Journal of Glaciology



Paper

Cite this article: Rounce DR, Khurana T, Short MB, Hock R, Shean DE, Brinkerhoff DJ (2020). 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. https://doi.org/10.1017/jog.2019.91

Received: 23 August 2019
Revised: 12 November 2019
Accepted: 13 November 2019
First published online: 27 January 2020

Key words:
Bayesian model; glaciers; mass change; High Mountain Asia; Markov chain Monte Carlo; parameter uncertainty

Author for correspondence:
David R. Rounce, E-mail: drounce@alaska.edu

© The Author(s) 2020. 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 in any medium, provided the original work is properly cited.

cambridge.org/jog

Quantifying parameter uncertainty in a large-scale glacier evolution model using Bayesian inference: application to High Mountain Asia

David R. Rounce[1], Tushar Khurana[1,2], Margaret B. Short[3], Regine Hock[1], David E. Shean[2] and Douglas J. Brinkerhoff[4]

[1]Geophyiscal Institute, University of Alaska Fairbanks, Fairbanks, AK, USA; [2]Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA; [3]College of Natural Science and Mathematics, University of Alaska Fairbanks, Fairbanks, AK, USA and [4]Department of Computer Science, University of Montana, Missoula, MT, USA

Abstract

The response of glaciers to climate change has major implications for sea-level change and water resources around the globe. Large-scale glacier evolution models are used to project glacier runoff and mass loss, but are constrained by limited observations, which result in models being over-parameterized. Recent systematic geodetic mass-balance observations provide an opportunity to improve the calibration of glacier evolution models. In this study, we develop a calibration scheme for a glacier evolution model using a Bayesian inverse model and geodetic mass-balance observations, which enable us to quantify model parameter uncertainty. The Bayesian model is applied to each glacier in High Mountain Asia using Markov chain Monte Carlo methods. After 10,000 steps, the chains generate a sufficient number of independent samples to estimate the properties of the model parameters from the joint posterior distribution. Their spatial distribution shows a clear orographic effect indicating the resolution of climate data is too coarse to resolve temperature and precipitation at high altitudes. Given the glacier evolution model is over-parameterized, particular attention is given to identifiability and the need for future work to integrate additional observations in order to better constrain the plausible sets of model parameters.

1. Introduction

Glacier mass loss in the next 100 years has critical implications for sea-level change and water resources across the globe. From 1961 to 2016, the rate of glacier mass loss, excluding the Greenland and Antarctic ice sheets, averaged 0.5 ± 0.4 mm sea-level equivalent (SLE) a^{-1} and recent rates of mass loss since 2010 now exceed 1.0 ± 0.4 mm SLE a^{-1} (Zemp and others, 2019). Future projections of mass loss from several glacier evolution models using output from an ensemble of General Circulation Models (GCMs) and Representative Concentration Pathways (RCPs) estimate the increase in sea level due to glaciers by 2100 relative to 2015 to range from 94 ± 25 mm SLE (RCP2.6) to 200 ± 44 mm SLE (RCP8.5) (Hock and others, 2019). While the rate of sea-level change is dominated by regions with the most glacier mass (Alaska and the arctic regions), glacier mass loss in other regions may fundamentally alter the quantity and timing of glacier runoff, thereby affecting water resources (Bliss and others, 2014; Huss and Hock, 2018).

Future projections of glacier mass loss for a given GCM and RCP vary considerably depending on the model, which is attributed to differences in the model physics, calibration data/methods, and input data (Hock and others, 2019). In recent decades, large-scale glacier evolution models have been hampered by a lack of observations. Models were typically calibrated with glacier-wide specific mass balances (Marzeion and others, 2012), mass-balance profiles from individual glaciers (Raper and Braithwaite, 2006; Radić and Hock, 2011; Slangen and others, 2012; Giesen and Oerlemans, 2013), regional-scale mass-balance estimates from glaciological, geodetic or gravimetric observations (Huss and Hock, 2015; Kraaijenbrink and others, 2017), or a combination of glacier and regional data (Hirabayashi and others, 2013; Radić and others, 2014). Excluding mass-balance sensitivity models (Slangen and others, 2012), these sparse observations are used to calibrate between two and seven parameters in each model, which affect the air temperature, precipitation, and mass balance (Table 1). Each model has at least one parameter affecting the mass balance (e.g., degree-day factors, temperature sensitivity, or mass-balance gradient), and seven of the nine models cited above use a precipitation correction factor to correct for potential biases in the climate data.

Given the lack of systematic observations, the resulting models are over-parameterized, i.e., there are not enough data to definitively calibrate each parameter, causing previous models to determine one 'best fit' solution accompanied by a sensitivity analysis. This is problematic as many combinations of parameter sets could yield results within the reported errors, and existing studies do not include rigorous analysis of the model parameter uncertainty. Furthermore, models that rely on limited mass-balance observations must transfer model parameters to glaciers without observations, which introduces another source of uncertainty. Fortunately,

Table 1. Calibration parameters and spatial domain of glacier mass balance data used for calibration in recently published large-scale glacier evolution models

Study	Study extent	Parameters	Spatial domain of data
Raper and Braithwaite (2006)	Global (excl. A&G)	T(2), P(1), MB(2)	>200 glaciers
Radić and Hock (2011)	Global	T(2), P(3), MB(2)	36 glaciers
Marzeion and others (2012)	Global (excl. A)	P(3), MB(3)	255 glaciers
Slangen and others (2012)	Global (excl. A&G)	MB(1)	12 glaciers
Giesen and Oerlemans (2013)	Global (excl. A&G)	T(1), P(2)	89 glaciers
Hirabayashi and others (2013)	Global	MB(2)	295 glaciers, RGI regions
Radić and others (2014)	Global	T(2), P(3), MB(2)	36 glaciers, RGI subregions
Huss and Hock (2015)	Global	T(1), P(1), MB(1)	RGI regions
Kraaijenbrink and others (2017)	HMA	P(1), MB(1)	HMA subregions

Parentheses show the number of parameters that were calibrated by each model that related to temperature (T: temperature bias, on-glacier lapse rate, off-glacier lapse rate), precipitation (P: precipitation factor, precipitation gradient, temperature threshold for liquid/solid precipitation), and/or mass balance (MB: degree-day factor of snow and/or ice, temperature sensitivity, bias correction, melt temperature threshold, mass balance gradient).
A, Antarctic periphery; G, Greenland periphery; HMA, High Mountain Asia; RGI, Randolph Glacier Inventory (RGI Consortium, 2017).

automated methods and archives of surface elevation measurements from satellite remote-sensing observations can now provide geodetic glacier mass-balance estimates on a regional to global scale (e.g., Brun and others, 2017; Shean and others, 2020). These unprecedented datasets can be used for systematic model calibration, with the ability to capture sub-regional spatial variability.

This study presents a new calibration scheme for a large-scale glacier evolution model that uses extensive geodetic mass-balance measurements with a Bayesian inverse model to determine the model parameters and their corresponding uncertainties for each glacier. The calibration scheme is applied to every glacier in High Mountain Asia (RGI Consortium, 2017). Properties of the posterior distribution, such as mean and standard deviation, are estimated using an iterative method, a form of Markov chain Monte Carlo (MCMC), to obtain samples from the distribution. Convergence diagnostics are used to assess the performance of the MCMC methods with a focus on determining an acceptable chain length that accurately estimates the model parameters while utilizing computational resources efficiently.

2. Glacier Evolution Model

The Python Glacier Evolution Model (PyGEM) (Rounce, 2019; Rounce and others, 2020) is used to model the mass change of glaciers in High Mountain Asia from 2000 to 2018. PyGEM is an open-source glacier evolution model that calculates the monthly climatic mass balance, b_{clim} (m w.e.), of each 10 m elevation bin for each glacier according to:

$$b_{\text{clim}} = a + c + R, \tag{1}$$

where a is ablation, c is accumulation, and R is refreeze (all have units of m w.e., mass losses are negative). The climatic mass balance for each elevation bin is aggregated to calculate the glacier-wide specific mass balance, B, according to:

$$B = \frac{\sum_{\text{bin}=0}^{n} b_{\text{clim,bin}}\ S_{\text{bin}}}{S}, \tag{2}$$

where S is the glacier area and *bin* refers to each elevation bin. Note the model does not account for frontal, internal, or basal ablation.

Ablation for each month, m, is calculated according to:

$$a_{\text{m,bin}} = \begin{cases} f_{\text{snow/ice/firn}} \cdot T_{\text{m,bin}} \cdot n & \text{if } T_{\text{m,bin}} \geq 0 \\ 0 & \text{if } T_{\text{m,bin}} < 0 \end{cases}, \tag{3}$$

where f is the degree-day factor of snow, ice or firn (m w.e. $\text{d}^{-1}\text{°C}^{-1}$), $T_{\text{m,bin}}$ is the monthly mean near-surface air temperature (°C), and n is the number of days in each month. $T_{\text{m,bin}}$ is calculated as a function of the temperature bias, T_{bias}, according to:

$$T_{\text{m,bin}} = T_{\text{m,GCM}} + lr_{\text{m,GCM}} \cdot (z_{\text{ref}} - z_{\text{GCM}}) + lr_{\text{m,glac}} \cdot (z_{\text{bin}} - z_{\text{ref}}) + T_{\text{bias}}, \tag{4}$$

where $T_{\text{m,GCM}}$ is the monthly temperature from the climate data based on the nearest neighbor; lr is the monthly temperature lapse rate used to account for elevation differences between the GCM's underlying topography and the glacier, $lr_{\text{m,GCM}}$, and elevation differences across the glacier, $lr_{\text{m,glac}}$; and z_{ref} is the glacier's reference (median) elevation, z_{GCM} is the elevation corresponding to $T_{\text{m,GCM}}$, and z_{bin} is the elevation of the elevation bin. In this study, $lr_{\text{m,GCM}}$ and $lr_{\text{m,glac}}$ are assumed to be equal and are derived from temperature data at various pressure levels.

Accumulation is calculated according to:

$$c_{\text{m,bin}} = \delta_{\text{m,bin}} P_{\text{m,bin}}, \tag{5}$$

where $\delta_{\text{m,bin}}$ is the monthly fraction of solid precipitation and $P_{\text{m,bin}}$ is the monthly precipitation (m w.e.). $\delta_{\text{m,bin}}$ is based on $T_{\text{m,bin}}$ and the temperature threshold (T_{snow}, °C) used to differentiate between liquid and solid precipitation as follows:

$$\delta_{\text{m,bin}} = \begin{cases} 1 & \text{if } T_{\text{m,bin}} \leq T_{\text{snow}} - 1 \\ 0 & \text{if } T_{\text{m,bin}} \geq T_{\text{snow}} + 1 \\ 0.5 + \dfrac{T_{\text{m,bin}} - T_{\text{snow}}}{2} & \text{if } T_{\text{snow}} - 1 < T_{\text{m,bin}} < T_{\text{snow}} + 1 \end{cases}. \tag{6}$$

Monthly precipitation is calculated as a function of the precipitation factor, k_{p}, according to:

$$P_{\text{m,bin}} = P_{\text{m,GCM}} \cdot k_{\text{p}} \cdot (1 + d_{\text{prec}} \cdot (z_{\text{bin}} - z_{\text{ref}})), \tag{7}$$

where $P_{\text{m,GCM}}$ is the monthly precipitation from the climate data based on the nearest neighbor and d_{prec} is the precipitation gradient (% m^{-1}).

Potential refreeze, $R_{\text{potential}}$, is calculated as a function of the weighted annual mean air temperature, T_{a} (°C), (Woodward and others, 1997) according to:

$$R_{\text{potential}} = -0.0069 \cdot T_{\text{a}} + 0.000096. \tag{8}$$

The potential refreeze is reset every October. Any melt that occurs is assumed to refreeze until the potential refreeze is exhausted. The model assumes refreeze occurs in the snow pack, as opposed to being superimposed ice, so refreeze cannot

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press

exceed the snow depth in any given month. Refreeze cannot be negative either.

Glacier dynamics, which allows the glacier to advance/retreat and alters the surface elevation, is accounted for annually by redistributing the glacier-wide mass change over each elevation bin using mass redistribution curves from Huss and Hock (2015). These curves, which were derived from 34 glaciers in the Swiss Alps (Huss and others, 2010), assume no surface lowering at the glacier's highest elevation and maximum surface lowering at the terminus.

The model requires glacier inventory and climate data. We use elevation-dependent binned glacier area and width from the RGIv6.0 (RGI Consortium, 2017) and glacier thickness data from Huss and Farinotti (2012) that have been updated to RGIv6.0. The model is forced with monthly air temperature and precipitation data from the ERA-Interim Climate Reanalysis project (Dee and others, 2011) from 2000 to 2018. ERA-Interim has a native resolution of 0.7°, which is bilinearly interpolated to a resolution of 0.5°.

2.1. Model parameters

The model parameters requiring calibration are the degree-day factor of snow (f_{snow}, mm w.e. $d^{-1}°C^{-1}$), precipitation factor (k_p, -), and temperature bias (T_{bias}, °C). The precipitation factor and temperature bias are a multiplier and additive correction that are used to correct the precipitation and air temperature data from ERA-Interim, respectively. These adjustment factors are needed to account for any biases in the climate data and downscale the data from its coarse resolution to each glacier. Additionally, these model parameters are used to account for any processes (e.g., debris cover, firn densification) that are not explicitly included in the glacier evolution model.

For the other model parameters, we either assume reasonable values or estimate them from climate data in order to reduce the number of model parameters. Specifically, the degree-day factor of snow is assumed to be 70% of the degree-day factor of ice, and the degree-day factor of firn is assumed to be the mean of the degree-day factor of snow and ice. We assume the precipitation gradient is 0.01% m^{-1} and the temperature threshold for the snow–rain transition is 1°C (Huss and Hock, 2015).

2.2. Calibration data

Geodetic mass-balance observations of 95,086 glaciers (99.6% of the total glacier area) (Shean and others, 2020) in High Mountain Asia were used to calibrate the glacier evolution model. These observations were derived from trends fit to at least five, and sometimes more than 50, digital elevation models per glacier derived from all available cloud-free Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and DigitalGlobe WorldView-1, WorldView-2, WorldView-3 and GeoEye-1 stereo images between 2000 and 2018 (Shean and others, 2016). Given the nearly complete coverage of glaciers in High Mountain Asia, the specific glacier-wide mass balances were computed as the mean elevation change rate for each glacier polygon, which was converted to mass balance using a density of 850 kg m^{-3} (Shean and others, 2020).

We quality controlled these mass balances by removing outliers using a 3-sigma filter. For 22 subregions defined by Bolch and others (2019), 954 glaciers (0.4% of the total glacier area) were identified as outliers based on a comparison of individual glacier mass-balance estimates and the regional mean. Additionally, the uncertainty associated with the mass-balance estimates of individual glaciers ranged from 0.02 to 13.2 m w.e. a^{-1}. Again, a 3-sigma filter was used to identify and remove 1401 glaciers (0.5% of the total glacier area) based on a comparison of the uncertainty associated with the individual glacier mass balance and the mean uncertainty of all the glaciers in High Mountain Asia. Glaciers that were not measured or removed as outliers were assumed to have a specific mass balance and uncertainty equal to the regional specific mass balance and the mean uncertainty associated with all the glaciers in their respective region.

The mean specific mass balance for all glaciers in High Mountain Asia from 2000 to 2018 was -0.20 ± 0.21 m w.e. a^{-1}, but varied greatly by region ranging from -0.65 ± 0.41 m w.e. a^{-1} in Hengduan Shan to 0.05 ± 0.18 m w.e. a^{-1} in Western Kunlun Shan. The uncertainty associated with the individual glacier mass-balance estimates ranged from 0.02 to 1.31 m w.e. a^{-1} with a median of 0.31 m w.e. a^{-1}. These results are consistent with Brun and others (2017) and emphasize the importance of using subregional data to calibrate large-scale glacier evolution models.

3. Bayesian inverse model

Bayesian inference provides a useful framework for assessing parameter uncertainty. Bayes' Theorem combines prior knowledge of the model parameters with observed data to create a posterior probability distribution, $p(\boldsymbol{\theta}|y)$:

$$p(\boldsymbol{\theta}|y) = \frac{p(\boldsymbol{\theta})p(y|\boldsymbol{\theta})}{p(y)}, \tag{9}$$

where $\boldsymbol{\theta}$ represents the parameter(s) of interest, and y is the data. $p(\boldsymbol{\theta})$ is the joint prior distribution for $\boldsymbol{\theta}$, which reflects what is known about $\boldsymbol{\theta}$ before any data are collected. $p(y|\boldsymbol{\theta})$ is the likelihood, which is the probability distribution of the data given a particular value of $\boldsymbol{\theta}$. $p(y)$ is the marginal distribution of the data, also called the prior predictive distribution (Gelman and others, 2014), defined as:

$$p(y) = \int p(\boldsymbol{\theta})p(y|\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}, \tag{10}$$

which can also be considered the normalization constant that makes $p(\boldsymbol{\theta}|y)$ integrate to 1. $p(\boldsymbol{\theta}|y)$ is the joint posterior distribution for $\boldsymbol{\theta}$ given the data y. The mean and variance of the joint posterior distribution provide information about the most plausible model parameters and the uncertainty associated with those parameters given the observations.

3.1. Inversion

Given the non-linear data-generating process:

$$B_{obs} = F(\boldsymbol{\theta}) + \varepsilon, \tag{11}$$

where B_{obs} is the observed specific mass balance, $F(\boldsymbol{\theta})$ is our forward model (PyGEM, see Section 2), $\boldsymbol{\theta}$ are the three model parameters: precipitation factor (k_p), temperature bias (T_{bias}), and degree-day factor of snow (f_{snow}), and ε is the error associated with the observation; we seek to determine the distribution of parameters, $\boldsymbol{\theta}$, consistent with observations. The error, ε, is assumed to be normally distributed with a mean of zero and a standard deviation, σ, derived from the processing of the geodetic mass-balance observations (Shean and others, 2020). Note that ε does not account for any uncertainty associated with how well the glacier evolution model actually represents reality as quantifying uncertainty associated with the model physics and input data is beyond the scope of this study.

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press

The likelihood function, $p(y|\boldsymbol{\theta})$, can therefore be described as:

$$p(y|\boldsymbol{\theta}) \propto \psi_1(\boldsymbol{\theta})\ \psi_2(\boldsymbol{\theta})\ \frac{1}{\sigma\sqrt{2\pi}} e^{-(B-F(\boldsymbol{\theta}))^2/2\sigma^2}, \quad (12)$$

where ψ_1 and ψ_2 are potential functions (Jordan, 2004) that impose two physical constraints. The first potential function, ψ_1, ensures the modeled mass balance is not less than the mass balance associated with completely melting the glacier, $B_{\text{max_loss}}$:

$$\psi_1(\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } F(\boldsymbol{\theta}) < B_{\text{max_loss}} \\ 1 & \text{if } F(\boldsymbol{\theta}) \geq B_{\text{max_loss}} \end{cases}. \quad (13)$$

The second potential function, ψ_2, ensures the climatic mass balance of the lowermost bin, $b_{\text{clim,bin}_{\text{lowest}}}$, is negative for at least 1 year, i.e., the glacier has an ablation area:

$$\psi_2(\boldsymbol{\theta}) = \begin{cases} 0 & b_{\text{clim,bin}_{\text{lowest}}}(\boldsymbol{\theta}) \geq 0 \text{ for every year} \\ 1 & b_{\text{clim,bin}_{\text{lowest}}}(\boldsymbol{\theta}) < 0 \text{ for at least one year} \end{cases}. \quad (14)$$

The constraints ensure the proposed sets of non-physical model parameters are rejected.

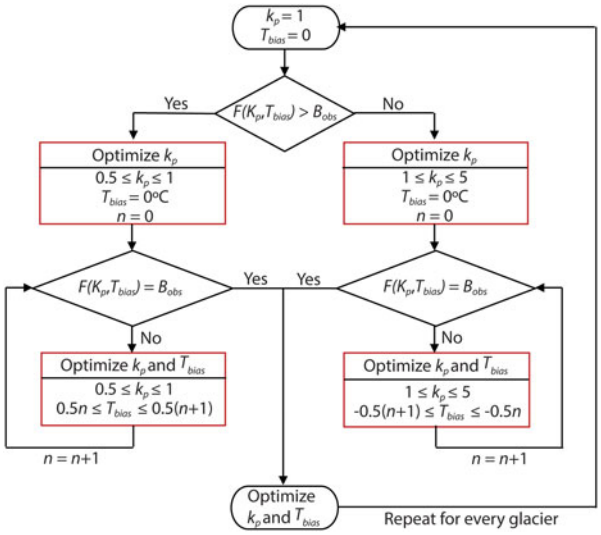


Fig. 1. Flow chart of the simple calibration scheme used to initially calibrate every glacier in order to generate the regional marginal prior distributions for the temperature bias (T_{bias}) and precipitation factor (k_{p}). The prior distribution for the degree-day factor of snow (f_{snow}) is based on Braithwaite (2008). B_{obs} refers to the observed mass balance.

3.2. Empirical Bayes and prior distributions

A preferable solution would be to employ a hierarchical model that jointly models glaciers in a (sub)region, which allows for the borrowing of information from nearby glaciers. Such a model would assume that the observations of each glacier are conditionally independent, but the parameters themselves would be assumed, through their prior distribution, to be drawn from a common underlying population (Carlin and Louis, 2008; Chapter 5). This hierarchical model could be applied to each of the 22 subregions (Bolch and others, 2019) as follows:

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\varphi}|y) &\propto p(y|\boldsymbol{\theta}, \boldsymbol{\varphi})p(\boldsymbol{\theta}, \boldsymbol{\varphi}) \\ &\propto \prod_{i=1}^{n_y}[p(y_i|\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i|\boldsymbol{\varphi})]p(\boldsymbol{\varphi}), \end{aligned} \quad (15)$$

where $\boldsymbol{\theta}_i$ is the set of model parameters for glacier i, y_i is the observation for glacier i, $\boldsymbol{\varphi}$ represents the regional prior parameters, and $p(\boldsymbol{\varphi})$ is therefore the regional joint prior distribution. Unfortunately, simultaneous inference over all glaciers is intractable.

We therefore approximate $p(\boldsymbol{\varphi})$ as $p(\boldsymbol{\varphi}|y)$ by computing a priori estimates for each glacier in each region (detailed below, Fig. 1), which is computationally cheap, and then fit normal and gamma distributions to the resulting histograms. This procedure is commonly used in mixed-effects models (such as the one in this study), where it is referred to as empirical Bayes (Efron, 2010).

The joint prior distribution reflects our knowledge of each parameter prior to any observations. Independent prior information for the degree-day factor of snow is available from glaciers from various regions around the world. Braithwaite (2008) compiled these data and found a mean ± standard deviation of 4.1 ± 1.5 mm w.e. d^{-1}°C^{-1}. Based on these data, we assume the prior distribution for the degree-day factor of snow is a truncated normal distribution with the given mean, μ, and standard deviation, σ, truncated at 0 and ∞ to ensure positivity:

$$p(f_{\text{snow}}) = N_{\text{T}}(\mu, \sigma, 0, \infty). \quad (16)$$

For the temperature bias and precipitation factor, no independent data are available due to a lack of meteorological observations, especially at high altitudes where the glaciers reside. In Central Asia, a comparison of air temperature observations with various reanalysis datasets, including ERA-Interim, found the air temperature agreed well with local observations (average absolute error between −0.6 and 1.6°C), but also identified minor differences between datasets in the air temperature trends from 1980 to 2011 in select subregions (Hu and others, 2014). Conversely, precipitation has been a great source of uncertainty in High Mountain Asia with many studies concluding that precipitation at high altitudes is highly inaccurate and that some regions need to be corrected by up to a factor of 10 (Immerzeel and others, 2015; Dahri and others, 2016; Wortmann and others, 2018).

Given this information, we first use a simple optimization scheme to fit the precipitation factor and temperature bias to the observed specific mass balance of every glacier in High Mountain Asia based on a modified version of Huss and Hock (2015) (Fig. 1). This initial optimization enables us to approximate the regional joint prior distribution, $p(\boldsymbol{\varphi})$, of the temperature bias and precipitation factor for each subregion. The optimizations are performed using the sequential least-squares minimization function from the open-source SciPy statistical package (Jones and others, 2001). Since precipitation is considered to be more uncertain than the temperature data, the scheme first attempts to optimize only the precipitation factor, and then incrementally adjusts the temperature bias bounds to ensure the precipitation factor remains within its bounds.

The optimized parameters are then regionally aggregated by the 22 subregions (Bolch and others, 2019) to develop prior distributions for each region. Specifically, the prior distribution for the temperature bias is assumed to be a normal distribution with a mean and standard deviation based on the regional mean, $\mu_{\text{reg},T_{\text{bias}}}$, and regional standard deviation, $\sigma_{\text{reg},T_{\text{bias}}}$, derived from the initial optimization of the temperature bias (Fig. S1):

$$p(T_{\text{bias}}) = \mathcal{N}(\mu_{\text{reg},T_{\text{bias}}}, \sigma_{\text{reg},T_{\text{bias}}}). \quad (17)$$

The prior distribution for the precipitation factor is assumed to be a gamma distribution, which ensures positivity, that has a

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press

shape parameter, α, and rate parameter, β, estimated from the regional mean, $\mu_{\mathrm{reg},k_\mathrm{p}}$, and standard deviation, $\sigma_{\mathrm{reg},k_\mathrm{p}}$, of the initial optimization of the precipitation factor (Fig. S2):

$$\beta = \frac{\mu_{\mathrm{reg},k_\mathrm{p}}}{\sigma_{\mathrm{reg},k_\mathrm{p}}} \tag{18}$$

$$\alpha = \beta\mu_{\mathrm{reg},k\mathrm{p}} \tag{19}$$

$$p(k_\mathrm{p}) = \Gamma(\alpha, \beta). \tag{20}$$

3.3. Markov chain Monte Carlo implementation

In practice, properties or characteristics of the posterior distribution $p(\boldsymbol{\theta}|y)$ are difficult to evaluate analytically. MCMC methods circumvent this issue by producing samples of the joint posterior distribution that can be used to estimate the center, shape, and spread of the marginal posterior distributions. Specifically, MCMC methods are based on theorems that prove if the Markov chain is run for enough steps, the chain converges to a unique stationary distribution, and the resulting samples are from the joint posterior distribution (Carlin and Louis, 2008).

The MCMC methods are implemented using the open source PyMC 2.3.7 statistical package (Fonnesbeck and others, 2014). Specifically, these methods are implemented using an adaptive Metropolis-Hastings sampling algorithm, which uses normal proposal distributions and single-component updating. To ensure the Markov chain mixes well, the chain can be tuned throughout. This is done by using a multiplicative factor to adjust the standard deviation associated with the proposal distribution of each parameter with the goal of adjusting the proposal distribution such that the acceptance rate is between 20 and 50% (Fonnesbeck and others, 2014). For our model, the Markov chains were tuned every 1,000 steps. Note these methods are computationally expensive because the likelihood function requires the forward model to be evaluated many times.

3.4. Identifiability

Each glacier has one observation and a corresponding estimate of uncertainty (i.e., mean and standard deviation of the glacier-wide specific mass balance), and three model parameters (precipitation factor, temperature bias, and degree-day factor of snow). Given how each model parameter affects the forward model (Fig. 2), the lack of observations results in there being an infinite number of parameter sets that will produce an exact match between the modeled and observed mass balance, i.e., the model is over-parameterized. For example, for glacier RGI60-13.26360, the modeled and observed mass balance agree if the temperature bias, precipitation factor, and degree-day factor of snow are −2°C, 1, 4.1 mm w.e. $\mathrm{d}^{-1}\mathrm{°C}^{-1}$; 0°C, 3.1, 2.6 mm w.e. $\mathrm{d}^{-1}\mathrm{°C}^{-1}$; or 0°C, 7.4, 5.6 mm w.e. $\mathrm{d}^{-1}\mathrm{°C}^{-1}$, respectively. Given our over-parameterized problem, we must consider identifiability and its implications for large-scale glacier evolution modeling.

Identifiability refers to how much information is gained about the individual components of $\boldsymbol{\theta}$ given the data y. Gelfand and Sahu (1999) provide a formal definition of identifiability for a simple Bayesian model with a likelihood function $p(y|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is partitioned into $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. $\boldsymbol{\theta}_2$ is said to be non-identifiable if

$$p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, y) = p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1). \tag{21}$$

In other words, $\boldsymbol{\theta}_2$ is non-identifiable if 'observing data y does not increase our prior knowledge about $\boldsymbol{\theta}_2$ given $\boldsymbol{\theta}_1$' (Gelfand and Sahu, 1999). Mathematically one can prove that $\boldsymbol{\theta}_2$ is non-identifiable if the likelihood function can be reparameterized such that the likelihood function is free of $\boldsymbol{\theta}_2$. In practice, analytically proving a parameter in a complex model is non-identifiable is difficult, sometimes nearly impossible, and instead non-identifiability may be detected intuitively (Eberly and Carlin, 2000) or empirically (Renard and others, 2010).

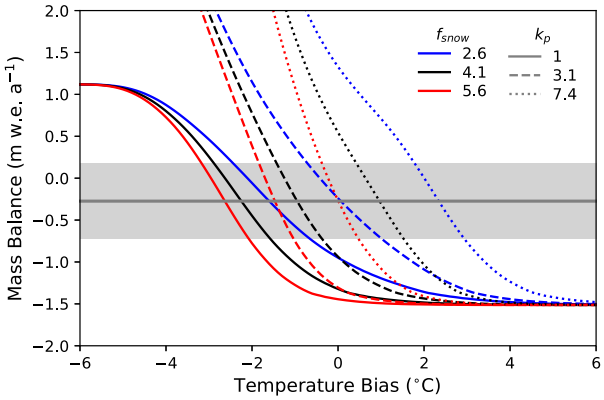


Fig. 2. The modeled mass balance for glacier RGI60-13.26360 as a function of the temperature bias (°C), using three different precipitation factors (k_p, -) and degree-day factors of snow (f_{snow}, mm w.e. $\mathrm{d}^{-1}\mathrm{°C}^{-1}$). The observed mass balance ± 3 standard deviations is shown by the horizontal gray line and shading, respectively.

Note that identifiability relies only on the likelihood function and is independent of the joint prior distribution (Renard and others, 2010). For our likelihood function, which includes the glacier mass balance computed using PyGEM, each model parameter clearly affects the mass balance (Fig. 2). Given that we only have a single mass-balance observation that provides no information as to which model parameters need to be modified, i.e., any model parameter or combination of model parameters can be adjusted such that the modeled mass balance agrees with the observation, we can intuitively determine that each model parameter is clearly non-identifiable (Renard and others, 2010).

3.5. Chain length and convergence diagnostics

One of the limitations of MCMC methods is the inability to know when the chain has converged. Therefore, we rely on three diagnostics (Gelman–Rubin statistic, Monte Carlo error, and effective sample size) to provide us with confidence that the chain has likely converged to its stationary distribution. This enables us to confidently sample from the joint posterior distribution and use ordinary estimators to evaluate the model parameters and their respective uncertainty. In an ideal setting, multiple long chains would be run for each glacier such that chain convergence could be assessed using a suite of diagnostics. However, this quickly becomes computationally prohibitive since each step in the chain requires the mass balance to be computed by PyGEM. Therefore, we need to determine the acceptable number of steps required for us to be confident that the chain has converged while using our computational resources efficiently.

For each of the three RGI regions in High Mountain Asia, 1,000 glaciers were randomly selected to determine the appropriate number of steps required for the Markov chain to converge. For each of these glaciers, three chains of 25,000 steps using over-dispersed starting points (one at the center and one at each of the end points of the 95% confidence interval) were generated. The overdispersed starting points helped determine if the chains converged to the unique stationary distribution. The amount of burn-in, i.e., the number of steps discarded at the start of the chain, was also investigated.

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press

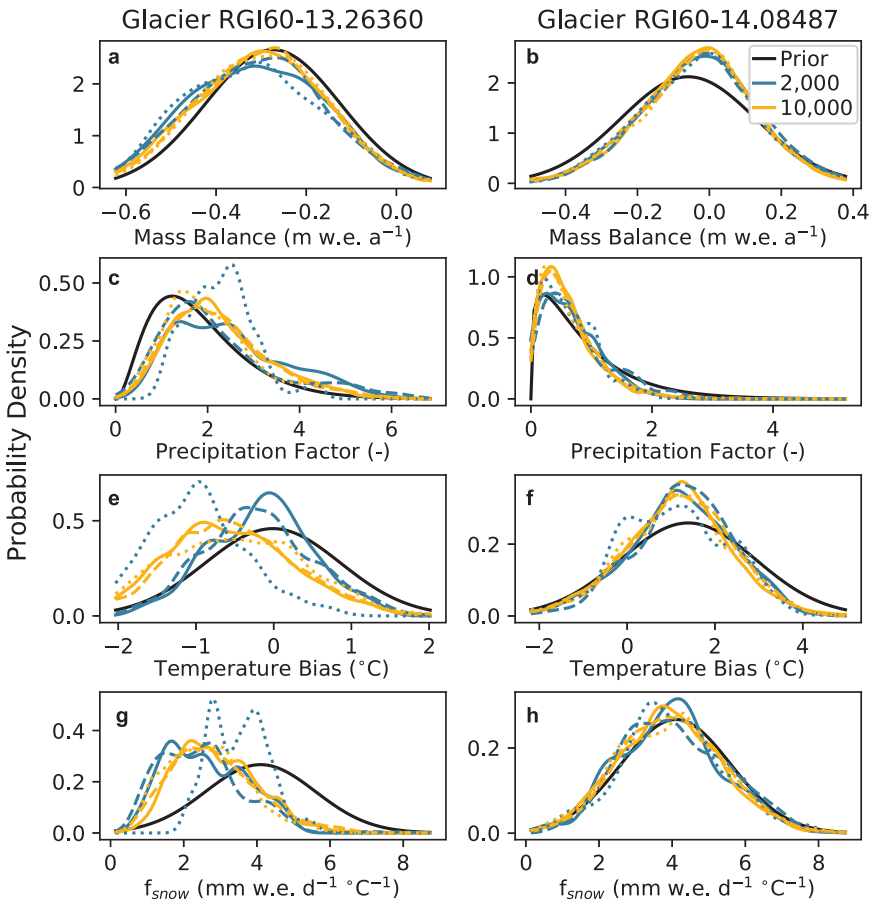


Fig. 3. Observed and predictive posterior distribution for the mass balance (a, b) along with prior and posterior distributions for the precipitation factor (c, d), temperature bias (e, f), and degree day factor of snow (f_{snow}) (g, h) for glaciers RGI60-13.26360 (left column) and RGI60-14.08487 (right column) showing convergence of chains with 2,000 and 10,000 steps (first 1,000 steps discarded as burn-in) and overdispersed starting points.

The Gelman–Rubin statistic uses multiple chains to compare the variance within each chain to the variance between chains for each parameter (Gelman and others, 2014). If the within-chain and between-chain variances are significantly different, the chain has likely not converged; while if they are similar, the chain has potentially converged. An upper threshold of 1.1 is used to assess possible convergence (Flegal and others, 2008; Gelman and others, 2014). Since successive steps from the Metropolis-Hastings algorithm are inherently correlated, the effective sample size is used to estimate the number of independent samples generated for each parameter. A threshold of 100 is used to ensure enough independent samples have been generated (Gelman and others, 2014) such that we can be confident in estimating 95% confidence intervals for the model output.

The Monte Carlo error is an estimate of the simulation error, which does not indicate chain convergence, but rather provides information regarding the quality of the reported ordinary estimators such as the mean and standard deviation (Flegal and others, 2008). We normalize the Monte Carlo error by the standard deviation of the posterior predictive distribution for the mass balance or the marginal posterior distribution for each model parameter. A high value indicates the uncertainty associated with the ordinary estimator is dominated by simulation error, while a low value indicates the uncertainty associated with the ordinary estimator is informed by the data and priors. We set a target threshold of 10%, but acknowledge that a high value is not necessarily cause for concern since it could reflect that the spread in the posterior is very small. The Monte Carlo error is computed using overlapping batch means with a batch size equal to the square root of the chain length.

The acceptable chain length is considered to be the value at which 90% of the glaciers pass the threshold associated with each diagnostic. A single chain with this acceptable number of steps is then run for each glacier in High Mountain Asia. These single chains are evaluated using only the Monte Carlo error and the effective sample size, since the Gelman–Rubin statistic requires multiple chains. This ensures that the subset of glaciers used to estimate the chain length is representative of the greater region.

4. Results

4.1. Chain length

Three chains of 25,000 steps for 1,000 glaciers in each of the three major RGI regions in High Mountain Asia were used to evaluate the performance of the MCMC method and determine the acceptable number of steps that should be used when calibrating all the glaciers. Figure 3 shows an example of how increasing the chain length improves convergence for two example glaciers. The posterior predictive distributions of the mass balance for both glaciers show that after 10,000 steps, the distributions agree well with the center, shape, and spread of the observations (Figs 3a, b). This good agreement is typical for most glaciers and expected since the model is over-parameterized. Glaciers with poor agreement indicate that the posterior predictive distribution gains more information from the joint prior distribution than the data (e.g., Figs S3–S5), which could be the result of an inappropriate joint prior distribution, poor model physics, and/or erroneous data (see Section 5.1). More importantly, the 10,000 step chains converge to the same distribution regardless of the overdispersed starting point, while the 2,000 step chains clearly have not converged. The posterior distributions of the model parameters also show the chains roughly converge after 10,000 steps and become smoother since the chains comprise more independent samples.

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press

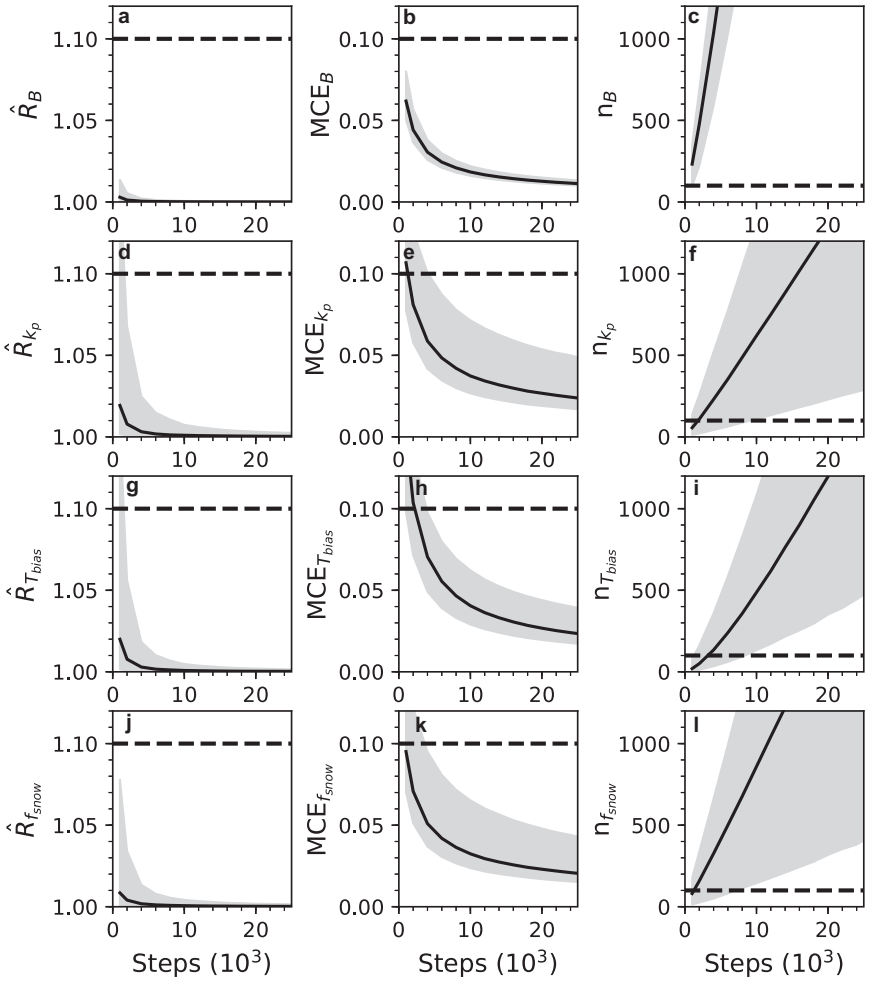


Fig. 4. The convergence metrics (Gelman–Rubin statistic, $\hat{R}$; Monte Carlo error, MCE; and effective sample size, n) as a function of the chain length for the mass balance, B (a–c), precipitation factor, k_p (b–f), temperature bias, T_{bias} (g–i), and degree-day factor of snow, f_{snow} (j–l). The line and fill represent the median and 80% credibility intervals (where 90% pass the given threshold) for the subset of 3,000 test glaciers. The dashed line shows the target value for each metric. The Monte Carlo error is normalized by the standard deviation of the posterior distribution.

Chain convergence for the 3,000 glaciers was quantitatively assessed using the Gelman–Rubin statistic, Monte Carlo error, and effective sample size. Figure 4 shows the 80% credibility intervals for each metric, i.e., the value for which 90% of the glaciers pass a given threshold. As the chain length increases, the Gelman–Rubin statistic and Monte Carlo error decrease while the effective sample size increases. The Gelman–Rubin statistic indicates the chains converge quickly. The Monte Carlo error is well below 10%, which indicates that the uncertainty associated with the mass balance and model parameters is dominated by the data as opposed to the numerical error. The effective sample size for the mass balance is an order of magnitude greater than the individual model parameters meaning the autocorrelation for the model parameters as a group is low, but much higher for each parameter independently. The precipitation factor and degree-day factor of snow generate the least number of independent samples and therefore dictate how long the chains should be.

For a chain length of 10,000 steps, 90% of the glaciers had an effective sample size >100 for the mass balance and all model parameters (Fig. 4). Therefore, a chain length of 10,000 steps was selected for all the glaciers in High Mountain Asia. The number of burn-in steps was found to have a negligible impact on the chain convergence metrics, and more-or-less only reduced the effective sample size by the percentage of steps discarded. Therefore, if using these model parameters to run simulations, the initial 2% of the chain should be discarded (Geyer, 1992).

For all the glaciers in High Mountain Asia, convergence was assessed using only the Monte Carlo error and effective sample size (Fig. S6). The normalized Monte Carlo error for which 90% of all the glaciers in High Mountain Asia were below was 2%, 7%, 6%, and 6% for the mass balance, precipitation factor, temperature bias, and degree-day factor for snow, respectively, thereby confirming that the uncertainty of the posterior distributions is driven by the data and not the numerical error. Furthermore, more than 90% of all glaciers had an effective sample size greater than 1915, 115, 177, and 167 for the mass balance, precipitation factor, temperature bias, and degree-day factor of snow, respectively, which confirms that a chain length of 10,000 steps generates a sufficient number of independent samples.

4.2. Spatial distribution of mass balance and model parameters

As expected for an over-parameterized problem, the mean mass balance of the predictive posterior distributions of -19.73 Gt a^{-1} agrees well with the observed mass balance of -19.36 Gt a^{-1} for all glaciers in High Mountain Asia from 2000 to 2018. The mean and observed mass balance agree well both spatially across High Mountain Asia and for larger glaciers (Fig. 5). We use the z-score, defined as the difference between the mean modeled and observed mass balance normalized by the standard deviation of the mass-balance observation, to assess the agreement between the modeled mass balance and the observation of each glacier. For 99% of larger glaciers (>5 km^2), the modeled mass balance is within 1 standard deviation of the observed mass balance (absolute z-score <1). For smaller glaciers (<5 km^2), 3% of them have larger discrepancies (absolute z-score

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press

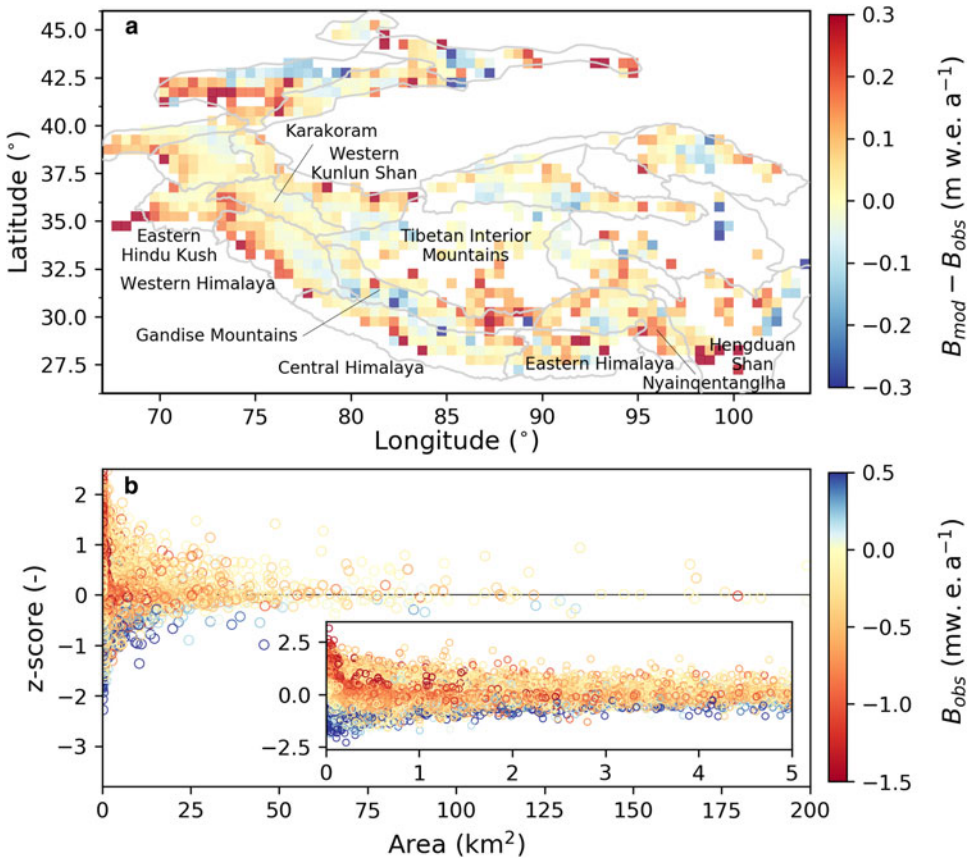


Fig. 5. (a) The difference between the mean of the posterior predictive distribution (B) and observed specific mass balance (B_{obs}) showing the spatial distribution aggregated to 0.5° grids, and (b) the z-score of the mean of the posterior predictive distribution as a function of glacier area for every glacier in High Mountain Asia. The inset shows a zoomed in view of the smaller glaciers. Gray outlines show 22 subregions from Bolch and others (2019).

>1) indicating that their posterior predictive distributions are being informed more by the joint prior distribution than the data. This is particularly apparent for small glaciers with very positive (>0.5 m w.e. a^{-1}) or negative (<−1.5 m w.e. a^{-1}) observed mass balances (Fig. 5b inset).

The spatial distribution of the mean model parameters shows significant regional differences along the front of the Hindu Kush Himalaya and in the interior around the Karakoram and Western Kunlun Shan (Fig. 6). The mean precipitation factor shows a strong orographic effect with precipitation factors much greater than 1 on the windward (south western) side of the Hindu Kush Himalaya and near or less than 1 on the leeward side. Precipitation factors are also high across the Nyainqentanglha, Gangdise Mountains, and Tibetan Interior Mountains. The mean temperature bias is near zero for most of High Mountain Asia with the exception of positive values in the Karakoram and Western Kunlun Shan and negative values in Nyainqentanglha. Similar to the precipitation factor, the degree-day factor of snow had lower values on the windward side of the Hindu Kush Himalaya and higher values on the leeward side; although the spatial variation across the rest of High Mountain Asia was fairly non-homogeneous.

The abrupt change in the precipitation factor at the front of the Hindu Kush Himalaya suggests the resolution of the ERA-Interim climate data (~0.7°) is too coarse to resolve spatial variations in climate caused by the steep topography. The lowest precipitation factors and highest temperature biases are found in the interior in the Karakoram and Western Kunlun Shan (Figs 6a, b). Interestingly, the Karakoram and Western Kunlun Shan have the most positive mass balances, so one might expect the precipitation factors to be higher in these regions. The low precipitation factors and high temperature biases suggest that either the climate data have significant biases or the model physics (e.g., debris cover, glacier dynamics, firn densification, response time to climate forcing) may not be properly accounted for in this region. Hence, caution should be used when interpreting spatial patterns in these model parameters.

4.3. Identifiability

Each of the three model parameters is non-identifiable because the model is over-parameterized and the data do not provide any information as to which parameter needs to be modified in order for the modeled mass balance to agree with the observations. However, the data do provide information as to how the parameters must be modified as a group in order for the modeled mass balance to agree (Fig. 2), so the mass balance is well-estimated. Histograms of the correlation between the mass balance and model parameters for the 10,000 step chains of each glacier support the fact that the data informed the model parameters as a group, since each model parameter was only weakly correlated with the mass balance (Fig. 7).

The correlation histograms also show the extent to which the various parameters compensate for one another. The precipitation factor and temperature bias were the most correlated parameters with a mean correlation coefficient of 0.39, followed by the precipitation factor and degree-day factor of snow with a mean of 0.33. The positive correlation reflects the redundancy caused by the over-parameterization, e.g., increasing the precipitation factor increases the accumulation, which can be compensated for by increasing the melt via increasing the temperature bias or degree-day factor of snow. Conversely, the negative correlation between

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press

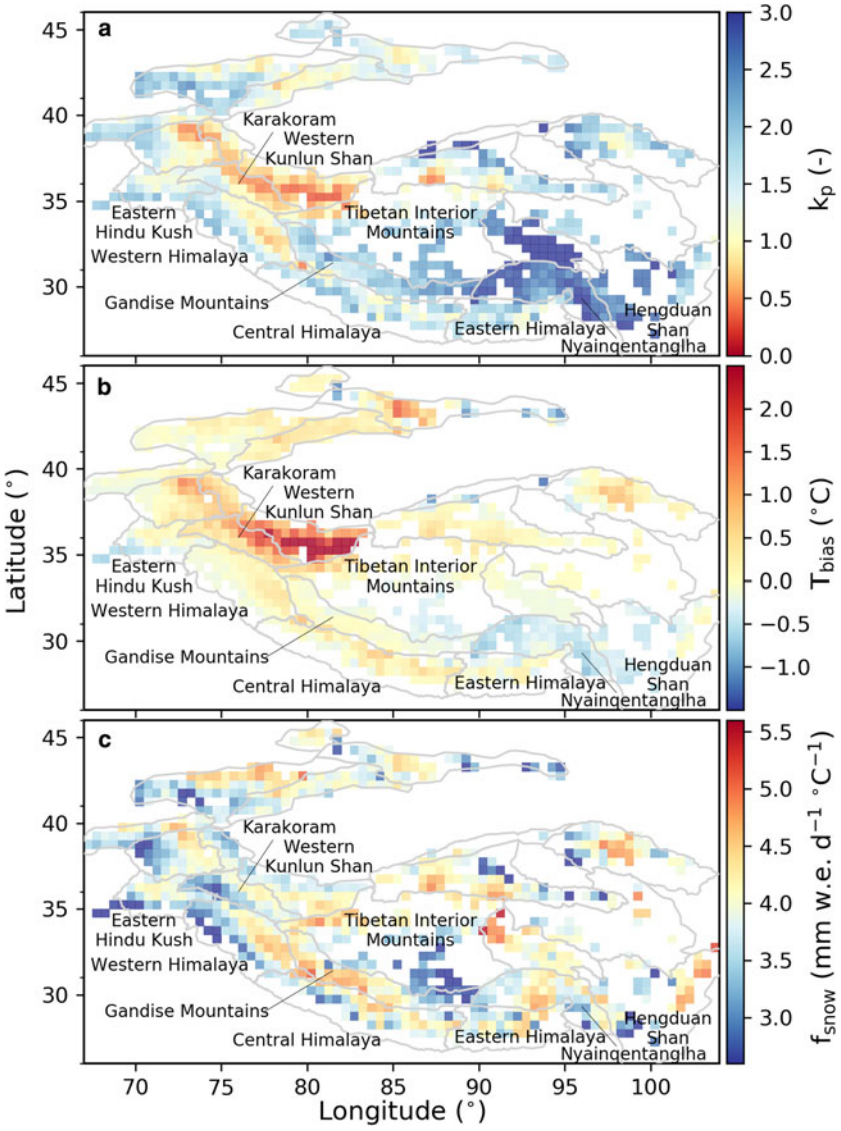


Fig. 6. Spatial distribution of the glacier area-weighted mean (a) precipitation factor (k_p), (b) temperature bias (T_{bias}), and (c) degree-day factor of snow (f_{snow}) over High Mountain Asia aggregated by 0.5°. Gray outlines show 22 subregions from Bolch and others (2019).

the temperature bias and degree-day factor of snow is expected as they both affect ablation.

The comparison of the marginal prior and posterior distributions shows how much the data informed each model parameter (Fig. 8). For all three parameters, the differences for the standard deviation were typically negative meaning the uncertainty associated with the model parameters was reduced. As for the mean, only 17%, 25%, and 17% of all glaciers were near zero for the precipitation factor, temperature bias, and degree-day factor of snow, respectively, which suggests that the center of the distributions was also informed by the data. The fact that there was no consistent positive or negative shift in the mean is not surprising, since regional prior distributions were used.

5. Discussion

5.1. Model performance for small glaciers

The posterior predictive distributions of the modeled mass balance derived from the MCMC methods are informed by a combination of the data and joint prior distribution. Given the lack of precipitation and temperature data at high altitude that could be used to develop prior distributions for the precipitation factor and temperature bias, we developed regional priors based on a modified version of the calibration scheme from Huss and Hock (2015) (Fig. 1). Overall, the Bayesian model generates posterior predictive distributions that more or less agree with the observed data (Figs 5, 8a, b), which indicates the parameters are being informed by the data. This is ideal because we typically have more confidence in the data than in the priors. However, for many small glaciers (<5 km^2) with very negative or positive mass balances, the posterior distribution is primarily informed by the priors. This also occurs for 1% of larger glaciers. Given the model is over-parameterized, these glaciers are particularly interesting because they indicate that (a) the priors are inappropriate, (b) model physics are poorly accounted for, or (c) there are issues with the data.

One of the issues with the smallest glaciers (<0.1 km^2) is that they approach the minimum threshold for the methods used to estimate the geodetic mass balance (Shean and others, 2020). Since there are fewer pixels, these glaciers are inherently more prone to larger errors, which is reflected in the uncertainty in the mass-balance observations, i.e., smaller glaciers have higher standard deviations. Furthermore, for 1,960 glaciers, all of which are <1.5 km^2, the observed mass balance exceeds the maximum modeled mass loss, i.e., the mass balance associated with completely melting the glacier based on the initial glacier area and ice thickness estimates. Given the limited glacier area and large uncertainties associated with the ice thickness estimates, which can underestimate the ice thickness by more than 60%

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press

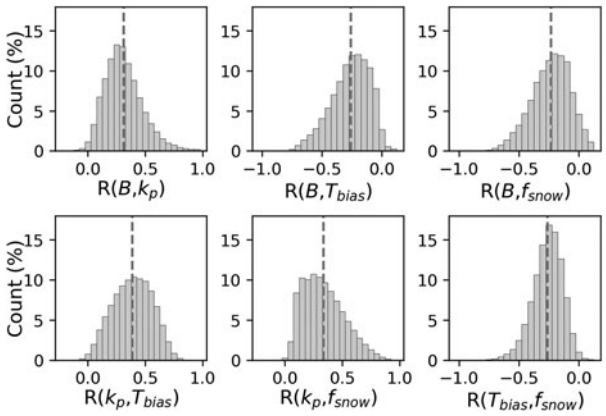


Fig. 7. Histograms of the correlation coefficient (R) between all combinations of the modeled mass balance (B), precipitation factor (k_p), temperature bias (T_{bias}), and degree-day factor of snow (f_{snow}) for all glaciers in High Mountain Asia. The dashed line shows the mean correlation coefficient.

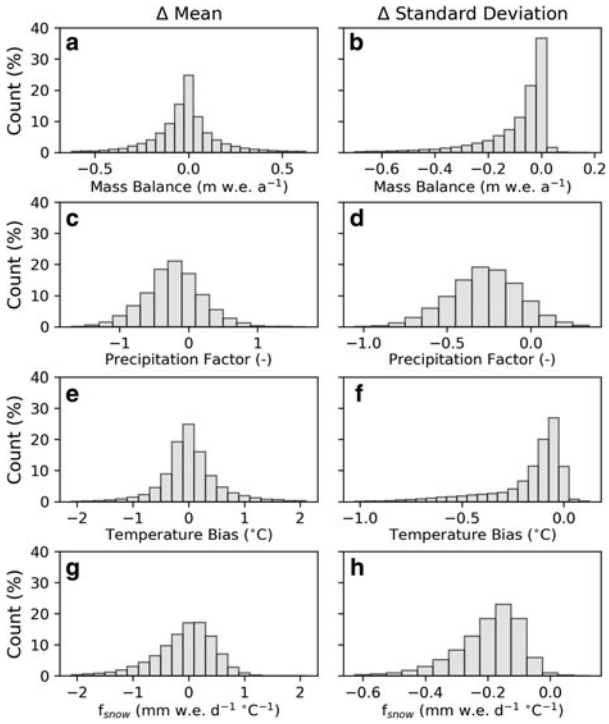


Fig. 8. Histograms of the change (Δ, posterior–prior) in the mean and standard deviation of the predictive posterior distribution and observed mass balance (a, b) and the marginal posterior and prior distributions for the precipitation factor (c, d), temperature bias (e, f), and degree-day factor of snow (f_{snow}) (g,h) for all glaciers in High Mountain Asia.

(Farinotti and others, 2017), this issue may be due to the total glacier volume being underestimated. Alternatively, the observed mass balance may have an error or some combination of both. Since there are limited measurements of ice thickness in High Mountain Asia, the cause of this discrepancy is difficult to assess. Nonetheless, this discrepancy is important to note because the potential function explicitly prevents the sets of model parameters that completely melt the glacier from being accepted, so these glaciers are forced to be informed by the prior distributions instead of the data (e.g., Fig. S3).

Small glaciers with very positive mass balance could also be the result of errors in the mass-balance observations, so having the posterior distributions be informed by the priors could be beneficial. However, Shean and others (2020) found that the smaller glaciers in each region typically had more positive mass balances, which they hypothesized was due to their higher elevations, occupying protected alcoves, or having higher amounts of accumulation due to wind redistribution and/or avalanching. Since PyGEM does not explicitly model solar radiation or account for redistribution of snow due to wind or avalanching, both of which would cause the mass balance to be more positive, the model could have a negative bias for these particular glaciers. Once again, the lack of information gained by the data could be due to errors with the data or model physics. Regardless, since the glacier model is meant for large-scale applications and trends in glacier mass change and runoff are primarily driven by the larger glaciers, the potential issues associated with smaller glaciers are minimal.

Another interesting case of glaciers to examine are those that may have inappropriate prior distributions, i.e., they may significantly deviate from the regional prior distributions. Glacier RGI60-15.10755 (32 km^2) is a good example of a glacier where agreement with the observed mass balance is unable to be reached because the regional prior distribution for the temperature bias prevents the temperature bias from accepting higher values (Fig. S4). Conversely, for glacier RGI60-15.12457 (5.4 km^2), the temperature bias needs to be more negative, but the regional prior distribution prevents this (Fig. S5). Once again, these glaciers call into question if the prior distribution is inappropriate, if the model physics are properly accounted for, or if there is some error with the data. Since these glaciers are larger and have more of an impact on the regional trends in mass balance and glacier runoff, these glaciers should be flagged to identify the potential causes of these discrepancies. This could be implemented by comparing the modeled and observed total mass change (Gt) and flagging glaciers that exceed a user-specified threshold.

5.2. Implications for glacier evolution models

The MCMC methods are an important advance for calibrating large-scale glacier evolution models because they generate distributions of viable parameters based on present-day mass-balance observations that can be used for prognostic simulations. Since the parameter sets were generated using Monte Carlo methods, any simulations must also be done in a Monte Carlo way as well, i.e., the joint posterior distribution should be used to generate a suite of viable model parameters. Using a suite of model parameters will enable the simulations to capture the uncertainty associated with the model parameters based on the observations and properly account for any issues regarding non-identifiability.

One important question to consider is how non-identifiability affects projections of glacier mass change and runoff. Since the model parameters are non-identifiable, the joint posterior distribution will contain different combinations of model parameters that result in equal (or near equal) values of the mass balance. For example, consider two viable sets of model parameters that cause the mass balance to agree with the observation: the first is a wetter and warmer set, i.e., a high precipitation factor compensated by a high temperature bias, and the second is a dryer and cooler set, i.e., a low precipitation factor compensated by a low temperature bias. Present-day glacier mass change will be the same and projections may also be similar, although there may be minor differences that are caused by how the glacier hypsometry impacts the glacier retreat. Conversely, the implications for glacier runoff are likely to be significant both for present-day and future simulations. The wetter and warmer set will generate more precipitation and melt resulting in more glacier runoff, while the dryer and cooler set will result in substantially less glacier runoff.

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press

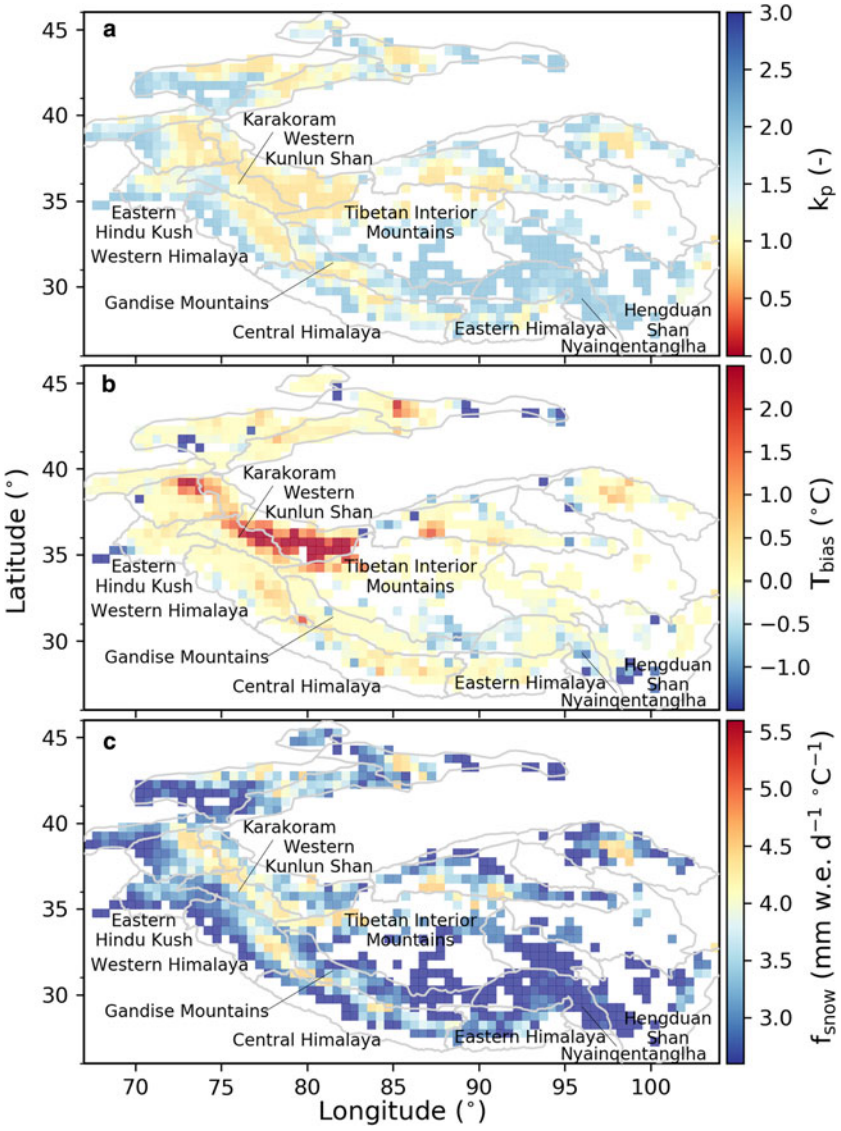


Fig. 9. Spatial distribution of the area-weighted mean (a) precipitation factor (k_p), (b) temperature bias (T_{bias}), and (c) degree-day factor of snow (f_{snow}) over High Mountain Asia aggregated by 0.5° using the calibration scheme of Huss and Hock (2015). Gray outlines show 22 subregions from Bolch and others (2019).

Similarly, non-identifiability is important to consider for studies that have used glacier models to infer biases in the temperature and precipitation data (e.g., Immerzeel and others, 2015). If the parameters in the model are non-identifiable, then caution must be used in interpreting the results. Given the range of applications and socioeconomic importance of projecting changes in glacier mass change and runoff, identifying these limitations is important for driving future work. Specifically, the issue of non-identifiability could be reduced by systematic observations of temperature and precipitation at high altitude, snowline altitudes, and/or elevation-dependent observations of the climatic mass balance, since these observations would provide parameter-specific information.

5.3. Comparison with previous studies

The MCMC method developed in this study showcases the amount of information that can be extracted from having a complete survey of all the glaciers in a region. Compared to previous studies that relied on 12–295 mass-balance observations to calibrate models globally (Table 1), the measurement of more than 95,000 glaciers in High Mountain Asia alone is unprecedented and effectively removes any uncertainty caused by the transfer of model parameters to glaciers without observations. This is also a marked improvement over studies that relied on regional mass-balance measurements (Huss and Hock, 2015; Kraaijenbrink and others, 2017) as the level of detail enables the model to be calibrated accurately on a sub-regional level.

The MCMC methods also enable the uncertainty associated with the model parameters to be quantified based on the uncertainty associated with the mass-balance observations. This is a major improvement compared to previous studies that relied on sensitivity tests or used the uncertainty associated with the GCMs and RCPs to broadly estimate the uncertainty associated with projections. Lastly, the MCMC methods enable the model to explore the solution space thereby providing insight into how the various parameters affect one another in this over-parameterized problem. While the methods developed here do not solve the over-parameterization issue that affects all large-scale glacier evolution models, they do provide a framework for understanding parameter uncertainty and for integrating more observations in the future.

Since previous studies were also over-parameterized, we provide an example of the spatial variation of model parameters derived using the calibration procedure from Huss and Hock (2015). The calibrated mass balances from Huss and Hock (2015) agree fairly well with observations (Fig. S7). Figure 9 shows the spatial trends for the precipitation factor and temperature bias are similar to those in our study (Fig. 6), which is unsurprising given the model physics and input data are nearly identical. However, the Huss and Hock (2015) methods assume

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press

the precipitation factor to be highly constrained (0.8–2.0). Since the degree-day factor of snow is the second parameter that is modified in their calibration scheme, this parameter gains a great deal of information from the data. If needed, they also adjust the temperature bias to address any remaining discrepancies between the model and the observations. This method has two major shortcomings: (1) the constraints placed on the precipitation factor do not reflect how underestimated the precipitation is in High Mountain Asia (Immerzeel and others, 2015; Dahri and others, 2016; Wortmann and others, 2018), and (2) glaciers with very positive mass balances that require precipitation factors >2 will not find agreement with the observations. However, given that both models are over-parameterized, one cannot definitively conclude that their sets of model parameters are more or less accurate than those generated by our calibration scheme. The major difference is that their set of model parameters is simply one viable combination, while our calibration scheme produces more than 100 viable combinations for each glacier.

6. Conclusions

This study developed a new calibration scheme using a Bayesian model (fitted using MCMC methods) and geodetic mass-balance observations of more than 95,000 glaciers to calibrate every glacier in High Mountain Asia. The convergence diagnostics, which include the Gelman–Rubin statistic, Monte Carlo error, and effective sample size, suggest that a chain length of 10,000 steps converges and produces an acceptable number of independent samples such that properties of each model parameter, e.g., mean and standard deviation, can be estimated from the joint posterior distribution. The resulting calibrated parameters provide an unprecedented level of spatial detail and can be used to quantify the uncertainty associated with model parameters in future projections.

The model parameters also suggest that the resolution of climate data is too coarse to resolve variations in the temperature and precipitation at high altitudes. However, given that the model is still over-parameterized, caution must be used to avoid over-interpreting spatial distributions of the model parameters since these parameters may also be accounting for the aspects of the model physics that are not properly accounted for. Observations of temperature and precipitation at high altitudes and higher resolution climate models could provide important information concerning biases in climate data. Future work should also seek to improve our understanding of various processes that are currently missing or poorly constrained in large-scale glacier evolution models, e.g., debris cover, firn densification, or glacier dynamics.

The acquisition of additional measurements that may provide temporal and/or elevation-dependent information could also be used to solve the over-parameterization issue. These observations could include snowline altitudes, gravimetric measurements, or improved geodetic mass-balance measurements that are able to provide additional temporal and/or elevation-dependent estimates of mass balance. Fortunately, the Bayesian model presented in this study is well-suited to ingest these additional observations and their respective uncertainties. These continued advances with respect to large-scale observations and model physics will greatly improve model projections and provide confidence that the spatial distribution of model parameters reflect true biases in the climate data.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/jog.2019.91

Acknowledgements. This work was supported in part by the high-performance computing and data storage resources operated by the Research Computing Systems Group at the University of Alaska Fairbanks Geophysical Institute. DR and RH received support from NASA-ROSES program grants NNX17AB27G and 80NSSC17K0566. TK was supported by Research Experience for Undergraduates NSF grant #1560372. DS was supported by NASA-ROSES program grant NNX16AQ88G. Matthias Huss provided ice thickness data.

Author contributions. DR led the study, performed the calculations and wrote the paper with support from an initial draft by TK. DR and TK made the figures. DR, MS, TK and RH developed the methodology and design of the study. MS and DB provided substantial support with Bayesian models. TK setup the initial code integrating PyMC with PyGEM and DR completed the code. RH initiated the study. DS provided the mass-balance data. All authors commented on the manuscript.

Data availability statement. The model code for this study is publicly available at https://github.com/drounce/PyGEM (PyGEMv0.1.0).

References

Bliss A, Hock R and Radić V (2014) Global response of glacier runoff to twenty-first century climate change. *Journal of Geophysical Research Earth* **119**, 717–730. doi: 10.1002/2013JF002931.

Bolch T (2019) Status and change of the cryosphere in the extended hindu kush himalaya region. In Wester P Mishra A Mukherji A and Shrestha AB ed. *The Hindu Kush Himalaya Assessment: Mountains, Climate Change, Sustainability and People*. ICIMOD, Cham, Switzerland, 209–255. doi: 10.1007/978-3-319-92288-1.

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.

Carlin BP and Louis TA (2008) *Bayesian Methods for Data Analysis*, 3rd Edn. Boca Raton: Taylor & Francis.

Dahri ZH and 5 others (2016) An appraisal of precipitation distribution in the high-altitude catchments of the Indus basin. *Science of the Total Environment* **548-549**, 289–306. doi: 10.1016/j.scitotenv.2016.01.001.

Dee D and 35 others (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137**(656), 553–597. doi: 10.1002/qj.828.

Eberly LE and Carlin BP (2000) Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *STAT MedCare* **19** (17-18), 2279–2294. doi: 10.1002/10970258(20000915/30)19:17/18<2279::AID-SIM569>3.0.CO;2-R).

Efron B (2010) *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press, Cambridge, United Kingdom.

Farinotti D and 36 others (2017) How accurate are estimates of glacier ice thickness? Results from ITMIX, the Ice Thickness Models Intercomparison eXperiment. *Cryosphere* **11**, 949–970. doi: 10.5194/tc-11-949-2017.

Flegal JM, Haran M and Jones GL (2008) Markov chain Monte Carlo: can we trust the third significant figure? *Statistical Science* **23**(2), 250–260. doi: 10.1214/08-STS257.

Fonnesbeck C, Patil A, Huard D and Salvatier J (2014) PyMC 2.3.7rc1 documentation. https://pymc-devs.github.io/pymc/index.html.

Gelfand EA and Sahu SK (1999) Identifiability, improper priors and Gibbs sampling in generalized linear mixed models. *Journal of the American Statistical Association* **94**(445), 247–253. doi: 10.1080/01621459.1999.10473840.

Gelman A and 5 others (2014) *Bayesian Data Analysis*, 3rd Edn. Boca Raton: Taylor & Francis, ISBN 9781439840955.

Geyer CJ (1992) Practical Markov chain Monte Carlo. *Statistical Science* **7**(4), 473–483.

Giesen RH and Oerlemans J (2013) Climate-model induced differences in the 21st century global and regional glacier contributions to sea-level rise. *Climate Dynamics* **41**(11–12), 3283–3300. doi: 10.1007/s00382-013-1743-7.

Hirabayashi Y, Koirala S, Kanae S, Watanabe S and Zang Y (2013) Projection of glacier mass changes under a high-emission climate scenario using the global glacier model HYOGA2. *Hydrological Research Letters* **7**(1), 6–11. doi: 10.3178/hrl.7.6.

Hock R and 7 others (2019) GlacierMIP – a model intercomparison of global-scale glacier mass-balance models and projections. *Journal of Glaciology* **65**(251), 453–467. doi: 10.1017/jog.2019.22.

Hu Z, Zhang C, Hu Q and Tian H (2014) Temperature changes in central Asia from 1979 to 2011 based on multiple datasets. *Journal of Climate* **27** (3), 1143–1167. doi: 10.1175/JCLI-D-13-00064.1.

Huss M and Farinotti D (2012) Distributed ice thickness and volume of all glaciers around the globe. *Journal of Geophysical Research-Earth* **117**(4), 1–10. doi: 10.1029/2012JF002523.

Huss M and Hock R (2015) A new model for global glacier change and sea-level rise. *Frontiers in Earth Science* **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**(2), 135–140. doi: 10.1038/s41558-017-0049-x.

Huss M, Jouvet G, Farinotti D and Bauder A (2010) Future high-mountain hydrology: a new parameterization of glacier retreat. *Hydrology and Earth System Sciences* **14**, 815–829. doi: 10.5194/hess-14-815-2010.

Immerzeel WW, Wanders N, Lutz AF, Shea JM and Bierkens MF (2015) Reconciling high-altitude precipitation in the upper Indus basin with glacier mass balances and runoff. *Hydrology and Earth System Sciences* **19**(11), 4673–4687. doi: 10.5194/hess-19-4673-2015.

Jones E and 32 others (2001) SciPy: Open source scientific tools for Python. Retrieved from http://www.scipy.org/.

Jordan MI (2004) Graphical models. *Statistical Science* **19**(1), 140–155.

Kraaijenbrink PD, Bierkens MF, Lutz AF and Immerzeel WW (2017) Impact of a global temperature rise of 1.5 degrees Celsius on Asia's glaciers. *Nature* **549**(7671), 257–260. doi: 10.1038/nature23878.

Marzeion B, Jarosch AH and Hofer M (2012) Past and future sea-level change from the surface mass balance of glaciers. *Cryosphere* **6**(6), 1295–1322. doi: 10.5194/tc-6-1295-2012.

Radić V and 5 others (2014) Regional and global projections of twenty first century glacier mass changes in response to climate scenarios from global climate models. *Climate Dynamics* **42**(1), 37–58. doi: 10.1007/s00382-013-1719-7.

Radić V and Hock R (2011) Regionally differentiated contribution of mountain glaciers and ice caps to future sea-level rise. *Nature Geoscience* **4**(2), 91–94. doi: 10.1038/ngeo1052.

Raper SCB and Braithwaite RJ (2006) Low sea level rise projections from mountain glaciers and icecaps under global warming. *Nature* **439** (January), 311–313. doi: 10.1038/nature04448.

Renard B, Kavetski D, Kuczera G, Thyer M and Franks SW (2010) Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. *Water Resources Research* **46**(5), 1–22. doi: 10.1029/2009WR008328.

RGI Consortium (2017) Randolph Glacier Inventory – a dataset of Global Glacier Outlines: Version 6.0. Technical report, Global Land Ice Measurements from Space, Colorado, USA. doi: 10.7265/N5-RGI-60.

Rounce DR (2019) Python Glacier Evolution Model 'PyGEM' Manual, Github repository. Available at https://github.com/drounce/PyGEM.

Rounce DR, Hock R and Shean DE (2020) 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.

Shean DE and 6 others (2016) An automated, open source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery. *ISPRS Journal of Photogrammetry* **116**, 101–117. doi: 10.1016/j.isprsjprs.2016.03.012.

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

Slangen AB, Katsman CA, van de Wal RS, Vermeersen LL and Riva RE (2012) Towards regional projections of twenty-first century sea-level change based on IPCC SRES scenarios. *Climate Dynamics* **38**(5-6), 1191–1209. doi: 10.1007/s00382-011-1057-6.

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**(1995), 186–190.

Wortmann M, Bolch T, Menz C, Tong J and Krysanova V (2018) Comparison and correction of high-mountain precipitation data based on glacio-hydrological modeling in the Tarim River Headwaters (High Asia). *Journal of Hydrometeorology* **19**(5), 777–801. doi: 10.1175/jhm-d-17-0106.1.

Zemp M (2019) Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **568**, 382–386. doi: 10.1038/s41586-019-1071-0.

https://doi.org/10.1017/jog.2019.91 Published online by Cambridge University Press