurs under the 4 °C scenario, where the 83rd percentile of the glacier volume projections increases by 28 % between the tuned and NROY ensemble projections (upper bound of ~48 % vs. 37 % of present-day volume). The largest relative difference, however, is observed under the 5 °C scenario: all members of the tuned ensemble project near-total volume losses (<1 % of present-day ice volume), whereas the NROY ensemble produces some simulations in which ice remains at stabilization, resulting in an uncertainty range that is an order of magnitude larger. Large relative differences also occur under low warming scenarios, where the likely range of the NROY ensemble spans 34 % of present-day volume under a 0 °C scenario – more than double the tuned ensemble's range of 17 %. Similarly, under the 1 °C scenario, the likely range of the NROY ensemble projections is 29 % of present-day volume, compared to 11 % from the tuned projections.

3.4. Sensitivity to Individual Model Parameters

To further investigate the drivers of uncertainty within the ensemble projections, we analyse the sensitivity of projected glacier volume change in the 0 °C, 2 °C and 4 °C binned scenarios to individual model parameters by fitting a LOWESS model to each parameter-projection relationship (Sect. 2.5).

Figure 9 shows the relationship between model parameters and glacier volume projections generated using the full (uncalibrated) ensemble. Across all temperature change scenarios, projected glacier volume shows a significant dependence on the temperature-dependent flux slope (c), with increasing values of c generally producing higher volume losses (Figure 9, second column). This is consistent with patterns observed in calibration (see Sect. 3.1, Figure 3), where c exerted a strong control on regional mean mass balance (Sect. 3.1, Figure 3). However, here the precipitation correction factor (p) and the precipitation phase partitioning temperature (T_{snow}) also have a strong influence (Figure 9, sixth column). This differs from the calibration period, during which p and T_{snow} had minimal impact on the regional mass balance. The increasing influence

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

of the p and T_{snow} largely reflect the differences in precipitation under the same temperature change scenario, thereby increasing the importance of parameters that directly control accumulation.

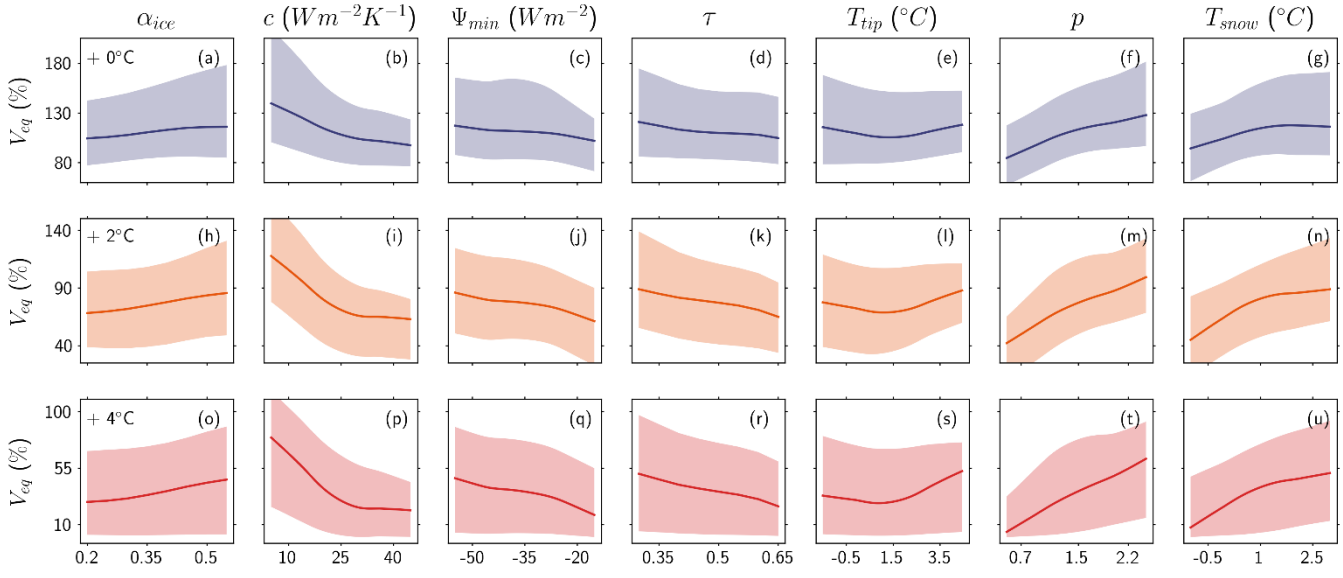

Figure 9. Relationship between model parameters and projected steady-state glacier volume (V_{eq}) as percentage of initial volume under + 0 °C (a-g, first row), + 2 °C (h-n, second row) and + 4 °C (o-u, third row) temperature change scenarios, generated using the full perturbed-parameter ensemble. The LOWESS median fit (bold line) and likely range (shaded region) are presented for each parameter-projection relationship. Note the different y-axis ranges for each row.

Different patterns emerge when we inspect the relationship between model parameters and projections in the NROY ensemble (Figure 10). In general, the model output is less sensitive to model parameters than in the full ensemble (note the smaller y-axis ranges), because history matching typically excludes the extremes of the parameter ranges. Under the 0 °C scenario, relationships are characterised by large uncertainties, reflecting the absence of dominant parameters. As in the uncalibrated ensemble, we find that the precipitation correction factor p exerts a strong control on the NROY projections under the 2 °C and 4 °C scenarios, with lower values generally resulting in greater volume losses (Figure 10m & 10t). A strong relationship between c and projected ice volume is also observed under these warming scenarios (Figure 10i & 10p). However, this relationship is counter-intuitive, since decreasing values of c generally lead to greater volume losses, despite it corresponding to higher values of the temperature-dependent flux. Given the calibrated values of c and p are correlated (Sect. 3.1.2), this reflects the dominance of parameters that directly affect accumulation in controlling projected volume change. Consistent with this picture, we also find that T_{snow} has a significant impact on projected ice volume, and this increases under higher warming scenarios (Figure 10u). Similarly, the parameter determining the threshold temperature for the temperature-dependent fluxes (T_{tip}) has a stronger influence under higher warming scenarios (Figure 10l & 10s). Certain parameters, including ice albedo (α_{ice}) and atmospheric transmissivity (τ), have a limited impact on projected glacier volume across all temperature change scenarios.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

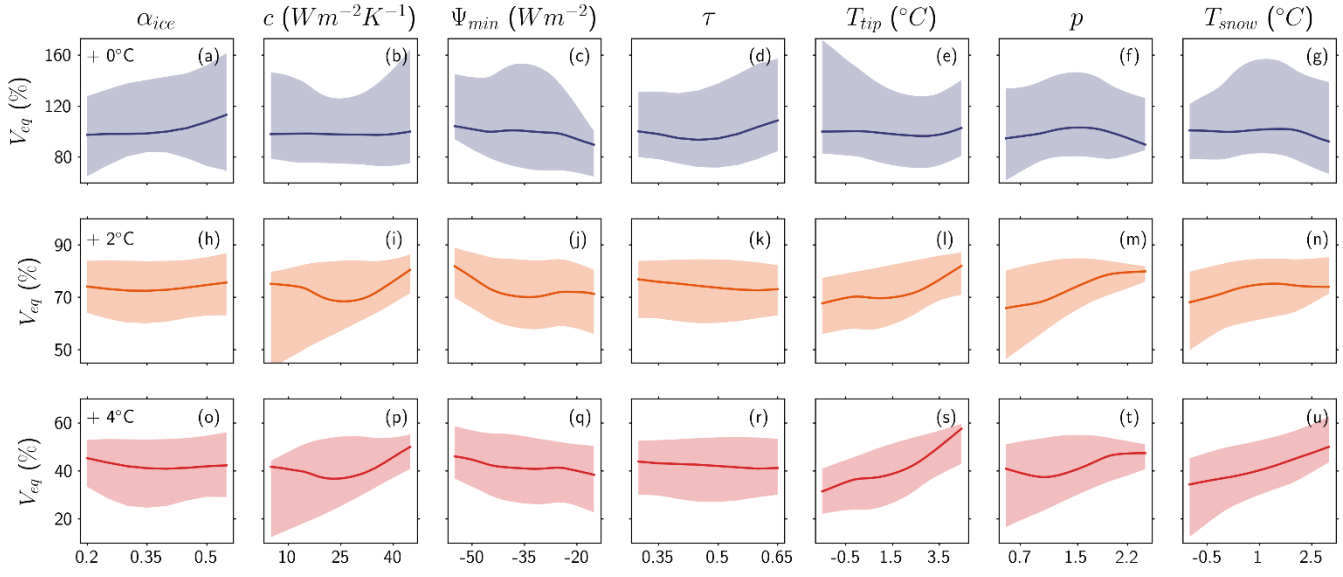

Figure 10. Same as Figure 9, but projections of steady-state glacier volume (V_{eq}) generated using the NROY ensemble. Note the different y-axis range for each row.

3.5. Comparison to the GlacierMIP3 Ensemble

To assess the relative size of model structural uncertainty and parametric uncertainty in projections of regional glacier volume change, we compare projections generated using the NROY ensemble with those from the GlacierMIP3 multi-model ensemble (MME) (Zekollari et al., 2025) for Iceland. Figure 11 shows the projected steady-state volume from the two ensembles as a function of global mean temperature change (as described in Sect. 3.4).

Across all temperature change scenarios, the GlacierMIP3 MME produces lower median steady-state volumes than the NROY ensemble. Under the 0 °C scenario, however, the MME yields a higher upper bound than the NROY ensemble. Here, the NROY ensemble projections lie within the uncertainty range of the MME projections, with a likely range of 84–125 % and 62–140 % for the NROY ensemble and GlacierMIP3 MME, respectively. Under the 1 °C and 2 °C scenarios, the spread of glacier volume projections produced by the two ensembles is similar, but the likely range of the MME is shifted towards lower steady-state ice volumes. In general, the MME projections become more constrained at higher warming levels as ensemble members converge towards strong volume losses.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

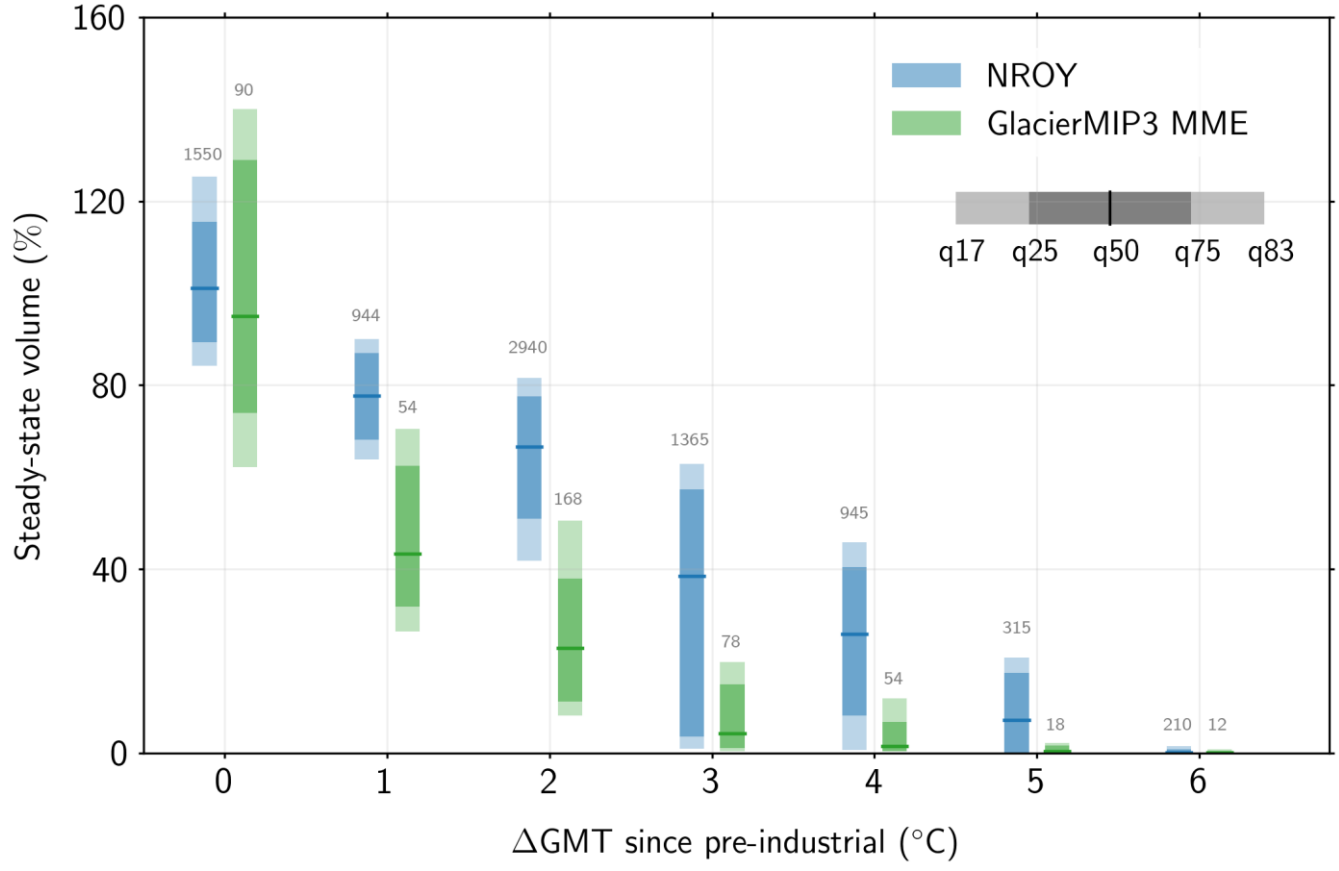

Figure 11. Projections of glacier volume change in Iceland relative to present-day generated using the not-ruled-out-yet ensemble (NROY, blue) and the GlacierMIP3 multi-model ensemble (GlacierMIP3 MME, green) under different scenarios of global mean temperature change (ΔGMT) relative to the pre-industrial period. For each bar, the bold line represents the median, the darker shading indicates the interquartile range, and the lighter shading shows the likely range. The number above each bar denotes the total number of regional-scale simulations aggregated within a given temperature change scenario. + 7 °C scenario not shown (n=6). Note different y-axis range compared to Figure 8.

For all temperature change scenarios exceeding 2 °C of warming, the range of glacier volume projections produced by the NROY ensemble exceeds that of the GlacierMIP3 MME. While the lower limits of the likely range from both ensembles converge towards near-total volume losses, differences arise primarily in the upper limits of the likely ranges. For these scenarios, the MME projections are increasingly skewed towards low steady-state volumes, whereas the NROY ensemble retains a broader range of potential ice volumes. The largest absolute difference in the likely range occurs under a 3 °C scenario, where the NROY ensemble projects a likely range of 2–77 % compared to 1–39 % from the MME. The largest relative difference, however, is found under the 5 °C scenario, where the MME projections are almost entirely concentrated at a total loss of ice volume (83rd percentile: 2 %), whereas the NROY ensemble projects a likely range of 0–21 %. Under a 6 °C scenario, all members of both the MME and NROY ensembles produce very low steady state volumes (<1 % of present-

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

day volume). At this warming level, projected steady-state volumes show little sensitivity to the ensemble construction, indicating that temperature forcing dominates over differences associated with structural or parametric uncertainty.

4 Discussion

In this study, we produced projections of glacier volume change for Iceland under the GlacierMIP3 constant climate scenarios using a large perturbed-parameter ensemble of the large-scale glacier evolution model GO-VA. We used two different calibration methods to constrain the ensemble: a traditional by-glacier tuning approach, and a history matching procedure that accounts for parametric uncertainty. Our results show that failing to account for parametric uncertainty can reduce the overall uncertainty in projections by 7–91 %. The size of this reduction depends on the climate forcing: for scenarios with high forcing uncertainty, parametric uncertainty plays a more minor role. However, for temperature change scenarios for which forcings are more narrowly constrained, parametric uncertainty plays a significant, or dominant, role. This is particularly relevant for the case study region, Iceland, where forcing uncertainties are large (Marzeion et al., 2020). In regions where forcing uncertainty is smaller, parametric uncertainty would likely dominate across a broader range of scenarios. Comparison with the GlacierMIP3 ensemble suggests that parametric uncertainty can be as large as, or even be larger than, the uncertainty arising from differences between models. This highlights the importance of considering both structural and parametric uncertainty to fully quantify model uncertainty in projections.

Our paper is one of the first to apply formal uncertainty quantification methods to the calibration of glacier models for large-scale applications. In a previous effort, Rounce et al. (2020a) used a Bayesian inverse model and glacier-specific mass balance observations for calibration, allowing model parameters to vary between glaciers. In our history matching calibration, we used regional-scale mass balance observations for the observational constraint to calibrate fixed parameter sets for every glacier in Iceland (though this fixed set is perturbed). In this procedure, errors in mass balance at the individual glacier-scale cancel out at a regional scale. One consequence, as in any calibration using aggregated quantities, is that some ensemble members pass calibration by simulating some glaciers with positive mass balance and others with negative (see Huss and Hock, 2015). This can introduce biases into simulations; in particular, the volume loss of a glacier with negative mass balance during calibration is limited to its initial volume, but there is no limit on how much a glacier with positive mass during calibration can grow (Rounce et al., 2020a). Consequently, regional glacier volume simulations may exhibit a positive bias in total volume change, since large growth experienced by some glaciers (i.e. those with positive mass balance during calibration) cannot be offset by equivalent volume losses elsewhere. This likely explains the systematic positive difference between our projections from history matching compared with those from tuning. This result is consistent with earlier studies (Huss and Hock, 2015; Rounce et al., 2020a), which found that regional-scale calibration tends to produce smaller volume losses than calibration at a glacier-scale.

In theory, this could be addressed by performing the history matching calibration at a glacier-scale, rather than regional-scale, using glacier-specific mass balance observations (see Rounce et al., 2020a). However, there are a number of

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

disadvantages to this approach. One is the additional complexity of considering the aggregation of uncertainties from the glacier to regional scale: treatment of parametric uncertainty as either spatially independent (varying by glacier) or spatially correlated (shared across glaciers, whether fully or partially) requires strong judgements and can substantially affect the degree of uncertainty in projections.

A more critical and general point is the risk of over-tuning. While the larger uncertainties in our regional history matching calibration might at first suggest that projections derived from tuning are more reliable or useful, such tuning typically compensates for structural deficiencies in models and errors in climate forcing. In other words, parameters can be tuned to improve agreement with glacier-specific mass balance observations over the calibration period, but this does not necessarily imply that the underlying mass balance processes are more accurately represented. For example, a model forced with systematically overestimated temperatures over the calibration period would require a low value of temperature sensitivity parameters to reproduce the observed mass balance. This parameter choice would then lead to an underestimation of glacier sensitivity to further warming. As such, tuning may lead to the right answer for the wrong reason. Physical model parameters should have the same physical meaning for every glacier, so tuning them individually underestimates the inherent model uncertainties that should be incorporated into projections.

Our sensitivity analysis provides further insight into these issues. We found that the parameter controlling the sensitivity of temperature-dependent fluxes to temperature change (c) exerts the dominant influence on mass balance during the calibration period, and subsequently on calibrated parameter space. While the other model parameters had a limited impact on mass balance compared to c, they were essential for constraining parameter space: only through co-variation with c could significant regions of parameter space be ruled out as plausible matches with regional mass balance observations. These interactions reflect the trade-off between the processes of accumulation and ablation, or competing processes of ablation, depending on the process each parameter controls. As a result, while different parameter combinations may yield similar mass balance values over the calibration period, they imply contrasting process representations (e.g. Schuster et al., 2023).

This underscores a key limitation with the calibration of large-scale glacier evolution models: the model parameters are non-identifiable, i.e. the number of model parameters exceeds the information available in the calibration data. It has been argued that the most effective strategy for addressing non-identifiability is to increase the parameter-specific information available in the calibration data (Gelman et al., 2014). Recent studies have shown that sub-annual mass balance observations can indeed help differentiate between winter and summer mass balance processes, enabling a more effective calibration of parameters controlling accumulation and ablation (Schuster et al., 2023; Phelps et al., 2025). However, these studies have focused on the calibration of temperature-index models. These simple models may absorb structural deficiencies through parameter tuning, i.e., improved calibration can reduce apparent parametric uncertainty without necessarily improving process representation. Importantly, more complex mass balance models with more meaningful parameters may be more amenable to constraint through improved observations. Thus, increasing model complexity can, in some cases, decrease parametric uncertainty provided the calibration data are sufficiently informative. The seven model parameters perturbed in this study can be broadly categorised based on their influence on key mass balance processes: temperature-dependent fluxes,

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

shortwave radiative fluxes, and accumulation. Consequently, observations that directly capture these individual processes might provide more targeted constraints on the associated parameters.

Currently, such data are only available for a limited number of glaciers, meaning that existing studies investigating the advantages of incorporating additional calibration data have generally been restricted to small-scale applications (7–95 glaciers). The temporal resolution of geodetic mass balance observations, which remain the primary source of calibration data for large-scale studies (e.g. Hugonnet et al., 2021), are limited by density assumptions when converting elevation change to mass change, which introduces large uncertainties over shorter timescales (Huss, 2013). Nevertheless, recent efforts to improve the temporal resolution of geodetic data through geostatistical modelling (Dussaillant et al., 2025) and machine learning (Sjursen et al., 2025) approaches represent promising advances towards improving the calibration data available for future large-scale modelling studies.

Alternatively, dimensionality reduction has been proposed as an approach to handle the non-identifiability problem (e.g. Zolles et al., 2019), also known as pre-calibration or screening. This involves fixing non-sensitive, or less sensitive, parameters to constant values to reduce the parameter space. Our analysis indicates that the parameters controlling the threshold temperatures for the temperature-dependent fluxes (T_{tip}) and precipitation phase partitioning (T_{snow}) have a limited impact on mass balance processes over the calibration period, so these parameters might have been considered candidates for fixing. However, our results caution against this approach, since we find T_{tip} and T_{snow} play a prominent role in controlling uncertainty in the glacier volume projections, especially under warming scenarios. Furthermore, because the calibration assumes static glacier geometry (i.e. constant area), certain parameters may appear less important than they would under conditions of evolving geometry (Rounce et al., 2020b; Aguayo et al., 2024; Schuster et al., 2023). Taken together, these findings highlight the risks of prematurely excluding parameters based on their relative lack of influence during the calibration period: sensitivity analysis should be performed not only under present-day conditions, but also for projected changes (e.g. Wieder et al., 2019), and always by perturbing all parameters at once.

Finally, we note the magnitude of parametric uncertainty may be influenced by our chosen parameter ranges and model error, as in any ensemble calibration. Parameter ranges were guided by literature and a decreasing success in reproducing observations, but alternative choices might be made. We defined model error as a multiple of observational uncertainty following Edwards et al. (2019) (see also for example Gladstone et al., 2012; Coulon et al., 2025). To test the sensitivity of this assumption, we show results for alternative multiplicative factors (Williamson et al., 2013; Coulon et al., 2024): differences in the median and likely ranges are small (Appendix A).

5 Conclusions

In this paper, we have presented projections of glacier volume change in Iceland from the GO-VA model in the framework of GlacierMIP3. We compared projections generated using two calibration approaches: a traditional tuning-by-glacier strategy and a history matching procedure. For the latter, we constructed an ensemble of parameter sets that were consistent

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

with regional mass balance observations, thus enabling an assessment of parametric uncertainty in projections. Our results showed that parametric uncertainty can be a major, and in some cases dominant, source of uncertainty in projections of glacier volume change. We found that failing to account for parametric uncertainty reduces total projection uncertainty by 7–91 % across scenarios of global mean temperature change, with the largest reductions occurring for scenarios where forcing uncertainty is highly constrained. Through comparison with the GlacierMIP3 ensemble, parametric uncertainty is shown to be comparable to structural model uncertainty and, depending on the strength of the forcing, can even produce larger projection uncertainties. Taken together, these findings demonstrate the importance of accounting for parametric uncertainty, alongside structural uncertainty, to more comprehensively characterise projection uncertainty in model intercomparison projects.

2 Appendices

Appendix A: Model Discrepancy Variance

Figure A1 shows the effect of using higher (10x) and lower (2x) multiples of observational uncertainty for the model error term than the default value (5x) in the history matching calibration (Sect. 2.4), illustrated using the NROY ensemble projections as a function of binned temperature scenarios (Sect. 3.3). The larger value represents a greater tolerance for model error, while the smaller value represents a greater expectation for the model to reproduce observations under the best parameter values. Differences in the median and likely ranges are small.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

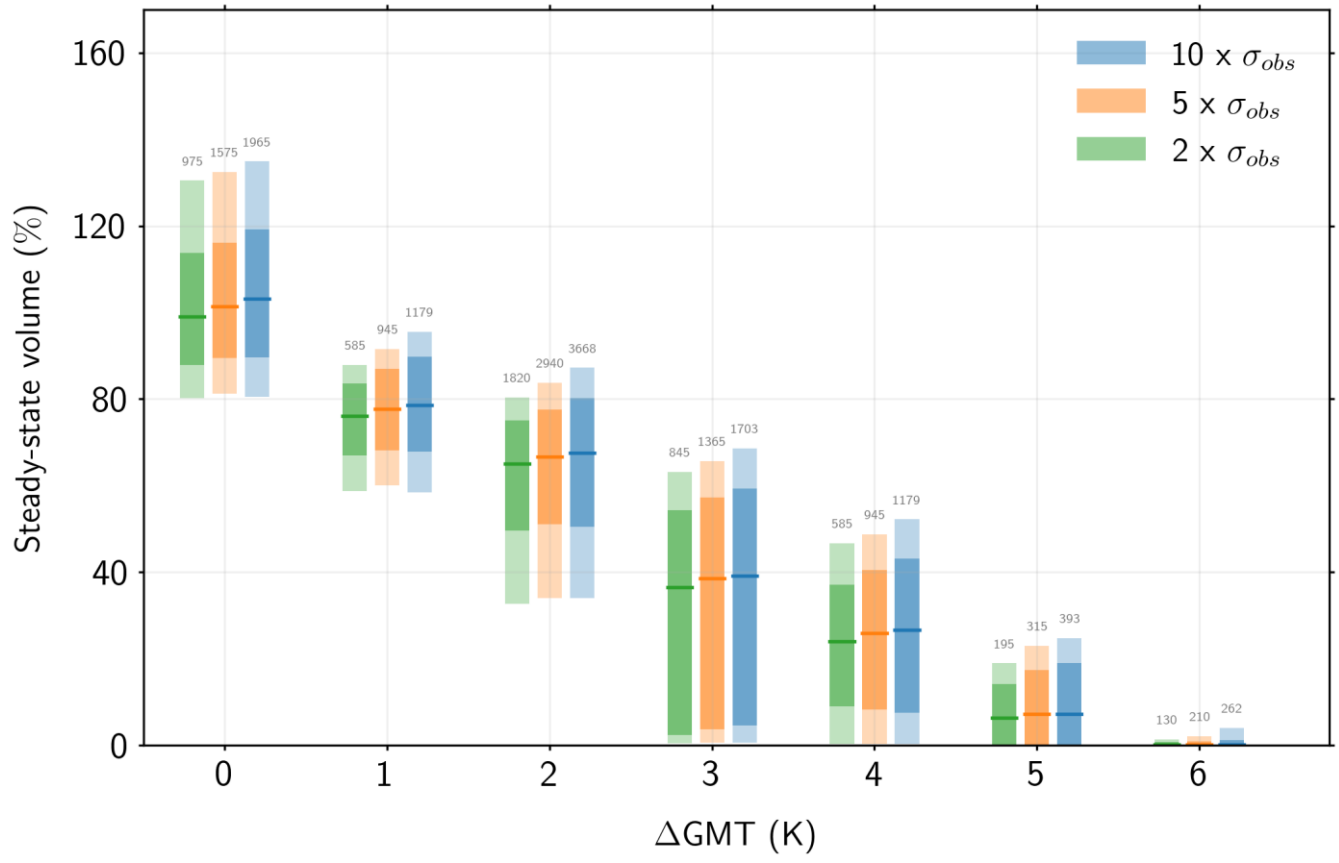

Figure A1. Steady-state volume projections in the NROY ensemble as a function of global mean temperature change for three different model error terms in the implausibility metric. Numbers above bars denote the number of regional simulations aggregated into each binned temperature change scenario.

Code and data availability

[to be added]

Supplement link

The link to the supplement will be included by Copernicus, if applicable.

Author contributions

MJ, TE and TM designed the study. MJ developed the model framework code, performed the simulations, and conducted analysis of results, with support and supervision from TE, TM and AB. HZ led the GlacierMIP3 project, and supported the

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

GO-VA GlacierMIP3 contribution. MJ wrote the original draft, created all figures, and all co-authors contributed to review and editing of the manuscript.

Competing interests

No conflicts of interest.

Disclaimer

Acknowledgements

[to be completed]

Financial support

[to be completed]

Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

References

Aguayo, R., Maussion, F., Schuster, L., Schaefer, M., Caro, A., Schmitt, P., Mackay, J., Ultee, L., Leon-Muñoz, J., and Aguayo, M.: Unravelling the sources of uncertainty in glacier runoff projections in the Patagonian Andes (40–56°S), The Cryosphere, 18, 5383–5406, https://doi.org/10.5194/tc-18-5383-2024, 2024.

Bahr, D. B., Meier, M. F., and Peckham, S. D.: The physical basis of glacier volume-area scaling, Journal of Geophysical Research: Solid Earth, 102, 20355–20362, https://doi.org/10.1029/97JB01696, 1997.

Bahr, D. B., Pfeffer, W. T., and Kaser, G.: A review of volume-area scaling of glaciers, Reviews of Geophysics, 53, 95–140, https://doi.org/10.1002/2014RG000470, 2015.

Bourn, A.: moepy – Merit-Order-Effect Python Implementation, Zenodo, https://doi.org/10.5281/zenodo.4642896, 2021.

Compagno, L., Zekollari, H., Huss, M., and Farinotti, D.: Limited impact of climate forcing products on future glacier evolution in Scandinavia and Iceland, Journal of Glaciology, 67, 727–743, https://doi.org/10.1017/jog.2021.24, 2021.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

Coulon, V., Klose, A. K., Kittel, C., Edwards, T., Turner, F., Winkelmann, R., and Pattyn, F.: Disentangling the drivers of future Antarctic ice loss with a historically calibrated ice-sheet model, The Cryosphere, 18, 653–681, https://doi.org/10.5194/tc-18-653-2024, 2024.

Coulon, V., Klose, A. K., Edwards, T. et al. From short-term uncertainties to long-term certainties in the future evolution of the Antarctic Ice Sheet, Nature Communications 16, 10385, https://doi.org/10.1038/s41467-025-66178-w, 2025.

Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., Rio, C., Audouin, O., Salter, J., Bazile, E., Brient, F., Favot, F., Honnert, R., Lefebvre, M.-P., Madeleine, J.-B., Rodier, Q., and Xu, W.: Process-Based Climate Model Development Harnessing Machine Learning: I. A Calibration Tool for Parameterization Improvement, Journal of Advances in Modeling Earth Systems, 13, e2020MS002217, https://doi.org/10.1029/2020MS002217, 2021.

Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A.: Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments, in: Case Studies in Bayesian Statistics, 37–93, https://doi.org/10.1007/978-1-4612-2290-3_2, 1997.

Edwards, T. L., Brandon, M. A., Durand, G., Edwards, N. R., Golledge, N. R., Holden, P. B., Nias, I. J., Payne, A. J., Ritz, C., and Wernecke, A.: Revisiting Antarctic ice loss due to marine ice-cliff instability, Nature, 566, 58–64, https://doi.org/10.1038/s41586-019-0901-4, 2019.

Farinotti, D., Huss, M., Fürst, J. J., Landmann, J., Machguth, H., Maussion, F., and Pandit, A.: A consensus estimate for the ice thickness distribution of all glaciers on Earth, Nat. Geosci., 12, 168–173, https://doi.org/10.1038/s41561-019-0300-3, 2019.

Felikson, D., Rounce, D. R., Fasullo, J., Rodriguez, A., Adhikari, S., Buzzanga, B., Dangendorf, S., Kopp, R. E., Lammers, R. B., Reager, J. T., Brinkerhoff, D., Csatho, B., Girotto, M., Hamlington, B., Ivins, E. R., Kumar, P., Larour, E., Nerem, R. S., Nowicki, S., Schlegel, N.-J., Tesdal, J.-E., and Weathers, M.: Progress and future directions in constraining uncertainties in sea-level projections using observations, Nat. Clim. Chang., 15, 1039–1051, https://doi.org/10.1038/s41558-025-02437-4, 2025.

Giesen, R. H. and Oerlemans, J.: Calibration of a surface mass balance model for global-scale applications, The Cryosphere, 6, 1463–1481, https://doi.org/10.5194/tc-6-1463-2012, 2012.

Gladstone, R. M., Lee, V., Rougier, J., Payne, A. J., Hellmer, H., Le Brocq, A., Shepherd, A., Edwards, T. L., Gregory, J., and Cornford, S. L.: Calibrated prediction of Pine Island Glacier retreat during the 21st and 22nd centuries with a coupled flowline model, Earth and Planetary Science Letters, 333–334, 191–199, https://doi.org/10.1016/j.epsl.2012.04.022, 2012.

Hersbach, H., Bell, B., Berrisford, P., Dahlgren, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Simmons, A., and Soci, C.: The ERA5 Global Reanalysis: achieving a detailed record of the climate and weather for the past 70 years., , https://doi.org/10.5194/egusphere-egu2020-10375, 2020.

Hock, R., Bliss, A., Marzeion, B., Giesen, R. H., Hirabayashi, Y., Huss, M., Radić, V., and Slangen, A. B. A.: GlacierMIP – A model intercomparison of global-scale glacier mass-balance models and projections, Journal of Glaciology, 65, 453–467, https://doi.org/10.1017/jog.2019.22, 2019.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

Hourdin, F., Ferster, B., Deshayes, J., Mignot, J., Musat, I., and Williamson, D.: Toward machine-assisted tuning avoiding the underestimation of uncertainty in climate change projections, Science Advances, 9, eadf2758, https://doi.org/10.1126/sciadv.adf2758, 2023.

Huss, M. and Hock, R.: A new model for global glacier change and sea-level rise, Front. Earth Sci., 3, https://doi.org/10.3389/feart.2015.00054, 2015.

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B. J., Elmore, A. C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J. S., Koppes, M., Kraaijenbrink, P. D. A., Kulkarni, A. V., Mayewski, P. A., Nepal, S., Pacheco, P., Painter, T. H., Pellicciotti, F., Rajaram, H., Rupper, S., Sinisalo, A., Shrestha, A. B., Viviroli, D., Wada, Y., Xiao, C., Yao, T., and Baillie, J. E. M.: Importance and vulnerability of the world's water towers, Nature, 577, 364–369, https://doi.org/10.1038/s41586-019-1822-y, 2020.

Intergovernmental Panel On Climate Change (IPCC): Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1st ed., Cambridge University Press, https://doi.org/10.1017/9781009157896, 2023.

Jennings, K. S., Winchell, T. S., Livneh, B., and Molotch, N. P.: Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere, Nat Commun, 9, 1148, https://doi.org/10.1038/s41467-018-03629-7, 2018.

Lecavalier, B. S. and Tarasov, L.: A history-matching analysis of the Antarctic Ice Sheet since the last interglacial – Part 1: Ice sheet evolution, https://doi.org/10.5194/egusphere-2024-1291, 15 May 2024.

Loeppky, J. L., Sacks, J., and Welch, W. J.: Choosing the Sample Size of a Computer Experiment: A Practical Guide, Technometrics, 51, 366–376, https://doi.org/10.1198/TECH.2009.08040, 2009.

Marzeion, B., Jarosch, A. H., and Hofer, M.: Past and future sea-level change from the surface mass balance of glaciers, The Cryosphere, 6, 1295–1322, https://doi.org/10.5194/tc-6-1295-2012, 2012.

Marzeion, B., Hock, R., Anderson, B., Bliss, A., Champollion, N., Fujita, K., Huss, M., Immerzeel, W. W., Kraaijenbrink, P., Malles, J.-H., Maussion, F., Radić, V., Rounce, D. R., Sakai, A., Shannon, S., van de Wal, R., and Zekollari, H.: Partitioning the Uncertainty of Ensemble Projections of Global Glacier Mass Change, Earth's Future, 8, e2019EF001470, https://doi.org/10.1029/2019EF001470, 2020.

Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., Gregor, P., Jarosch, A. H., Landmann, J., Oesterle, F., Recinos, B., Rothenpieler, T., Vlug, A., Wild, C. T., and Marzeion, B.: The Open Global Glacier Model (OGGM) v1.1, Geoscientific Model Development, 12, 909–931, https://doi.org/10.5194/gmd-12-909-2019, 2019.

McKay M. D., Beckman, R. J., and W. J. Conover, W. J.: A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics, 21:2, 239–245, https://doi.org/10.2307/1268522, 1979

Meehl G.A. et al. (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models, Sci. Adv. 6:26, eaba1981, https://doi.org/10.1126/sciadv.aba1981, 2020.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

Mölg, T., Hardy, D. R., and Kaser, G.: Solar-radiation-maintained glacier recession on Kilimanjaro drawn from combined ice-radiation geometry modeling, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2003JD003546, 2003.

Oerlemans, J. and Knap, W. H.: A 1 year record of global radiation and albedo in the ablation zone of Morteratschgletscher, Switzerland, Journal of Glaciology, 44, 231–238, https://doi.org/10.3189/S0022143000002574, 1998.

Pukelsheim, F.: The Three Sigma Rule, The American Statistician, 48, 88–91, https://doi.org/10.1080/00031305.1994.10476030, 1994.

Radić, V., Hock, R., and Oerlemans, J.: Analysis of scaling methods in deriving future volume evolutions of valley glaciers, Journal of Glaciology, 54, 601–612, https://doi.org/10.3189/002214308786570809, 2008.

Raoult, N., Beylat, S., Salter, J. M., Hourdin, F., Bastrikov, V., Ottlé, C., and Peylin, P.: Exploring the potential of history matching for land surface model calibration, Geoscientific Model Development, 17, 5779–5801, https://doi.org/10.5194/gmd-17-5779-2024, 2024.

Rounce, D. R., Hock, R., and Shean, D. E.: Glacier Mass Change in High Mountain Asia Through 2100 Using the Open-Source Python Glacier Evolution Model (PyGEM), Front. Earth Sci., 7, https://doi.org/10.3389/feart.2019.00331, 2020a.

Rounce, D. R., Khurana, T., Short, M. B., Hock, R., Shean, D. E., and Brinkerhoff, D. J.: Quantifying parameter uncertainty in a large-scale glacier evolution model using Bayesian inference: application to High Mountain Asia, Journal of Glaciology, 66, 175–187, https://doi.org/10.1017/jog.2019.91, 2020b.

Rounce, D. R., Hock, R., Maussion, F., Hugonnet, R., Kochtitzky, W., Huss, M., Berthier, E., Brinkerhoff, D., Compagno, L., Copland, L., Farinotti, D., Menounos, B., and McNabb, R. W.: Global glacier change in the 21st century: Every increase in temperature matters, Science, 379, 78–83, https://doi.org/10.1126/science.abo1324, 2023.

Schuster, L., Rounce, D. R., and Maussion, F.: Glacier projections sensitivity to temperature-index model choices and calibration strategies, Annals of Glaciology, 64, 293–308, https://doi.org/10.1017/aog.2023.57, 2023.

Schuster, L., Zekollari, H., Maussion, F., Hock, R., Marzeion, B., Rounce, D. R., Compagno, L., Fujita, K., Huss, M., James, M., Kraaijenbrink, P. D. A., Lipscomb, W. H., Minallah, S., Oberrauch, M., and Van Tricht, L.: Data from Glacier Model Intercomparison Project Phase 3 (GlacierMIP3), https://doi.org/10.5281/zenodo.15046588, 2025.

Shean, D. E., Bhushan, S., Montesano, P., Rounce, D. R., Arendt, A., and Osmanoglu, B.: A Systematic, Regional Assessment of High Mountain Asia Glacier Mass Balance, Front. Earth Sci., 7, https://doi.org/10.3389/feart.2019.00363, 2020.

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., and Yamazaki, K.: History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble, Clim Dyn, 41, 1703–1729, https://doi.org/10.1007/s00382-013-1896-4, 2013.

Williamson, D., Blaker, A. T., Hampton, C., and Salter, J.: Identifying and removing structural biases in climate models with history matching, Clim Dyn, 45, 1299–1324, https://doi.org/10.1007/s00382-014-2378-z, 2015.

https://doi.org/10.5194/egusphere-2026-2069
Preprint. Discussion started: 8 May 2026

Williamson, D. B., Blaker, A. T., and Sinha, B.: Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model, Geoscientific Model Development, 10, 1789–1816, https://doi.org/10.5194/gmd-10-1789-2017, 2017.

Zekollari, H., Schuster, L., Maussion, F., Hock, R., Marzeion, B., Rounce, D. R., Compagno, L., Fujita, K., Huss, M., James, M., Kraaijenbrink, P. D. A., Lipscomb, W. H., Minallah, S., Oberrauch, M., Van Tricht, L., Champollion, N., Edwards, T., Farinotti, D., Immerzeel, W., Leguy, G., and Sakai, A.: Glacier preservation doubled by limiting warming to 1.5°C versus 2.7°C, Science, 388, 979–983, https://doi.org/10.1126/science.adu4675, 2025.

Zemp, M., Jakob, L., Dussaillant, I., Nussbaumer, S. U., Gourmelen, N., Dubber, S., A, G., Abdullahi, S., Andreassen, L. M., Berthier, E., Bhattacharya, A., Blazquez, A., Boehm Vock, L. F., Bolch, T., Box, J., Braun, M. H., Brun, F., Cicero, E., Colgan, W., Eckert, N., Farinotti, D., Florentine, C., Floricioiu, D., Gardner, A., Harig, C., Hassan, J., Hugonnet, R., Huss, M., Jóhannesson, T., Liang, C.-C. A., Ke, C.-Q., Khan, S. A., King, O., Kneib, M., Krieger, L., Maussion, F., Mattea, E., McNabb, R., Menounos, B., Miles, E., Moholdt, G., Nilsson, J., Pálsson, F., Pfeffer, J., Piermattei, L., Plummer, S., Richter, A., Sasgen, I., Schuster, L., Seehaus, T., Shen, X., Sommer, C., Sutterley, T., Treichler, D., Velicogna, I., Wouters, B., Zekollari, H., Zheng, W., and The GlaMBIE Team: Community estimate of global glacier mass changes from 2000 to 2023, Nature, 639, 382–388, https://doi.org/10.1038/s41586-024-08545-z, 2025.

Zolles, T., Maussion, F., Galos, S. P., Gurgiser, W., and Nicholson, L.: Robust uncertainty assessment of the spatio-temporal transferability of glacier mass and energy balance models, The Cryosphere, 13, 469–489, https://doi.org/10.5194/tc-13-469-2019, 2019.