Alexandra Pulwicki
December 9, 2016

Results
Basin Scale

Calculating DEM - compiling and correcting DEMS
Params - different rasters - get equations for topo params
MLR Results - MLR coeffs table with R2 values

Overview

This is a document that shows the sampled and full ranges of topographic parameters and then goes into the multiple linear regression and Bayesian model averaging that was used to explain SWE with topographic parameters.

Contents

1 Topographic parameters of study sites

1.1 Creating DEM

1.2 Calculating topographic parameters

One method of interpolating snow water equivalent (SWE) is to relate the observed SWE with topographic parameters derived from a digital elevation model (DEM) of the study area. The imagery we were using was from the SPOT-5 DEMs and were provided at no cost by the French Space Agency (CNES) through the SPIRIT International Polar Year project (Korona and others, 2009). The DEM has a cell size of 40x40 m. The following topographic parameters were calculated from the DEM in QGIS:

Elevation values were taken from the SPOT-5 DEMs directly.

Distance from centreline was calculated as the minimum distance between the Easting and Northing of the northwest corner of each cell and a centreline that was drawn by hand in QGIS. This was completed in Matlab using the script 'CentrelineDistance.m'.

Slope is the second derivative of the elevations.

Tangential Curvature represents the curvature in the direction of the contour tangent.

Profile Curvature represents the curvature in the direction of the the steepest slope.

"Northness" is defined as the product of the cosine of aspect and sine of slope (Molotch and others, 2005). A value of -1 represents a steep, south facing slope, a value of +1 represents a steep, north facing slope, and a flat surfaces yield 0.

Aspect represent what direction a slope is facing with 0° defined as north. These were calculated using Terrain Analysis in QGIS.

Sx represents wind exposure/shelter and is based on selecting a cell within a certain angle and distance from the cell of interest that has the greatest upward slope relative to the cell of interest (Winstral and others, 2002). This can be referred to as the maximum upwind slope. Negative Sx values represent exposure relative to the shelter-defining pixel, which means that the cell of interest was higher than the cell with greatest upward slope. Conversely, positive values represent sheltered cells. To determine Sx values, we use the equation

$$Sx_{A,dmax}(x_i, y_i) = \max\left[\tan^{-1}\left(\frac{z(x_v, y_v) - z(x_i, y_i)}{[(x_v - x_i)^2 + (y_v - y_i)^2]^{1/2}}\right)\right], \quad (1)$$

where A is the azimuth of the search direction, (x_i, y_i) are the coordinates of the cell of interest, and (x_v, y_v) are the set of all cell coordinates located along the search vector defined by (x_i, y_i), the azimuth (A), and maximum search distance ($d$max). Code for this calculation was provided by Adam Winstral (2016, personal

Table 1: Values of azimuth (A) and maximum search distance($d$max), which define the Sx that had the highest correlation to observed SWE.

	A (° from North)	***d*max (m)**
Glacier 4	75	200
Glacier 2	55	200
Glacier 13	325	200

communication). As done by McGrath and others (2015), we computed Sx at 5° azimuth increments for $d$max distances of 100, 200, and 300 m. These values were then correlated with observed values of SWE and the A and $d$max values with the highest correlation were used for subsequent analysis (see Table 1).

1.3 Topographic parameters from QGIS to Matlab

The values of each topographic parameters at the sampling locations was determined in QGIS. The sampling locations were imported to QGIS and the Point Sampling Tool was used to determine the value of the cell of the raster for each topographic parameter at that location. This set of parameters that corresponded to each location was then exported to a .csv file and imported into Matlab with the script 'Import_Topo.m'. Note that selection of Sx values was completed first (see above) and only the Sx data with highest correlation to SWE was exported. After importing to Matlab, the sets of parameter values (x_i) were standardized (x_s) using the equation $x_s = \frac{x_i - \mu}{\sigma}$. The resulting structure is called `topo_sampled` and is used in the MLR to be able to compare the explanatory power of each parameter (Section 2.1). A non-standardized copy of the topographic parameters at the sampling locations was also kept for plotting purposes and is called is called `topo_sampled_ns`. Both structures are organized as a vector of values corresponding to the vectors in the `SWE` variable.

The raster data for each topographic parameter was also exported from QGIS as a .csv file and then imported into Matlab with the script 'Import_Topo.m'. The values are stored in the structure `topo_full` and are a matrix, where each cell corresponds to one DEM cell. Cells outside of the glacier outline have no value (`NaN`).

1.4 Parameter correlation

The correlation between topographic parameters at sampling locations on each glacier is shown in Table 2. Correlation values were generally low, with the exception of the correlation between northness and aspect on Glacier 4 and northness and elevation on Glacier 2, which were both negative and larger than 0.7. Since there is little correlation and correlation that varies between glaciers, the use of a multiple linear regression with these topographic parameters as predictor variables is warranted.

Table 2: Pearson correlation coefficients between topographic parameters at sampled locations. The parameters include aspect (α), elevation (z), northness (N), profile curvature (κ_P), slope (m), tangent curvature (κ_T), wind redistribution (Sx) and distance from centreline (d_C).

Glacier		α	z	N	κ_P	m	κ_T	Sx	d_C
Glacier 4	α	1	0.52	-0.78	0.24	-0.2	0.12	0.30	-0.12
	z	0	1	-0.53	0.16	<0.01	0.15	0.21	0.20
	N	0	0	1	-0.33	0.07	-0.13	-0.49	0.09
	κ_P	0	0	0	1	-0.20	0.44	-0.30	-0.16
	m	0	0	0	0	1	<0.01	0.20	0.36
	κ_T	0	0	0	0	0	1	-0.28	-0.01
	Sx	0	0	0	0	0	0	1	0.07
	d_C	0	0	0	0	0	0	0	1
Glacier 2	α	1	-0.02	0.04	0.13	-0.16	-0.02	-0.02	-0.12
	z	0	1	-0.62	0.06	-0.46	0.02	0.14	0.05
	N	0	0	1	-0.03	0.5	0.09	-0.15	0.32
	κ_P	0	0	0	1	-0.01	0.45	0.01	0.08
	m	0	0	0	0	1	-0.02	-0.07	0.12
	κ_T	0	0	0	0	0	1	-0.03	0.19
	Sx	0	0	0	0	0	0	1	-0.18
	d_C	0	0	0	0	0	0	0	1
Glacier 13	α	1	-0.24	-0.07	0.02	-0.1	0.02	0.08	0.06
	z	0	1	0.14	0.08	0.01	0.01	-0.24	0.13
	N	0	0	1	0.06	0.24	0.08	-0.01	0.07
	κ_P	0	0	0	1	-0.02	0.37	-0.04	0.07
	m	0	0	0	0	1	-0.15	-0.02	0.09
	κ_T	0	0	0	0	0	1	<0.01	<0.01
	Sx	0	0	0	0	0	0	1	-0.06
	d_C	0	0	0	0	0	0	0	1

1.5 Maps of topographic parameters and range of parameters sampled

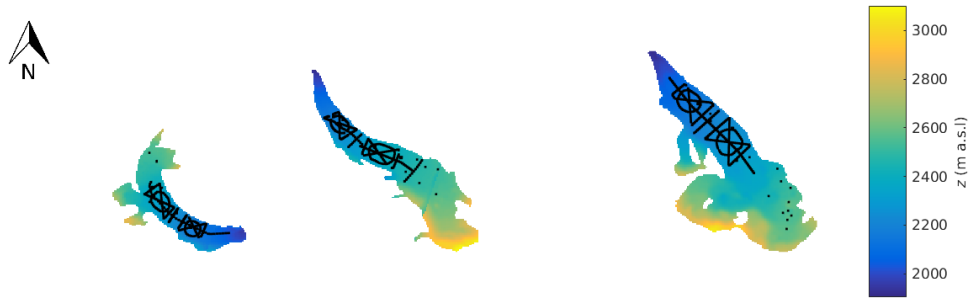

Figure 1: Values of elevation used in the topographic regressions for three study glaciers. This DEM is derived from a SPOT5 satellite image and has a grid size of 40x40 m. Subsequent topographic parameters were derived from this DEM.

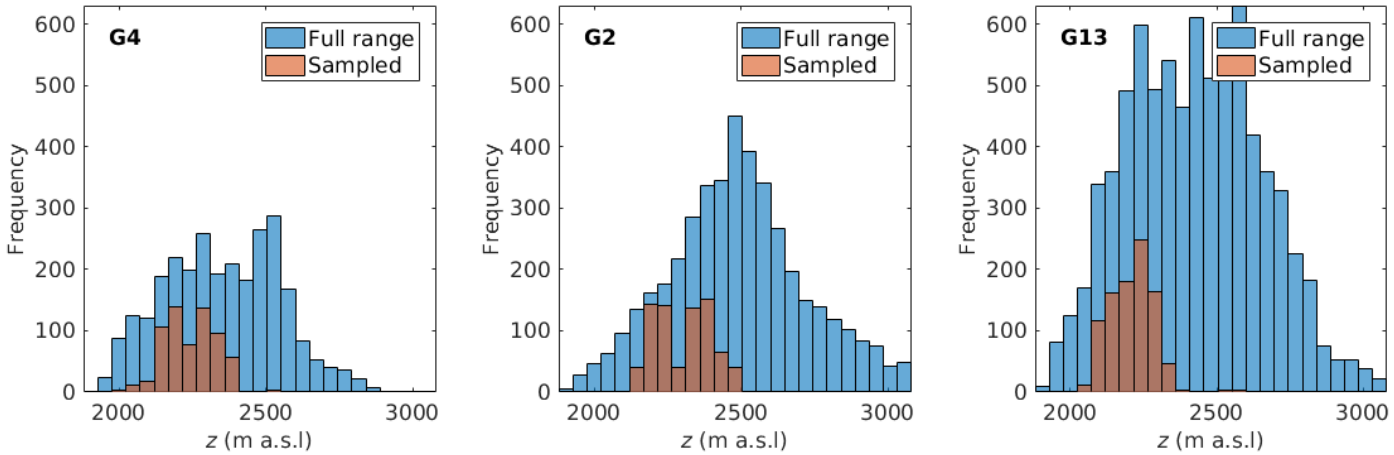

Figure 2: Range of elevation sampled as compared to total range of elevation of study glaciers.

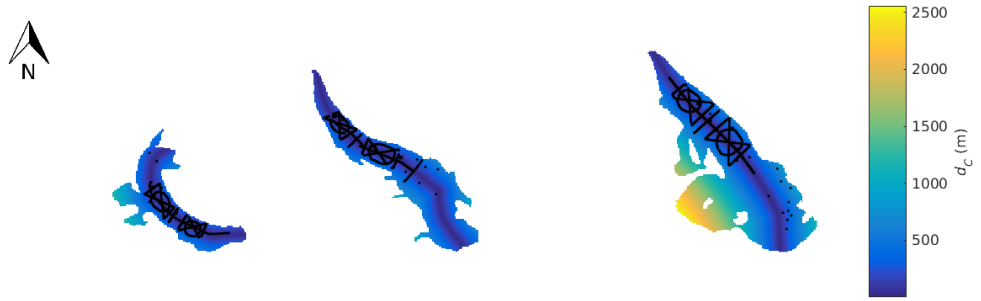

Figure 3: Values of distance from centreline used in the topographic regressions for three study glaciers. Centreline was drawn by hand in QGIS.

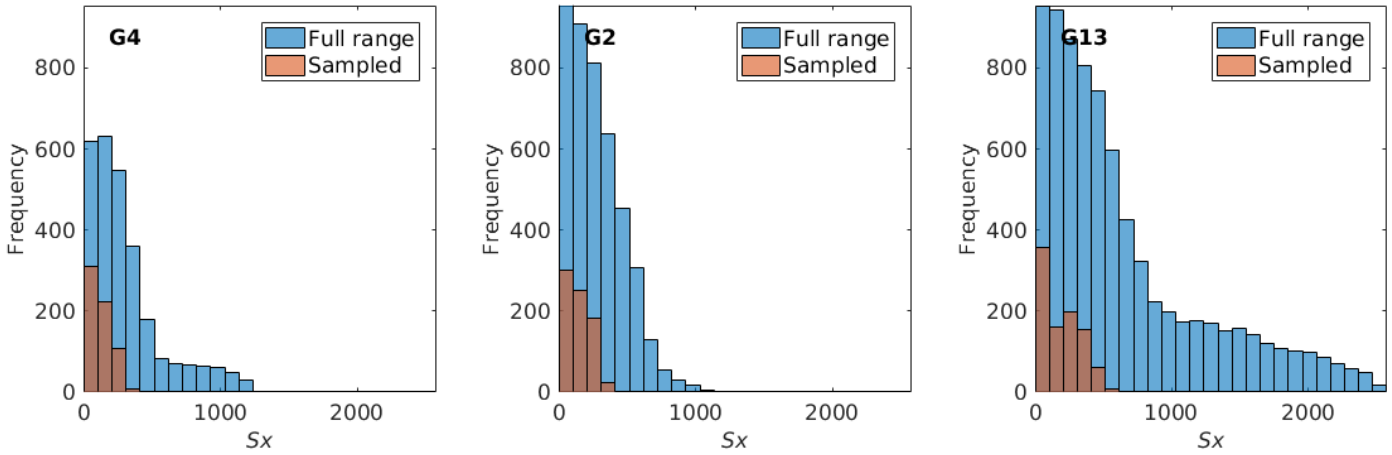

Figure 4: Range of distance from centreline sampled as compared to total range of distance from centreline of study glaciers.

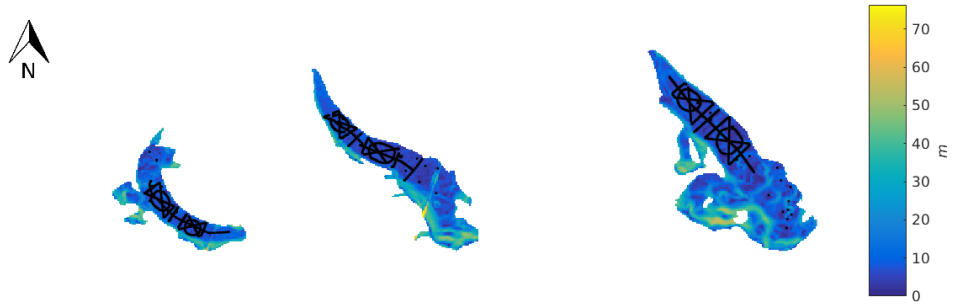

Figure 5: Values of slope used in the topographic regressions for three study glaciers. Values were derived from a SPOT5 satellite derived DEM (grid size of 40x40 m) in QGIS.

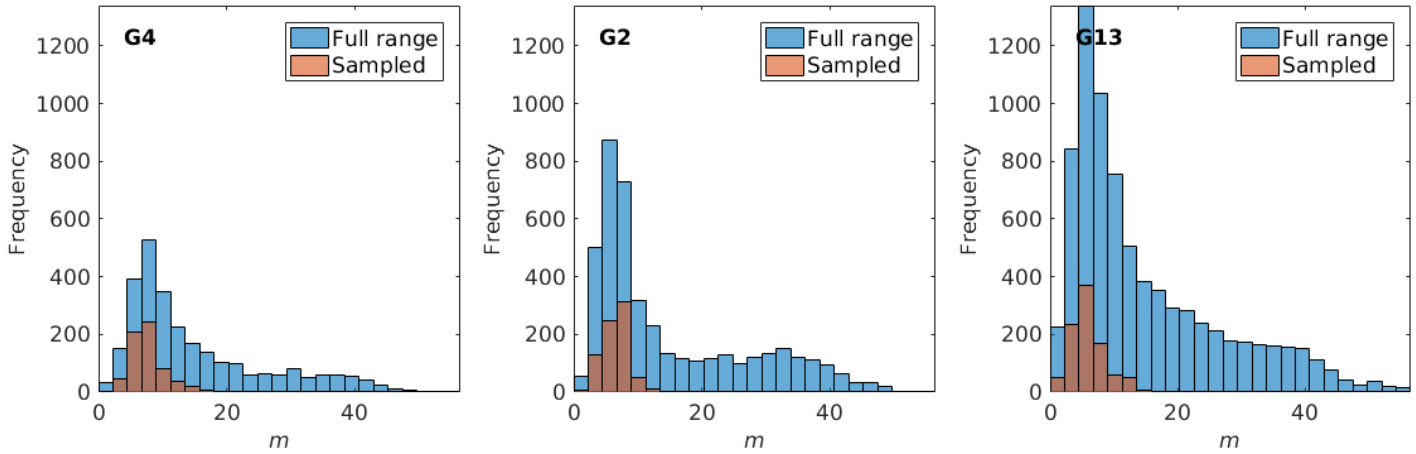

Figure 6: Range of slope sampled as compared to total range of slope of study glaciers.

Figure 7: Values of tangential curvature used in the topographic regressions for three study glaciers. Values were derived from a SPOT5 satellite derived DEM (grid size of 40x40 m) in QGIS.

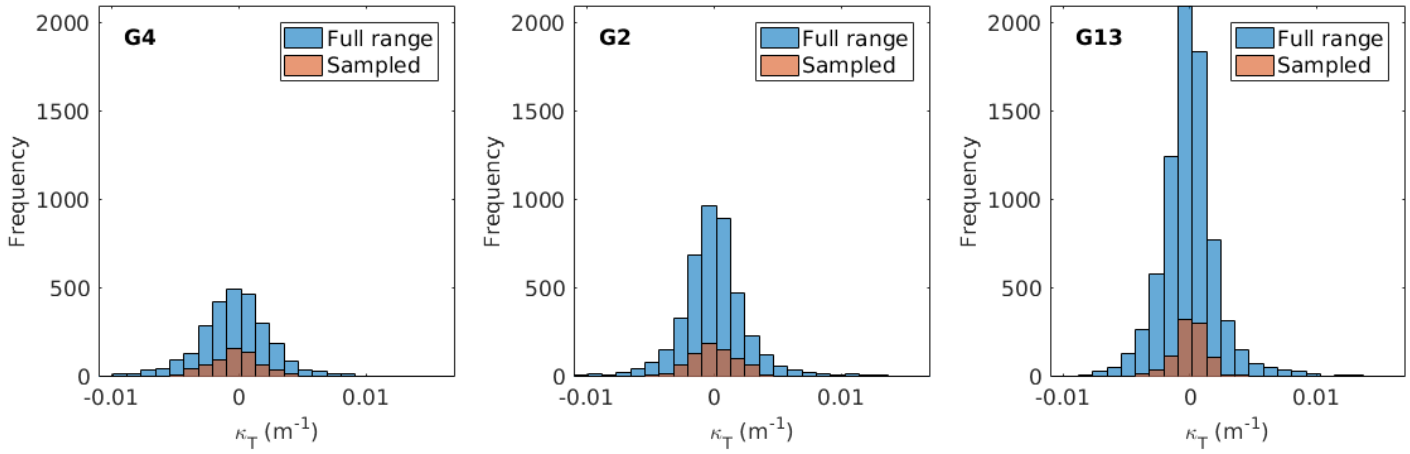

Figure 8: Range of tangential curvature sampled as compared to total range of tangential curvature of study glaciers.

Figure 9: Values of distance from centreline used in the topographic regressions for three study glaciers. Values were derived from a SPOT5 satellite derived DEM (grid size of 40x40 m) in QGIS.

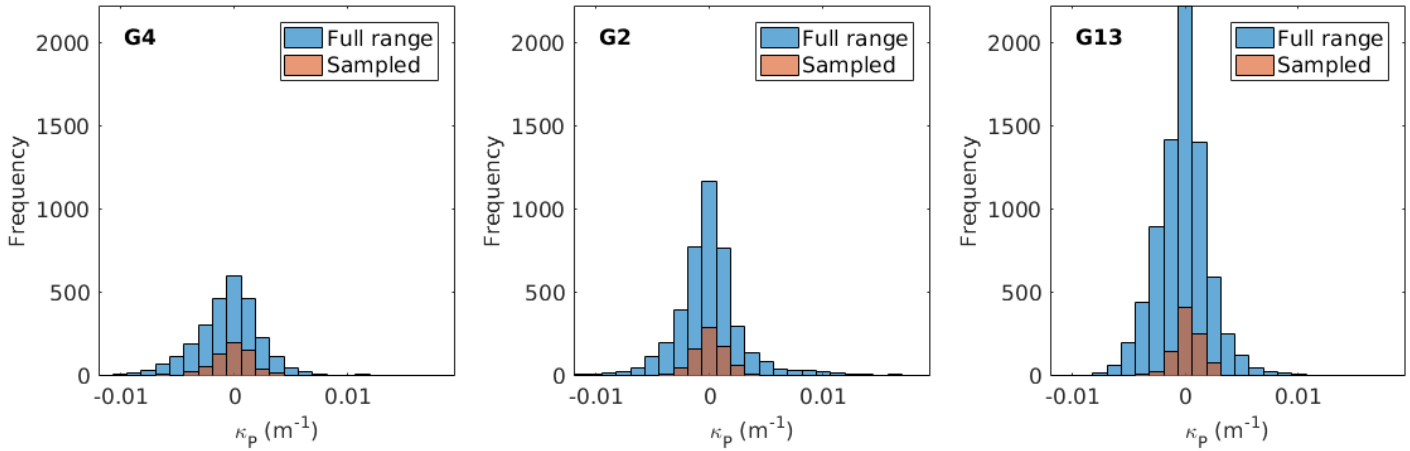

Figure 10: Range of profile curvature sampled as compared to total range of profile curvature of study glaciers.

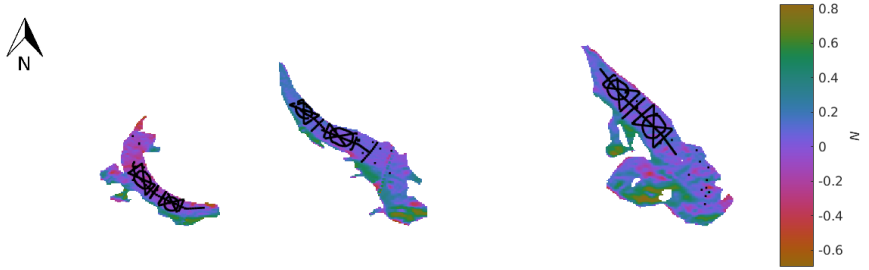

Figure 11: Values of "northness" used in the topographic regressions for three study glaciers. "Northness" is defined as the product of the cosine of aspect and sine of slope. A value of -1 represents a steep, south facing slope, a value of +1 represents a steep, north facing slope, and flat surfaces yield 0. Values were derived from a SPOT5 satellite derived DEM (grid size of 40x40 m) in QGIS.

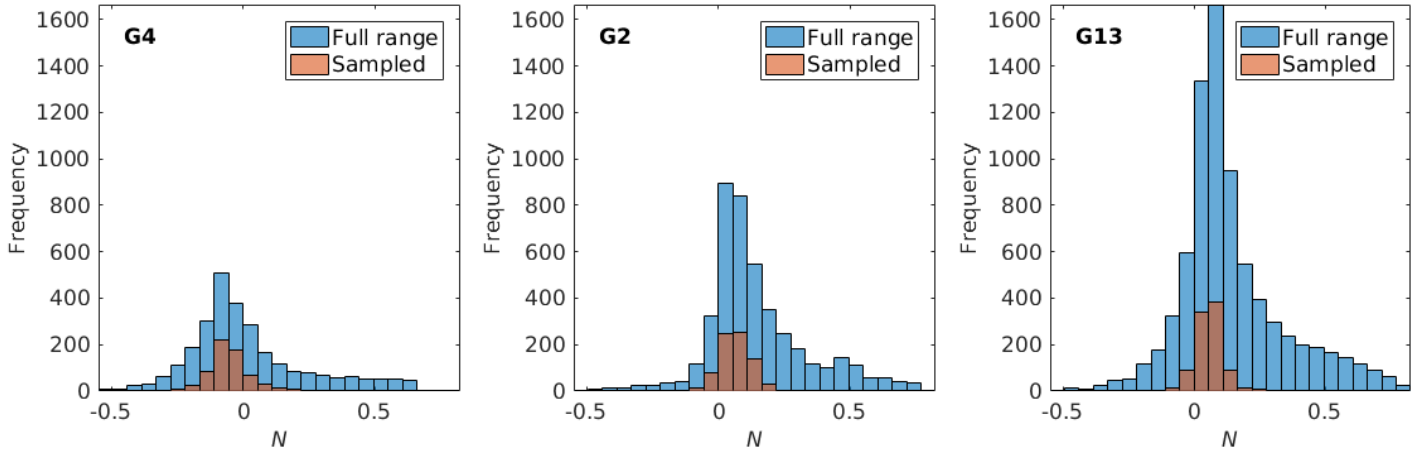

Figure 12: Range of "northness" sampled as compared to total range of "northness" of study glaciers.

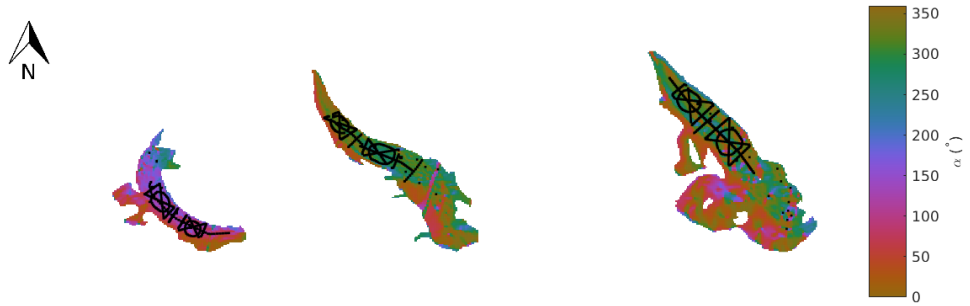

Figure 13: Values of aspect, which represent what direction a slope is facing (0° defined as north), used in the topographic regressions for three study glaciers.

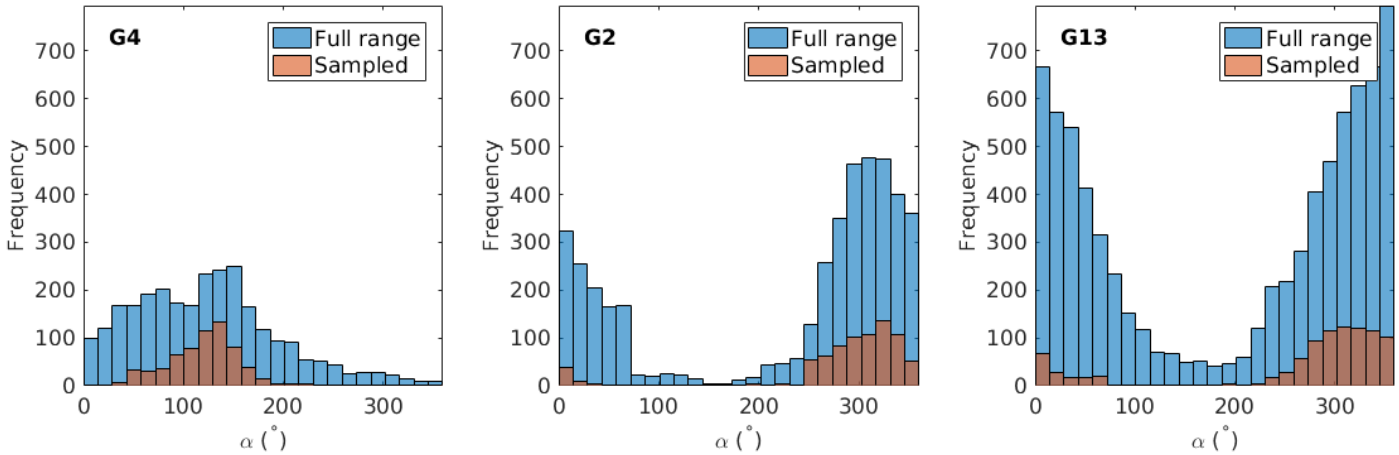

Figure 14: Range of aspect sampled as compared to total range of aspect of study glaciers.

Figure 15: Values of Sx, which is a wind redistribution parameter, used in the topographic regressions for three study glaciers. See section **??** and the original paper by Winstral and others (2002) for more details on calculation .

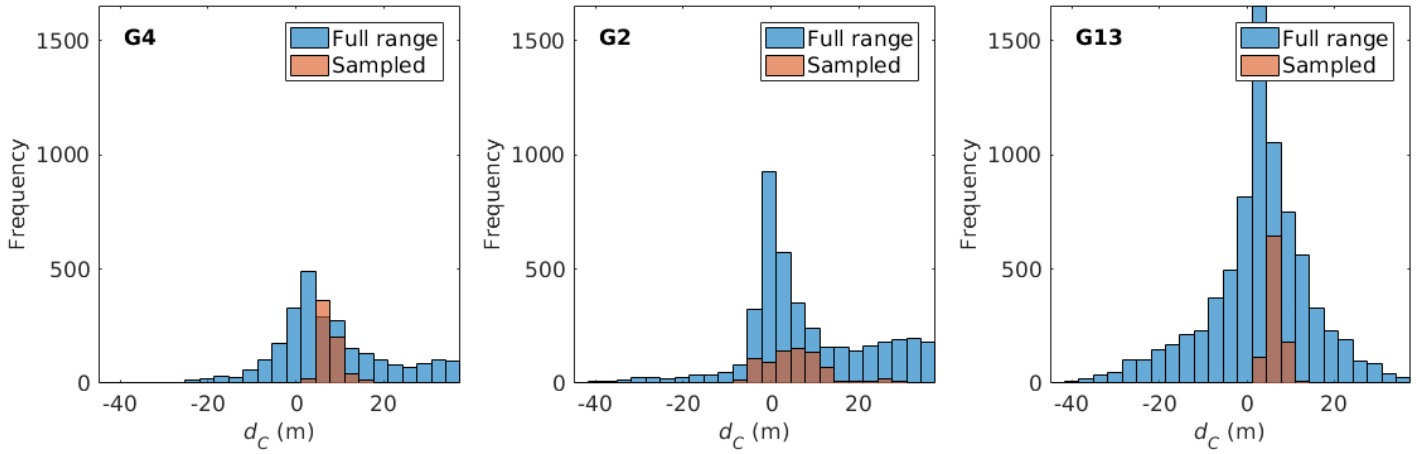

Figure 16: Range of Sx sampled as compared to total range of Sx of study glaciers.

2 Linear Regressions

Relating snow accumulation and terrain parameters to better predict accumulation within a basin has been employed for decades (e.g. **?**Molotch and others, 2005; McGrath and others, 2015). The most common type of relation between topographic parameters and accumulation is a linear regression, where the observed snow water equivalent (SWE) is related to a linear combination of topographic parameters at each measurement location.

A linear regression takes the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where the matrix $\mathbf{X}$ contains the set of independent regressors $\mathbf{x}_i$ used to explain the dependant variable $\mathbf{y}$. The regression coefficients for each regressor is given by $\boldsymbol{\beta}$ and the error or noise of the system is given by $\boldsymbol{\varepsilon}$. Applied to this study, the matrix of independent regressors ($\mathbf{X}$) are the topographic parameters at the sampling locations, the dependant variable $\mathbf{y}$ is the observed SWE, and the $\boldsymbol{\beta}$ values are determined using a fitting model. Two types of fitting models are the multiple linear regression (MLR) and Bayesian model averaging (BMA).

The $\boldsymbol{\beta}$ values can then be applied to other locations where regressors are known to calculate values of the dependant variable. For example, known values of topographic parameters can be multiplied by their respective $\boldsymbol{\beta}$ coefficients and added together to obtain the modelled or predicted value of SWE for that DEM grid cell.

2.1 Multiple Linear Regression

2.1.1 Background

Perhaps the most basic and well used method for relating SWE and topographic parameters is a multiple linear regression (MLR) (e.g. **?**). The best fit line of an MLR is the one described by coefficients that minimize the sum of squares of the vertical deviations of each data point from the line. Note that if a point falls on the line then the deviation is zero and that the positive and negative deviations from the line do not cancel because the values are first squared and then summed. The residuals are simply the difference between the modelled and observed data.

To reduce the chance of over fitting the data, two additional methods were applied. The first method was model averaging over all possible models that allow for linear combinations of the chosen topographic parameters. An MLR was estimated for each model (total number of models is 2^n, where n is the number of topographic parameters) and the Bayesian information criterion (BIC) value was computed for the model according to

$$\mathrm{BIC} = -2\ln L(\hat{\theta}_k|y) + k\ln(n), \tag{3}$$

where the values of $\hat{\theta}_k$, which are the model parameters, maximizes the likelihood function for data y. The number of data points is n and the number of regressors is k. BIC values are used to assess the relative predictive success of models while penalizing for overfitting

of data. While the absolute BIC value is meaningless, models can be selected or averaged using the relative BIC value, with lower values indicating a better model.

In this study, there was no reason to favour any of the models so a weighted average of all models based on their BIC values was conducted. The BIC values for each model (BIC_i) were used to determine the normalized weight of each model (w_i) relative to the best model (lowest BIC value BIC_{min}) according to (Burnham and Anderson, 2004)

$$w_i = \frac{e^{-0.5(BIC_i - BIC_{min})}}{\sum_{i=1}^{R} e^{-0.5(BIC_i - BIC_{min})}}. \tag{4}$$

When parameters were excluded from a model, a coefficient of zero was assigned to that parameter. The sum of the weighted coefficients gave the final $\boldsymbol{\beta}$ values.

The second method used to prevent over fitting was to calculate the MLRs within a cross validation framework. This meant that for each model, a randomly selected portion of the data was used to estimate regression coefficients and these were used to predict values that correspond to the remaining data. The root mean squared error (RMSE) between the modelled and observed data was then calculated. This process was done one thousand times and the set of coefficients that resulted in the lowest RMSE were then chosen for that model.

Together, cross validation allowed for reputable choosing of regression coefficients for each possible model and model averaging with BIC weights allowed for stability of final coefficient values and favouring of simpler models, which would help to identify main topographic drivers of snow accumulation.

2.1.2 Methods

A multiple linear regression (MLR) between snow water equivalent (SWE) and all linear combinations of the calculated topographic parameters was completed. This methodology for calculating the MLR also involved using cross validation, where a portion of the data is withheld from the MLR fit and these points are then used to assess the ability of the MLR to predict values. The MLR fit with the best prediction ability is then carried forward.

The MLR was completed with the following steps (executed using the function 'MLRcalval.m'):

1. The topographic parameters
2. The `topo_sampled` structure for one glacier as well as the `SWE` strcture is passed into the function.
3. A set of initializations is completed. This includes 1) creating a logical matrix to choose all linear combination of topographic parameters, 2) selecting the number of runs, 3) creating a matrix of random numbers for selecting data point in the cross validation procedure, 4) initializing matrices, and 5) converting the input structure to a table.

4. For each linear combination of topographic parameters, one thousand runs of a cross validation MLR are then executed. Three-quarters of the total data is randomly selected to use for calculating the regression (using the function `regress()` which is a basic regression function with fast execution). The MLR equations is used to predict the SWE using the remaining one-quarter of the topographic parameters. The root-mean-squared error (RMSE) between the predicted and observed SWE values is then calculated and the MLR equation with the lowest RMSE is then chosen for that combination of topographic parameters. The function `fitlm()` is then used to calculate the MLR from the set of data that gave the lowest RMSE. This function is slower but calculates a number of additional values that characterize the fit of the model. One of these values is the Bayesian information criterion (BIC), which allows for model selection among a finite set of models (Burnham and Anderson, 2004). The BIC from the best model for each combination of topographic parameters is saved.

5. A weighted sum of all models found using linear combinations of topographic parameters was then found. The BIC values for each model (BIC_i) were used to determine the normalized weight of each model (w_i) relative to the best model (lowest BIC value BIC_{min}) according to Eq. 4.

6. The percent variance ($var_{\%}$) explained by each parameter was calculated using the equation $var_{\%} = \frac{SSr}{SSt} \times 100$, where SSr is the sum of squares of the residual (fitted topographic parameters) and SSt is the total sum of squares (SWE observations). The final coefficients and the percent varience explained by each one can be found in the `coeffs_final` table within the function and in the `mlr` structure when run for all density options and glaciers in the main script 'TopoRegression.m'.

7. The residuals of the fit have also been calculated as a separate variable than can be returned when the function is called. The residual was calculated as the difference between the predicted and observed value.

2.1.3 Results

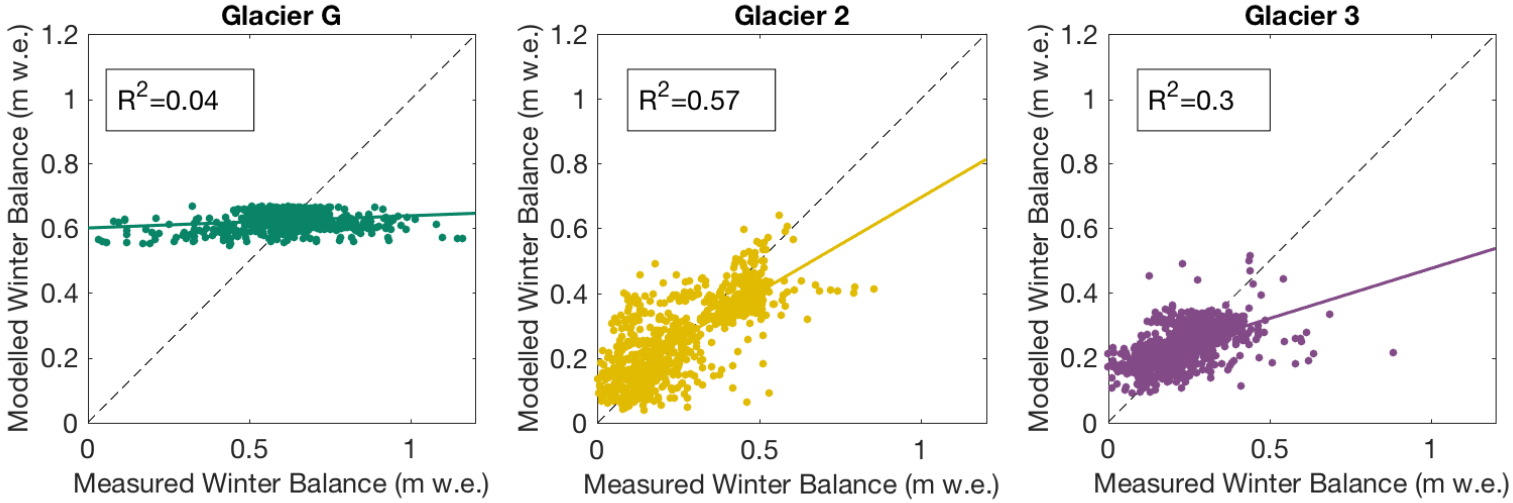

Figure 17: Comparison of predicted (MLR) and observed (original) snow water equivalent (SWE) for three study glaciers. The SWE values were calculated inverse distance weighted snowpit densities (Option 8).

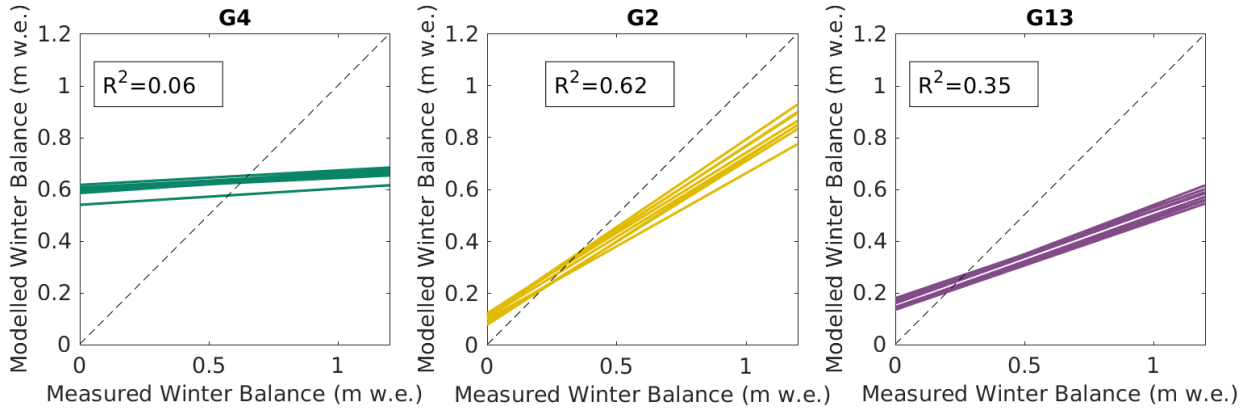

Figure 18: Plot of all linear fits between modelled and observed SWE using eight options for calculating density. Mean R^2 value is shown for each sub-plot and a reference 1:1 line is also provided. See Figure 17 for a plot of the data.

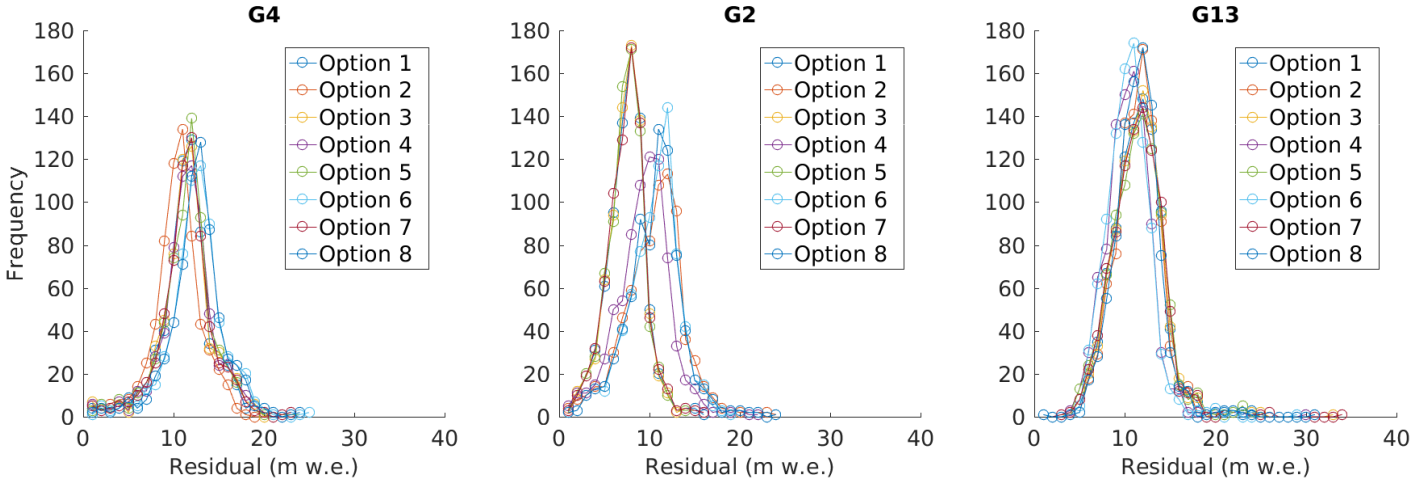

Figure 19: Residuals of SWE predicted using MLR for all options of estimating density.

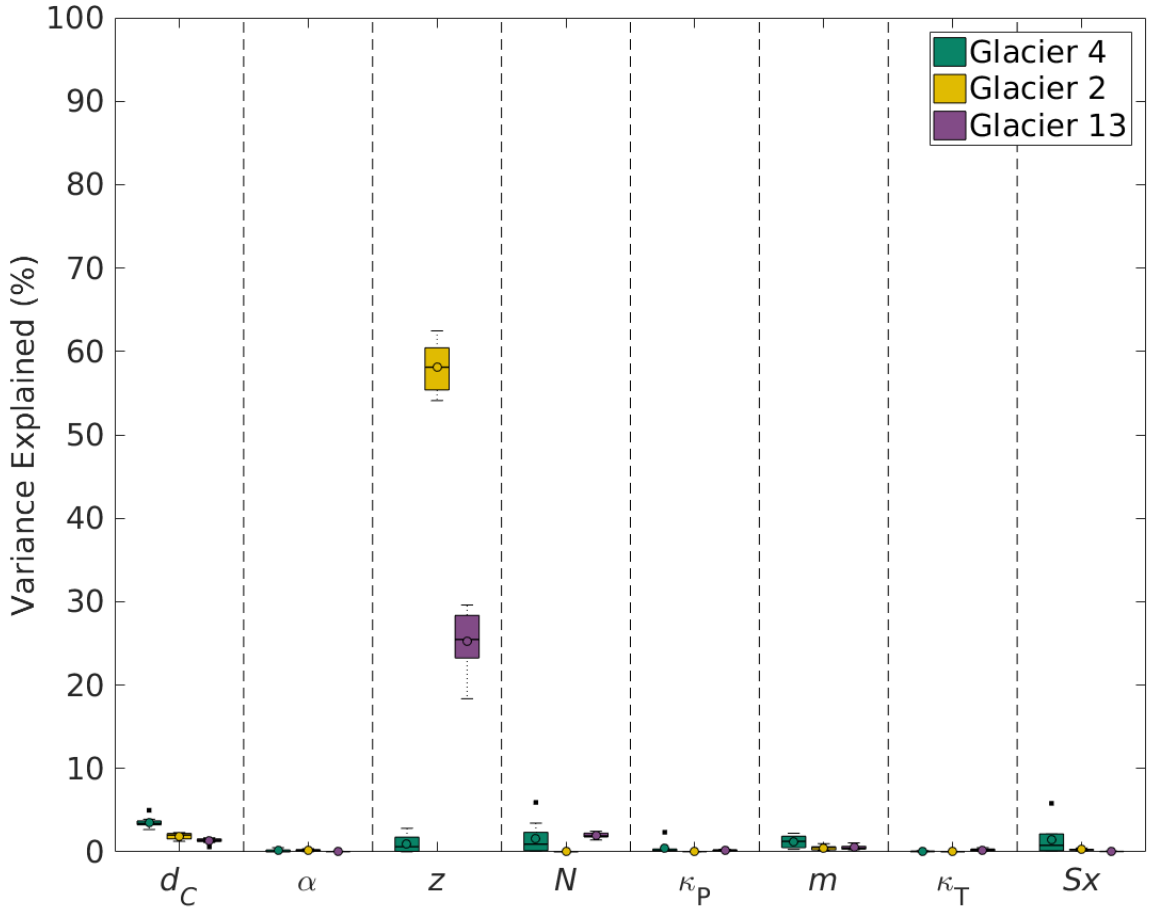

Figure 20: Boxplot showing the range of percent variance explained by each topographic parameter for each option of estimating snow water equivalent (SWE). Topographic parameters include aspect (α), elevation (z), northness (N), profile curvature (κ_P), slope (m), tangent curvature (κ_T), wind redistribution (Sx) and distance from centreline (d_C).

2.2 Bayesian Model Averaging

2.2.1 Background

Bayesian model averaging (BMA) is a method of estimating all possible linear combinations of predictor variables and then averaging over all models (**???**). This method is based on Bayesian principals where we wish to determine the probability of an outcome based on an initial "guess" as well as the data provided. Given that the predictive outcome has a probability distribution of x given y, written as $P(x|y)$, we use Bayes' rule to to write this as

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}. \tag{5}$$

$P(x|y)$ is often called the posterior model probability (PMP). The quantity $P(y|x)$ is the likelihood function, which determines unknown parameters from a known outcome (i.e. observed data). The term $P(x)$ is an observer determined prior probability distribution (typically just called a *prior*) and it reflects the prior knowledge of the system. Choosing an appropriate prior is one of the most challenge components of Bayesian probability theory. The$P(y)$ term can be obtained by integrating $P(y|x)P(x)$ over all x and is thus a constant for all models that is typically discarded.

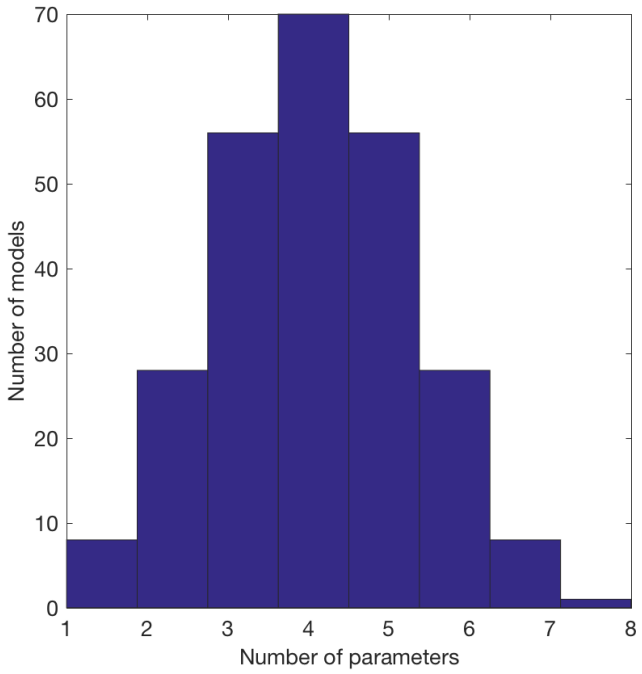

Figure 21: Uniform model prior for eight topographic regressors used in BMA.

Together then, the posterior model probability is a function of both the model prior, specified by the researcher, as well as the distribution of the observed data — it is the transformation of the prior as a result of considering collected data. This can be loosely expressed as

$$\text{posterior} \propto \text{prior} \times \text{likelihood} \tag{6}$$

If the prior is uninformative then the posterior will be strongly influded by the data. An informative prior will result in a posterior that is a mix of the prior and the data as as the prior becomes more informative, the amount of data needed to transform the distribution increases. If there is a large amount of data then the prior will have little effect on the posterior. The final value of the desired output is often reported as the posterior distribution mean or the value that maximizes the log-likelihood.

Within BMA, Bayes' theorem is used to find the posterior model probability and this is used as a weight when averaging over all models. The model weighted posterior distribution for the coefficients β of K number of models after renormalization is given by (**?**)

$$P(\beta\ |y, X) = \sum_{i=1}^{2^K} P(M_i|X, y)P(\beta|M_i, y, X), \tag{7}$$

where the responding variable is given by y and the matrix of variables is given by X. Here, the model prior is $P(M_i|X, y)$ and the likelihood for the β coefficient is $P(\beta|M_i, y, X)$.

There are a number of different priors to describe model size distribution that have been applied in BMA. A commonly used prior is the 'uniform' prior, which assumes a

normal model distribution with a total of 2^n models, where n is the number of regressors. This model prior states that the observer has no knowledge of the system and all models are equally likely. The uniform prior has a prior model probability of the form $P(M_i) = 2^{-n}$, as shown in Figure 21, which is symmetric about the mean $n/2$. This prior inherently favours models of an intermediate size.

Other types of priors include those that are skewed to favour smaller models, ones with equal probability for all model sizes, and ones with varying probability for individual regressors. In this case, a uniform prior was chosen because there was no knowledge of what the model distribution would look like and to match the distribution of models sizes in the MLR analysis.

With a small number of regressors, the posterior of all possible models can be determined. As the number of regressors increases, this computation becomes increasingly expensive since the total number of models is 2^n. To overcome this, BMA can use Markov chain Monte Carlo (MCMC) model composition to directly approximate the posterior distribution. In our study, there are eight regressors so $2^8 = 256$models. It is possible to visit all models and to obtain an exact solution so a (MCMC) model composition was not employed.

BMA allows for the calculation of a metric called the posterior inclusion probability (PIP), which is used to evaluate the importance of a regressor in explaining the observed data. PIP is the sum of all posterior model probabilities (PMP) where the variable was included in the model. A higher PIP indicates that the regressor is more important in the regression.

2.2.2 Methods

The BMA process was implemented in R, using the Bayesian model statistics (BMS) package developed by **?**. The function `BMS_R()` computes the β coefficients for all topographic parameters for all study glaciers as well as the percent variance explained by each parameter suing the following steps:

1. The SWE observations and associated topographic parameters are called in the function.
2. A portion of the data (3/4) is randomly chosen as the calibration set and saved as a .mat file.
3. The R script is called through the terminal
   (a) The R script imports the .mat file with SWE and topographic parameter values. It then creates a data frame with the SWE values as the first column and the topographic parameters as the remaining columns.
   (b) The BMS package is used to complete BMA for the imported values and the mean coefficient value, coefficient standard deviation, PIP, and the posterior probability of a positive coefficient (how probable it is for the sign of the coefficient to be positive).
   (c) The coefficients are saved as a .mat file.

4. Values calculated in R are loaded into Matlab and a data table is created with the coefficients.
5. The remaining portion of the topographic parameter values (1/4) are then used to calculate a modelled value of SWE at those locations. These are compared to the observed SWE values and a RMSE value is determined.
6. The above step are completed 1000 times and the coefficients associated with the lowest RMSE are chosen.
7. Percent variance explained by each parameter is the calculated in the same way as for the MLR (see Section 2.1.2).
8. The final table of values includes the coefficients and percent variance explained for all topographic parameters associated with the lowest RMSE. It also includes the intercept and the actual RMSE value. This table is returned from the function.
9. The residuals of the best BMA fit are also calculated and can be return from the function.

References

Burnham KP and Anderson DR (2004) Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, **33**(2), 261–304 (doi: 10.1177/0049124104268644)

Korona J, Berthier E, Bernard M, Rémy F and Thouvenot E (2009) SPIRIT. SPOT 5 stereoscopic survey of Polar Ice: reference images and topographies during the fourth International Polar Year (2007–2009). *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**(2), 204–212

McGrath D, Sass L, O'Neel S, Arendt A, Wolken G, Gusmeroli A, Kienholz C and McNeil C (2015) End-of-winter snow depth variability on glaciers in Alaska. *Journal of Geophysical Research: Earth Surface*, **120**(8), 1530–1550 (doi: 10.1002/2015JF003539)

Molotch N, Colee M, Bales R and Dozier J (2005) Estimating the spatial distribution of snow water equivalent in an alpine basin using binary regression tree models: the impact of digital elevation data and independent variable selection. *Hydrological Processes*, **19**(7), 1459–1479 (doi: 10.1002/hyp.5586)

Winstral A, Elder K and Davis RE (2002) Spatial snow modeling of wind-redistributed snow using terrain-based parameters. *Journal of Hydrometeorology*, **3**(5), 524–538