Journal of Glaciology



Article

Cite this article: Sjursen KH, Dunse T, Tambue A, Schuler TV, Andreassen LM (2023). Bayesian parameter estimation in glacier mass-balance modelling using observations with distinct temporal resolutions and uncertainties. *Journal of Glaciology* **69**(278), 1803–1822. https://doi.org/10.1017/jog.2023.62

Received: 17 December 2022
Revised: 19 July 2023
Accepted: 21 July 2023
First published online: 31 August 2023

Keywords:
Glacier mass balance; glacier modelling; surface mass budget

Corresponding author:
Kamilla H. Sjursen;
Email: kamilla.hauknes.sjursen@hvl.no

© The Author(s), 2023. Published by Cambridge University Press on behalf of International Glaciological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

cambridge.org/jog

Bayesian parameter estimation in glacier mass-balance modelling using observations with distinct temporal resolutions and uncertainties

Kamilla H. Sjursen[1], Thorben Dunse[1], Antoine Tambue[2], Thomas V. Schuler[3] and Liss M. Andreassen[4]

[1]Department of Environmental Sciences, Western Norway University of Applied Sciences, Sogndal, Norway; [2]Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway; [3]Department of Geosciences, University of Oslo, Oslo, Norway and [4]Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway

Abstract

Empirical glacier mass-balance models are commonly used in assessments of glacier and runoff evolution. Recent satellite-borne geodetic mass-balance observations of global coverage facilitate large-scale model calibration that previously relied on sparse in situ observations of glacier mass change. Geodetic observations constitute temporally aggregated mass-balance signals with significant uncertainty, raising questions about the role of observations with different temporal resolutions and uncertainties in constraining model parameters. We employ a Bayesian approach and demonstrate the sensitivity of parameter values to commonly used mass-balance observations of seasonal, annual and decadal resolution with uncertainties characteristic to in situ and satellite-borne observations. For glaciers along a continentality gradient in Norway, the use of annual mass balances results in around 20% lower magnitude of modelled ablation and accumulation (1960–2020), compared to employing seasonal balances. Decadal mass balance also underestimates magnitudes of ablation and accumulation, but parameter values are strongly influenced by the prior distribution. The datasets yield similar estimates of annual mass balance with different margins of uncertainty. Decadal observations are afflicted with considerable uncertainty in mass-balance sensitivity due to high parameter uncertainty. Our results highlight the importance of seasonal observations when model applications require accurate magnitudes of ablation, e.g. to estimate meltwater runoff.

1. Introduction

Glacier mass-balance models are important tools for assessing glacier evolution and associated meltwater runoff, and to describe mass and energy exchange between the glacier surface and the atmosphere. Mass-balance models can be used to analyse past trends and interannual and seasonal variations in glacier mass change (e.g. Engelhardt and others, 2013; van Pelt and others, 2019; Eidhammer and others, 2021; Geck and others, 2021), to explore glacier sensitivity to climate perturbations (e.g. Schuler and others, 2005; Andreassen and Oerlemans, 2009; Giesen and Oerlemans, 2010; Engelhardt and others, 2015) and to provide predictions of glacier and runoff evolution with climate change on local, regional and global scales (e.g. Huss and Hock, 2018; Zekollari and others, 2019; Rounce and others, 2020a; Compagno and others, 2021). Temperature index models use empirical relationships to describe accumulation and ablation on the glacier surface. While more complex physically based energy-balance approaches may require a large amount of temporally and spatially distributed meteorological data and significant computational resources, temperature index models are less data intensive, computationally inexpensive and often used in large-scale assessments of glacier change (e.g. Huss and Hock, 2015; Maussion and others, 2019; Rounce and others, 2020b).

The performance of temperature index models relies on the availability of observations of glacier mass change to calibrate model parameters. However, in situ observations of glacier mass change exist for less than 0.3% of the world's glaciers (WGMS, 2022). Long-term series are rare, especially of seasonal mass balances. Only 42 glaciers worldwide have ongoing (2019/2020) long-term reference series (30 years or more) of both annual and seasonal mass-balance measurements (WGMS, 2021). The scarcity of observations poses a challenge in temperature index modelling of unmonitored glaciers. Approaches adopted in regional and global models include calibrating parameters based on regional average mass-balance estimates (Huss and Hock, 2015) or employing mass-balance observations from a nearby glacier of similar size (Zekollari and others, 2019). Although neighbouring glaciers experience similar climatic conditions, their mass balances can differ significantly due to topographic controls or lake-enhanced melt (Andreassen and others, 2020). Another calibration strategy for modelling unmonitored glaciers is to employ a uniform and constant set of parameters for a large number of glaciers, constrained by relatively few mass-balance observations (e.g. Engelhardt and others, 2013). This may be problematic because the transferability of parameter values in space and time is not necessarily given (e.g. Gabbi and others, 2014).

In recent years, the increasing availability of multi-temporal digital elevation models (DEMs) from satellite remote sensing has enabled regional and global-scale geodetic mapping of glacier changes (e.g. Brun and others, 2017; Shean and others, 2020; Hugonnet and others, 2021). For most of the world's glacier population, geodetic observations constitute the only source of information on individual glacier change and have thus been increasingly applied to determine glacier-specific parameter values in calibration of large-scale glacier evolution models (e.g. Zekollari and others, 2019; Rounce and others, 2020a; Compagno and others, 2021, 2022). Satellite-borne geodetic observations provide mass-balance estimates over multi-year periods and at higher uncertainties compared to geodetic observations based on airborne surveys. The relatively high uncertainty in satellite-borne geodetic observations is mainly attributed to the glacier-wide extrapolation of variable spatio-temporal sampling and uncertainties in densities used to convert glacier volume to mass change (Hugonnet and others, 2021).

Temperature index models are known to suffer from *equifinality* (Beven, 2006); multiple equally plausible parameter sets produce results that are in line with observations, due to the combination of a simple, parameterised model structure that exhibits compensating effects, and a lack of observations to constrain representations of mass change. The typically low temporal resolutions and high uncertainties of geodetic mass-balance estimates may hamper constraining model parameters and promote equifinality. Although the calibration of parameter values for individual glaciers using satellite-borne geodetic observations has been shown to improve the spatial and temporal resolution of mass-balance estimates in large-scale modelling (Rounce and others, 2020a), it is still unclear how the use of such observations influences parameter values and thus, model sensitivity.

Various calibration techniques have been employed to determine equifinal parameter sets that satisfy some model performance criteria, e.g. stepwise adjustment of parameter values (e.g. Schuler and others, 2007; Compagno and others, 2021), Monte Carlo simulations (e.g. Østby and others, 2013; Engelhardt and others, 2014) and automatic multi-objective optimisation schemes (e.g. Rye and others, 2012; Zolles and others, 2019). However, these techniques do not provide a formal quantification of parameter uncertainty in probabilistic terms, and the effect of the design of the calibration on parameter estimates is largely unexplored (van Tiel and others, 2020).

In recent years, Bayesian approaches have been increasingly applied in glacier mass-balance studies (e.g. Eckert and others, 2011; Martín-Español and others, 2016, 2017; Zemp and others, 2019; Rounce and others, 2020a, 2020b; Landmann and others, 2021). Rounce and others (2020b) introduced Bayesian inference in temperature index mass-balance modelling to perform rigorous assessment of parameter uncertainty. A Bayesian framework allows to infer a joint posterior probability distribution of parameter values from observations, conditioned on a joint prior distribution that represents our prior knowledge of plausible parameter values. The marginal posterior distributions, i.e. the posterior distributions of each individual parameter, can be estimated from the joint posterior distribution. Comparison of the statistics of the marginal prior and posterior distributions can be used to assess the degree to which different datasets and their uncertainties constrain parameter values. Analysing the marginal posterior distributions can also give indications of shortcomings in the model structure and bias in the input data, e.g. through poorly constrained parameters or high correlation between distributions (Rounce and others, 2020b), or a high probability of 'unphysical' parameter values.

In this work, we apply a Bayesian framework to investigate the impact of employing different mass-balance observations commonly used in calibration of parameters in glacier mass-balance modelling, each with their characteristic temporal resolution and uncertainty. We apply a temperature index model to simulate surface mass balance (SMB) of seven glaciers in Norway that have long-term series (>30 years) of in situ mass-balance observations and are situated along a maritime-continental climate gradient. For each glacier, we estimate three sets of parameters based on three sets of data derived from in situ observations: (1) glacier-wide seasonal SMB, (2) glacier-wide annual SMB and (3) 10-year averages of glacier-wide annual SMB. For seasonal and annual SMB, we employ uncertainties reported in glaciological records. The 10-year averages of annual SMB represent our analogue to satellite-based decadal geodetic observations, and we thus assign uncertainties characteristic to such observations to this dataset. The observational datasets differ in temporal resolutions and uncertainties but maintain a consistent mass-balance signal. We quantify the uncertainty associated with the parameters inferred from the datasets and assess the degree to which posterior parameter distributions are informed by the observations in each of the three experiments. Our aim is to assess how the use of different observational datasets with distinct temporal resolutions and uncertainties affects model performance and influences modelled seasonal and annual SMB.

2. Study sites

The glaciers selected for mass-balance modelling in this study are situated along a climate gradient in southern Norway, from the maritime Ålfotbreen and Hansebreen glaciers in the west to the continental Gråsubreen glacier in the interior part of the country (Fig. 1). The glaciers are part of the monitoring program of the Norwegian Water Resources and Energy Directorate (NVE) and have long glaciological SMB records (>30 years; Table 1). We employ the reanalysed SMB series (available at http://glacier.nve.no/viewer/CI), which are partly homogenised (all seven glaciers) and partly calibrated (Ålfotbreen, Hansebreen, and Nigardsbreen) where discrepancies between glaciological and geodetic balances were significant (Andreassen and others, 2016; Kjøllmoen, 2022a, 2022b). Over the past 60 years, the SMB records show large inter-annual and decadal fluctuations. All seven glaciers gained mass in the early 1990s, a period of particularly high snow accumulation on maritime glaciers in southern Norway. Since the early 2000s, all seven glaciers have been experiencing mass loss and retreat (Andreassen and others, 2020).

3. Glacier mass-balance model

3.1. Model description

The glacier mass-balance model computes SMB $b_{\mathrm{mod},i,t}$ as the sum of accumulation $a_{i,t}$, ablation $m_{i,t}$, and refreezing $r_{i,t}$ in each model

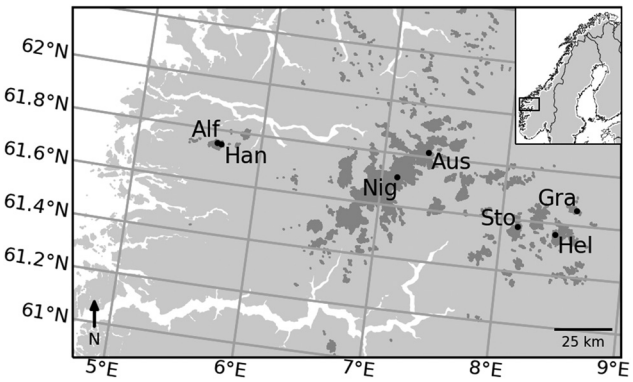


Figure 1. Overview map of the seven study glaciers in southern Norway: Ålfotbreen (Alf), Hansebreen (Han), Nigardsbreen (Nig), Austdalsbreen (Aus), Storbreen (Sto), Helstugubreen (Hel) and Gråsubreen (Gra). Glacierized areas are shown in dark grey.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

Table 1. Characteristics of glaciers selected for this study and overview of surface mass balance (SMB) records

Glacier name	Abbreviation	Area (km^2)	Altitude range (m a.s.l.)	Slope (°)	Start year	Outlines	$\overline{B_w}$ (m w.e. a^{-1})	$\overline{B_s}$ (m w.e. a^{-1})	$\overline{B_a}$ (m w.e. a^{-1})
Ålfotbreen	Alf	3.5	1000 – 1360	10	1963	1968, −88, −97, 2010, −19	3.74	−4.07	−0.34
Hansebreen	Han	2.5	927 – 1303	9	1986	1968, −88, −97, 2010, −19	3.48	−4.12	−0.64
Nigardsbreen	Nig	44.9	389 – 1955	8	1962	1964, −74, −84, 2009, −13, −20	2.32	−2.29	0.03
Austdalsbreen	Aus	10.1	1200 – 1740	5	1988	1966, −88, 2009, −19	2.17	−2.63	−0.46
Storbreen	Sto	4.9	1420 – 2091	14	1949	1951, −68, −84, −97, 2009, −19	1.46	−1.98	−0.52
Hellstugubreen	Hel	2.7	1482 – 2229	13	1962	1962, −68, −80, −97, 2009, −19	0.75	−1.30	−0.56
Gråsubreen	Gra	1.7	1854 – 2277	12	1962	1968, −84, −97, 2009, −19	1.10	−1.67	−0.57

Glaciers are sorted from west to east along a continentality gradient from maritime towards continental conditions. Area and altitude range are taken from Kjøllmoen and others (2022c), and slope refers to the latest glacier inventory from 2018/2019 (Andreassen and others, 2022). Start year refers to the start year of SMB measurements. Outline refers to years of glacier outlines applied in the homogenised SMB record of each glacier (Kjøllmoen, 2022b). $\overline{B_w}$, $\overline{B_s}$ and $\overline{B_a}$ refer to average winter, summer and annual balance, respectively, over the common period of SMB measurements 1988–2020 (http://glacier.nve.no/viewer/CI).

grid cell i (1×1 km in this study) on a daily time step t:

$$b_{\mathrm{mod},i,t} = a_{i,t} + m_{i,t} + r_{i,t}. \quad (1)$$

The daily mean temperature $T_{i,t}$ (°C) and daily total precipitation $P_{i,t}$ (mm) are calculated from the corresponding climate forcing data $T_{f,i,t}$ and $P_{f,i,t}$ as follows:

$$T_{i,t} = T_{f,i,t} + T_{\mathrm{corr}}, \quad (2)$$

$$P_{i,t} = P_{\mathrm{corr}} \times P_{f,i,t}, \quad (3)$$

where P_{corr} (–) is a precipitation correction factor and T_{corr} (°C) is a temperature bias correction. Daily accumulation $a_{i,t}$ (mm w.e. d^{-1}) is calculated as the fraction of precipitation falling as snow assuming a linear transition interval ΔT (°C) around a snow/rain temperature threshold T_s (°C) as follows:

$$a_{i,t} = \begin{cases} P_i & T_{i,t} \leq T_s - \frac{\Delta T}{2} \\ 0 & T_{i,t} \geq T_s + \frac{\Delta T}{2} \\ P_i\left(\frac{T_s - T_{i,t}}{\Delta T} + \frac{1}{2}\right) & T_s - \frac{\Delta T}{2} < T_{i,t} < T_s + \frac{\Delta T}{2} \end{cases}. \quad (4)$$

Daily melt of snow and ice in each grid cell $m_{\mathrm{snow/ice},i,t}$ (mm w.e. d^{-1}) is modelled using a temperature index approach:

$$m_{\mathrm{snow/ice},i,t} = \begin{cases} MF_{\mathrm{snow/ice}}(T_{i,t} - T_m) & T_{i,t} > T_m \\ 0 & T_{i,t} \leq T_m \end{cases}, \quad (5)$$

where $MF_{\mathrm{snow}}/_{\mathrm{ice}}$ (mm w.e.°$C^{-1}d^{-1}$) are melt factors for snow and ice, relating the amount of melt of snow and ice to the temperature difference above the melt threshold temperature T_m (°C). Melt factors partly reflect differences in surface albedo of snow and ice. Since characteristic values of the albedo of firn are between those of snow and ice (e.g. Cuffey and Paterson, 2010), daily melt of firn $m_{\mathrm{firn},i,t}$ (mm w.e. d^{-1}) is estimated as the average of daily melt of snow and ice:

$$m_{\mathrm{firn},i,t} = \frac{m_{\mathrm{snow},i,t} + m_{\mathrm{ice},i,t}}{2}. \quad (6)$$

At the onset of melt, refreezing of meltwater in a grid cell is assumed to occur until a maximum refreezing potential in the cell is reached. Potential refreeze in each cell $R_{\mathrm{pot},i}$ (mm w.e. d^{-1}) is calculated based on the study by Woodward and others (1997) as a function of the mean annual air temperature in the cell $\overline{T}_{a,i}$ (°C) according to:

$$R_{\mathrm{pot},i} = -0.69\overline{T}_{a,i} - 0.0096. \quad (7)$$

Refreezing is assumed to occur in the snowpack, such that refrozen meltwater is added to the snow reservoir in the given cell. Refreezing is thus assumed to occur above the summer surface, such that it contributes to SMB and not to internal accumulation within the glacier. Potential refreeze is reset at the end of each hydrological year (defined as 1 October–30 September). In each grid cell, the snow/firn reservoir is required to melt away before melting the underlying firn/ice. At the end of the hydrological year, a fraction (25%, same as in Engelhardt and others (2014) for glaciers in Norway) of firn is assumed to become ice, and the remaining snow is added to the firn reservoir. The specific glacier-wide SMB B_{mod} (m w.e.) over a time period τ is calculated over the glacier area A as follows:

$$B_{\mathrm{mod}} = \frac{1}{A}\sum_i b_{\mathrm{mod},i,\tau}A_i, \quad (8)$$

where A_i is the glacier area in cell i. Glacier-wide annual SMB is the sum of winter and summer balance, which are calculated from the maximum and minimum glacier mass over the accumulation and ablation seasons, respectively.

3.2. Input data

To force the model, we employ gridded daily temperature and precipitation from seNorge_2018 (Lussana and others, 2019), provided at 1 km resolution over mainland Norway from 1957 to present. SeNorge ('see Norway') is a collection of long-term observational datasets based on spatial interpolation of in situ observations from a large network of weather stations and is developed and maintained by the Norwegian Meteorological Institute (MET Norway). Since the release of the first seNorge version, seNorge1 (Mohr, 2008), several studies have demonstrated the usefulness of the datasets in glacio-hydrological modelling in Norway (e.g. Engelhardt and others, 2012, 2013; Li and others, 2015). Engelhardt and others (2012) found good agreement between modelled and measured winter balance at stake locations on Ålfotbreen and Nigardsbreen using seNorge1, while results for Storbreen indicated a 20% underestimation of winter balance. The latest version (seNorge_2018; Lussana and others (2019)) applies monthly precipitation reference fields derived from a regional climate simulation based on dynamical downscaling of ERA-Interim. The use of reference fields allows to better capture spatial precipitation patterns over data-sparse regions. In addition, the in situ precipitation measurements used in the interpolation scheme are corrected for wind-induced undercatch. While precipitation estimates in seNorge_2018 are expected to be superior to previous versions, they are still prone to uncertainties in mountainous regions with low station density and complex terrain. For intense precipitation (observed values greater than 10 mm d^{-1}), the probability of large errors in data-sparse areas is estimated to around 30% (Lussana and others, 2019). Winter (October–April) precipitation in seNorge_2018 is 44%, 32% and 18%

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

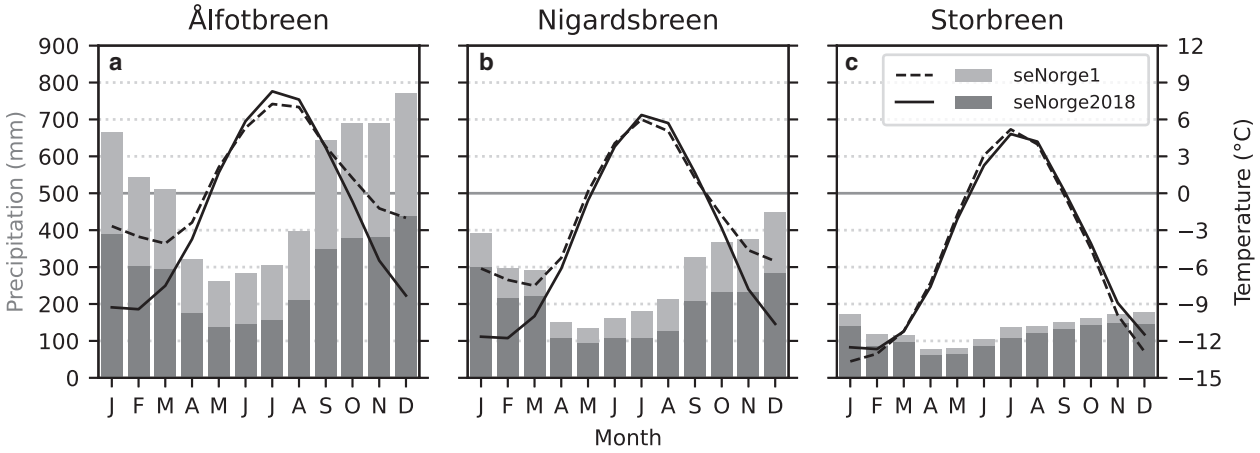


Figure 2. Mean monthly precipitation sums (bars) and mean monthly temperature (lines) for Ålfotbreen (a), Nigardsbreen (b), and Storbreen (c) from seNorge1 (light grey bars and dashed lines) and seNorge_2018 (dark grey bars and solid lines) for the period 1960–2019. seNorge_2018 is used as forcing in this study.

lower than seNorge1 for Ålfotbreen, Nigardsbreen and Storbreen, respectively (Fig. 2).

The glacier mass-balance model employs the seNorge DEM (1 km resolution), on which the seNorge_2018 data are provided, such that no downscaling of the climate data is required. As input to the mass-balance model, we produce a sequence of glacier masks for each glacier by intersecting multi-temporal glacier outlines with the seNorge DEM. Glacier outlines are provided by NVE and based on the area employed in reanalysis of glaciological records (Kjøllmoen, 2022b). Between four and six outlines are available for each glacier over the period 1957–2020 (Table 1). In accordance with the procedure applied in reanalysis of the SMB record, the mass-balance model updates the glacier mask such that each mask is used for half of the period before and after each update (Andreassen and others, 2016).

3.3. Model parameters

The mass-balance model contains several parameters that control accumulation, melt and refreezing: precipitation and temperature corrections (P_{corr} and T_{corr}), temperature thresholds for snow and melt (T_s and T_m) and melt factors for snow and ice (MF_{snow} and MF_{ice}). In our comparative analysis, the dataset of 10-year average annual SMB is expected to impose the weakest constraints on parameter values and is thereby the limiting case that determines the number of parameters that can be reasonably estimated. Rounce and others (2020b) found that, for most glaciers in High Mountain Asia, the joint posterior distribution of three parameters could be informed by a single 18-year geodetic mass-balance estimate. Following their findings, we focus our investigation on the estimation of precipitation and temperature bias corrections, P_{corr} and T_{corr}, and the melt factor for snow MF_{snow} and define $MF_{\text{ice}} = MF_{\text{snow}}/0.7$, based on the assumption that the efficiency of snow melt is 70% of that of ice (e.g. Singh and others (2000), same as assumed by Rounce and others (2020b)). For each glacier, we define the set of model parameters to be estimated as $\theta = \{P_{\text{corr}},\ T_{\text{corr}},\ MF_{\text{snow}}\}$. We assign plausible values for the remaining parameters based on what is commonly reported in the literature: $T_m = 0°\text{C}$, $T_s = 1°\text{C}$ and $\Delta T = 2°\text{C}$ (e.g. Engelhardt and others, 2013; Huss and Hock, 2015; Rounce and others, 2020b).

4. Parameter estimation

4.1. Experiment set-up

We estimate parameters for each glacier based on three representations of the reanalysed glaciological SMB records (Kjøllmoen and others, 2022c). The three datasets are assigned uncertainties associated with observations commonly used in calibration of mass-balance model parameters (Fig. 3). The three experiments, which we henceforth refer to by their respective abbreviation, are as follows: (1) glacier-wide seasonal SMB, $B_{w/s}$; and (2) glacier-wide annual SMB, B_a, both with uncertainties characteristic to glaciological observations; and (3) 10-year averages of glacier-wide annual SMB, B_{10yr}, with uncertainties characteristic to decadal geodetic observations derived from satellite remote sensing. The B_{10yr} experiment is thus our analogue to satellite-borne decadal geodetic observations (e.g. Shean and others, 2020; Hugonnet and others, 2021), but it is derived from the glaciological records to avoid any influence on parameter estimates that result from inconsistencies in mass-balance signals between different datasets. For each glacier, we employ 20 years of the respective SMB records for parameter estimation and remaining years between 1960 and 2020 (minimum 13 years) for validation of modelled SMB. We use the period 1990–2009 for parameter estimation because (1) all seven study glaciers have SMB records that cover this period, and (2) it spans years of variability in climate and mass-balance trends (Andreassen and others, 2020), reducing parameter bias towards specific meteorological conditions. To summarise, for the 20-year calibration period, the three datasets contain (1) 40 observations that express seasonal variability in

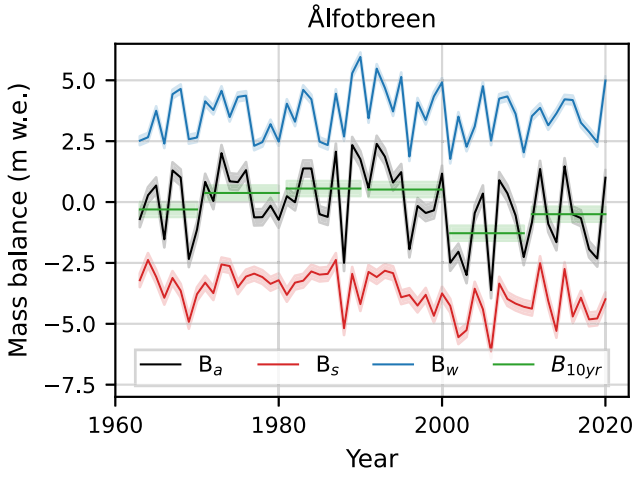


Figure 3. Observational record for Ålfotbreen showing winter (blue, top), summer (red, bottom), annual (black, middle) and 10-year average annual (green, horizontal) surface mass balance (SMB). Shaded areas represent estimated one standard deviation uncertainties in observations. For the 10-year average annual SMB, uncertainties reflect those of decadal geodetic observations from satellite remote sensing.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

SMB; (2) 20 observations that describe annual variability in SMB and (3) two observations (1990–99 and 2000–09) that describe decadal variability in SMB.

4.2. Bayesian inference

We are interested in estimating the probability distributions of the set of mass-balance model parameters $\theta = \{P_{\text{corr}},\ T_{\text{corr}},\ MF_{\text{snow}}\}$, given a set of $n = 1{:}N$ observations of glacier mass change $B_{\text{obs},1:N}$. We assume that the observed SMB $B_{\text{obs},n}$ over the incremental time period τ (e.g. season, year, decade) can be explained by the modelled SMB $B_{\text{mod},n}(X_n, \theta)$ and the error associated with the observations ϵ_n:

$$B_{\text{obs},n} = B_{\text{mod},n}(X_n, \theta) + \epsilon_n, \tag{9}$$

where X_n is the set of model input data over τ.

In a Bayesian framework, we can express the joint posterior distribution of the set of parameters θ, given a set of N observations, $p(\theta|B_{\text{obs},1:N}, X_{1:N})$ in terms of Bayes' theorem (see e.g. Gelman and others, 2014):

$$p(\theta|B_{\text{obs},1:N}, X_{1:N}) = \frac{p(\theta)p(B_{\text{obs},1:N}|\theta, X_{1:N})}{p(B_{\text{obs},1:N})}. \tag{10}$$

The term $p(\theta)$ is the joint prior distribution of θ and represents the beliefs about the value of the parameters before the observed data are considered. $p(B_{\text{obs},1:N}|\theta, X_{1:N})$ is termed the likelihood and is the probability of observing the data given a particular set of parameters and $p(B_{\text{obs},1:N})$ is the probability of observing the data, also called evidence. We can use Markov chain Monte Carlo (MCMC) simulations that make use of the following proportionality:

$$p(\theta|B_{\text{obs},1:N}, X_{1:N}) \propto p(\theta)p(B_{\text{obs},1:N}|\theta, X_{1:N}), \tag{11}$$

to obtain samples from the joint posterior distribution given appropriate formulations of the joint prior distribution and the likelihood.

4.3. Likelihood

We assume that errors ϵ_n in Eqn (9) are independent and normally distributed ($\mathcal{N}$) with mean of zero and constant variance $\sigma^2_{B_{\text{obs}}}$, such that the likelihood $L_{B_{\text{obs}}} = p(B_{\text{obs},1:N}|\theta, X_{1:N})$ can be expressed as follows:

$$L_{B_{\text{obs}}} = \prod_{n=1}^{N} \mathcal{N}(B_{\text{mod,n}}(X_n, \theta), \sigma^2_{B_{\text{obs}}}). \tag{12}$$

For computational efficiency and stability, we employ the log likelihood function $l_{B_{\text{obs}}} = \ln(L_{B_{\text{obs}}})$ which can be expressed as:

$$\begin{aligned} l_{B_{\text{obs}}} = &-\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_{B_{\text{obs}}}) \\ &- \frac{1}{2\sigma^2_{B_{\text{obs}}}}\sum_{n=1}^{N}(B_{\text{obs},n} - B_{\text{mod},n}(X_n, \theta))^2. \end{aligned} \tag{13}$$

For the parameter estimation experiments B_a and B_{10yr}, we express the log likelihood function as defined in Eqn (13). For $B_{w/s}$, we assume that observations of glacier-wide winter and summer balance ($B_{w,1:N}$ and $B_{s,1:N}$, respectively) are conditionally independent given the set of parameters θ, such that:

$$\begin{aligned} p(B_{w,1:N}, B_{s,1:N}|X_{1:N}, \theta) = &\ p(B_{w,1:N}|X_{1:N}, \theta) \\ &\times p(B_{s,1:N}|X_{1:N}, \theta). \end{aligned} \tag{14}$$

The full log likelihood function $l_{B_{w/s}}$ can be expressed as the sum of log likelihood functions for each of the winter (l_{B_w}) and summer (l_{B_s}) balance:

$$l_{B_{w/s}} = l_{B_w} + l_{B_s}. \tag{15}$$

Uncertainty in annual SMB (σ_{B_a}) for each glacier is assessed in the reanalysis of the glaciological SMB records (Andreassen and others, 2016). For each glacier, we assume that the reported uncertainty is constant throughout their respective period of observations. We derive uncertainties in winter and summer balance (σ_{B_w} and σ_{B_s}, respectively) given that observations of annual SMB are the sum of winter and summer balances and under the assumption that observations of seasonal balances are independent (Dyurgerov and Meier, 1999) and with normally distributed errors. The distribution of the error in observed annual SMB $\epsilon_{n,a}$ can thus be represented as follows:

$$\epsilon_{n,a} \sim \mathcal{N}(0,\ \sigma^2_{B_w} + \sigma^2_{B_s}), \tag{16}$$

which entails that the variance in observed annual SMB can be expressed as the sum of variances in observed seasonal balances:

$$\sigma^2_{B_a} = \sigma^2_{B_w} + \sigma^2_{B_s}. \tag{17}$$

We derive σ_{B_w} and σ_{B_s} from σ_{B_a} for each glacier (Table 2), assuming that uncertainty in observed summer balance accounts for two-thirds of the uncertainty in the annual balance based on experience and typical uncertainty estimates in the SMB records (e.g. Kjøllmoen and others, 2022c).

Since B_{10yr} represents our analogue to satellite-based geodetic mass balance, the uncertainty in these observations should reflect those typical to satellite-based decadal geodetic estimates. We estimate the uncertainty in B_{10yr} for each glacier as the average of the reported standard deviations in two glacier-specific decadal geodetic estimates (2000–09 and 2010–19) from Hugonnet and others (2021) and assign the resulting uncertainty to each time period 1990–99 and 2000–09. To illustrate, the uncertainty in geodetic mass balance for Storbreen is 0.24 m w.e. a^{-1} and 0.28 m w.e. a^{-1} for the decades 2000–09 and 2010–19 (Hugonnet and others, 2021), respectively, and we thus ascribe an uncertainty of 0.26 m w.e. a^{-1} to the B_{10yr} estimates (Table 2).

4.4. Prior distributions

The joint posterior distribution represents a belief of the parameter values inferred from the data (likelihood), conditioned on prior knowledge (joint prior distribution). The chosen form of the prior, in terms of strength and properties of the distribution, influences how much information the posterior distribution can gain from the observations. We choose weakly informative

Table 2. Uncertainties (standard deviation; m w.e. a^{-1}) of observations used in parameter estimation

Glacier name	$\sigma^{a}_{B_w}$	$\sigma^{a}_{B_s}$	$\sigma^{a}_{B_a}$	$\sigma^{b}_{B_{10yr}}$
Ålfotbreen	0.18	0.28	0.33	0.31
Hansebreen	0.18	0.28	0.33	0.33
Nigardsbreen	0.19	0.28	0.34	0.19
Austdalsbreen	0.17	0.25	0.30	0.23
Storbreen	0.10	0.15	0.18	0.26
Hellstugubreen	0.08	0.13	0.15	0.28
Gråsubreen	0.06	0.08	0.10	0.31

[a]Derived from Andreassen and others (2016).
[b]Derived from Hugonnet and others (2021).
Glaciers are sorted from west to east along a maritime to continental climate gradient.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

(wide) priors since we are interested in the effect of the observational datasets on the posterior distribution.

Priors are constructed such that they reflect our knowledge of physically plausible parameter values and prohibit sampling of nonphysical values. Braithwaite (2008) modelled accumulation at the equilibrium line altitude for 66 mid-latitude and polar glaciers and found that 76% of the accumulation variance could be represented by a temperature index model using a melt factor for snow of 4.1 ± 1.5 mm w.e.$^{\circ}$C^{-1}d^{-1}. This is in line with melt factors derived for glaciers in southern Norway (Engelhardt and others, 2012, 2014). Thus, we represent the prior distribution for the melt factor for snow as a truncated normal distribution ($\mathcal{N}_T$) with mean 4.1 mm w.e.$^{\circ}$C^{-1}d^{-1} and standard deviation 1.5 mm w.e.$^{\circ}$C^{-1}d^{-1}, truncated at 0 to ensure positivity (same as in the study by Rounce and others (2020b)):

$$p(MF_{snow}) = \mathcal{N}_T(4.1, 1.5^2, 0). \quad (18)$$

To inform our choice of prior distribution for the precipitation correction factor, we contrast precipitation and temperature in seNorge1 and seNorge_2018 (Fig. 2) in light of previous comparisons of precipitation sums in seNorge1 and winter mass balances (Engelhardt and others, 2012). We roughly estimate a 25% underestimation of precipitation in seNorge_2018 and choose the following truncated normal distribution as prior for the precipitation correction factor:

$$p(P_{corr}) = \mathcal{N}_T(1.25, 0.8^2, 0). \quad (19)$$

Although winter temperature has been highlighted as more prone to large errors in data-void areas (Lussana and others, 2019), there is no clear indication that seNorge_2018 under- or overestimates temperatures in mountainous regions. We therefore choose a normal distribution with mean 0°C and standard deviation 1.5°C as the prior for the temperature bias correction:

$$p(T_{corr}) = \mathcal{N}(0, 1.5^2). \quad (20)$$

We apply the same joint prior distribution in all parameter estimation experiments. We acknowledge that posterior estimates could be improved by adapting priors to site-specific conditions, e.g. by adopting different melt factors for maritime and continental glaciers or including information from alternative data sources where possible. However, the goal of our analysis is not to produce the best possible mass-balance estimates for each glacier, but to investigate data-driven differences in posterior distributions across glaciers.

4.5. Markov chain Monte Carlo sampling

MCMC methods (see e.g. Gelman and others, 2014, Chapter 11) provide techniques for generating random samples from the joint probability distribution of continuous random variables that is proportional to a known function (Eqn (11)). Through a sequential process, proposed samples of the joint posterior distribution are either accepted or rejected, depending on how likely a sample is to explain the observations. Given enough steps in a Markov chain, the sample distribution converges towards the joint posterior distribution from which statistics of the marginal posterior distributions can be estimated.

The Bayesian model is implemented using the probabilistic programming framework *PyMC3* for Python (Salvatier and others, 2016), which provides a suite of MCMC sampling algorithms. We employ an adaptive Metropolis-type algorithm, DEMetropolisZ (based on ter Braak and Vrugt (2008)), which offers increased efficiency compared to a standard random-walk Metropolis (Metropolis and others, 1953) through an adaptive mechanism for which proposed samples of the posterior distribution are produced (ter Braak, 2006). In accordance with the recommendation of Vehtari and others (2021), we run four parallel MCMC chains to allow for reliable assessment of convergence. We apply a similar modelling approach to Rounce and others (2020b) and estimate the same number of parameters. On the basis of their evaluation of diagnostics and chain convergence, we consider that running four MCMC chains for each glacier with 2000 tune and 10 000 sampling iterations each is sufficient to ensure chain convergence and an adequate number of independent samples of the posterior distribution. We assess convergence and the stability of posterior estimates using the metrics rank-normalised $\widehat{R}$ and rank-normalised effective sample size (bulk and tail) recommended by Vehtari and others (2021), and the Monte Carlo standard error (bulk and tail). Although we can never conclusively claim convergence, visual and numerical inspection does not indicate nonconvergence in any simulation cases (see Supplementary Material for details). Chains display good mixing and stationarity, and provide a sufficient number of independent samples across all quantiles, such that we are confident that the MCMC simulations provide satisfactory approximations of the marginal posterior distributions of each parameter across all experiments.

5. Results

5.1. Posterior distributions of model parameters

The difference between marginal posterior and prior parameter distributions in each experiment reflects the influence of the observations (through the likelihood) on the posterior distribution. In all experiments, the spread of the marginal posterior distributions of the parameters (hereafter referred to as parameter uncertainty) is lower than that of the corresponding priors, implying that posterior distributions are to some degree informed by the observations (Fig. 4). In the $B_{w/s}$ and B_a experiments, posteriors are well constrained. Posteriors are less constrained in the B_{10yr} experiment and appear to largely reflect the priors, although there is some shift in the mode and reduction in the spread of the posteriors compared to the priors.

B_a generally yields lower values for P_{corr} and T_{corr} than $B_{w/s}$ and B_{10yr}, but MF_{snow} is on average slightly higher (see Tables S1–S3 in Supplementary Material for details). The highest values of P_{corr} and MF_{snow} are inferred for the most maritime glaciers, Ålfotbreen and Hansebreen. For most glaciers and parameters, the interquartile range of posterior distributions for B_{10yr} contains the median of posteriors for $B_{w/s}$ and B_a (Fig. 4). This reflects that high probability parameter values for $B_{w/s}$ and B_a are also considered likely in the B_{10yr} experiment. The posterior distributions for $B_{w/s}$ and B_a, however, do not show agreement. This is especially visible for P_{corr}, for which posterior distributions for B_a yield lower values than $B_{w/s}$ across all glaciers. The difference in parameter uncertainty between B_{10yr} and $B_{w/s}$ is typically less than one order of magnitude for MF_{snow} and T_{corr} and greater than one order of magnitude for P_{corr}. Parameter uncertainty is higher for B_a than $B_{w/s}$, but significantly lower than for B_{10yr}.

5.2. Model performance

We evaluate the performance of posterior distributions in modelling seasonal and annual balances in two main ways: (1) in terms of root mean squared error (RMSE) and mean bias between modelled (posterior predictive median) and observed mass balances over the calibration and validation years (Fig. 5); and (2) by

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

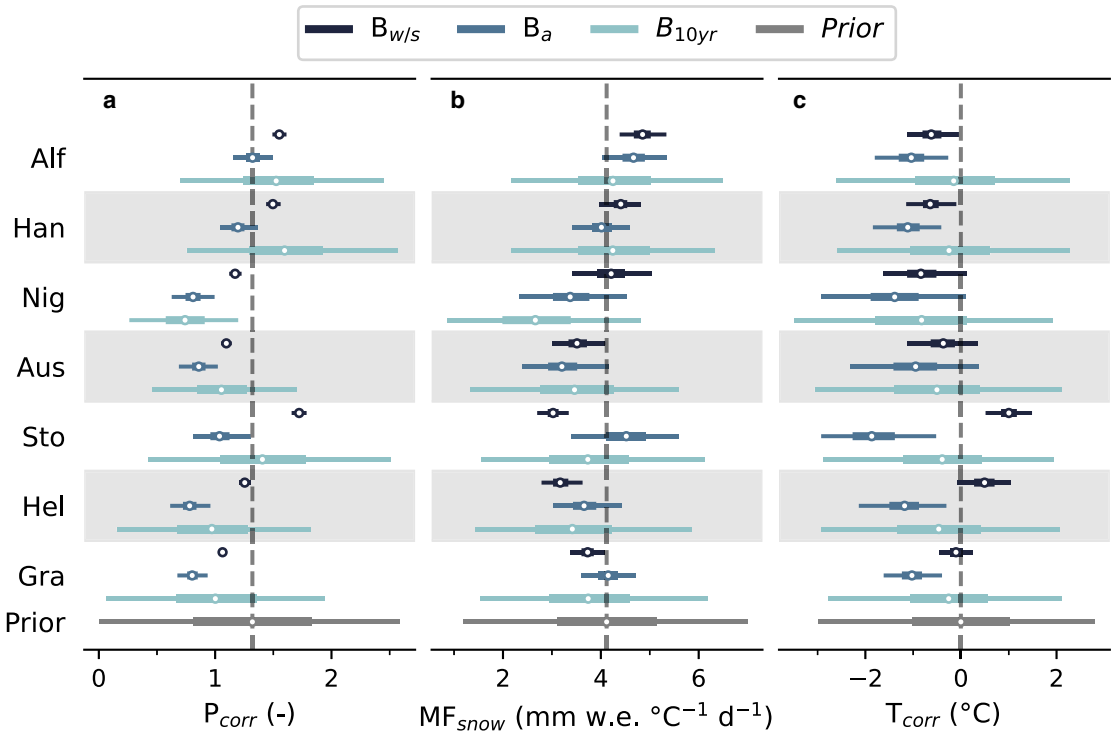


Figure 4. Marginal posterior probability distributions for P_{corr} (a), MF_{snow} (b) and T_{corr} (c) for all glaciers and experiments $B_{w/s}$, B_a, and B_{10yr}. Marginal prior distribution for each parameter is shown at the bottom in grey. Markers indicate median of posterior distributions, while bold and thin lines indicate interquartile range and credible interval (in terms of the 95% highest posterior density interval), respectively. Grey dashed lines indicate prior medians. Glaciers are sorted from west to east along a maritime to continental climate gradient.

comparing probability density functions (PDFs) of modelled (posterior predictive) and observed mass balances over the period of respective observations for each glacier ($B_{w/s}$; Fig. 6, B_a; Fig. A1, and B_{10yr}; Fig. A2). To complement (1), we also evaluated the median absolute deviation and median bias between modelled and observed mass balances (Fig. A3), which gives similar results.

When considering the ensemble of glaciers, the three experiments display similar performance in modelling annual SMB over the calibration and validation years in terms of RMSE and

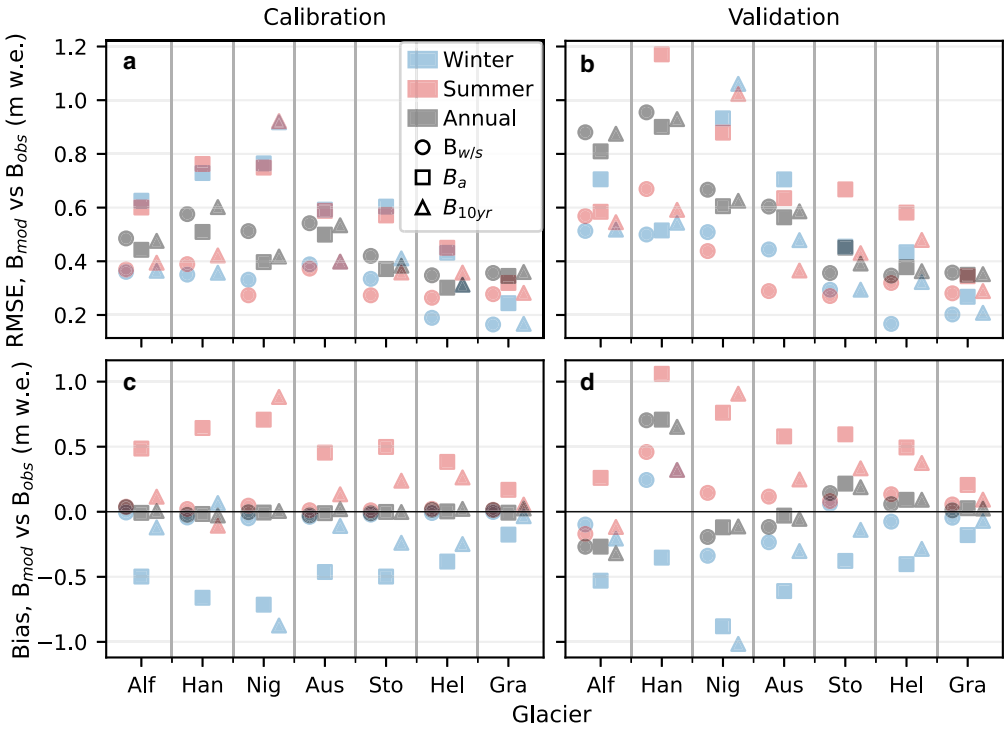


Figure 5. Model performance for each glacier in terms of RMSE (a, b) and mean bias (c, d) between modelled (posterior predictive median) and observed seasonal (winter in blue, summer in red) and annual (in grey) surface mass balance over calibration (a, c) and validation (b, d) periods for parameter estimation experiments $B_{w/s}$ (circle), B_a (triangle) and B_{10yr} (square). Glaciers are sorted from west to east along a maritime to continental climate gradient.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

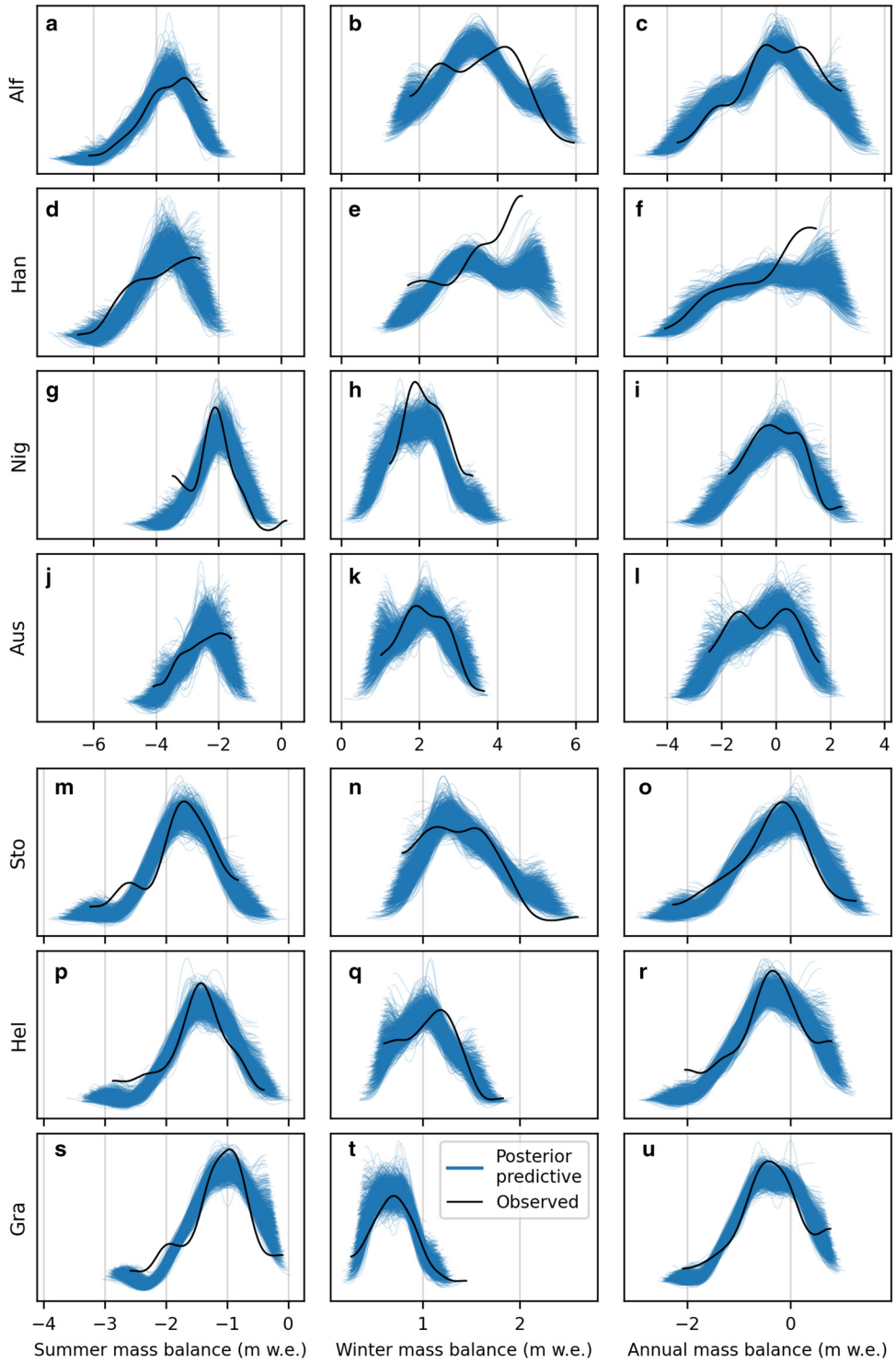


Figure 6. Probability density functions for mass balances in the $B_{w/s}$ experiment for 1000 posterior predictive samples (blue lines) and observations (black lines). Summer, winter and annual mass balance (m w.e.) are shown from left to right for each glacier. Glaciers are sorted from west to east along a maritime to continental climate gradient.

bias between modelled and observed SMB (Fig. 5). Posterior distributions for $B_{w/s}$ give the lowest RMSE and bias for seasonal SMB. Overall, B_a and B_{10yr} show a bias in the magnitudes of modelled seasonal balances, with higher (less negative) summer balances and lower (less positive) winter balances than observed. The underestimation of the magnitudes of seasonal SMB for B_a and B_{10yr} is in line with parameter values of posterior distributions that promote lower ablation and accumulation compared to $B_{w/s}$. B_a shows a larger bias in seasonal SMB compared to B_{10yr}, and higher RMSE for most glaciers. B_a appears to systematically favour mass-balance scenarios characterised by low ablation and low accumulation, in line with the lower parameter values of the posteriors in this experiment (Fig. 4).

RMSE is typically lower for the calibration period than the validation period (Fig. 5), which is not unexpected given that posteriors are informed by observations in the calibration period. Overall, the lowest RMSE is found for the three most continental glaciers (Storbreen, Hellstugubreen and Gråsubreen), and the largest deterioration in RMSE from the calibration to the validation period is found for the most maritime glaciers (Ålfotbreen and

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

Hansebreen). For Ålfotbreen, observed seasonal and annual SMB is generally within the range of predicted values for the calibration years, while the correspondence is poor for some years of the validation period (e.g. too low winter balance in 1960s and 70 s, too negative summer balance in the 1970s and too high winter balance in the 2010s; Fig. A4).

The PDFs of posterior predictive samples of modelled mass balances (Figs. 6, A1 and A2) describe the probability distribution of mass balances over the observational period. For a well-performing and robust model, the bulk of PDFs of the posterior predictive samples should reflect the PDF of the observations. For $B_{w/s}$, posterior predictive PDFs encompass the PDF of observations to a degree for most glaciers and observations, although discrepancies are visible in all cases (Fig. 6). For some ranges of mass balance, PDFs of posterior predictive samples do not encompass the PDF of observations. This implies that the model under- or overestimates SMB such that the observed SMB is not within the distribution of modelled SMB (e.g. as for Ålfotbreen for many years in the 1960s and 70s; Fig. A4). For instance, winter balance of Nigardsbreen shows predicted values around 0.5–1 and 3.5–4 m w.e., while the PDF of observations is constrained between 1–3.5 m w.e. and shows a higher density than posterior predictive PDFs in this range (Fig. 6h). For Hansebreen, correspondence between posterior predictive and observed PDFs is particularly poor, e.g. posterior predictive samples show a high frequency of large winter balances (>5 m w.e.), while the frequency of winter balances in the range 3.5–4.5 m w.e. is lower than observed (Fig. 6e). The consequence is that annual SMB (Fig. 6f) is underrepresented in the range 0-1.5 m w.e. and overrepresented in the range 1.5–3 m w.e., which indicates that annual SMB for Hansebreen is overestimated in several years. Generally, $B_{w/s}$ shows better correspondence for Austdalsbreen and the continental glaciers (Figs. 6j-l) than for Ålfotbreen, Hansebreen and Nigardsbreen (Figs. 6a–i).

B_a (Fig. A1) yields similar results as $B_{w/s}$ in terms of modelled annual SMB, but with slightly higher spread such that PDFs of observed annual SMB are to a larger degree encompassed by posterior predictive samples (e.g. Ålfotbreen and Hansebreen). However, PDFs of posterior predictive samples for winter and summer balances are shifted towards low magnitudes and do not show correspondence with observations. This is in line with the low mass-turnover parameter values for B_a and systematic bias in modelled seasonal balances (Fig. 5). PDFs of posterior predictive samples of B_{10yr} (Fig. A2) show an extremely large spread in comparison to $B_{w/s}$ and B_a and quite high densities towards low values of seasonal balances.

We assess the performance improvement achieved through posterior inference in the three experiments by evaluating the difference between RMSE of the posterior predictive median (predictions using joint posterior distributions) and RMSE of the prior predictive median (predictions using joint prior distributions; Fig. 7). A positive performance improvement indicates that the posterior distribution improves modelled SMB compared to using the prior distribution and reflects the added benefit of using the observational data in parameter estimation as opposed to no data at all. Over the calibration period, all experiments show a performance improvement (reduction in RMSE) in terms of modelling annual SMB across all glaciers (Figs. 7a–c). The same is true for the validation period (Figs. 7d–f), except for Hansebreen. $B_{w/s}$ is the only experiment where posterior inference consistently improves modelling of seasonal SMB over the calibration period, and largely also over the validation period. Although both B_a and B_{10yr} give an overall performance improvement in modelling annual SMB, in many instances, the performance in modelling underlying seasonal balances has worsened compared to using the prior distribution.

5.3. Modelled surface mass balance, 1960–2020

The modelled SMB indicates a net mass loss for all glaciers over the period 1960–2020 (Figs. 8 and 9). Nigardsbreen is the only glacier that is in near balance. Modelled (in terms of posterior predictive median) average annual SMB is similar across

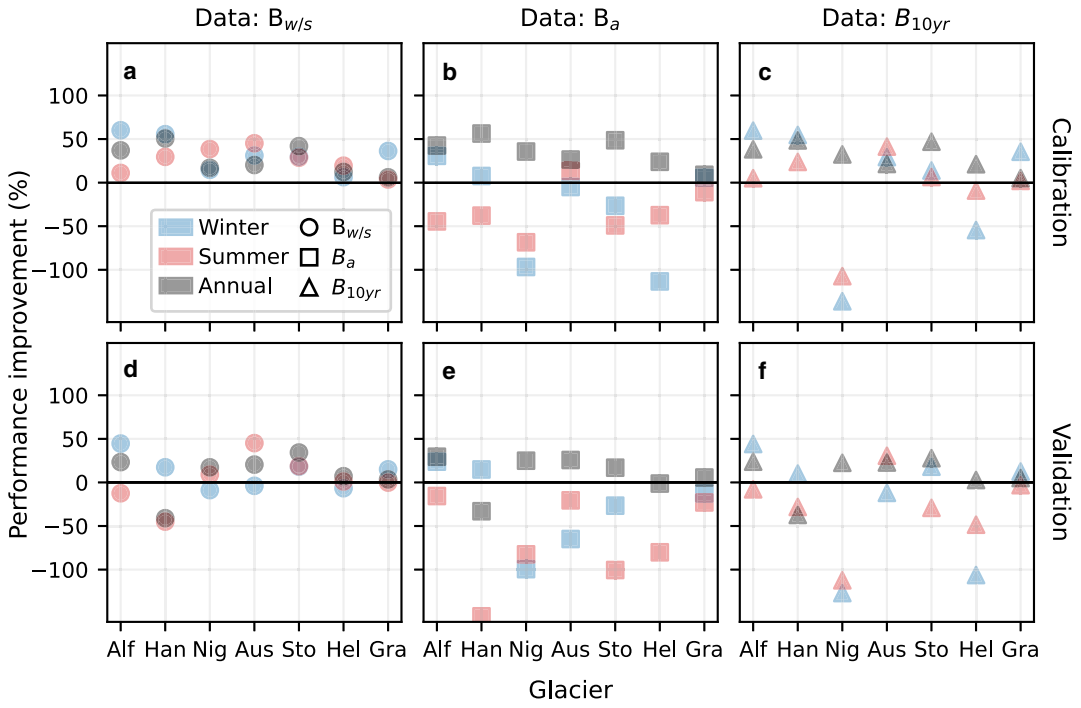


Figure 7. Performance improvement in terms of percentage reduction in RMSE of posterior predictive median compared to RMSE of prior predictive median for each glacier and experiments $B_{w/s}$ (a, d), B_a (b, e) and B_{10yr} (c, f) in calibration (a–c) and validation (d–f) periods. Positive performance improvement (reduction in RMSE, up to 100%) reflects improvement in model performance when using posterior compared to prior distributions to model mass balances. Negative performance improvement (increase in RMSE, up to −∞%) reflects deterioration in model performance.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

Glacier	Winter (m w.e. a^{-1})			Summer (m w.e. a^{-1})			Annual (m w.e. a^{-1})		
	$B_{w/s}$	B_a	B_{10yr}	$B_{w/s}$	B_a	B_{10yr}	$B_{w/s}$	B_a	B_{10yr}
Alf	3.50	3.05	3.39	-3.81	-3.37	-3.75	-0.31	-0.31	-0.35
Han	3.20	2.64	3.29	-3.74	-3.13	-3.88	-0.54	-0.49	-0.58
Nig	1.95	1.38	1.23	-2.05	-1.41	-1.26	-0.10	-0.04	-0.03
Aus	1.65	1.33	1.59	-2.40	-1.94	-2.27	-0.75	-0.62	-0.68
Sto	1.44	0.99	1.24	-1.75	-1.24	-1.50	-0.30	-0.25	-0.27
Hel	1.03	0.69	0.81	-1.43	-1.07	-1.19	-0.40	-0.38	-0.38
Gra	0.69	0.55	0.67	-1.12	-0.97	-1.08	-0.43	-0.42	-0.42

Figure 8. Average rates of modelled (posterior predictive median) winter, summer and annual mass balance over the period 1960–2020 for each glacier and parameter estimation experiment $B_{w/s}$, B_a and B_{10yr}. Glaciers are sorted from west to east along a maritime to continental climate gradient.

experiments (Fig. 8). B_a and B_{10yr} generally give lower magnitudes of summer and winter balances than $B_{w/s}$, with B_a showing the lowest magnitudes overall. All experiments illustrate a large difference in mass turnover from the maritime to continental glaciers.

The three experiments give similar results in terms of modelled cumulative SMB, but with different margins of uncertainty (Fig. 9). Uncertainty in modelled SMB (credible region in Fig. 9, which is given in terms of the 90% highest posterior density

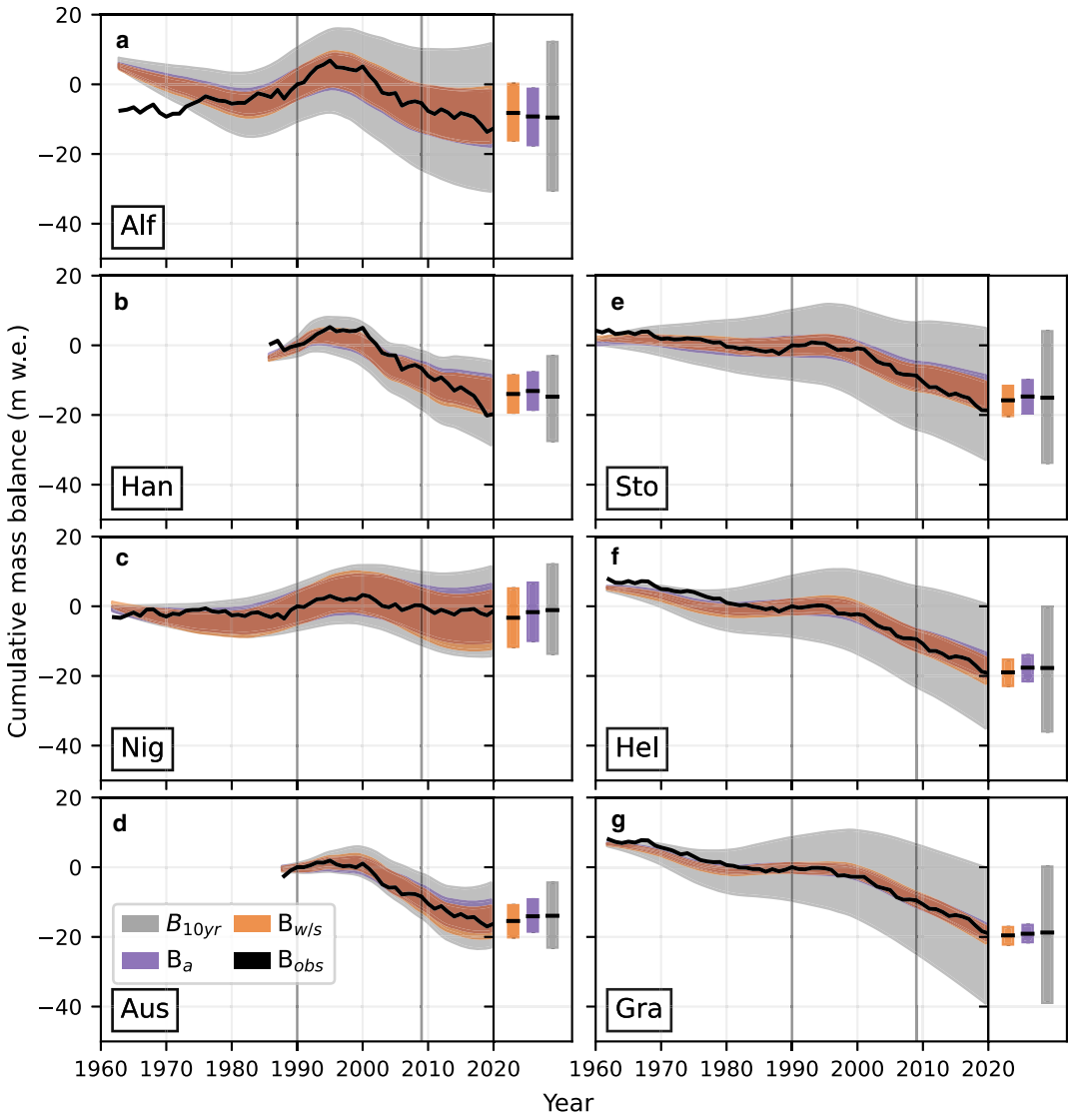


Figure 9. Cumulative annual mass balance for each glacier and experiments $B_{w/s}$, B_a and B_{10yr} over their respective period of glaciological mass-balance observations. Shaded areas represent credible intervals in terms of the 90% highest posterior density interval. Black lines indicate medians. Zero is set to the first year of the calibration period (1990). Grey vertical lines indicate the start (1990) and end (2009) of the calibration period.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

interval; the shortest interval that spans 90% of the density of the distribution) from propagation of parameter uncertainty for $B_{w/s}$ and B_a is comparable. Considering the glaciers with observation records beginning in the 1960s, uncertainty in modelled SMB for $B_{w/s}$ and B_a is generally higher for Ålfotbreen and Nigardsbreen, compared to Storbreen, Hellstugubreen and Gråsubreen. Across glaciers, B_{10yr} gives the highest uncertainty in modelled SMB. The largest difference between experiments is found for the small continental glacier Gråsubreen, where the 90% credible interval of cumulative annual SMB is an order of magnitude larger for B_{10yr} compared to $B_{w/s}$ and B_a.

Modelled and observed annual SMB correspond well over the period 1960s–2020 for Storbreen, Hellstugubreen and Gråsubreen, and over the shorter records (1980s–2020) of Hansebreen and Austdalsbreen (Fig. 10). The agreement between modelled and observed annual SMB is more variable over time for Ålfotbreen and Nigardsbreen, e.g. annual SMB is generally underestimated (more negative) for Ålfotbreen from the early 1960s to the late 1970s but shows better correspondence in the subsequent decades (Figs. 9a, 10a and A4). B_a shows a more or less consistent underestimation of the magnitude of summer balance over the period of available observations (Fig. 10), in line with the systematic bias found for modelled seasonal balances (Figs. 5c and d). A similar underestimation is visible for B_{10yr} for some glaciers (Nigardsbreen, Storbreen, Hellstugubreen and the shorter record of Austdalsbreen). For Storbreen, Hellstugubreen and Austdalsbreen, the magnitudes of seasonal balances for B_{10yr} are in general higher than for B_a (Fig. 8), such that the underestimation is less severe. However, for Nigardsbreen, B_{10yr} displays a greater underestimation of modelled ablation than B_a, which is also reflected in the bias between modelled and observed seasonal balances over the calibration and validation periods (Figs. 5c and d). For both B_a and B_{10yr}, underestimation of the magnitude of summer balance is counteracted by an underestimation of the magnitude of winter balance, such that the resulting modelled annual SMB is comparable across experiments (Figs. 8 and 5c and d).

6. Discussion

6.1. Prior and likelihood influence on posterior

Observations influence the posterior distribution through the likelihood function. We highlight three aspects that distinguish the datasets used in each experiment and that influence the effect of the likelihood (observations) on the posterior distribution: (1) the number of observations, (2) the mass-balance information and (3) the observational uncertainty. For $B_{w/s}$ and B_a, posterior distributions are strongly constrained compared to B_{10yr}, which indicates that posteriors are more influenced by the likelihood

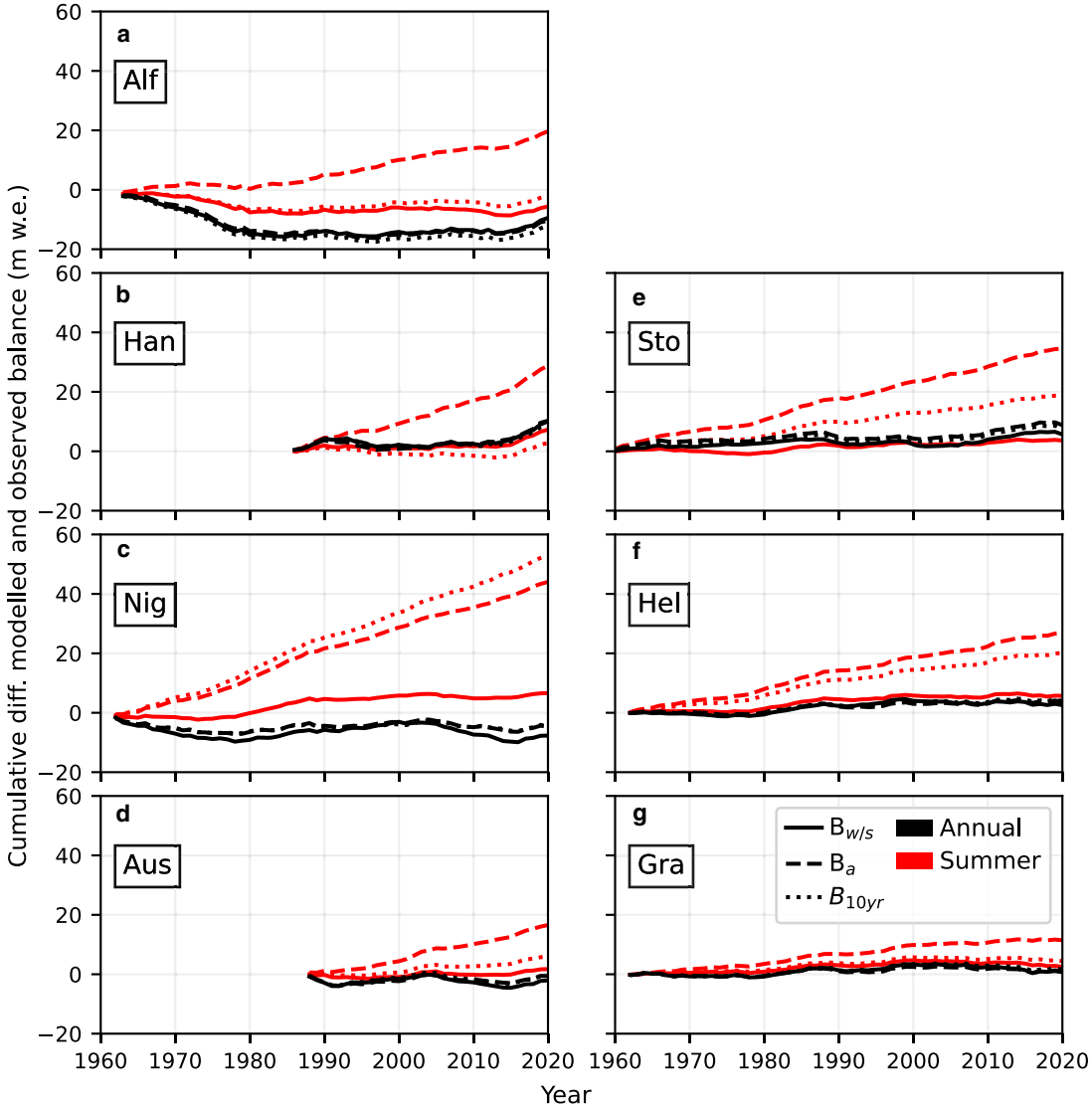


Figure 10. Cumulative difference between median of modelled (posterior predictive) and observed summer (red) and annual (black) mass balance for each glacier for experiments $B_{w/s}$ (solid lines), B_a (dashed lines) and B_{10yr} (dotted lines).

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

(observations) than in the B_{10yr} experiment. $B_{w/s}$ and B_a show lower posterior uncertainty compared to B_{10yr} due to the combination of the three factors: (1) a larger number of observations which increases the relative importance of the likelihood in posterior inference (Gelman and others, 2014), (2) lower equifinality in the mass-balance information and (3) lower observational uncertainty. In all experiments, accounting for observational uncertainty aggravates equifinality because we are searching for parameter values that explain a wider range of probable observations. Posterior distributions for B_{10yr} show high uncertainty, not only as a result of the limited number of observations but also due to the combination of significant equifinality in the mass-balance information and high observational uncertainty.

Due to the combination of the three mentioned factors, B_{10yr} posteriors show a wide range of probable parameter values (Fig. 4). Although posterior distributions for $B_{w/s}$ and B_a are more constrained than B_{10yr}, equifinality is still reflected in the correlation between parameter values (Figs. A5 and A6) and in the difference in the spread of distributions between parameters. Overall, posteriors of P_{corr} are more constrained than posteriors of T_{corr} and MF_{snow}. P_{corr} alone largely controls accumulation of snow over the glaciers since winter temperatures are mostly below the snow/rain threshold (Fig. 2). Ablation, however, is controlled by both T_{corr} and MF_{snow}, such that there is a strong correlation between these two parameters (Figs. A5 and A6) and thus higher uncertainty in posteriors. For $B_{w/s}$, distributions of P_{corr} are especially narrow because the posterior is informed by observations of winter balance which are assigned lower uncertainties than summer balance observations.

Since posteriors of $B_{w/s}$ and B_a seem to be highly constrained by the likelihood (observations) in our formulation of the Bayesian model, we would expect that our choice of prior has little influence in posterior inference in these experiments. We investigate posterior sensitivity to the choice of prior distribution for three glaciers across the maritime-continental climate gradient (Ålfotbreen, Nigardsbreen and Gråsubreen) by performing inference with low and high priors for parameters P_{corr} and MF_{snow} in each experiment. The sensitivity experiment illustrates the influence on the posterior for a low or high mass-turnover prior. The low and high priors were constructed from the original priors, but with means shifted up or down by one standard deviation of the original priors, e.g. such that the high (low) prior for MF_{snow} was a truncated normal distribution with a mean of 5.6 (2.6) mm w.e.$^{\circ}C^{-1}d^{-1}$.

For $B_{w/s}$ and B_a, posteriors display similar properties regardless of the choice of prior (e.g. for Nigardsbreen in Fig. 11) which implies that the posterior distribution is highly influenced by the likelihood (observations). Our choice of weakly informative priors and the assigned low observational uncertainties for $B_{w/s}$ and B_a produce posterior distributions that approach standard maximum likelihood estimates and thus reflect a narrow range of values that minimise the difference between observed and modelled SMB (Eqn (13)). For B_a, the model seemingly achieves a better fit between observed and modelled annual SMB with smaller magnitudes of accumulation and ablation. This may be due to limitations in the model, e.g. simplified process representation, lack of resolution and changes in hypsometry and incremental updates of glacier areas. Since posteriors for B_a are largely driven by the likelihood, we could expect that the bias of B_a towards low mass-turnover parameter values would be similar in standard maximum likelihood estimation.

In contrast to $B_{w/s}$ and B_a, prior influence on the posterior for B_{10yr} is pronounced, with marginal posteriors displaying a shift towards low or high values depending on the choice of prior. B_{10yr} could essentially have the same effect of favouring low mass-turnover parameter values as B_a. This indication is supported by the results for Nigardsbreen, where posteriors for B_{10yr} and B_a adopt values in the low range of the prior, in contrast to the posterior for $B_{w/s}$. The low accumulation, low ablation parameter values for Nigardsbreen achieve a significantly lower RMSE for annual SMB than in the $B_{w/s}$ experiment (Figs. 5a and b), but give a large bias in seasonal balances (Figs. 5c and d, 10c and A1). However, the tendency to favor low mass-turnover scenarios is generally less pronounced for B_{10yr} due to the comparatively stronger prior influence and weaker influence of the likelihood (observations). The high parameter uncertainty and sensitivity to choice of prior distribution in the B_{10yr} experiment implies that the number of unknown parameters is at the upper limit of what can be reliably estimated using decadal resolution satellite-borne geodetic observations with their characteristic uncertainty. Increasing the number of unknown parameters could aggravate equifinality such that the inferred posterior distributions would largely reflect the priors.

6.2. Parameter values

We compare parameter values for $B_{w/s}$ across glaciers as this experiment most correctly represents accumulation and ablation.

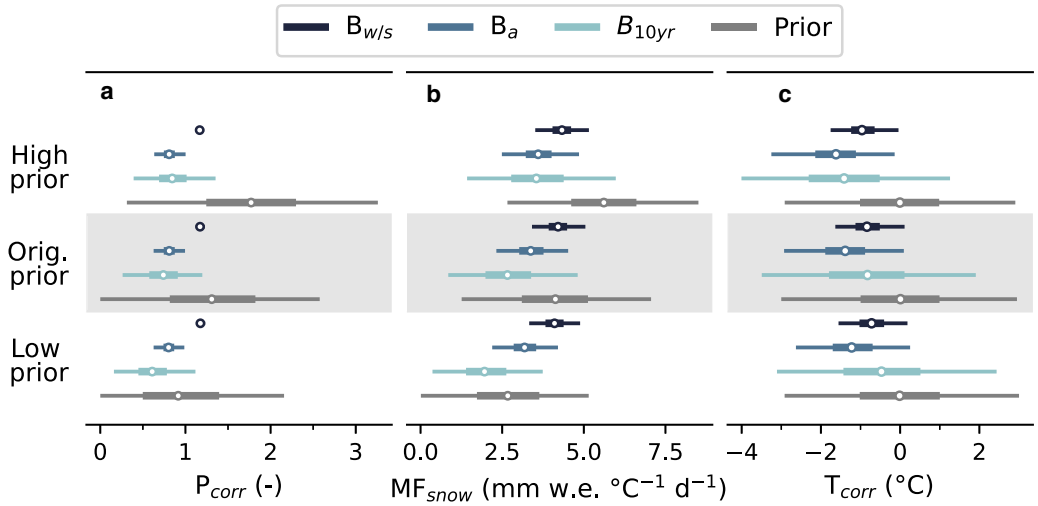


Figure 11. Marginal posterior probability distributions for P_{corr} (a), MF_{snow} (b) and T_{corr} (c) for Nigardsbreen for experiments $B_{w/s}$, B_a and B_{10yr} using different priors (high, original and low). Prior distributions for each parameter are shown in grey. Markers indicate medians of posterior distributions, while bold and thin lines indicate interquartile range and credible interval (in terms of the 95% highest posterior density interval), respectively.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

For all glaciers, values of P_{corr} for $B_{w/s}$ (medians range from 1.06 to 1.55, see Tables S1–S3) indicate that seNorge_2018 may underestimate winter precipitation. To our knowledge, no formal evaluation of seNorge_2018 in glacio-hydrological modelling in remote data-void regions has been performed. However, considering previous evaluation of winter precipitation in seNorge1 (Engelhardt and others, 2012), our results are in line with lower precipitations sums in seNorge_2018 compared to seNorge1 (Fig. 2). P_{corr} shows a trend from higher to lower values from coastal (Ålfotbreen and Hansebreen) to more inland glaciers (Nigardsbreen and Austdalsbreen) across experiments. This indicates that seNorge_2018 may not adequately capture the strong longitudinal precipitation gradient in southern Norway. A similar pattern in values of P_{corr} from west to east over the continental glaciers also suggests that orographic effects over the central mountain ranges may not be accurately represented. Although parameter values suggest underestimation of precipitation in seNorge_2018, strong conclusions regarding the magnitude of underestimation should not be drawn based on these values. Parameter values are influenced by the model limitations discussed earlier, as well as potential biases in observations and choice of calibration period. Other processes that may influence accumulation, such as redistribution of snow by wind or avalanching, are also not included in the model.

Values of T_{corr} for $B_{w/s}$ indicate a slight overestimation of temperature in seNorge_2018 for maritime glaciers and an underestimation for the westernmost continental glaciers. However, in addition to the limitations in model structure and potential influences on parameter values mentioned earlier, general statements about the quality of seNorge_2018 temperature are even more precarious due to compensating effects between T_{corr} and MF_{snow} on modelled ablation. Considering MF_{snow}, maritime glaciers generally exhibit higher values than continental glaciers, in agreement with the findings of Zhang and others (2006) for degree-day factors of maritime and continental glaciers in western China. Engelhardt and others (2014) found a similar pattern in parameters controlling melt across Åfotbreen, Nigardsbreen and Storbreen, but parameter values are not directly comparable to this study due to differences in the model structure. For Storbreen and Hellstugubreen, which have lowest values of MF_{snow}, we would expect higher values of MF_{snow} without the compensating effect of T_{corr}.

6.3. Effect of observations on modelled mass balance

The magnitudes of seasonal balances are generally underestimated (too low) for B_a (Figs. 5, 10 and A1), due to parameter values promoting lower accumulation and ablation (Fig. 4). Underestimation of the magnitude of ablation implies that also summer meltwater runoff would be underestimated for B_a. B_{10yr} also shows underestimation of magnitudes of summer and winter balances, which is in line with findings of Rounce and others (2020a) for glaciers in High Mountain Asia with parameters calibrated with 20-year geodetic mass-balance observations. $B_{w/s}$ is the only experiment that adequately resolves seasonal balances (Figs. 5 and 6) and consistently improves modelled seasonal balances (Fig. 7) because posteriors are informed by winter and summer balance observations. The overall better performance in terms of median values of B_{10yr} in modelling seasonal SMB compared to B_a (Figs. 5 and 10) could be attributed to a choice of priors that produce posteriors centered around the same values as $B_{w/s}$, and thus give decent estimates of accumulation and ablation.

B_a and B_{10yr} (which is essentially an aggregation of annual SMB observations) give decent mass-balance estimates (Figs. 5, A1 and A2) and improvement of annual SMB (Fig. 7) since posteriors are informed by annual SMB observations. However, this does not necessarily improve performance in modelling seasonal balances since parameter values that give the best estimates of annual SMB do not necessarily reproduce seasonal variations. Nevertheless, due to the equifinality in the mass-balance model, B_a and B_{10yr} give estimates of annual SMB that are in line with those achieved by resolving seasonal balances in the $B_{w/s}$ experiment.

Overall, all experiments improve estimates of annual SMB compared to employing parameter values based on prior knowledge (Fig. 7). A performance improvement does not necessarily imply a good model performance, but rather reflects the added value of using the observations in calibration for a given prior. Across experiments and glaciers, deterioration in RMSE reflects that a posterior centred around the median of the prior distribution would give lower RMSE than the inferred posterior. The deterioration in RMSE in seasonal and annual SMB over the validation period for some glaciers (e.g. Hansebreen; Figs. 7d–f) reflects that posterior parameter values are less representative over this period than the prior values. B_{10yr} posteriors are clearly sensitive to the choice of prior (Fig. 11) but improve modelling of annual SMB for most glaciers and priors we evaluated (Fig. 12). Although B_{10yr} represents a mass-balance signal of low temporal resolution and high uncertainty, posterior distributions informed by such observations yield improved estimation of annual SMB compared to employing parameter values based on prior knowledge alone. In the low and original prior cases for Ålfotbreen and Gråsubreen, B_{10yr} also improves modelled seasonal SMB because posterior parameter values also produce better representations of accumulation and ablation.

Although B_{10yr} gives decent estimates of annual SMB in terms of median values, the modelled uncertainty is significantly higher

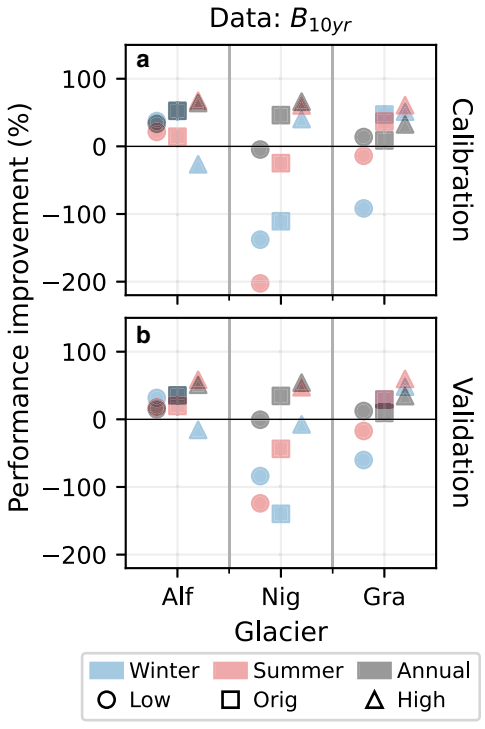


Figure 12. Performance improvement for B_{10yr} in terms of percentage reduction in RMSE of posterior predictive mean compared to RMSE of prior predictive mean for calibration (a) and validation (b) periods for Ålfotbreen (Alf), Nigardsbreen (Nig) and Gråsubreen (Gra) using high, original and low priors. Positive performance improvement (reduction in RMSE, up to 100%) reflects improvement in model performance when using posterior compared to prior distributions to model mass balances. Negative performance improvement (increase in RMSE, up to −∞%) reflects deterioration in model performance.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

than for $B_{w/s}$ and B_a (Fig. 9). The high parameter uncertainty in the B_{10yr} experiment leads to a considerably higher uncertainty (standard deviation) in mass-balance sensitivity compared to the other experiments (by a factor of 1.2–3.1 compared to $B_{w/s}$, see Supplementary Material for details). This uncertainty reflects the parameter uncertainty, which is again determined by the influence of the likelihood (given our prior formulations). High temporal resolution, low uncertainty observations ($B_{w/s}$ and B_a) correspond to low uncertainties in modelled SMB, while low temporal resolution, high uncertainty observations (B_{10yr}) yield high modelled uncertainties. For $B_{w/s}$ and B_a, uncertainty in modelled SMB is higher for Ålfotbreen and Nigardsbreen compared to continental glaciers. This reflects that glaciological observations of high mass-turnover glaciers are generally associated with higher uncertainty (Table 2). For B_{10yr}, the difference in uncertainty in modelled SMB between glaciers generally corresponds to the difference in magnitudes of observation uncertainties (Table 2 and Fig. 9). Gråsubreen, which shows the largest difference in modelled uncertainties between $B_{w/s}$/B_a and B_{10yr}, has particularly low uncertainty in glaciological observations and large uncertainty in geodetic observations. For satellite-borne geodetic observations, uncertainty is in general higher for smaller (e.g. Gråsubreen) compared to larger (e.g. Nigardsbreen) glaciers due to the combination of limited spatial coverage of elevation change estimates, high uncertainty in the temporal evolution of the glacier area, and uncertainty in the volume-to-mass density conversion (Hugonnet and others, 2021).

Modelled mass balances for Hansebreen and Austdalsbreen are validated on fewer observations than for the other glaciers. The difference in bias over the validation period between the similar (in terms of hypsometry and area) neighboring glaciers Ålfotbreen and Hansebreen (Fig. 5d) could be partly explained by their validation periods. For these two glaciers, the late 1980s to mid-1990s was a period of predominant mass gain, followed by two decades of predominantly large mass losses from 2001 (Fig. 3). While the validation period for Ålfotbreen covers years of moderate positive and negative balances from 1963 to 1990, Hansebreen is mainly validated on a shorter period of more extreme mass-balance years (end of 1980s and 2010s). This could also explain the particularly poor correspondence between modelled and observed mass-balance distributions for Hansebreen (Figs. 6d–f) and deterioration of RMSE in the validation period (Fig. 7).

6.4. Parameter and model robustness

An interesting result of posterior inference with our model is that posterior distributions of B_a and $B_{w/s}$ are not equifinal solutions of each other. We would expect high probability regions of posteriors of $B_{w/s}$ and B_a to show overlap as both posteriors produce decent estimates of annual SMB (e.g. Figs. 5 and 10). We believe that the discrepancy is due to a combination of two factors: (1) limitations in the model structure (e.g. lack of complexity, processes representation and limited spatial resolution) and the number of parameters estimated, such that simultaneous reproduction of the most accurate representation of annual and inter-annual variations is not possible; and (2) that the assigned uncertainties in $B_{w/s}$ and B_a are too low, such that the influence of the likelihood in our Bayesian model is too strong.

We only estimate three parameters due to the low temporal resolution and high uncertainty in B_{10yr}, which limits the variability in SMB that can be achieved with our simple model. Although we find decent results in terms of median values, PDFs of $B_{w/s}$ and B_a suggest that observed values are in many cases outside the spread of predictions, which indicates that these models are not particularly robust. This suggests that the spread in the error term (Eqn (9)) to which we assign the observation uncertainty may be too low to produce robust predictions for $B_{w/s}$ and B_a given our mass-balance model. Options to increase robustness of $B_{w/s}$ and B_a could be to explore the use of larger variance than the reported observation uncertaintyor whether other distributions than the normal distribution are more appropriate for the likelihood or prior. We investigated the use of Student's t distributions, which are more heavy tailed than the normal distribution, for the likelihood in the $B_{w/s}$ and B_a experiments for some glaciers (see Supplementary Material for details). We found that using Student's t distributions for the likelihood gives posterior distributions centred around the same values, but with larger spread, thus giving a broader range of posterior predictive samples. However, we still found the same bias in B_a towards low mass-turnover parameter values in the tests we performed.

$B_{w/s}$ yields a decent estimation of seasonal and annual SMB, but the overall reduction in performance from calibration to validation period (Fig. 5) reflects a lack of robustness of parameter values over time. The deterioration in performance is especially visible for maritime glaciers that have high mass-balance gradients (e.g. for Ålfotbreen; Fig. A4). This highlights a general issue of temperature index models concerning the representativeness of constant parameter values, e.g. those of melt factors, over longer time scales where meteorological conditions may vary. Melt factors implicitly account for the relative contribution of the various energy-balance components and therefore vary in both space (e.g. aspect, elevation) and time (e.g. change in albedo with snow/ice surface properties; Hock, 2003, 2005). Calibrated melt factors may be representative of climatic conditions specific for the calibration period, but not necessarily beyond (Gabbi and others, 2014). Although we attempt to reduce bias in parameter values to specific climatic conditions by employing a calibration period that encompasses years of mass-balance variability, results nevertheless indicate a lack of representativeness of parameter values in certain time periods. The underestimation of SMB in the 1960s and 1970 s is not only prominent for Ålfotbreen, but also visible for the other glaciers with records dating back to the 1960s (Figs. 9 and 10). This bias could also be due to unresolved geometry changes in the model (constant hypsometry and incremental area updates), but we would not expect this to be the main influence as all glaciers have outlines from the 1960s (Table 1) and surface elevation changes over the glaciers are limited (−2.2 m to −15.8 m for Nigardsbreen and Hellstugubreen, respectively, from the 1960s to the 2010s Andreassen and others (2020)). Alternatively, there could be biases in the climate dataset specific to certain time periods, e.g. due to varying quality of meteorological observations used in the reanalysis, that are not captured by the constant temperature and precipitation corrections. Still, we argue that capturing parameter uncertainty from uncertainty in observations is an indirect way to, at least partly, reflect the variability of parameter values over time. Employing more robust formulations of the Bayesian model could make the model predictions more robust to these variations, at a cost of increased uncertainty in modelled mass balances.

Our results show that parameter estimation employing annual observations with low uncertainty or decadal observations with high uncertainty could produce decent estimates of annual SMB, and we would thus expect such observations to be useful in constraining net glacier mass changes, e.g. in determining future sea-level rise. However, the underlying seasonal mass changes may not be resolved using these observations. To improve estimates of accumulation and ablation, annual and decadal mass-balance observations could be used in combination with observations that constrain accumulation and/or ablation, e.g by using independent data on snow distribution (e.g. Geck and others,

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

2021). However, such observations are generally lacking, particularly at the scale of individual glaciers. High-quality, high-resolution climate data in complex terrain, or widespread, higher resolution mass-balance observations are needed to properly resolve interannual and interseasonal variations in mass balance.

7. Conclusions

We modelled SMB of seven glaciers in southern Norway over the period 1960–2020 using a temperature index model. The glaciers are located along a maritime-to-continental climate gradient and have long-term glaciological SMB records. To explore how the use of distinct mass-balance observations affects model calibration, we applied a Bayesian framework to estimate three sets of parameters for each glacier based on three representations of the respective glaciological records: seasonal, annual and 10-year average SMB. For seasonal and annual mass balances, we employed associated uncertainties from the glaciological records. Ten-year average SMB represented our analogue to geodetic mass balance based on satellite remote sensing and was thus assigned uncertainties characteristic to such observations.

Posterior parameter distributions are strongly dependent on the observations used in parameter estimation and, in the experiment using 10-year average SMB, also on the choice of prior distribution. The marginal posterior distributions of the precipitation correction factor indicates that the seNorge_2018 climate dataset may underestimate winter precipitation over the westernmost glaciers.

All datasets demonstrated decent performance in modelling annual SMB and generally improved estimates of annual SMB compared to using parameter values based on prior knowledge. However, using 10-year average SMB displayed higher uncertainty in modelled mass balance and in mass-balance sensitivity to changes in climate forcing due to considerable parameter uncertainty.

Reliable estimates of the magnitude of accumulation and ablation were only achieved using seasonal balances to constrain parameters. However, model performance varied over time due to the application of constant parameter values and limited robustness of the Bayesian model. Using annual mass-balance observations consistently favoured parameter values that promote low accumulation and ablation. Consequently, magnitudes of summer and winter balances were on average around 20% lower over the period 1960–2020 using annual compared to seasonal balances in parameter estimation.

Similar to the use of annual SMB, the use of 10-year average SMB also tended to underestimate magnitudes of accumulation and ablation but was restricted by the influence of the prior distribution. Nevertheless, our results demonstrate that high-uncertainty geodetic observations, which typically represent decadal snapshots of glacier mass change, can be used in parameter calibration to improve model results of annual mass balance. Since our representation of decadal mass-balance observations are constructed from glaciological records, the actual performance of satellite-borne geodetic observations will depend on the quality of these observations.

Propagation of parameter uncertainty revealed large differences in uncertainty in modelled SMB between the observational datasets, up to an order of magnitude for the small, continental glacier Gråsubreen. However, for the larger, more maritime glacier Nigardsbreen, uncertainty in modelled SMB was comparable between datasets due to higher mass-balance variability and similar observational uncertainty in glaciological and geodetic observations.

We underline that our findings are based on a limited sample of glaciers and are conditional on the specific model structure in this study. Generalisation to other model structures and glaciers should thus be made with caution. Nevertheless, our findings highlight that the purpose of modelling should be considered when choosing the type of mass-balance data to be used in model calibration. For reliable predictions of the timing and magnitude of runoff, e.g. in assessment of climate change impacts on freshwater availability, evaluating the performance of models on seasonal mass balances is important. Therefore, more extensive and robust seasonal mass-balance data are needed to both improve and evaluate model performance on seasonal scales.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/jog.2023.62.

Data. The source code of the model is available in the GitHub repository https://github.com/khsjursen/BI_glacier_mb_model. seNorge_2018 is available for download at https://thredds.met.no/thredds/catalog/senorge/seNorge_2018/catalog.html. Mass balance observations can be found at http://glacier.nve.no/glacier/viewer/ci/en/ and time series of glacier outlines used in this study are available in the model GitHub repository.

Acknowledgements. This study is a contribution to the JOSTICE project funded by the Norwegian Research Council (RCN grant #302458). We thank Hallgeir Elvehøy and Bjarne Kjøllmoen at NVE for providing mass-balance observations and time series of glacier outlines of Ålfotbreen, Hansebreen, Nigardsbreen and Austdalsbreen, and Markus Engelhardt for providing MATLAB code that the implementation of the mass-balance model is partly based upon. We would also like to recognise the Western Norway University of Applied Sciences (HVL) for providing computational resources and the managers of the HVL cluster for technical support. In addition, we thank Gregoire Guillet, David Rounce and Lilian Schuster for evaluating our manuscript and providing constructive feedback that has both improved the quality and clarity of the paper.

Author contribution. KHS developed the code for the mass-balance model and Bayesian parameter estimation routine, performed the MCMC simulations and initial analysis, and prepared the figures, tables and initial draft of the article. TD, AT, TVS and LMA read and edited the article, and provided input for additional analyses. LMA provided mass-balance observations and time series of glacier outlines for Storbreen, Hellstugubreen and Gråsubreen.

References

Andreassen LM, Elvehøy H, Kjøllmoen B and Belart JM (2020) Glacier change in Norway since the 1960s—an overview of mass balance, area, length and surface elevation changes. *Journal of Glaciology* **66**(256), 313–328. doi: 10.1017/jog.2020.10

Andreassen LM, Elvehøy H, Kjøllmoen B and Engeset RV (2016) Reanalysis of long-term series of glaciological and geodetic mass balance for 10 Norwegian glaciers. *The Cryosphere* **10**, 535–552. doi: 10.5194/tc-10-535-2016

Andreassen LM, Nagy T, Kjøllmoen B and Leigh JR (2022) An inventory of Norway's glaciers and ice-marginal lakes from 2018-19 Sentinel-2 data. *Journal of Glaciology* **68**(272), 1085–1106. doi: 10.1017/jog.2022.20

Andreassen LM and Oerlemans J (2009) Modelling long-term summer and winter balances and the climate sensitivity of Storbreen, Norway. *Geografiska Annaler: Series A, Physical Geography* **91**(4), 233–251. doi: 10.1111/j.1468-0459.2009.00366.x

Beven K (2006) A manifesto for the equifinality thesis. *Journal of Hydrology* **320**(1), 18–36. doi: 10.1016/j.jhydrol.2005.07.007

Braithwaite RJ (2008) Temperature and precipitation climate at the equilibrium-line altitude of glaciers expressed by the degree-day factor for melting snow. *Journal of Glaciology* **54**(186), 437–444. doi: 10.3189/002214308785836968

Brun F, Berthier E, Wagnon P, Kääb A and Treichler D (2017) A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016. *Nature Geoscience* **10**(9), 668–673. doi: 10.1038/ngeo2999

Compagno L and 7 others (2022) Modelling supraglacial debris-cover evolution from the single-glacier to the regional scale: an application to High Mountain Asia. *The Cryosphere* **16**(5), 1697–1718. doi: 10.5194/tc-16-1697-2022

Compagno L, Zekollari H, Huss M and Farinotti D (2021) Limited impact of climate forcing products on future glacier evolution in Scandinavia and Iceland. *Journal of Glaciology* **67**(264), 727–743. doi: 10.1017/jog.2021.24

Cuffey KM and Paterson WSB (2010) *The Physics of Glaciers*, 4th Ed. Burlington (MA): Elsevier.

Dyurgerov M and Meier M (1999) Analysis of winter and summer glacier mass balances. *Geografiska Annaler: Series A, Physical Geography* **81**(4), 541–554. doi: 10.1111/1468-0459.00082

Eckert N, Baya H, Thibert E and Vincent C (2011) Extracting the temporal signal from a winter and summer mass-balance series: application to a six-decade record at Glacier de Sarennes, French Alps. *Journal of Glaciology* **57**(201), 134–150. doi: 10.3189/002214311795306673

Eidhammer T and 9 others (2021) Mass balance and hydrological modeling of the Hardangerjøkulen ice cap in south-central Norway. *Hydrology and Earth System Sciences* **25**(8), 4275–4297. doi: 10.5194/hess-25-4275-2021

Engelhardt M, Schuler T and Andreassen LM (2012) Evaluation of gridded precipitation for Norway using glacier mass-balance measurements. *Geografiska Annaler: Series A, Physical Geography* **94**(4), 501–509. doi: 10.2307/23360734

Engelhardt M, Schuler TV and Andreassen LM (2013) Glacier mass balance of Norway 1961-2010 calculated by a temperature-index model. *Annals of Glaciology* **54**(63), 32–40. doi: 10.3189/2013AoG63A245

Engelhardt M, Schuler TV and Andreassen LM (2014) Contribution of snow and glacier melt to discharge for highly glacierised catchments in Norway. *Hydrology and Earth System Sciences* **18**(2), 511–523. doi: 10.5194/hess-18-511-2014

Engelhardt M, Schuler TV and Andreassen LM (2015) Sensitivities of glacier mass balance and runoff to climate perturbations in Norway. *Annals of Glaciology* **56**(70), 79–88. doi: 10.3189/2015AoG70A004

Gabbi J, Carenzo M, Pellicciotti F, Bauder A and Funk M (2014) A comparison of empirical and physically based glacier surface melt models for long-term simulations of glacier response. *Journal of Glaciology* **60**(224), 1140–1154. doi: 10.3189/2014JoG14J011

Geck J, Hock R, Loso MG, Ostman J and Dial R (2021) Modeling the impacts of climate change on mass balance and discharge of Eklutna Glacier, Alaska, 1985–2019. *Journal of Glaciology* **67**(265), 909–920. doi: 10.1017/jog.2021.41

Gelman A and 5 others (2014) *Bayesian Data Analysis*, 3rd Ed. Boca Raton (FL): Chapman and Hall/CRC.

Giesen RH and Oerlemans J (2010) Response of the ice cap Hardangerjøkulen in southern Norway to the 20th and 21st century climates. *The Cryosphere* **4**(2), 191–213. doi: 10.5194/tc-4-191-2010

Hock R (2003) Temperature index melt modelling in mountain areas. *Journal of Hydrology* **282**, 104–115. doi: 10.1016/S0022-1694(03)00257-9

Hock R (2005) Glacier melt: a review of processes and their modelling. *Progress in Physical Geography* **29**(3), 362–391. doi: 10.1191/0309133305pp453ra

Hugonnet R and 10 others (2021) Accelerated global glacier mass loss in the early twenty-first century. *Nature* **592**(7856), 726–731. doi: 10.1038/s41586-021-03436-z

Huss M and Hock R (2015) A new model for global glacier change and sea-level rise. *Frontiers in Earth Sciences* **3**(54), 1–22. doi: 10.3389/feart.2015.00054.

Huss M and Hock R (2018) Global-scale hydrological response to future glacier mass loss. *Nature Climate Change* **8**, 135–140. doi: 10.1038/s41558-017-0049-x

Kjøllmoen B (2022a) Reanalysing a glacier mass balance measurement series - Nigardsbreen 2014–2020. NVE Rapport no. 7/2022, Norwegian Water Resources and Energy Directorate.

Kjøllmoen B (2022b) Reanalysing a glacier mass balance measurement series - Ålfotbreen 2010–2019. NVE Rapport no. 6/2022, Norwegian Water Resources and Energy Directorate.

Kjøllmoen B, Andreassen LM, Elvehøy H and Storheil S (2022c) Glaciological investigations in Norway. NVE Rapport no. 27/2022, Norwegian Water Resources and Energy Directorate.

Landmann JM and 5 others (2021) Assimilating near-real-time mass balance stake readings into a model ensemble using a particle filter. *The Cryosphere* **15**(11), 5017–5040. doi: 10.5194/tc-15-5017-2021

Li H and 5 others (2015) Integrating a glacier retreat model into a hydrological model—case studies of three glacierised catchments in Norway and Himalayan region. *Journal of Hydrology* **527**, 656–667. doi: 10.1016/j.jhydrol.2015.05.017

Lussana C, Tveito O, Dobler A and Tunheim K (2019) seNorge_2018, daily precipitation, and temperature datasets over Norway. *Earth System Science Data* **11**(4), 1531–1551. doi: 10.5194/essd-11-1531-2019

Martín-Español A and 11 others (2016) Spatial and temporal Antarctic Ice Sheet mass trends, glacio-isostatic adjustment, and surface processes from a joint inversion of satellite altimeter, gravity, and GPS data. *Journal of Geophysical Research: Earth Surface* **121**(2), 182–200. doi: 10.1002/2015JF003550

Martín-Español A, Bamber JL and Zammit-Mangion A (2017) Constraining the mass balance of East Antarctica. *Geophysical Research Letters* **44**(9), 4168–4175. doi: 10.1002/2017GL072937

Maussion F and 14 others (2019) The open global glacier model (OGGM) v1.1. *Geoscientific Model Development* **12**(3), 909–931. doi: 10.5194/gmd-12-909-2019

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH and Teller E (1953) Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**(6), 1087–1092. doi: 10.1063/1.1699114ISSN 0021-9606.

Mohr M (2008) New routines for gridding of temperature and precipitation observations for 'seNorge. no'. Technical Report 08/2008, Norwegian Meteorological Institute, Oslo, Norway.

Østby TI, Schuler TV, Hagen JO, Hock R and Reijmer CH (2013) Parameter uncertainty, refreezing and surface energy balance modelling at Austfonna ice cap, Svalbard, 2004–08. *Annals of Glaciology* **54**(63), 229–240. doi: 10.3189/2013AoG63A280

Rounce DR and 5 others (2020b) Quantifying parameter uncertainty in a large-scale glacier evolution model using Bayesian inference: application to High Mountain Asia. *Journal of Glaciology* **66**(256), 175–187. doi: 10.1017/jog.2019.91

Rounce DR, Hock R and Shean DE (2020a) Glacier mass change in high mountain Asia through 2100 using the open-source python glacier evolution model (PyGEM). *Frontiers in Earth Science* **7**, 331. doi: 10.3389/feart.2019.00331

Rye C, Willis I, Arnold N and Kohler J (2012) On the need for automated multiobjective optimization and uncertainty estimation of glacier mass balance models. *Journal of Geophysical Research (Earth Surface)* **117**(F2), 1–21. doi: 10.1029/2011JF002184.

Salvatier J, Wiecki TV and Fonnesbeck C (2016) Probablistic programming in Python using PyMC3. *PeerJ Computer Science* **2**(e55), 1–24. doi: 10.7717/peerj-cs.55.

Schuler T and 6 others (2005) Distributed mass balance modelling on Engabreen (Norway). *Annals of Glaciology* **42**, 395–401. doi: 10.3189/172756405781812998

Schuler T and 5 others (2007) Calibrating a surface mass-balance model for Austfonna ice cap, Svalbard. *Annals of Glaciology* **46**, 241–248. doi: 10.3189/172756407782871783

Shean DE and 5 others (2020) A systematic, regional assessment of high mountain Asia glacier mass balance. *Frontiers in Earth Science* **7**(363), 1–19. doi: 10.3389/feart.2019.00363.

Singh P, Kumar N and Arora M (2000) Degree-day factors for snow and ice for Dokriani Glacier, Garhwal Himalayas. *Journal of Hydrology* **235**, 1–11. doi: 10.1016/S0022-1694(00)00249-3

ter Braak CJF (2006) A Markov Chain Monte Carlo version of the genetic algorithm Differential evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing* **16**(3), 239–249. doi: 10.1007/s11222-006-8769-1

ter Braak CJF and Vrugt JA (2008) Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing* **18**(4), 435–446. doi: 10.1007/s11222-008-9104-9

van Pelt W and 10 others (2019) A long-term dataset of climatic mass balance, snow conditions, and runoff in Svalbard (1957–2018). *The Cryosphere* **13**(9), 2259–2280. doi: 10.5194/tc-13-2259-2019

van Tiel M, Stahl K, Freudiger D and Seibert J (2020) Glacio-hydrological model calibration and evaluation. *WIREs Water* **7**(6), 1–51. doi: 10.1002/wat2.1483.

Vehtari A, Gelman A, Simpson D, Carpenter B and Bürkner PC (2021) Rank-normalization, folding, and localization: an improved $\widehat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis* **16**(2), 667–718. doi: 10.1214/20-BA1221

WGMS(2021) *Global Glacier Change Bulletin No. 4 (2018-2019)*. ISC(WDS)/IUGG(IACS)/UNEP/UNESCO/WMO, Zemp M, Nussbaumer SU, Gärtner-Roer I, Paul F and Hoelzle M (eds.), World Glacier Monitoring

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

Service, Zurich, Switzerland, pp. 1–278, publication based on database version: doi:10.5904/wgms-fog-2021-05.

WGMS (2022) *Fluctuations of glaciers database*. World Glacier Monitoring Service (WGMS), Zurich, Switzerland.

Woodward J, Sharp M and Arendt A (1997) The influence of superimposed-ice formation on the sensitivity of glacier mass balance to climate change. *Annals of Glaciology* **24**, 186–190. doi: 10.3189/S0260305500012155

Zekollari H, Huss M and Farinotti D (2019) Modelling the future evolution of glaciers in the European Alps under the EURO-CORDEX RCM ensemble. *The Cryosphere* **13**(4), 1125–1146. doi: 10.5194/tc-13-1125-2019

Zemp M and 14 others (2019) Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **568**(7752), 382–386.

Zhang Y, Liu S and Ding Y (2006) Observed degree-day factors and their spatial variation on glaciers in western China. *Annals of Glaciology* **43**, 301–306. doi: 10.3189/172756406781811952

Zolles T, Maussion F, Galos SP, Gurgiser W and Nicholson L (2019) Robust uncertainty assessment of the spatio-temporal transferability of glacier mass and energy balance models. *The Cryosphere* **13**(2), 469–489. doi: 10.5194/tc-13-469-2019

Appendix A. Additional figures

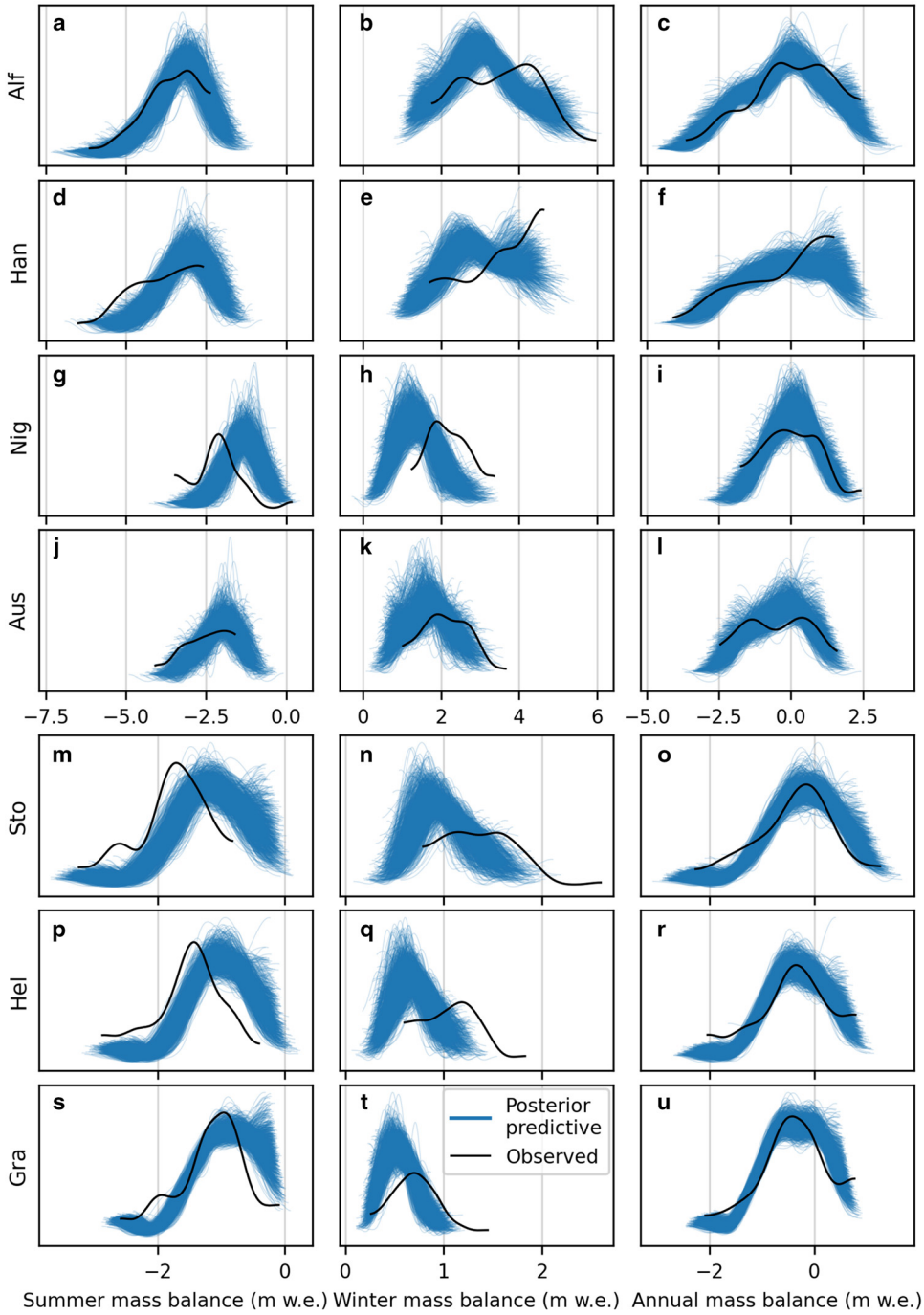


Figure A1. Probability density functions for mass balances in the B_a experiment for 1000 posterior predictive samples (blue lines) and observations (black lines). Summer, winter and annual mass balance (m w.e.) is shown from left to right for each glacier. Glaciers are sorted from west to east along a maritime to continental climate gradient.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

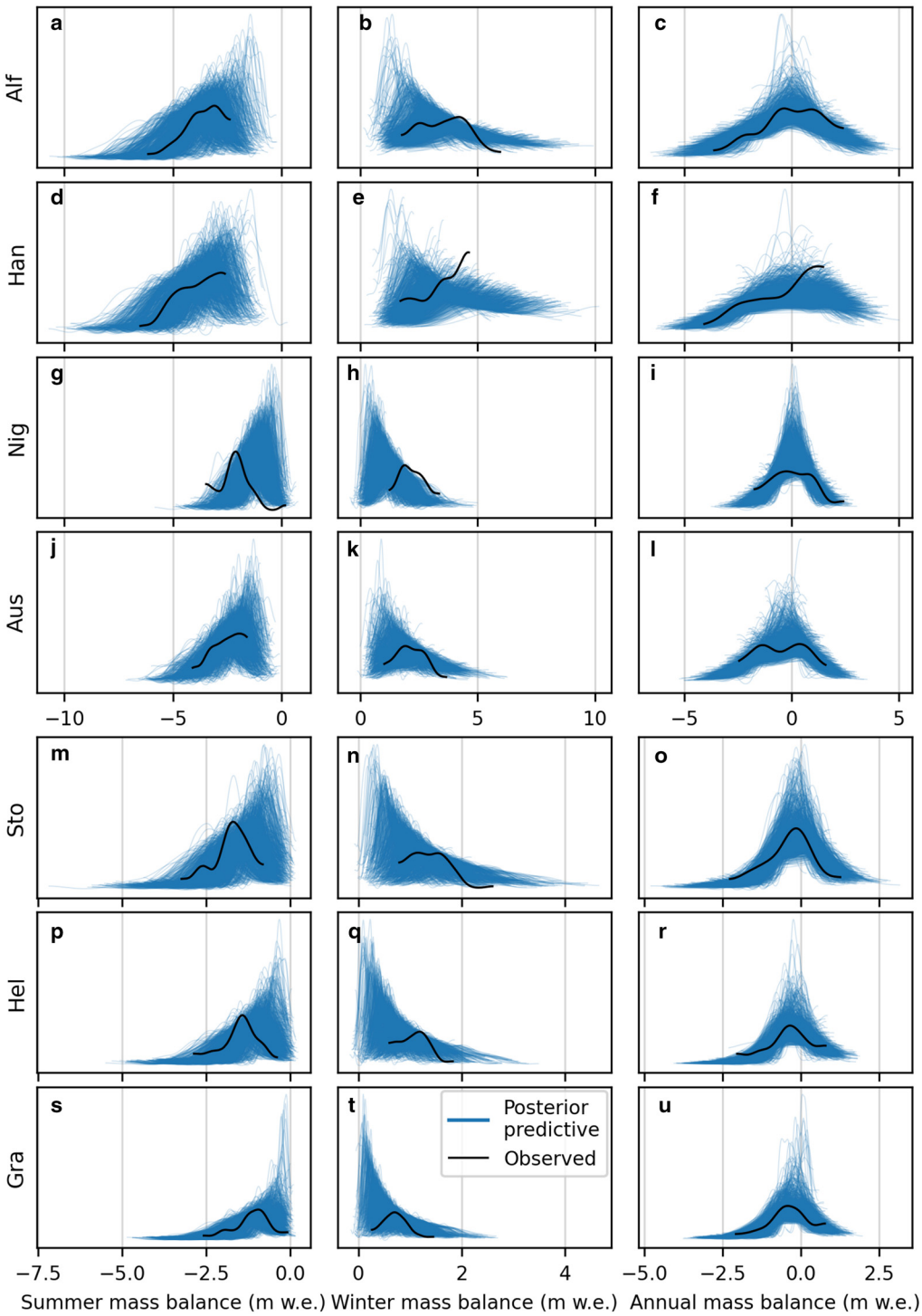


Figure A2. Probability density functions for mass balances in the $B_{a,10yr}$ experiment for 1000 posterior predictive samples (blue lines) and observations (black lines). Summer, winter and annual mass balance (m w.e.) is shown from left to right for each glacier. Glaciers are sorted from west to east along a maritime to continental climate gradient.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

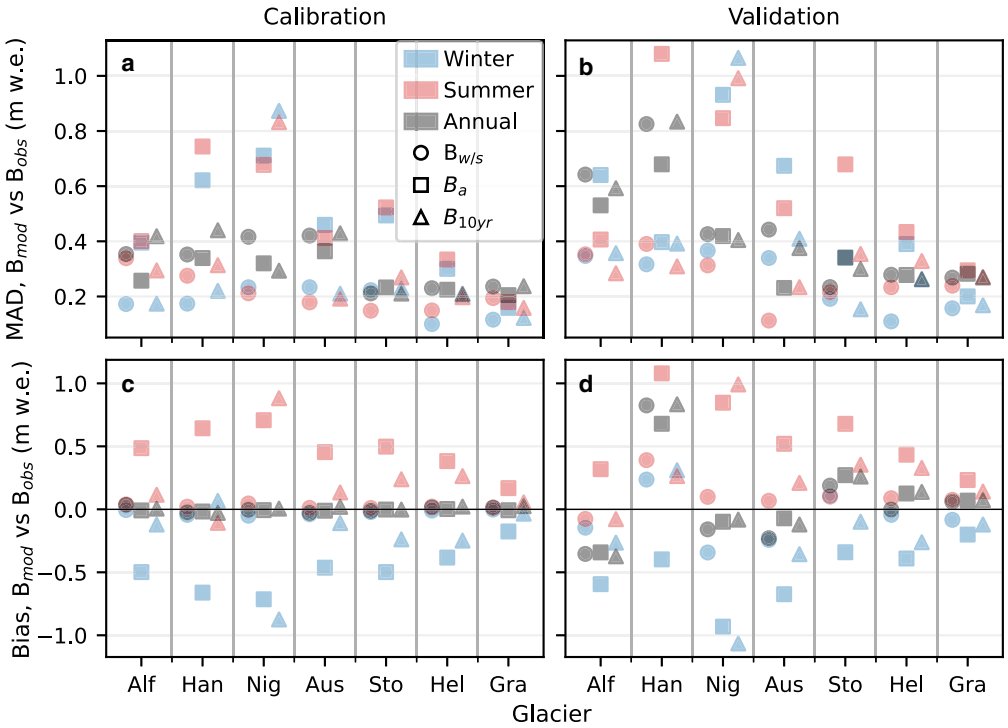


Figure A3. Model performance per glacier in terms of median absolute deviation (MAD) (a, b) and median bias (c, d) between modelled (posterior predictive median) and observed seasonal (winter in blue and summer in red) and annual (grey) surface mass balance over calibration (a, c) and validation (b, d) periods for experiments $B_{w/s}$ (circle), B_a (triangle) and B_{10yr} (square). Glaciers are sorted from west to east along a maritime to continental climate gradient.

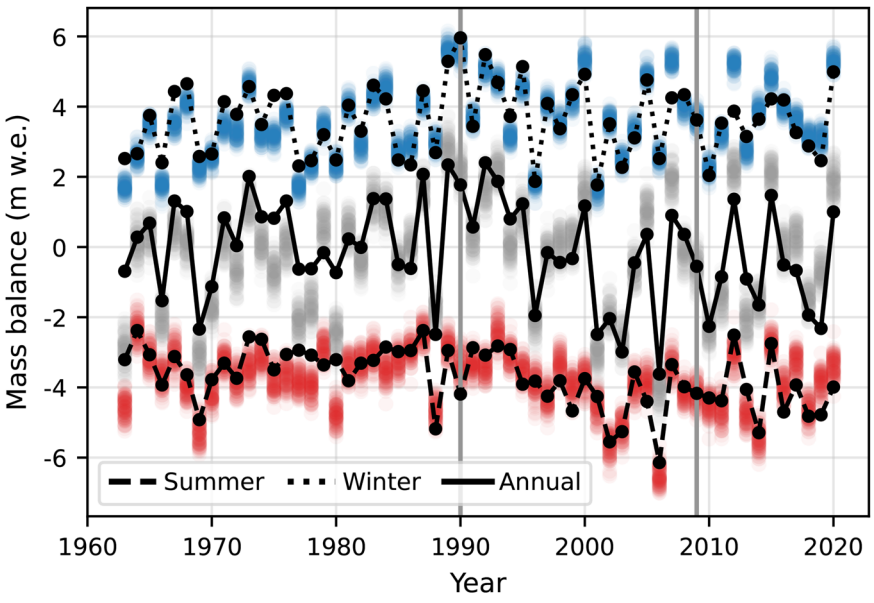


Figure A4. Posterior predictive samples (100 samples) in $B_{w/s}$ experiment of winter (blue), summer (red) and annual (grey) mass balances for Ålfotbreen 1963–2020. Observations of winter (dotted line), summer (dashed line) and annual (solid line) mass balances are shown as black dots. Grey vertical lines indicate the start (1990) and the end (2009) of the calibration period.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press

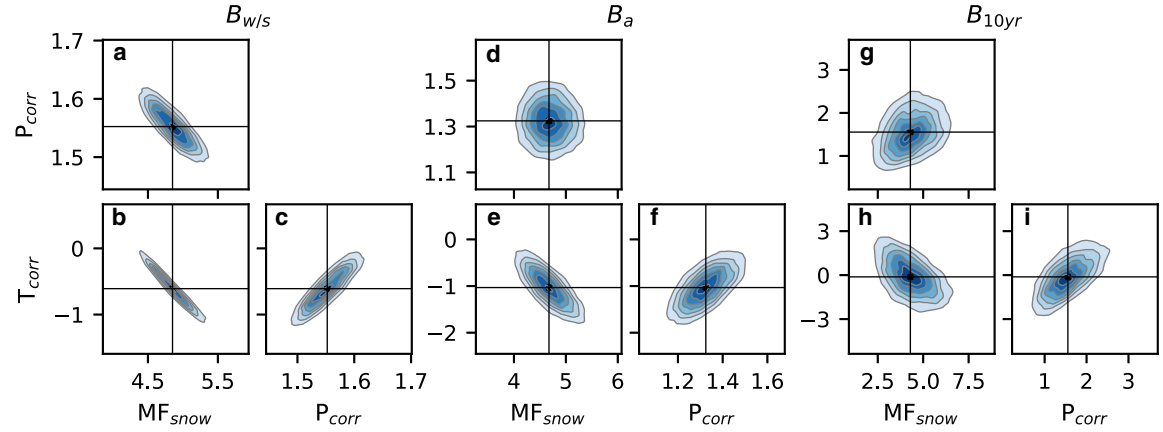


Figure A5. Joint posterior distributions of pairs of parameters for Ålfotbreen for the three experiments $B_{w/s}$ (a–c), B_a (d–f) and B_{10yr} (g–i). Parameters are melt factor for snow (MF_{snow}; mm w.e.$^\circ$C^{-1}d^{-1}), temperature correction factor (T_{corr}; $^\circ$C) and precipitation correction factor (P_{corr}; -). Lines and points indicate medians of distributions.

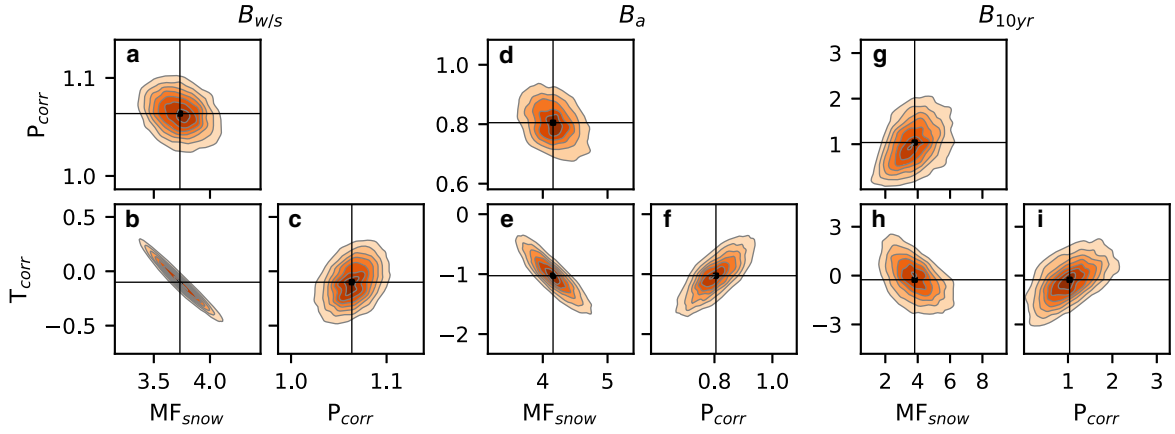


Figure A6. Joint posterior distributions of pairs of parameters for Gråsubreen for the three experiments $B_{w/s}$ (a–c), B_a (d–f), and B_{10yr} (g–i). Parameters are melt factor for snow (MF_{snow}; mm w.e.$^\circ$C^{-1}d^{-1}), temperature correction factor (T_{corr}; $^\circ$C) and precipitation correction factor (P_{corr}; -). Lines and points indicate medians of distributions.

https://doi.org/10.1017/jog.2023.62 Published online by Cambridge University Press