

Bayesian approach to estimate proglacial lake volume (BE-GLAV)

Prateek Gantayat¹, Ashim Sattar², Umesh K Haritashya¹, C. Scott Watson^{3,4} and Jeffrey S. Kargel⁵

¹Department of Geology and Environmental Geosciences, University of Dayton, Dayton, OH, United States

²School of Earth, Ocean and Climate Sciences, Indian Institute of Technology, Bhubaneswar, Odisha, India

³COMET, School of Earth and Environment, University of Leeds, Leeds, United Kingdom

⁴Department of Hydrology and Atmospheric Science, The University of Arizona, Tucson, AZ, United States

⁵Planetary Science Institute, Tucson, AZ, United States

Corresponding author: Prateek Gantayat (pgantayat1@udayton.edu)

Key Points:

- Bayesian calibration based estimation of proglacial lake volume
- Method applied to sixty-six lakes located in the High Mountain Asia
- Reduction of maximum error in modelled lake volume by >5 times as compared to existing scaling models
- Requires minimal calibration with observations as compared to scaling models

Abstract

We present a new model called Bayesian Estimated Glacial Lake Volume (BE-GLAV) to estimate the volume of proglacial lakes. Presuming the lake cross-section as trapezoidal, BE-GLAV uses a Bayesian calibration approach to adjust the cross-sectional geometry to match modelled and observed lake surface widths. We validated our model using bathymetric measurements from lakes spread across High Mountain Asia (specifically, the Himalaya and Tien-Shan), with aerial extents ranging from 0.01 to 5.5 km². The modelled lake volumes agreed with the measured lake volume with a root-mean-square absolute uncertainty of ~14%. With minimum and maximum errors of ~0.3% and ~61.2%, BE-GLAV performed well compared to ten other models in a model inter-comparison experiment. Using the measured set of volumes, our model can constrain both the root mean square (RMS) error and the maximum percentage error in modelled lake volume, unlike other models, some of which can compute just the RMS uncertainty.

1 Introduction

Glaciers have experienced extensive recession and thinning throughout the 20th and 21st century, primarily due to global warming. In many regions, glacier retreat has accelerated over the past few decades (Bolch et al., 2012; Azam et al., 2018; Hock et al., 2019; Zemp et al., 2019), causing the formation and (or) expansion of proglacial lakes (Shugar et al., 2020; Wang et al., 2020; Chen et al. 2020; Haritashya et al., 2018). Some studies suggest a ~50% increase in the number of lakes and a ~53% increase in the total lake area worldwide (Shugar et al., 2020).

The High Mountain Asia (HMA) is home to ~15000 to ~25000 proglacial lakes (e.g., Furian et al., 2021). Some of these lakes are also the sources of Glacial lake outburst floods (GLOFs), which occur when the lake water is discharged catastrophically, causing significant damage to communities and infrastructure and the geomorphology of downstream regions (e.g., Wilson et al., 2019; Zheng et al., 2021; Sattar et al., 2022). Veh et al. (2019) showed that the flood peaks from GLOFs can rival monsoon-fed discharges in rivers up to thousands of kilometers downstream. Recently, Emmer et al. (2020) have shown that several thousand lives have been claimed by a small number of GLOFs and associated mass movements, and many more people suffer when major infrastructure is damaged by them (e.g., Lutzow et al., 2023).

The volume of a glacial lake is an important input parameter for many hydrological applications such as modelling the severity of GLOF events (e.g., Rinzin et al., 2023; Emmer et al., 2022); estimating glacier retreat when coupled with various minimalistic ice-flow models (Oerlemans and Nick, 2006); and local small-scale hydropower production (e.g., Sattar et al., 2019). Unfortunately, glacial lakes are often located in some of the most remote and challenging terrains, making the in-situ measurement of their volumes extremely difficult. Consequently, the in-situ estimates of lake volume exist only for a handful of lakes in the HMA (e.g., Haritashya et al., 2018; Zhang et al., 2023; Sharma et al., 2018; Watson et al., 2020; Kapitsa et al., 2017).

Several studies have proposed power-law (non-linear) relationships between lake volume and other lake parameters (i.e., area and/or mean depth) to estimate lake volume. These relationships (scaling models) are derived by fitting an exponential/ polynomial curve to data representing volume-area/ volume-mean depth data for multiple lakes (e.g., O'Connor et al., 2001; Zhang et al., 2023; Huggel et al., 2002). Some also utilise mean lake surface width, volume and area (e.g., Munoz et al., 2020). Scaling models have been popular because of their simplistic formulation and

ease of implementation on large, spatial scales. The scaling models proposed by O'Connor et al. (2001), Huggel et al. (2002), and Fujita et al. (2013) were calibrated against the lake volume and area from the Central Oregon Cascade Range, Swiss Alps, and Himalaya, respectively, and Cook and Quincey (2015) derived a scaling relationship using a global dataset of lake volume and area. Some scaling models have been derived for a specific lake using time series data of the lake volume and area (e.g., Yao et al., 2012; Sharma et al., 2018). Wang et al. (2012) developed a model using the volume and area data from twenty lakes in the Longbasaba and Pida regions (in China), and Emmer and Vilimek (2013) calibrated their scaling model with volume and area data from thirty-five lakes from the Peruvian Andes. Further, Zhang et al. (2023) calibrated their model against the areas and volume data from 47 lakes across the Indian, Nepal, Bhutan and Tibetan Himalaya. The scaling models that Munoz et al. (2020) and Qi et al. (2022) developed were among the first to use lake surface width as an input parameter, along with volume and area, calibrating them against data from glacial lakes in the Andes and the Tibetan Himalaya, respectively.

The uncertainty associated with the lake volume estimated using scaling models exponentially increases as the lake area increases (e.g., Zhang et al., 2023) due to the non-linear nature of the empirical relationship between lake volume and lake area. Therefore, the scaling models have been reported to give unrealistic estimates of lake volume (i.e., more than three times the measured lake volume in some cases) for individual lakes (e.g., Zhang et al., 2023). Additionally, Cook and Quincey (2015) conclusively showed that lake volume and lake area are auto-correlated, which can give a false impression of accuracy in the estimated lake volume. Hence, the use of lake area in both the variables of Volume- Area (V-A) regression for proglacial lakes may create a misleading impression about the quality of the data used in the relationship, and the high correlation coefficients may misleadingly indicate reduced uncertainty in the V-A relation (e.g.,

Cook and Quincey (2015); Haeberli et al., 2016). With over 5000 existing glacial lakes at present and more predicted to form in the future, e.g., in the Karakoram region (Furian et al., 2021; Zheng et al., 2021), it is important to develop a comprehensive modelling framework that addresses concerns with the empirical volume-area scaling approaches and also constrains the level of uncertainty. In section 4.2 we address the auto-correlation issue.

The Bayesian Calibration based models have been successfully used in many glaciological applications, such as the estimation of ice thickness (Brinkerhoff et al., 2017; Werder et al., 2020); estimation of supraglacial debris heat conductivity (Laha et al., 2022); optimisation of firn densification models (Verjans et al., 2020); and inferring basal sliding parameters (Gopalan et al., 2021). The calibration process is based on Bayes' theorem of probability and allows combining prior knowledge of the system variable(s) with the fitting of a physical model to the observations (e.g., Werder et al., 2020).

This study presents a new Bayesian statistics-based numerical model (Bayesian Estimated Glacial Lake Volume - BE-GLAV) to estimate proglacial lake volume. To our knowledge, ours is the first of its kind among the existing glacial lake volume estimation models.

2 Materials and Methods

Our model (BE-GLAV) assumes that at any point along the lake's centreline, the cross-sectional profile of the bed elevation of a proglacial lake can be approximated as parabolic or trapezoidal in nature (Figure 1a). This is because the present lake bed was formerly under a glacier whose cross-sectional geometry resembles a trapezium (e.g., Oerlemans, 1997). This geometry works particularly well for lakes with thick sediment deposits on their beds and nearly

constant sloping side walls. It is also a useful approximation for lakes that have U-shaped cross-sections. We expect that most of the statistical uncertainty associated with this model—which we determine directly for our sample of measured lakes— will occur for common U-shaped lake cross sections, where there are anomalous extreme overdeepenings, or exceptional thicknesses of deposited sediment (hence, shallowing), and other sharp departures from trapezoids. Another source of the anomaly is the ice-ramps that extend under the water line from the connected glacier’s terminus (Robertson et al., 2012). These ramps have been found to extend up to >500 m from the calving termini of the connected glacier into the respective proglacial lake. The surfaces of these ice ramps are undulating and covered with a thick layer of unsorted sediment derived from supraglacial and englacial debris, lateral moraines and fluvial deltaic deposition (Robertson et al., 2012). The sediment on the ice ramps reduces melting and, therefore, thinning of the ramps, and can load the ramp to hinder flotation. Consequently, the ice ramps can remain intact a few hundreds of metres in front of the calving termini (Robertson et al., 2012).

Throughout the paper, the phrase ‘present lake bed’ refers to the deglaciated area under the lake at the time of bathymetric measurements. Consequently, for every cross-section, we have three variables, namely lake base width (W_b), lake depth (H), and side slope ($\lambda = \cot\alpha_0 + \cot\alpha_1$) (Figure 1a). The lake surface width at a particular cross-section (W_s) can be expressed in terms of the variables as (Figure 1a):

$$W_s = W_b + \lambda H \quad (1)$$

Therefore, the volume (V) of a proglacial lake with n cross-sections is computed as:

$$V = \frac{1}{3} 0.5(W_{b1} + W_{s1})H_1x + \sum_{i=2}^{n-2} 0.5(W_{si} + W_{bi})H_i x + \frac{1}{3} 0.5(W_{bn-1} + W_{sn-1})H_{n-1}x$$

(2)

Where, W_{si} , W_{bi} and H_i are the lake surface width, base width and lake depth at cross-section i , respectively. x represents 50 m grid resolution.

Our model estimates the physically plausible values of W_b , λ and H for a cross-section by minimising the difference between observed and modelled surface width for each grid point along the lake's centreline (Figure 1d). The minimization procedure was performed using Bayesian calibration.

2.1 Bayesian Calibration

The Bayesian calibration approach identifies W_b , λ and H as the free parameters and defines this parameter set as θ . The calibration process relies on Bayes' theorem (Eq. 2) that allows updating a probability distribution $P(\theta)$ based on observed data D :

$$P(\theta/D) = \frac{P(D/\theta)P(\theta)}{P(D)} \quad (3)$$

Where $P(\theta)$ is called the prior distribution, $P(D/\theta)$ is called the likelihood distribution that quantifies the difference between modelled and measured lake surface width, i.e., W_s and $P(\theta/D)$ is the posterior probability distribution and is the quantity that our model calculates. The term $P(D)$ is the marginal probability distribution of the data; is a constant term, independent of θ , that does not affect $P(\theta/D)$ and is therefore disregarded in our analyses (e.g., Brinkerhoff et al., 2017).

2.2 Estimation of $P(\theta)$

2.2.1 For Type A lakes

By Type A lakes, we refer to those lakes where the ice thickness map of the corresponding connected glacier (i.e., from a time when part of the lake's current bed was beneath the glacier) is available.

We use normal and weakly informative prior distributions about the initial values of θ so that the constraint of $P(\theta)$ on $P(D/\theta)$ is minor (e.g., Verjans et al., 2020). To calculate $P(\theta)$ for the grid points along the lake's centreline (Figure 1d), we first estimate the distribution for each of the free parameters (i.e., W_b , λ and H) at the grid points using an ice thickness map of a glacier from a time when either the whole or a part of the current lake bed was glaciated (i.e., the light-green coloured region in Figure 1d). Using the ice thickness and the corresponding surface elevation maps of the glacier (Figure 1b), we estimate the maximum extent and the corresponding depth distribution of the glacier bed overdeepening of which the lake is a part (i.e., the aqua coloured region in Figure 1b). We then delineate the centreline of the overdeepening and at every 50 m interval along the centreline we estimate W_s , W_b , λ , H , W_b/W_s ($R1$) and H/W_s ($R2$) for the corresponding cross-sections of the bed overdeepening, assuming that the depression (i.e., the aqua coloured region in Figure 1c) is completely filled. We then estimate the mean (μ_λ , μ_{R1} , μ_{R2}) and standard deviation (σ_λ , σ_{R1} , σ_{R2}) of λ , $R1$ and $R2$ at the 50 m apart grid points for the area shown in Figure 1c. We also estimate the correlation coefficients for the pairs (W_b, H) , (H, λ) and (W_b, λ) , namely ρ_1 , ρ_2 and ρ_3 , respectively. ρ_1 , ρ_2 and ρ_3 are considered non-zero only when the total number of grid points in the overdeepening area exceeds ten. Because the overdeepening and the lake are both parts of a single, larger glacier bed depression, our model assumes that the mean, standard deviations, and

correlation coefficients calculated above can be applied to the corresponding entities in the present lake bed, which is the light-green coloured region shown in Figure 1d. The means and standard deviations of λ , $R1$, $R2$ for all grid points (except for the extremes; see Figure 1d) located along the centreline of the present lake bed (Figure 1d) are μ_λ , μ_{R1} , μ_{R2} and σ_λ , σ_{R1} , σ_{R2} , respectively. The minimum and maximum values of $R2$ (i.e., m_{R2} and M_{R2} , respectively) are also calculated for use in the posterior distribution, as explained in section 2.4. Similarly, within the present lake-bed, the correlation coefficients for the pairs (W_b, H) , (H, λ) and (W_b, λ) are ρ_1 , ρ_2 and ρ_3 , respectively. At the two grid points where the lake centreline ends, we assume $H = 0$, $W_b = 0$ and an undefined λ . H is then calculated for each grid point along the centreline of the present lake-bed (Figure 1d) by inferring depth values from the distributed depth map of the glacier bed overdeepening. This map is produced by using the ice thickness and glacier surface elevation for the previously glaciated region, which is now part of the lake (i.e., the green region in Figure 1d). By linearly interpolating the depth values from the grid point that is closest to the former glacier front and is located in the light-green region (i.e., X in Figure 1d), H is calculated for the remaining grid points (i.e., those that are in the blue zone of Figure 1d).

The mean and standard deviation for λ (i.e., μ_λ and σ_λ) stay constant at each grid point along the lake's centreline (aside from the first and last) (Figure 1d). The mean and standard deviation for W_b is formulated as $\mu_W = W_s \mu_{R1}$ and $\sigma_W = \sigma_{R1} W_s$, where W_s is the lake surface width at the grid point (Figure 1d). The mean for parameter H (i.e., μ_H) at every grid point is calculated as $\mu_H = H$. The standard deviation (σ_H) of H at the grid point is assumed to be of the form:

$$\sigma_H = kH \quad (4)$$

where, the value of the parameter k is set as 0.5 (i.e., we assume 50% margins around lake depth). Finally, for every grid point lying along the lake's centreline (Figure 1d), the prior

200 probability distribution of θ was formulated as a multivariate normal distribution with mean μ_θ
 201 $[\mu_\lambda, \mu_W, \mu_H]$ and the covariance (Σ_θ) a 3-by-3 matrix as shown below:

$$\begin{bmatrix} \sigma_W^2 & \sigma_W \sigma_H \rho_1 & \sigma_\lambda \sigma_W \rho_3 \\ \sigma_W \sigma_H \rho_1 & \sigma_H^2 & \sigma_\lambda \sigma_H \rho_2 \\ \sigma_\lambda \sigma_W \rho_3 & \sigma_\lambda \sigma_H \rho_2 & \sigma_\lambda^2 \end{bmatrix}$$

202 The above procedure was followed for seven lakes (i.e., the lakes named in bold fonts) in Table
 203 S1.

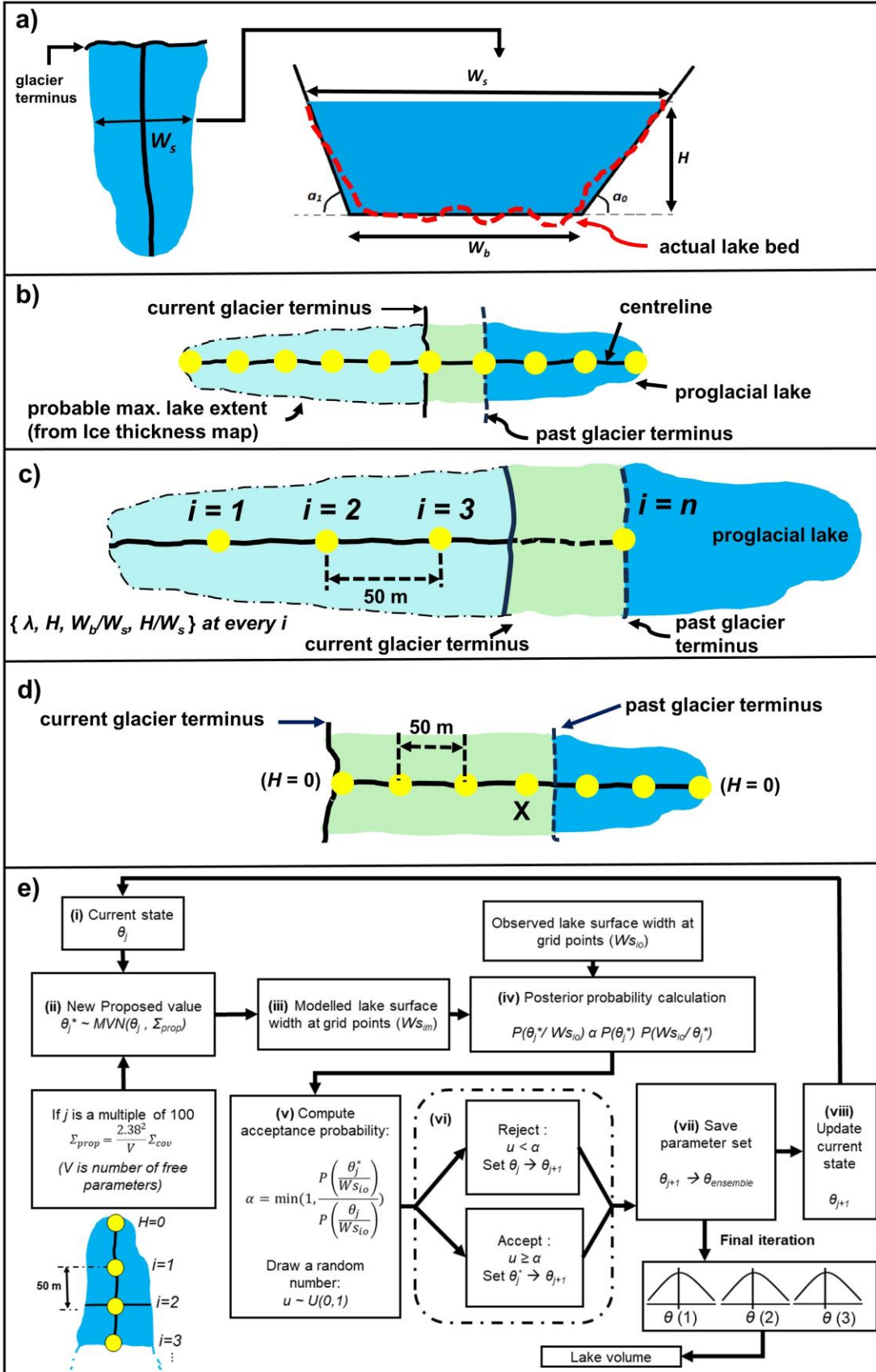


Figure 1. Schematic representation outlining the process to estimate $P(\theta)$ for Type A lakes (a-d); aqua coloured regions represent the currently glaciated areas, the blue region represents the region that was filled with water in the past as well as during the time of bathymetric surveys, green regions represent the portion of the current lake bed that was glaciated in the past but is currently part of the lake bed; a) assumption of trapezoidal-shaped lake cross-sections where dashed red line represents the actual lake bed, free parameters are W_b (base width), λ (sidewall slope parameter), H (lake depth). The lake surface width is W_s . Here, λ is formulated as a function of the slope angles (i.e., α_0 and α_1) as $\lambda = \cot\alpha_0 + \cot\alpha_1$; b) top view of the lake's 2-D extent and the past ice thickness map of the region; yellow dots represent grid points at 50 m intervals; c) the procedure to derive the marginal and the joint prior distribution of the free parameters; d) schematic showing the necessary boundary conditions i.e., $H=0$, $W_b=0$ and undefined λ ; e) flowchart of the MCMC block of our model (see section 2.4). For Type A lakes, variable $V=3$, else it is set to 2 (Eqn. 7); the Monte Carlo module is solved at every grid point (yellow circle) located along the lake's centreline at 50m intervals (Bottom left corner). Identification of the flow in panel (e) MCMC block can be determined from the Roman numerals.

2.2.2 For type B lakes

By Type B lakes, we refer to those lakes where the corresponding glacier ice thickness map is unavailable, unlike the Type A counterparts. Our approach to calculating $P(\theta)$ at grid points differs slightly when historical ice thickness maps are unavailable. In this case, we use two free parameters

(W_b and H) instead of three, with λ being treated as a constant equal to two (i.e., the lake's sidewalls are assumed to have 45-degree slopes). Therefore, for these lakes, the free parameter set θ consists of two parameters, W_b and H . To avoid limiting $P(\theta)$ on $P(D/\theta)$, we assume normal and weakly informative priors for the initial values of θ (i.e., W_b and H) (e.g., Verjans et al., 2020). Except for the first and last grid points, the prior distribution for H was modelled as a normal distribution with mean $\mu_H = H_{lake}$ and standard deviation $\sigma_H = 0.5H$ (i.e., we assumed 50% margins about the initial estimate of H). Where H_{lake} is equal to the mean lake depth as determined by the two volume-area scaling models of Zhang et al. (2023) for the Himalaya and Kapista et al. (2017) for the Tien Shan. These models were chosen because they were calibrated against the available *in-situ* data (i.e., volume and area) in the respective regions. We establish a normal prior for W_b with a mean μ_W and standard deviation σ_W . The standard deviation of the observed lake surface widths at the grid points is also set to be equal to σ_W . Since there is no prior information on the lake geometry, we assume that μ_W equals the grid spacing of 50m. We further assumed that the free parameters W_b and H are independent of each other for lakes where there is a scarcity of past ice-thickness maps of their corresponding glaciers. As a result, for each grid point along the lake's centreline, the prior probability distribution of θ , i.e., $P(\theta)$, was defined as a multivariate normal distribution with a mean $\mu_\theta [\mu_W, \mu_H]$ and a 2-by-2 covariance matrix (Σ_θ) as shown below:

$$\begin{bmatrix} \sigma_W^2 & 0 \\ 0 & \sigma_H^2 \end{bmatrix}$$

Finally, the prior distribution of θ is formulated as a multivariate normal distribution:

$$P(\theta) \sim MVN(\mu_\theta, \Sigma_\theta) \quad (5)$$

2.3 Estimation of $P(D/\theta)$

Our model uses a normal likelihood function $P(D/\theta)$ that quantifies the match between modelled and observed lake surface widths:

$$P(D/\theta) \propto e^{-\frac{1}{2}\left(\frac{W_{sm}-W_{so}}{\sigma}\right)^2} \quad (6)$$

Where, W_{sm} and W_{so} represent modelled and observed lake surface widths, respectively, and σ^2 is the variance. The variance sets the spread allowed for the model outputs when compared to measured values and are calculated by taking 10% margins around the corresponding observed surface widths (e.g., Verjans et al., 2020). Allowing for such spread is necessary because model–observation discrepancy can result from various factors, including errors in measuring surface widths from satellite imagery, uncertainty in estimating the mean and standard deviation of free parameters, and uncertainty in modelled past ice thickness map of the glacier that covered a portion of the lake’s current bed. Errors are also likely to be introduced by the measured volume, depending on the survey style of the bathymetric measurements, track density, and interpolation technique used in preparing 3-D bathymetric maps from point-based, single beam sonar survey (Watson et al., 2020).

2.4 Estimation of $P(\theta/D)$

The posterior distribution $P(\theta/D)$ gives our model’s probability distribution over the parameter space (i.e., θ) given the observed lake surface widths. In our model, with a weak informative $P(\theta)$, the distribution $P(\theta/D)$ is substantially governed by $P(D/\theta)$. Due to the paucity of any

analytical form of $P(\theta/D)$ for studied lakes, we investigate the parameter space to generate an ensemble of θ_j approximating $P(\theta/D)$ (e.g., Brinkerhoff et al. 2017; Verjans et al., 2020). This is achieved efficiently by applying a gradient-free Markov Chain Monte Carlo (MCMC) algorithm, Random Walk Metropolis (RWM) (Hastings, 1970). The posterior distribution of the free parameter set is calculated at every grid point. We first estimate lake surface width for all grid points using the grid point-specific initial estimates of μ_θ . The difference between modelled and observed lake surface width is quantified by the likelihood, i.e., $P(D/\theta)$. The posterior probability (i.e., $P(\theta/D)$) is calculated following Eq. 3. At this point, the RWM algorithm starts, and the state of the chain, θ_j (Figure 1e part (i)), is set to the original values of the free parameters, and the corresponding posterior probability is saved as $P(\theta_j/D)$. Note that the j subscript denotes the iteration number, which is equal to 0 at this initial step. The RWM algorithm then proposes a new θ_j^* from a proposal distribution (Figure 1e part (ii)). Here, we assume the proposal distribution to be the symmetric multivariate normal distribution centred on θ_j . Consequently, the random selection of θ_j^* depends only on the current state θ_j and the proposal variance (Σ_{prop} in Figure 1e). For the parameter set θ_j^* , the model recalculates the lake surface width for the grid point and $P(\theta_j^*/D)$ is computed (Figure 1e part (iii) & part (iv)). From there, the proposed θ_j^* is either accepted or rejected in the ensemble approximating $P(\theta/D)$ (vi; Figure 1e part (vi)). This decision depends on the ratio $\alpha = P(\theta_j^*/D)/P(\theta_j/D)$ (Figure 1e part (v)). The RWM then selects a random number u from the uniform distribution $U(0,1)$ and θ_j^* is accepted in the ensemble only if $u > \alpha$ (Figure 1e part (vi)). The saved set becomes the updated current status for the next iteration θ_{i+1} with its associated $P(\theta_{i+1}/D)$ (Figure 1e part (vii) & part (viii)). The algorithm iterates this process and reaches a final posterior distribution over θ . The RWM algorithm has the property that the chain

will ultimately converge to a stationary distribution representing the posterior $P(\theta/D)$. Thus, after a sufficiently high number of iterations of the algorithm, the ensemble of parameter sets is representative of $P(\theta/D)$. In our study, we discovered that 5000 iterations are sufficient to achieve convergence. The proposal covariance Σ_{prop} must account for inter-dependence among the different components of θ ; i.e., the value of one free parameter can influence the value of another free parameter for the model to reach a good match with the observed data. Σ_{prop} can capture this dependence between parameters and, for optimality, it is updated every given number of iterations (100 in our study) using Eq. (7) (Rosenthal, 2011):

$$\Sigma_{prop} = \frac{2.38^2}{V} \Sigma_{cov} \quad (7)$$

Where Σ_{cov} is the covariance matrix between the model's free parameters at this stage of the iterative chain, and V is the number of free parameters, which is three for lakes with a historical ice thickness map of connected glaciers, and two for those that do not.

The ensemble of $P(\theta/D)$ was then checked for the presence of physically unrealistic parameter sets before computing the posterior probability distribution. This was carried out following a two-step approach wherein we first removed all the θ s with non-positive values for any free parameters (i.e., W_b , λ , H). We then separated the lakes where the historical map of the associated glacier's ice thickness was available and those without. For lakes with ice thickness map, we eliminated θ s where the relationship between the simulated lake depth and the associated observed lake surface width was outside the range $[m_{R2}, M_{R2}]$ (see section 2.2 for more information on m_{R2} and M_{R2}). For the rest of the lakes with no ice thickness map, we first estimated the means of m_{R2} and M_{R2} (i.e., μm_{R2} and μM_{R2} , respectively) and rejected θ s where the ratio of modelled lake depth to observed lake surface width was not in the interval $[\mu m_{R2}, \mu M_{R2}]$.

The remaining θ s in the ensemble after the above filtering process were used to estimate the

posterior distribution. After obtaining the posterior distributions of the free parameter set at every grid point along the lake's centreline, we generate an ensemble of lake volume estimates using Eq. (2). Finally, the mean of the ensemble is our modelled estimate of lake volume.

3 Data and Validation sites

We validated our model, assessing its effectiveness across 66 proglacial lakes in the central and eastern Himalaya and the Tien Shan (Figure S1a and Figure S1b). Table S1 provides information on the lake parameters of all 66 lakes. The glacier ice thickness and corresponding surface elevation were inferred from the Global consensus ice thickness dataset of Farinotti et al. (2019) and Millan et al. (2022).

4 Results and Discussion

4.1 Modelled lake volumes

Modelled and measured estimates of all 66 lake's volume agree well with each other, with a mean, absolute, and relative uncertainty of ~14% (Table S1). Lakes Bechung and Aksu 1 had the largest (~61.2 %) and the smallest (0.3%) error in the predicted lake volume, respectively. The inventory of 66 lakes was approximately equally divided into four categories based on their lake area (Figure 2). Except for Lake Bechung, the percentage deviation of our modelled lake volumes from measured is consistently within the range of ~ -36% to ~ 36% (Figure 2b; Table

S1). In addition to that, the results were unchanged when a coarse grid resolution of 50-100 m was implemented.

In our further treatment, we use the term error to mean the percentage departure of an individual measured lake volume from its modeled volume. Uncertainty then carries a statistical significance of how the actual (but unmeasured) volumes of an unknown population of lakes may vary from the modelled mean for a given lake area, whereas the term “maximum error” could, for instance, refer to the possible greatest outliers of the shallowest and deepest lakes for a given lake area in any comparable population of about 66 unmeasured lakes in any location where glacial valley hypsometry closely resembles that in the Himalaya and Tien Shan.

4.2 Comparison with other models

Figure 2 and Table S3 show the performance of our model with respect to ten previously published models (listed in Table S2) that are essentially power-law relations between lake volume and lake area.

The minimum error in lake volume simulated by each scaling model ranged from ~0.1% to 92%, while the maximum error ranged from ~92% to ~1779% (Table S3). The large value of the maximum bound of the error indicates the unreliability of the scaling models for individual lakes. Lake volumes modelled using Fujita et al. (2013) and Yao et al. (2012) approaches, both calibrated using volume-area data from Himalayan lakes, showed a systematic uncertainty of at least 130% and 250%, respectively, for all lakes in the Tien Shan (Table S3). This demonstrates

that using some scaling models in places where they have not been calibrated with corresponding observed lake volume-area data is dubious.

The scaling model of Zhang et al. (2023) gave the overall third-best estimate of proglacial lake volume in the Himalayan region with a minimum and maximum error of 0.1% and 233%, respectively, while the scaling model developed by Kapista et al. (2017) gave the best estimate of lake volume for all the lakes located in the Tien Shan region (i.e., maximum error < ~10%) (Table S3). This is likely due to the inclusion of in-situ volume-area data from 47 lakes in the case of Zhang et al. (2023) and 50 for Kapista et al. (2017).

The scaling approach of Emmer and Vilimek (2013) performed second best by being among the top five models for 61 out of 66 lakes (i.e., 48 lakes in Tien Shan and 13 in the Himalaya) (Figure S3). We were surprised by this outcome because that model was calibrated using volume and area data from lakes in the Peruvian Andes (Emmer and Vilimek, 2013). This suggests morphometric similarities between the Andes and the HMA, which could be attributed to active, receding glaciers and concurrent intense tectonism in both regions. However, our analysis is restricted due to the poor quality of existing knowledge about depth erosion by glaciers and the overdeepened topography of their rock beds (Haeberli et al., 2016). We believe that more research is needed to support the above hypothesis, such as running our model on Peruvian Andes lakes and comparing its performance to that of Emmer and Vilimek (2013).

Our model was among the top five performing models at 65 out of 66 lakes with minimal prior

353 tuning against measured lake area-volume data, unlike scaling models. In our volume
estimates, the maximum and minimum errors were ~1% and ~61%, respectively (Figure 2 and
Table S3; Supplementary). Compared to all the scaling models, our model's projected lake
volume was among the two closest to the observed volume for Lake Bechung (~61% deviation
from measured volume), even with the high magnitude of error. The outcome was the same for
model runs at higher spatial resolutions of 30m and 20m, suggesting that sometimes a lake's
morphometry might be too complicated to simulate for any model despite its capability.

Besides volume, BE-GLAV can estimate depth along the lake's centreline. Bathymetry was
available for only 17 lakes located in the Himalaya. Unlike volume, modelled and measured lake
depth was in good agreement with each other only for eight lakes (Figure S2). Out of these eight
lakes, Thulagi, Imja, Lower Barun, South Lhonak, Bechung, Luggye and Ranzeriaco were
classified as Type A lakes, and Maqiongco was classified as Type B lake (Figure S2; Table S1).
For these eight lakes, the RMS uncertainty between modelled and measured lake depth along the
lake centreline ranges between 4m to 17m (Figure S2). Discrepancies between modelled and
measured lake depth (i.e., along the lake centerline) were found to be large near the calving front.
However, the lake depth near the glacier terminus is not only difficult to measure but, at many
times, is also marred with uncertainties due to the presence of ice ramps that extend laterally up
to at least a hundred meters from the calving front (Haeberli et al., 2016). Additionally, due to
the dynamic nature of the region near the glacier terminus, the lake depth is estimated via spatial
interpolation of point-based lake depth values that were measured relatively further away from
the glacier terminus. These factors contribute to the uncertainty in measured lake depth near the

glacier terminus. However, this does not pose an issue, because the main aim of our paper is to estimate lake volume and not lake bathymetry along the lake centerline.

Results demonstrate that, unlike scaling models where the level of uncertainty in the lake volume is unpredictable, BE-GLAV can restrict the upper limit of the error in volume (likely due to the Bayesian calibration procedure) and do that with minimal pre-calibration against the regional, observed lake volume, and area data.

For every Type B lake, we formulated the prior distribution of lake-depth at every gridpoint (i.e., along the lake centerline) as a Gaussian distribution centered around H_{lake} (where, H_{lake} is the mean depth that is derived using the V-A scaling models proposed by Zhang et al. (2023) and Kapitsa et al. (2017) for lakes located in the Himalaya and the Tien Shan, respectively). Interestingly, despite using auto-correlated variables (i.e., lake volume and lake area) in calculating H_{lake} , the Bayesian calibration part of our model was able to restrict the maximum level of error for individual lake volumes as well as the 95% confidence level of global uncertainty in the modelled lake volume. This shows that BE-GLAV provides a robust way-forward to estimate proglacial lake volume by using mean lake depth derived from a V-A scaling model. Such a feat was earlier not possible by solely using V-A scaling techniques because lake area and lake volume are autocorrelated (discussed in section 1), which can give a false impression of accuracy in the corresponding estimated lake volume (e.g., Cook and Quincey, 2015; Haeberli et al. (2016)).

424 It is crucial to highlight that BE-GLAV, as the geometry is presently developed, cannot be
425 applied to supraglacial, ice-marginal, or extremely shallow lakes (depth $<5\text{-}10$ m) since our
426 assumption regarding the trapezoidal or parabolic structure of the lake cross-section may not
427 hold true for such lakes.

428

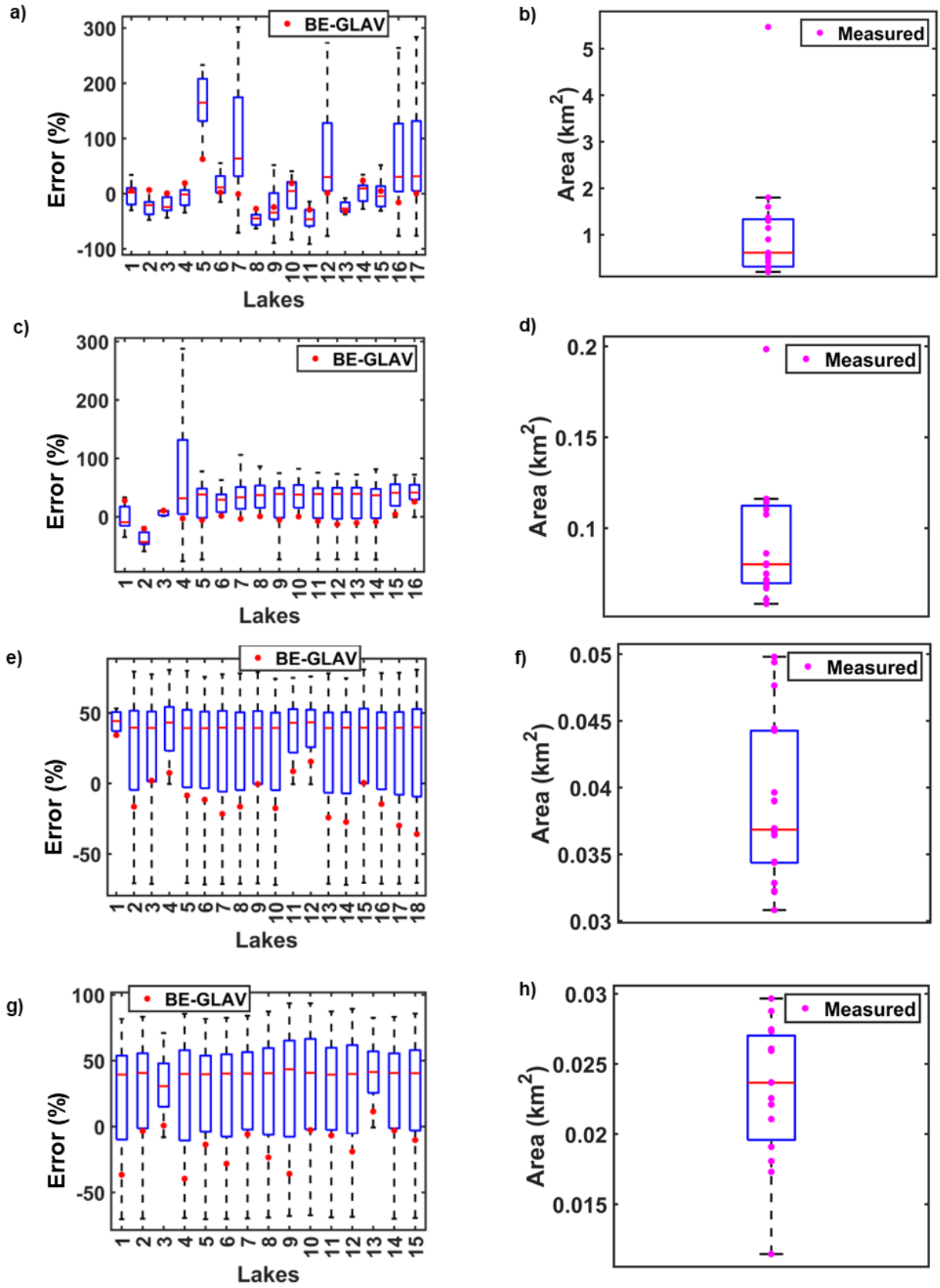


Figure 2. For simplicity, the set of 66 lakes was divided into 4 parts based on lake area (a,c,e,g). Boxplot of % error (i.e., % deviation from measured) of modelled lake volumes for lakes with area $> 0.2 \text{ km}^2$ (a), $0.05 \text{ km}^2 \leq \text{area} < 0.2 \text{ km}^2$ (c), $0.03 \text{ km}^2 \leq \text{area} < 0.05 \text{ km}^2$ (e), area $> 0.03 \text{ km}^2$ (g). The red dots in a, c, e, g represent the % error in lake volumes modelled by BE-GLAV; Box plot of measured lake area for lakes with area $> 0.2 \text{ km}^2$ (b), $0.05 \text{ km}^2 \leq \text{area} < 0.2 \text{ km}^2$ (d), $0.03 \text{ km}^2 \leq \text{area} < 0.05 \text{ km}^2$ (f), area $> 0.03 \text{ km}^2$ (h). In all the box and whisker plots, the ends of the whiskers mark the $\pm 3\sigma$ limits beyond which any data point was treated as an outlier. The vertical ends of the boxes represent the Interquartile range (IQR).

5 Conclusions

We present a proof of concept where a new Bayesian calibration-based numerical model (BE-GLAV) is used to estimate the volume of a proglacial lake. Assuming the lake cross-section to be trapezoidal, BE-GLAV minimises the difference between modelled and measured lake surface width. Our modelled volumes were in good agreement with that measured with a mean absolute uncertainty of $\sim 14\%$, and it was able to limit the maximum error to $\sim 61\%$, unlike the large errors of the scaling models. A model intercomparison experiment showed that all models (scaling and our method) showed acceptable performance (about average) for lakes with extremely shallow bathymetry ($< 10 \text{ m}$) for more than half of the current lake extent (e.g., Bechung).

BE-GLAV could be a good—we think better— alternative to scaling models for regions where there is a paucity of bathymetric surveys. It is easy to implement with minimal calibration, unlike the scaling counterparts. We have indirect evidence that BE-GLAV may perform well for the

Peruvian Andes, but we would need to test it there as well as in areas where active tectonics and valley hypsometry could differ from HMA.

6 Acknowledgments

PG and UKH acknowledge time funded by the University of Dayton's Mann Endowed Chair in the Natural Sciences. UKH and JSK were also supported by the NASA High Mountain Asia Grant 80NSSC19K0653 and the NASA Interdisciplinary Research in Earth Science grant 80NSSC18K0432. AS acknowledge DST Inspire Faculty Award for the support.

Open Research

The model code and the input data used for Type A lakes can be found at (Gantayat, 2023a). The model code and the input data used for Type B lakes can be found at (Gantayat, 2023b). Both the model codes have been written in MATLAB. The figures were prepared using ArcMap and MATLAB.

References

Azam, M.F., Wagnon, P., Berthier, E. & Kargel, J.S. (2018), Review and status of mass changes of Himalayan-Karakoram glaciers, *Journal of Glaciology*, doi: 10.1017/jog.2017.86.

Bolch, T., Kulkarni, A., Kaab, A., Huggel, C., Paul, F., Cogley, J.G., Frey, H., Kargel, J.S., Fujita, K., Scheel, M., Bajracharya, S. & Stoffel, M. (2012), The State and Fate of Himalayan Glaciers. *Science*, 336, 310-314.

Brinkerhoff, D.J, Aschwanden, A., & Truffer, M. (2016), Bayesian inference of subglacial topography using mass conservation. *Frontiers in Earth Science*, 4. doi:10.3389/feart.2016.00008

Chen, F., Zhang, M., Guo, H., Allen, S., Kargel, J.S., Haritashya, U.K., & Watson, C. S. (2020), Annual 30-meter Dataset for Glacial Lakes in High Mountain Asia from 2008 to 2017. *Earth System Science Data*, 13, 741–766. <https://doi.org/10.5194/essd-13-741-2021>

Cook, S. J. & Quincey, D. J. (2015), Estimating the volume of Alpine glacial lakes. *Earth Surface Dynamics*, 3, 559–575, <https://doi.org/10.5194/esurf-3-559-2015>

Emmer, A., & Vilímek, V. (2013), Review article: lake and breach hazard assessment for moraine-dammed lakes: an example from the cordillera blanca (Peru). *Natural Hazards Earth System Science*, 13 (6), 1551–1565. doi:10.5194/nhess-13-1551-2013

Emmer, A., Harrison, S., Mergili, M., Allen, S., Frey, H., & Huggel, C. (2020), 70 years of lake evolution and glacial lake outburst floods in the Cordillera Blanca (Peru) and implications for the future. *Geomorphology*, 365, 107178, ISSN 0169-555X, <https://doi.org/10.1016/j.geomorph.2020.107178>

Emmer, A., Allen, S. K., Carey, M., Frey, H., Huggel, C., Korup, O., Mergili, M., Sattar, A., Veh, G., Chen, T. Y., Cook, S. J., Correias-Gonzalez, M., Das, S., Diaz Moreno, A., Drenkhan, F., Fischer, M., Immerzeel, W. W., Izagirre, E., Joshi, R. C., Kougkoulos, I., Kuyakanon Knapp, R., Li, D., Majeed, U., Matti, S., Moulton, H., Nick, F., Piroton, V., Rashid, I., Reza, M., Ribeiro

de Figueiredo, A., Riveros, C., Shrestha, F., Shrestha, M., Steiner, J., Walker-Crawford, N.,
Wood, J. L., & Yde, J. C. (2022), Progress and challenges in glacial lake outburst flood research
(2017–2021): a research community perspective. *Natural Hazards Earth System Science*, 22,
3041–3061, <https://doi.org/10.5194/nhess-22-3041-2022>

Farinotti, D., Huss, M., Fürst, J. J., Landmann, J., Machguth, H., Maussion, F., & Pandit, A.
(2019), A consensus estimate for the ice thickness distribution of all glaciers on Earth. *Nature
Geoscience*, 12, 168–173. doi:10.1038/s41561-019-0300-3

Fujita, K., Sakai, A., Takenaka, S., Nuimura, T., Surazakov, A. B., Sawagaki, T., &
Yamanokuchi, T. (2013), Potential flood volume of Himalayan glacial lakes. *Natural Hazards
Earth System Science*, 13, 1827–1839. <https://doi.org/10.5194/nhess-13-1827-2013>

Furian, W., Maussion, F., & Schneider, C. (2021), Projected 21st-Century Glacial Lake
Evolution in High Mountain Asia. *Frontiers in Earth Science*, 10, doi:10.3389/feart.2022.821798

Gopalan, G., Hrafnkelsson, B., Aðalgeirsdóttir, G., & Pálsson, F. (2021), Bayesian Inference of
Ice Softness and Basal Sliding Parameters at Langjökull. *Frontiers in Earth Science*, 9.
<https://www.frontiersin.org/articles/10.3389/feart.2021.610069>

Haeberli, W., Linsbauer, A., Cochachin, A., Salazar, C., & Fischer, U. H. (2016), On the
morphological characteristics of overdeepenings in high-mountain glacier beds. *Earth Surface
Process and Landforms*, 41, 1980–1990. doi: 10.1002/esp.3966

- Haritashya, U.K., Kargel, J.S., Shugar, D.H., Leonard, G.J., Strattman, K., Watson, C.S., Shean, D., Harrison, S., Mandli, K.T., Regmi, D. (2018), Evolution and Controls of Large Glacial Lakes in the Nepal Himalaya. *Remote Sensing*, 10, 798. <https://doi.org/10.3390/rs10050798>
- Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97–109
- Huggel, C., Kääb, A., Haeblerli, W., Teyssie, P., & Paul, F. (2002), Remote sensing-based assessment of hazards from glacier lake outbursts: A case study in the Swiss alps. *Canadian Geotechnical Journal*, 39 (2), 316–330. doi:10.1139/t01-099
- Kapitsa, V., Shahgedanova, M., Machguth, H., Severskiy, I., & Medeu, A. (2017), Assessment of evolution of mountain lakes and risks of glacier lake outbursts in the djungarskiy (Jetysu) Alatau, central Asia, using landsat imagery and Glacier Bed topography modelling. *Natural Hazards Earth System Science*, 17, 1837–1856. doi:10.5194/nhess-2017-134
- Laha, S., Winter-Billington, A., Banerjee, A., Shankar, R., Nainwal, H., & Koppes, M. (2023), Estimation of ice ablation on a debris-covered glacier from vertical debris-temperature profiles. *Journal of Glaciology*, 69(273), 1-12. doi:10.1017/jog.2022.35

Loriaux, T., & Casassa, G. (2013), Evolution of glacial lakes from the Northern Patagonia Icefield and terrestrial water storage in a sea-level rise context. *Global Planetary Change*, 102, 33–40. doi:10.1016/j.gloplacha.2012.12.012

Lützow, N., Veh, G., & Korup, O. (2023), A global database of historic glacier lake outburst floods. *Earth System Science Data*, 15, 2983–3000, <https://doi.org/10.5194/essd-15-2983-2023>

Millan, R., Mouginot, J., Rabatel, A., & Morlighem, M. (2022), Ice velocity and thickness of the world's glaciers. *Nat. Geosci.* **15**, 124–129. <https://doi.org/10.1038/s41561-021-00885-z>

Muñoz, R., Huggel, C., Frey, H., Cochachin, A., & Haeberli, W. (2020), Glacial Lake depth and volume estimation based on a large bathymetric dataset from the Cordillera Blanca, Peru. *Earth Surface Processes and Landforms*, 45 (7), 1510–1527. doi:10.1002/esp.4826

Nick, F. & Oerlemans, J. (2006), Dynamics of calving glaciers: comparison of three models. *Journal of Glaciology*, 52, 177.

O'Connor, J. E., Hardison, J. H., & Costa, J. E. (2001), Debris Flows from Failures of Neoglacial – Age Moraine Dams in the Three Sisters and Mount Jefferson Wilderness Areas, Oregon, Professional Paper 1606, U.S. Geological Survey, Reston, 93.

Oerlemans, J. (1997), A flowline model for Nigardsbreen, Norway: Projection of future glacier length based on dynamic calibration with the historic record. *Annals of Glaciology*, 24, 382-389. doi:10.3189/S0260305500012489

Gantayat, P. (2023a). Model code for type A lakes [Software]. Zenodo. <https://doi.org/10.5281/zenodo.10155265>

Gantayat, P. (2023b). Model code for type B lakes [Software]. Zenodo. <https://doi.org/10.5281/zenodo.10157362>

Rinzin, S., Zhang, G., Sattar, A., Wangchuk, S., Allen, S.K., Dunning, S., & Peng, M. (2023), GLOF hazard, exposure, vulnerability, and risk assessment of potentially dangerous glacial lakes in the Bhutan Himalaya. *Journal of Hydrology*, Volume 619, 2023, 129311, ISSN 0022-1694, <https://doi.org/10.1016/j.jhydrol.2023.129311>

Robertson, C.M., Benn, D.I., Brook, M.S., Fuller, IC., & Holt, K.A. (2012), Subaqueous calving margin morphology at Mueller, Hooker and Tasman glaciers in Aoraki/Mount Cook National Park, New Zealand. *Journal of Glaciology*, 58(212), 1037-1046, doi:10.3189/2012JoG12J048

Qi, M., Liu, S., Wu, K., Zhu, Y., Xie, F., Jin, H. (2022), Improving the accuracy of glacial lake volume estimation: A case study in the poiqu basin, central Himalayas. *J. Hydrol.* 610, 127973. doi:10.1016/j.jhydrol.2022.127973

Rosenthal, J. (2011), Optimal Proposal Distributions and Adaptive MCMC. *Handbook of Markov Chain Monte Carlo*, edited by: Brooks, S., Gelman A., Jones G., and Meng X., Chapman and Hall, Boca Raton, USA, 93–112

Sattar, A., Goswami, A. & Kulkarni, A.V. (2019), Application of 1D and 2D hydrodynamic modelling to study glacial lake outburst flood (GLOF) and its impact on a hydropower station in Central Himalaya. *Natural Hazards*, 98, 817. <https://doi.org/10.1007/s11069-019-03747-5>

Sattar, A., Haritashya, U. K., Kargel, J. S., & Karki, A. (2022), Transition of a small Himalayan glacier lake outburst flood to a giant transborder flood and debris flow. *Scientific reports*, 12(1), 12421.

Sharma, R.K., Pradhan, P., Sharma, N.P., & Shreshta, D.G. (2018), Remote sensing and in situ-based assessment of rapidly growing South Lhonak glacial lake in eastern Himalaya, India. *Natural Hazards*, 93, 393–409. <https://doi.org/10.1007/s11069-018-3305-0>

Shugar, D.H., Burr, A., Haritashya, U.K., Kargel J. S., Watson, C.S., Kennedy, M.C., Bevington, A.R., Betts R.A., Harrison, S. & Strattman, K. (2020), Rapid worldwide growth of glacial lakes since 1990. *Nature Climate Change*, 10, 939–945. <https://doi.org/10.1038/s41558-020-0855-4>

Veh, G., Korup, O., & Walz, A. (2019), Hazard from Himalayan glacier lake outburst floods. *PNAS*, 117(2), 907-912. <https://doi.org/10.1073/pnas.191489811>

Verjans, V., Leeson, A. A., Nemeth, C., Stevens, C. M., Kuipers Munneke, P., Noël, B., & van Wessem, J. M. (2020), Bayesian calibration of firn densification models. *The Cryosphere*, 14, 3017–3032. <https://doi.org/10.5194/tc-14-3017-2020>

Wang, X., Liu, S., Ding, Y., Guo, W., Jiang, Z., Lin, J., & Han, Y. (2012), An approach for estimating the breach probabilities of moraine-dammed lakes in the Chinese Himalayas using remote-sensing data. *Natural Hazards Earth System Science*, 12(10), 3109–3122. doi:10.5194/nhess-12-3109-2012

Wang, X., Guo, X., Yang, C., Liu, Q., Wei, J., Zhang, Y., Liu, S., Zhang, Y., Jiang, Z., and Tang, Z. (2020), Glacial lake inventory of high-mountain Asia in 1990 and 2018 derived from Landsat images. *Earth System Science Data*, 12, 2169–2182, <https://doi.org/10.5194/essd-12-2169-2020>

Werder, M., Huss, M., Paul, F., Dehecq, A., & Farinotti, D. (2020), A Bayesian ice thickness estimation model for large-scale applications. *Journal of Glaciology*, 66(255), 137-152. doi:10.1017/jog.2019.93

Wilson, R., Harrison, S., Reynolds, J., Hubbard, A., Glasser, N.F., Wünderlich, O., Anaconda, P.I., Mao, L., & Shannon, S. (2019), The 2015 Chileno Valley glacial lake outburst flood, Patagonia. *Geomorphology*, 332, Pages 51-65, ISSN 0169-555X, <https://doi.org/10.1016/j.geomorph.2019.01.015>

Yao, X., Liu, S., Sun, M., Wei, J., and Guo, W. (2012), Volume calculation and analysis of the changes in moraine-dammed lakes in the north Himalaya: a case study of Longbasaba lake, *Journal of Glaciology*, 58, 753–760

Zemp, M., Huss, M., Thibert, E. *et al.* 2019, Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature*, 568, 382–386. <https://doi.org/10.1038/s41586-019-1071-0>

Zhang, G., Bolch, T., Yao, T., Rounce, D. R., Chen, W., Veh, G. (2023), Underestimated mass loss from lake-terminating glaciers in the greater Himalaya. *Nature Geoscience*, 16, 333–338. doi:10.1038/s41561-023-01150-1

Zheng, G., Mergili, M., Emmer, A., Allen, S., Bao, A., Guo, H., & Stoffel, M. (2021), The 2020 glacial lake outburst flood at Jinwuco, Tibet: causes, impacts, and implications for hazard and risk assessment. *The Cryosphere*, 15, 3159–3180, <https://doi.org/10.5194/tc-15-3159-2021>