This is supplementary material for the article **Bayesian parameter estimation in glacier mass-balance modelling using observations with distinct temporal resolutions and uncertainties**

Kamilla H. Sjursen, Thorben Dunse, Antoine Tambue, Thomas V. Schuler, Liss M. Andreassen

CONVERGENCE OF MARKOV CHAIN MONTE CARLO (MCMC) SIMULATIONS

For each glacier and simulation case we run four Markov chains with 10,000 samples, giving a total of 40,000 samples from the joint posterior distribution of each parameter set. To facilitate the comparison of parameter distributions across the observational experiments, we must be confident that chains have converged to stationary posterior distributions in each case and that the number of independent samples is sufficient to produce stable estimates of the statistics of the marginal posterior distributions. Nearby samples in a Markov chain are inherently correlated, meaning that a substantial number of steps are needed to generate a sufficient number of independent samples from the target distribution. In addition, the initial part of a Markov chain is often discarded as *burn-in* (our 2,000 tuning iterations) as these samples rely heavily on the starting point of the chain and cannot be assumed to be drawn from the target distribution. Arbitrarily increasing the chain length to ensure convergence comes at the expense of additional computational resources.

Several empirical diagnostic tools have been developed to assess convergence of Markov chains and to evaluate the quality of estimators (e.g., mean and quantiles) of the posterior distribution (see e.g. Gelman and others, 2014, Chapter 11). As no single diagnostic can be used to conclusively establish the convergence of an MCMC sampler, evaluation should rely on multiple tools that enable the detection of convergence problems. In our analysis, we employ three main diagnostics of convergence and accuracy available in the *ArviZ* statistical package (Kumar and others, 2019): the rank-normalized $\widehat{R}$ statistic, the effective sample size, ESS, and the Monte Carlo standard error, MCSE. In addition to numerical convergence diagnostics (Tables S1–S3), we apply diagnostic visualizations to assess MCMC performance (Figs. S1 and S2).

By running multiple, independent chains we can compute the $\widehat{R}$ diagnostic, which compares the variance within and between chains (Gelman and Rubin, 1992). We use the rank-normalized $\widehat{R}$ proposed by Vehtari and others (2021), which offers improved divergence detection by identifying cases of poor chain mixing that cannot be uncovered by traditional $\widehat{R}$ (e.g. Gelman and Rubin, 1992). The $\widehat{R}$ statistic is below the upper limit of 1.01 (Vehtari and others, 2021) for all simulation cases (Tables S1–S3) implying equal variance within and between chains and thus the absence of convergence issues.

Rank plots can be used to assess the relative amount of time that each chain spends exploring a region (see Vehtari and others, 2021). Samples from the four chains are ranked from lowest to highest value and the frequency of the ranks in each chain are shown relative to where a uniform distribution would lie. Ranks are close to uniform across chains for all cases (Fig. S1), which indicates good mixing and that chains are targeting the same stationary distribution.

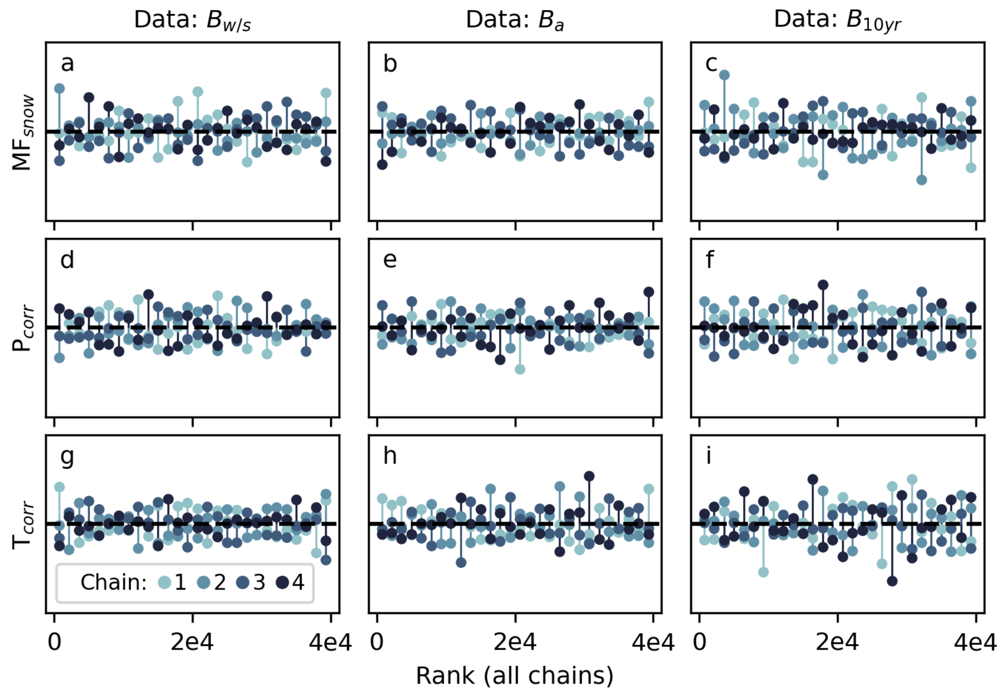

Fig. S1. Rank plot of Markov chains for Nigardsbreen showing simulation experiments $B_{w/s}$ (a, d, g), B_a (b, e, f) and B_{10yr} (c, f, i) and parameters MF_{snow} (a–c), P_{corr} (d–f) and T_{corr} (g–i). The horizontal axis shows rank, from lowest sample (1) to largest sample (40,000) in the four chains. The vertical axis represents the frequency of ranks in each bin of a histogram of the ranks, relative to where a uniform distribution would lie (dashed line).

ESS is a measure of the number of independent samples of the posterior distribution in an MCMC simulation (Gelman and others, 2014). Generally, the quality of the inference increases with ESS. Vehtari and others (2021) proposes a rank-normalized version of ESS, which offers improved detection of non-convergence across several cases where traditional ESS estimates fail. Since chain convergence is not uniform across the distribution, they argue that ESS should be assessed not only for the bulk of the distribution but also for the tails (extreme quantiles), to ensure robust estimates of how well both the center and quantiles of the distribution are resolved. A threshold of ESS $>$ 100 has been considered sufficient for reasonable accuracy in most cases (Gelman and others, 2014). More samples could be needed if the goal of the analysis requires increased stability and higher precision of the posterior summaries. A minimum threshold of 400 for the rank-normalized ESS (bulk and tail) is recommended by Vehtari and others (2021) to ensure that enough independent samples are generated to produce a reliable estimate of $\widehat{R}$. We follow their advice, while also considering that increased precision is favorable since comparison of marginal posterior distributions across cases is a central component of our assessment. ESS is well above the recommended threshold of 400 for all simulation experiments (Tables S1–S3). For the melt factor for snow, ESS increases approximately linearly with the number of iterations (Fig. S2a–c), with a minimum of ESS bulk (tail) of 1892 (3354), 2509 (4043) and 2320 (2292) for the cases $B_{w/s}$, B_a, and B_{10yr}, respectively. The stable increase in ESS with chain length indicates that chains have converged to a stationary distribution, such that running a longer chain would result in more samples from the same distribution. Tails of the distributions are efficiently explored with ESS > 2000 for all quantiles (Fig. S2d–f).

MCSE is a measure of the quality, or precision, of the estimators (mean and quantiles) of the posterior distribution. MCSE of a point estimate does not indicate convergence but is a measure of the error associated with estimating the quantify from a finite number of samples of the distribution. A high degree of precision (low MCSE) comes at the expense of computational cost and is not necessarily required in practical inference (Gelman and others, 2014). MCSE limits should be assessed on a case-by-case basis (Vehtari and others, 2021) and reported to allow for objective assessment of the accuracy of simulations (Flegal and others, 2008).

MCSE is mostly uniform across quantiles of the distribution (Fig. S2g–i), which is indicative of efficient exploration across the distribution. MCSE is below 0.025 for the simulation experiments $B_{w/s}$ and B_a, and below 0.05 across quantiles for B_{10yr}. In the latter case, MCSE is generally higher, particularly for lower and higher quantiles of the distribution. This could be a consequence of a relatively wide and long-tailed posterior distribution, which would be as expected given the limited constraints imposed by B_{10yr}.

MCSE for the mean (standard deviation) of the marginal posterior distributions is $\leq$ 5% ($\leq$ 2%) for $<$ 90% (100%) of all parameters and glaciers (Tables S1–S3). For the current analysis, we consider the simulations to provide reasonable accuracy and sufficiently stable estimates of posterior quantities. As an example, the marginal posterior distribution for MF_{snow} for Gråsubreen using B_{10yr} has an estimated mean (standard deviation) of 3.813 (1.215) mm w.e.$^{\circ}$C^{-1}d^{-1} and associated

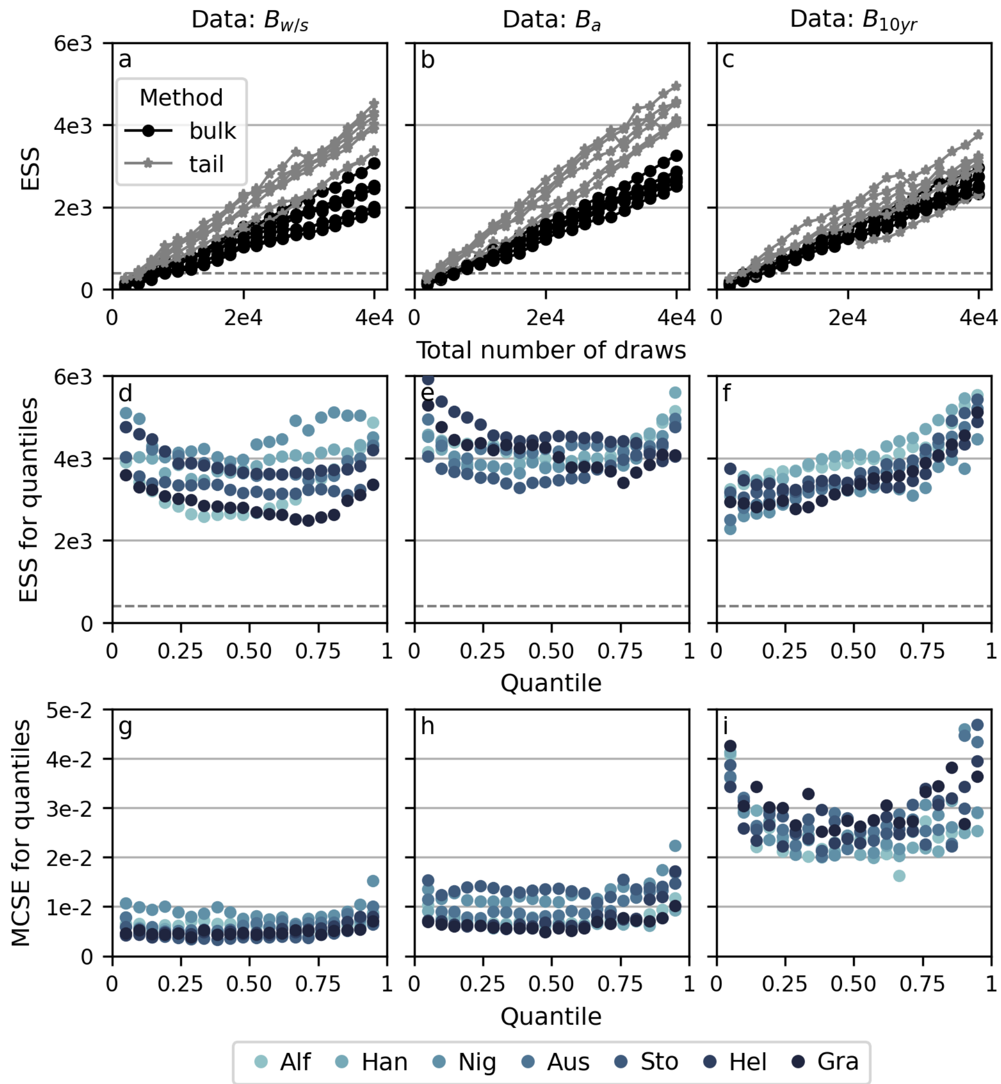

Fig. S2. Convergence diagnostics for the melt factor for snow (MF_{snow}) for all glaciers and MCMC simulation cases $B_{w/s}$ (a, d, g), B_a (b, e, f) and B_{10yr} (c, f, i). Top row (a–c) shows evolution of effective sample size (ESS), bulk and tail, with increasing chain length. Middle row (d–f) and bottom row (g–i) shows ESS and Monte Carlo standard error (MCSE) for different quantiles of the distributions, respectively. In panels a-f the grey dashed line shows the recommended threshold of ESS = 400.

MCSE of 0.024 (0.017) mm w.e.°C^{-1}d^{-1}. We would thus be comfortable reporting an estimate of 3.8 ± 1.2 mm w.e.°C^{-1}d^{-1} for this parameter.

Table S1. Summary statistics of marginal posterior distributions and Markov chain Monte Carlo simulations (4 chains of 10,000 steps) for each glacier and parameters: precipitation correction factor, P_{corr} (–), melt factor for snow, MF_{snow} (mm w.e.°C^{-1}d^{-1}), and temperature bias correction, T_{corr} (°C), for the $B_{w/s}$ experiment. SD, HDI, MCSE, and ESS refer to standard deviation, high density interval, Monte Carlo standard error, and effective sample size, respectively. $\widehat{R}$ (–) refers to rank-normalized $\widehat{R}$ (Vehtari and others, 2021). Units for mean, SD, HDI, and MCSE correspond to parameter units, while unit for ESS is number of samples.

Parameter	Glacier	Mean	Median	SD	$HDI_{2.5\%}$	$HDI_{97.5\%}$	$MCSE_{Mean}$	$MCSE_{SD}$	ESS_{bulk}	ESS_{tail}	$\widehat{R}$
	Alf	4.852	4.854	0.249	4.385	5.340	0.0056	0.0039	1989	3903	1.003
	Han	4.404	4.408	0.212	3.972	4.815	0.0043	0.0030	2451	4027	1.001
	Nig	4.213	4.212	0.417	3.413	5.048	0.0075	0.0053	3070	4519	1.001
MF_{snow}	Aus	3.534	3.513	0.284	2.997	4.092	0.0057	0.0040	2507	4302	1.001
	Sto	3.028	3.024	0.164	2.699	3.344	0.0037	0.0026	2021	3354	1.003
	Hel	3.187	3.170	0.223	2.788	3.627	0.0044	0.0031	2534	4202	1.002
	Gra	3.732	3.731	0.182	3.383	4.084	0.0042	0.0030	1892	3363	1.001
	Alf	1.552	1.551	0.032	1.492	1.613	0.0007	0.0005	1995	4445	1.004
	Han	1.498	1.496	0.031	1.438	1.563	0.0006	0.0005	2374	3210	1.002
	Nig	1.173	1.171	0.027	1.121	1.227	0.0005	0.0003	3307	5141	1.001
P_{corr}	Aus	1.096	1.096	0.024	1.050	1.143	0.0005	0.0004	2352	4098	1.001
	Sto	1.722	1.722	0.033	1.657	1.786	0.0007	0.0005	2291	3673	1.001
	Hel	1.255	1.254	0.026	1.207	1.308	0.0005	0.0004	2695	4333	1.001
	Gra	1.064	1.064	0.019	1.027	1.100	0.0005	0.0003	1501	2951	1.001
	Alf	−0.609	−0.621	0.280	−1.122	−0.050	0.0060	0.0043	2177	4216	1.003
	Han	−0.627	−0.643	0.264	−1.146	−0.097	0.0054	0.0039	2373	3835	1.002
	Nig	−0.808	−0.838	0.449	−1.636	0.112	0.0080	0.0057	3165	5073	1.001
T_{corr}	Aus	−0.382	−0.370	0.378	−1.129	0.336	0.0076	0.0054	2499	4470	1.001
	Sto	1.005	1.002	0.246	0.515	1.487	0.0054	0.0038	2048	3596	1.003
	Hel	0.484	0.494	0.298	−0.073	1.047	0.0059	0.0042	2545	4297	1.002
	Gra	−0.101	−0.104	0.185	−0.453	0.258	0.0042	0.0029	1973	3522	1.001

Table S2. Summary statistics of marginal posterior distributions and Markov chain Monte Carlo simulations (4 chains of 10,000 steps) for each glacier and parameters: precipitation correction factor, P_{corr} (–), melt factor for snow, MF_{snow} (mm w.e.°C^{-1}d^{-1}), and temperature bias correction, T_{corr} (°C), for the B_a experiment. SD, HDI, MCSE, and ESS refer to standard deviation, high density interval, Monte Carlo standard error, and effective sample size, respectively. $\widehat{R}$ (–) refers to rank-normalized $\widehat{R}$ (Vehtari and others, 2021). Units for mean, SD, HDI, and MCSE correspond to parameter units, while unit for ESS is number of samples.

Parameter	Glacier	Mean	Median	SD	HDI$_{2.5\%}$	HDI$_{97.5\%}$	MCSE$_{Mean}$	MCSE$_{SD}$	ESS$_{bulk}$	ESS$_{tail}$	$\widehat{R}$
	Alf	4.675	4.668	0.341	4.031	5.358	0.0064	0.0045	2870	4152	1.001
	Han	4.016	4.013	0.302	3.411	4.590	0.0058	0.0041	2721	4567	1.001
	Nig	3.413	3.373	0.565	2.322	4.520	0.0108	0.0077	2690	4526	1.001
MF_{snow}	Aus	3.235	3.204	0.451	2.399	4.154	0.0084	0.0060	2847	4942	1.000
	Sto	4.522	4.520	0.583	3.389	5.589	0.0116	0.0082	2509	4043	1.002
	Hel	3.685	3.657	0.363	3.025	4.430	0.0064	0.0046	3263	4047	1.001
	Gra	4.159	4.151	0.288	3.609	4.718	0.0056	0.0040	2616	4078	1.001
	Alf	1.325	1.322	0.087	1.152	1.492	0.0016	0.0012	2896	4702	1.001
	Han	1.196	1.196	0.083	1.041	1.364	0.0015	0.0011	2959	4330	1.002
	Nig	0.811	0.810	0.095	0.626	0.995	0.0017	0.0012	2959	3995	1.002
P_{corr}	Aus	0.860	0.859	0.086	0.688	1.023	0.0016	0.0012	2770	4109	1.001
	Sto	1.045	1.038	0.127	0.807	1.302	0.0026	0.0018	2368	3654	1.002
	Hel	0.782	0.781	0.088	0.613	0.959	0.0017	0.0012	2791	4821	1.001
	Gra	0.805	0.803	0.067	0.675	0.936	0.0012	0.0008	3212	5345	1.001
	Alf	−1.032	−1.039	0.396	−1.804	−0.266	0.0072	0.0051	3025	4580	1.001
	Han	−1.102	−1.114	0.362	−1.843	−0.411	0.0068	0.0048	2857	4189	1.002
	Nig	−1.395	−1.387	0.767	−2.929	0.093	0.0142	0.0101	2958	4345	1.001
T_{corr}	Aus	−0.961	−0.949	0.687	−2.307	0.380	0.0129	0.0091	2840	3893	1.001
	Sto	−1.809	−1.867	0.633	−2.925	−0.519	0.0128	0.0091	2477	3663	1.002
	Hel	−1.199	−1.180	0.464	−2.137	−0.304	0.0085	0.0061	2982	4027	1.001
	Gra	−1.026	−1.025	0.315	−1.618	−0.396	0.0059	0.0042	2854	4596	1.001

Table S3. Summary statistics of marginal posterior distributions and Markov chain Monte Carlo simulations (4 chains of 10,000 steps) for each glacier and parameters: precipitation correction factor, P_{corr} (–), melt factor for snow, MF_{snow} (mm w.e.°C^{-1}d^{-1}), and temperature bias correction, T_{corr} (°C), for the B_{10yr} experiment. SD, HDI, MCSE, and ESS refer to standard deviation, high density interval, Monte Carlo standard error, and effective sample size, respectively. $\widehat{R}$ (–) refers to rank-normalized $\widehat{R}$ (Vehtari and others, 2021). Units for mean, SD, HDI, and MCSE correspond to parameter units, while unit for ESS is number of samples.

Parameter	Glacier	Mean	Median	SD	HDI$_{2.5\%}$	HDI$_{97.5\%}$	MCSE$_{Mean}$	MCSE$_{SD}$	ESS$_{bulk}$	ESS$_{tail}$	$\widehat{R}$
	Alf	4.294	4.248	1.096	2.174	6.492	0.0207	0.0146	2748	3245	1.002
	Han	4.274	4.245	1.076	2.166	6.331	0.0195	0.0138	2962	3085	1.002
	Nig	2.738	2.661	1.037	0.849	4.812	0.0208	0.0147	2320	2292	1.001
MF_{snow}	Aus	3.535	3.462	1.100	1.339	5.601	0.0215	0.0152	2504	2509	1.001
	Sto	3.784	3.734	1.182	1.556	6.132	0.0228	0.0161	2534	3179	1.001
	Hel	3.500	3.417	1.151	1.432	5.864	0.0216	0.0153	2740	3756	1.001
	Gra	3.813	3.741	1.215	1.540	6.180	0.0239	0.0169	2478	2942	1.003
	Alf	1.555	1.524	0.452	0.697	2.449	0.0083	0.0059	2843	3238	1.003
	Han	1.627	1.597	0.470	0.756	2.565	0.0087	0.0062	2788	2929	1.001
	Nig	0.749	0.741	0.244	0.262	1.198	0.0049	0.0034	2464	2984	1.000
P_{corr}	Aus	1.068	1.054	0.315	0.461	1.699	0.0062	0.0044	2433	2694	1.001
	Sto	1.425	1.406	0.542	0.422	2.508	0.0106	0.0075	2479	2451	1.002
	Hel	0.999	0.972	0.440	0.154	1.819	0.0086	0.0061	2438	2542	1.002
	Gra	1.037	1.002	0.503	0.068	1.943	0.0118	0.0084	1584	1441	1.002
	Alf	−0.119	−0.153	1.252	−2.603	2.286	0.0226	0.016	3073	4491	1.001
	Han	−0.216	−0.250	1.245	−2.591	2.275	0.022	0.0156	3208	4760	1.001
	Nig	−0.831	−0.824	1.395	−3.493	1.910	0.0248	0.0176	3146	4344	1.002
T_{corr}	Aus	−0.480	−0.503	1.318	−3.059	2.102	0.0242	0.0172	2959	4894	1.001
	Sto	−0.368	−0.393	1.227	−2.886	1.936	0.0223	0.0158	3031	4302	1.001
	Hel	−0.435	−0.465	1.277	−2.926	2.062	0.0241	0.0170	2821	4502	1.002
	Gra	−0.256	−0.257	1.228	−2.777	2.115	0.025	0.0177	2408	4113	1.003

SENSITIVITY TO CHANGES IN CLIMATE FORCING

We investigated the effect of posterior distributions on sensitivities of SMB to changes in climate forcing by running posterior predictive simulations over the period 1960–2020 with +1 K increase in temperature and +10% increase in precipitation. Average sensitivities (Table S4) are comparable across experiments, generally higher for maritime than continental glaciers, and in line with values found in previous studies (e.g. Rasmussen and Conway, 2005; Schuler and others, 2005; De Woul and Hock, 2005). However, uncertainty (standard deviation) in mass-balance sensitivity in the B_{10yr} experiment is considerably higher than the other experiments (by a factor of 1.2 – 3.1 compared to $B_{w/s}$) due to higher parameter uncertainty.

Table S4. Surface mass balance (SMB) sensitivities (mean ± standard deviation) from posterior predictive simulations (1960–2020) with +1 K temperature increase (dSMB/dT) and +10% precipitation increase (dSMB/dP) for each glacier and experiment $B_{w/s}$, B_a, and B_{10yr}. Glaciers are sorted from west to east along a maritime to continental climate gradient.

Experiment	Glacier	SMB (m w.e. a^{-1})	dSMB/dT (m w.e. a^{-1} K^{-1})	dSMB/dP (m w.e. a^{-1} $10\%^{-1}$)
	Alf	−0.31	−1.37 ± 0.49	0.44 ± 0.49
	Han	−0.54	−1.33 ± 0.49	0.42 ± 0.49
	Nig	−0.10	−0.78 ± 0.50	0.24 ± 0.50
$B_{w/s}$	Aus	−0.75	−0.78 ± 0.44	0.22 ± 0.44
	Sto	−0.30	−0.70 ± 0.26	0.19 ± 0.26
	Hel	−0.40	−0.60 ± 0.23	0.15 ± 0.22
	Gra	−0.43	−0.54 ± 0.15	0.10 ± 0.15
	Alf	−0.31	−1.22 ± 0.50	0.39 ± 0.50
	Han	−0.49	−1.12 ± 0.50	0.35 ± 0.50
	Nig	−0.04	−0.57 ± 0.51	0.17 ± 0.51
B_a	Aus	−0.62	−0.65 ± 0.47	0.18 ± 0.46
	Sto	−0.25	−0.63 ± 0.28	0.14 ± 0.28
	Hel	−0.38	−0.51 ± 0.23	0.10 ± 0.23
	Gra	−0.42	−0.49 ± 0.15	0.08 ± 0.15
	Alf	−0.35	−1.35 ± 0.86	0.43 ± 0.78
	Han	−0.58	−1.40 ± 0.84	0.43 ± 0.77
	Nig	−0.03	−0.49 ± 0.62	0.15 ± 0.59
B_{10yr}	Aus	−0.68	−0.76 ± 0.71	0.21 ± 0.65
	Sto	−0.27	−0.67 ± 0.52	0.17 ± 0.48
	Hel	−0.38	−0.54 ± 0.47	0.12 ± 0.43
	Gra	−0.42	−0.52 ± 0.45	0.10 ± 0.41

INFERENCE WITH STUDENT-T DISTRIBUTIONS FOR LIKELIHOOD

We investigated the use of Student-t distributions for the likelihood in the $B_{w/s}$ and B_a experiments to investigate the effect on inferred posterior parameter distributions and robustness of modelled mass balances (Gelman and others, 2014, Chapter 17). We performed two tests with Student-t distributions for the likelihood both with four degrees of freedom and three and five times the reported observation variance used in the original $B_{w/s}$ and B_a experiments (e.g. as shown for the example of Ålfotbreen annual balances in Fig. S3). The tests were performed for the three glaciers that showed the worst performance in the original experiments: Ålfotbreen, Hansebreen, and Nigardsbreen.

For both experiments $B_{w/s}$ (Fig. S4 and Fig. (S6) and B_a (Fig. S5 and Fig. (S7) posterior parameter distributions show a larger spread, but are roughly centred around the same values as in the original experiments. For some experiments, especially those with the largest variance for B_a (Fig. (S7), posteriors of MF_{snow} show some shift towards lower values. This shift is compensated by a similar shift in the posterior of T_{corr} toward more positive values, such that the decrease in melt from applying a lower value of MF_{snow} is compensated by higher temperatures. The mode of the distributions of P_{corr} are generally aligned with posteriors in the original cases, but the spread in the distributions is greater.

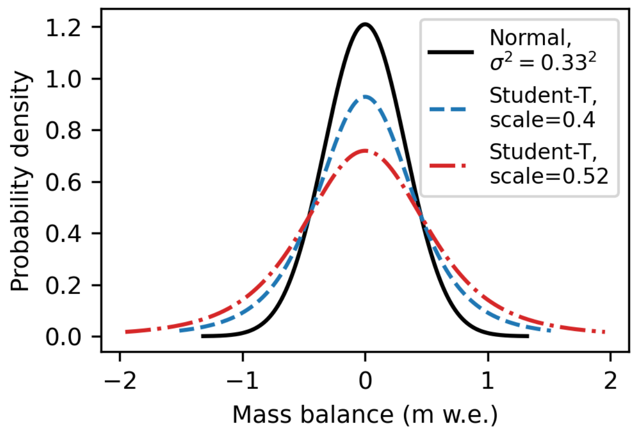

Fig. S3. Example of normal distribution (black solid line) and Student-T distributions with four degrees of freedom and three (blue dashed line) and five (red dash-dotted line) times reported observation variance.

Posterior predictive probability density functions (PDFs) of modelled mass balances reflect the increased spread of the posterior distributions (Figs. S8, S9, S10, and S11). In some cases (e.g. Ålfotbreen; Fig. S8c and Fig. S9c) modelled mass balances are somewhat improved in terms of the PDFs of posterior predictive samples to a greater degree capturing the PDF of observations. We would thus expect predictions to be somewhat more robust. However, the long tails of the posterior distributions also result in a greater number of extreme predicted values, e.g. positive summer balances and negative winter balances (e.g. Fig. S10h,g and Fig. S11h,g). This indicates that additional restrictions in our likelihood formulation (e.g. as those used in Rounce and others (2020)) may be necessary to constrain modelled mass balances within plausible ranges. The underestimation of the magnitude of seasonal balances for B_a is still found using Student-t distributions for the likelihood (Fig. S9 and Fig. S11).

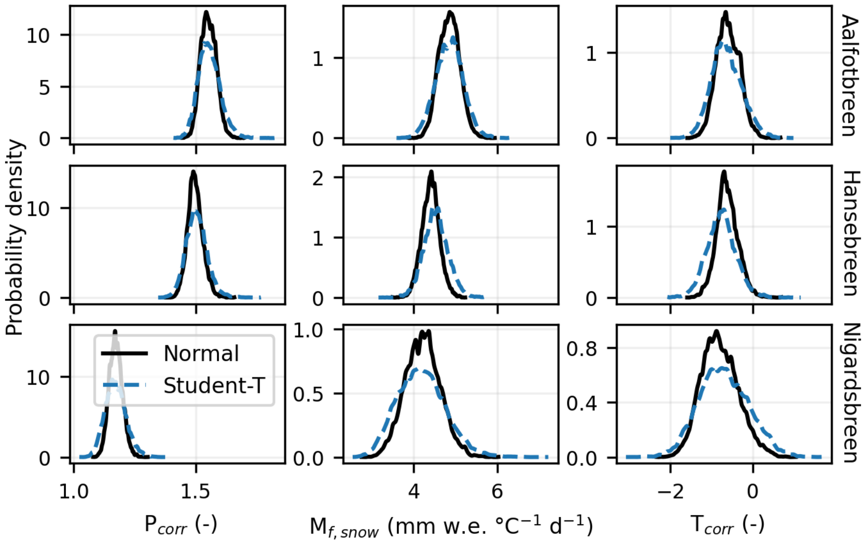

Fig. S4. Comparison of posterior parameter distributions for $B_{w/s}$ using Student-T distributions with scale parameter computed from three times reported observation variance (blue dashed line) and normal distributions with reported observation variance (black solid line).

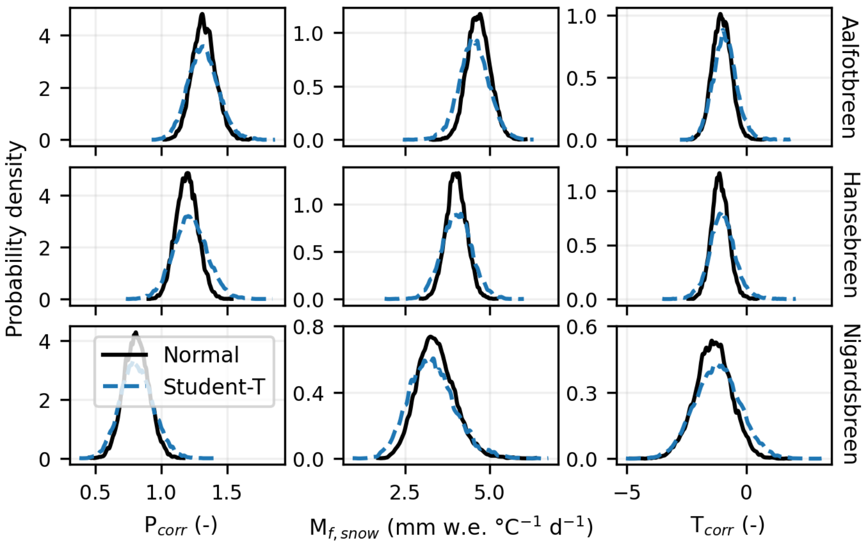

Fig. S5. Comparison of posterior parameter distributions for B_a using Student-T distributions with scale parameter computed from three times reported observation variance (blue dashed line) and normal distributions with reported observation variance (black solid line).

112

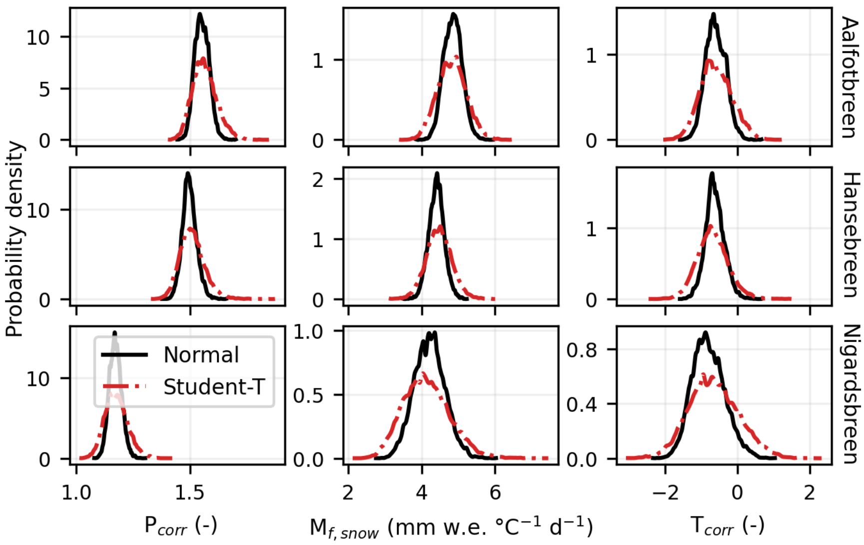

Fig. S6. Comparison of posterior parameter distributions for $B_{w/s}$ using Student-T distributions with scale parameter computed from five times reported observation variance (red dash-dotted line) and normal distributions with reported observation variance (black solid line).

113

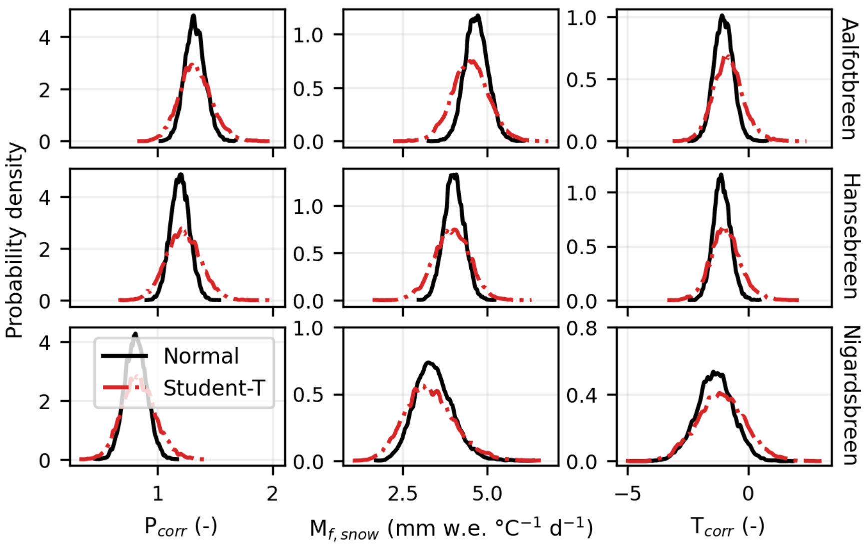

Fig. S7. Comparison of posterior parameter distributions for B_a using Student-T distributions with scale parameter computed from five times reported observation variance (red dash-dotted line) and normal distributions with reported observation variance (black solid line).

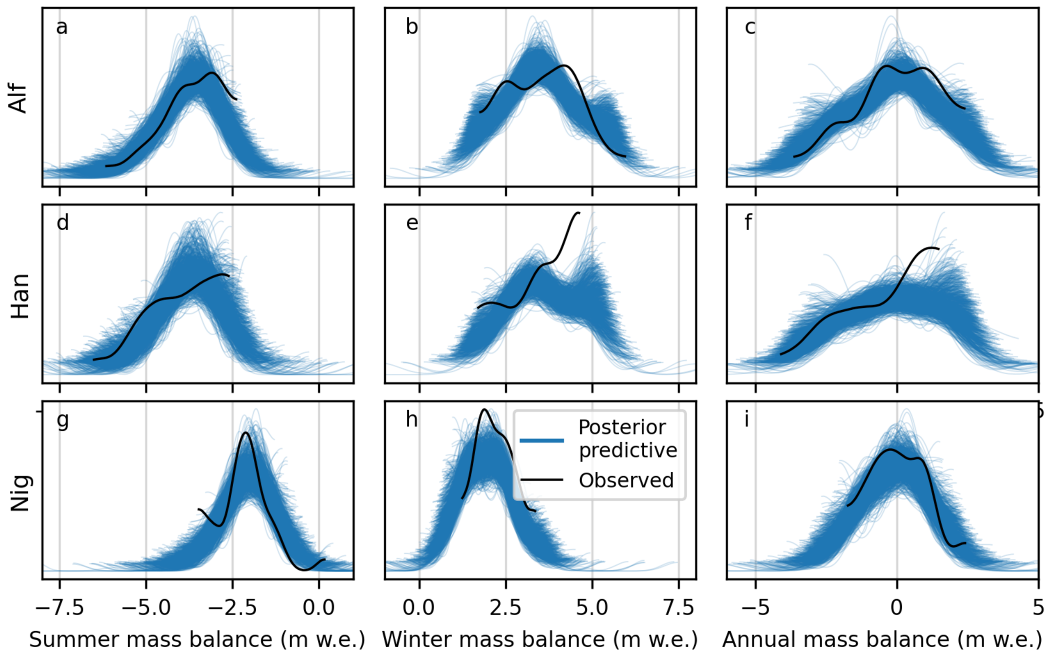

Fig. S8. Probability density functions (PDFs) for mass balances in the $B_{w/s}$ experiment for 1000 posterior predictive samples (blue lines) using Student-T distributions with scale parameter computed from three times reported observation variance for the likelihood.

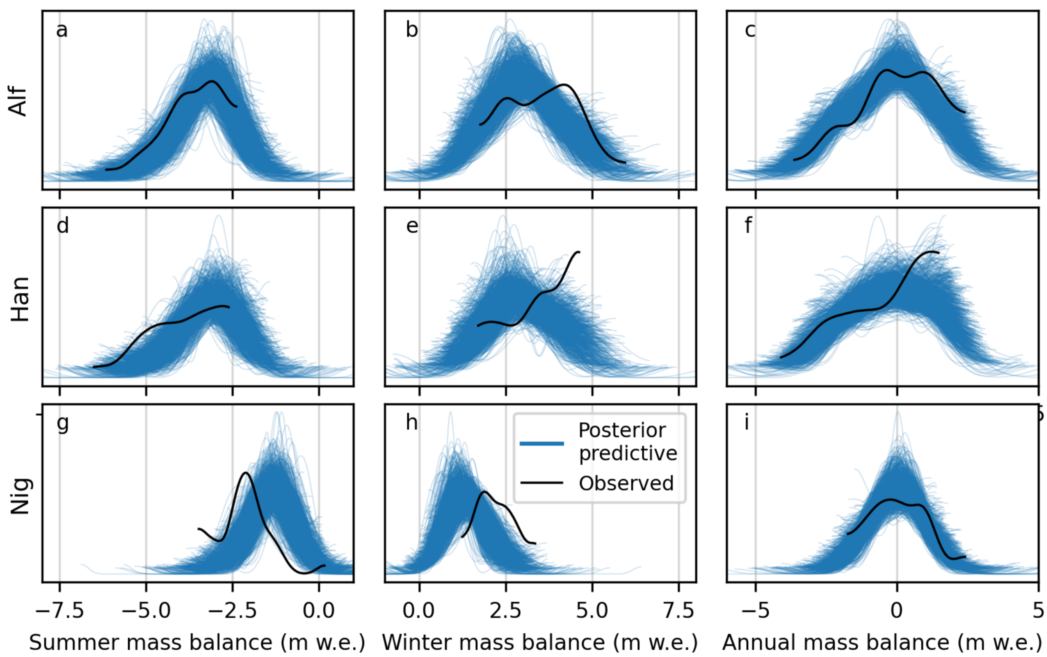

Fig. S9. Probability density functions (PDFs) for mass balances in the B_a experiment for 1000 posterior predictive samples (blue lines) using Student-T distributions with scale parameter computed from three times reported observation variance for the likelihood.

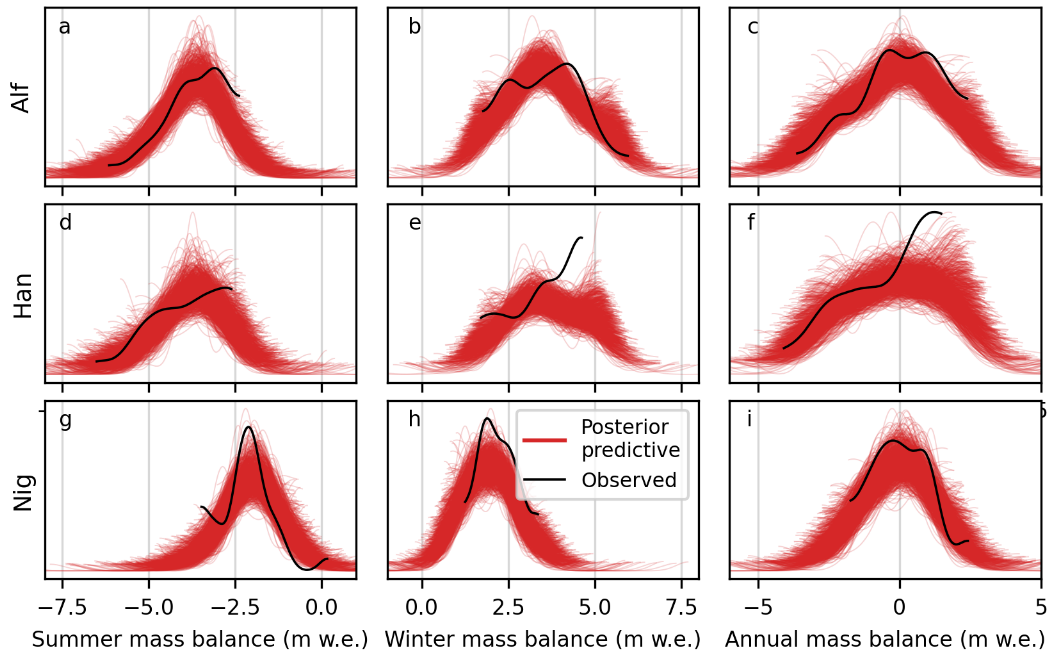

Fig. S10. Probability density functions (PDFs) for mass balances in the $B_{w/s}$ experiment for 1000 posterior predictive samples (red lines) using Student-T distributions with scale parameter computed from five times reported observation variance for the likelihood.

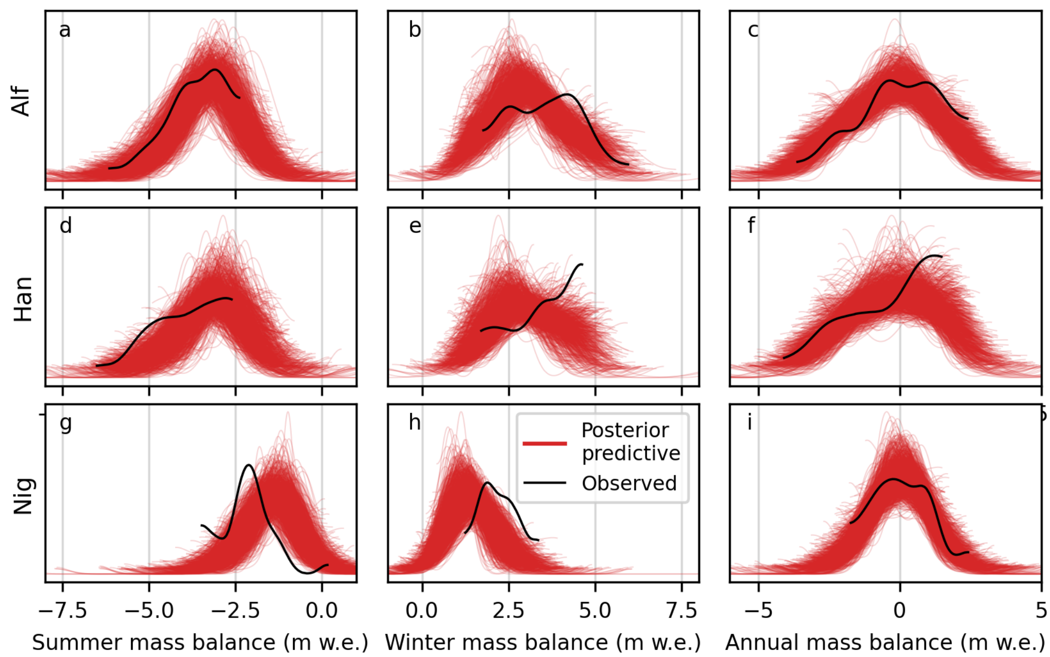

Fig. S11. Probability density functions (PDFs) for mass balances in the B_a experiment for 1000 posterior predictive samples (red lines) using Student-T distributions with scale parameter computed from five times reported observation variance for the likelihood.

REFERENCES

De Woul M and Hock R (2005) Static mass-balance sensitivity of Arctic glaciers and ice caps using a degree-day approach. *Annals of Glaciology*, **42**, 217–224 (doi: 10.3189/172756405781813096)

Flegal J, Haran M and Jones G (2008) Markov Chain Monte Carlo: Can We Trust the Third Significant Figure? *Statistical Science*, **23** (doi: 10.1214/08-STS257)

Gelman A and Rubin DB (1992) Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, **7**(4), 457–472 (doi: 10.1214/ss/1177011136)

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB (2014) *Bayesian Data Analysis*. Chapman and Hall/CRC, third edition, ISBN 978–1–4398–4095–5 (doi: 10.1201/b16018)

Kumar R, Carroll C, Hartikainen A and Martin O (2019) ArviZ a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software*, **4**(33), 1143 (doi: 10.21105/joss.01143)

Rasmussen L and Conway H (2005) Influence of upper-air conditions on glaciers in Scandinavia. *Annals of Glaciology*, **42**, 402–408 (doi: 10.3189/172756405781812727)

Rounce DR, Khurana T, Short MB, Hock R, Shean DE and Brinkerhoff DJ (2020) Quantifying parameter uncertainty in a large-scale glacier evolution model using Bayesian inference: application to High Mountain Asia. *Journal of Glaciology*, 1–13 (doi: 10.1017/jog.2019.91)

Schuler T, Hock R, Jackson M, Elvehøy H, Braun M, Brown I and Hagen J (2005) Distributed mass balance modelling on Engabreen (Norway). *Annals of Glaciology*, **42**, 395–401 (doi: 10.3189/172756405781812998)

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)