

MASTER THESIS

Influence of Glacier Morphology on the Specific Mass Balance Variability in High Mountain Asia

A Comparison Between Homo- and Heterogeneous Climate Settings



Utrecht University

Author:

Bo VAN DEN BOSCH
Student no. 5718791

Supervised by:

dr. Philip KRAAIJENBRINK
prof. dr. Walter IMMERZEEL

June 29, 2022

Abstract

Globally, climate change causes glaciers to retreat. Driving mechanisms at local scales are poorly understood. This study aims to discover which climatic and morphological variables most contribute to explaining the Specific Mass Balance (SMB) variability of 9098 individual glaciers in High Mountain Asia (HMA). We separate the data into 15 regions assumed climatically homogeneous. We calculated correlations between the selected variables and the SMB. Additionally, we calculated a multivariate linear model and assessed a linear combination of the variables that best explain the SMB variability. We exclude all predictor-predictor correlations $> |0.7|$. We concluded that in the heterogeneous climate setting (HMA), the selected variables explain 30.7% of the glaciers' SMB variability, with the most important predictors being the presence of a lake, the slope and the mean temperature and precipitation in 2000-2020. All predictors, except for the slope, are found to be associated with negative SMB. In the subregions, we conclude that the selected variables explain 18.1% to 50.0% of the SMB variability, with generally the most important predictors being the morphological variables: the presence of a Lake, the slope, and the median elevation. The slope and the median elevation are found to be associated with positive SMB. In all analyses, we observed a large influence of glacial lakes on the SMB. We expect this partly results from unobserved correlations between the variables lake and slope. Only the subregion West Kun Lun was associated with positive SMB. Here, we expect this contrasting behaviour to result from the dry and cold climate settings. A major limitation of this study is the use of a linear model for non-linear data. This resulted in relatively low model performance in the climatically heterogeneous region HMA.

Contents

1	Introduction	3
1.1	Relevance	3
1.2	Spatial Variability in Glacier Mass Balance	3
1.3	Research Question	4
2	Data	5
2.1	Documentation of Variables	5
2.2	Data Prepossessing and Outlier Detection	8
2.3	Missing Data Analyses	8
3	Methods	9
3.1	Pearson Correlations	9
3.2	Best Subset Selection with LASSO	9
3.3	Ordinary Least Squares Multivariate Regression with LOOCV	10
4	Results	11
4.1	Data Exploration Results	11
4.2	Predictor-Predictor Pearson Correlations	13
4.3	Predictor-SMB Pearson Correlations	14
4.4	LASSO Regression	15
4.5	OLS Multivariate Regression	16
5	Discussion	18
5.1	Contrasting Behaviour of West Kun Lun glaciers	18
5.2	Influence of Lakes and Slopes on Glacier Mass Balance	18
5.3	Influence of Debris on Glacier Mass Balance	18
5.4	Non-linear Glacier Response to Climate Variables	18
5.5	Limitation of the Analysis	19
6	Conclusion and Outlook	20
A	Acknowledgements	24
B	Code Availability	24
C	Supplementary Tables and Figures	25
C.1	Distribution of predictors variables	25
C.2	SMB as a function of glacier median elevation	26
C.3	SMB as a function of glaciers slope	27
C.4	Pearson Correlations per Subregion	28
C.5	Regression Residuals per Subregion	33

1 Introduction

1.1 Relevance

Globally, glaciers are extensively impacted by climate change. Glacier retreat has accelerated in recent decades (Viviroli et al., 2011; Bolch et al., 2012). Evidence suggests that the rate of warming is amplified in higher mountain environments (Pepin et al., 2015). The threat of global warming is even more apparent in High Mountain Asia (HMA), where the warming rates are significantly higher than in the Northern Hemisphere and global averages (Yang et al., 2014). Climate change affects glaciers and their crucial role in the global water cycle (Viviroli et al., 2007) as glacier ablation contributes to global sea-level changes and regional water supplies (Bojinski et al., 2014). Accumulation and ablation dynamics are being disturbed, and precipitation and evaporation patterns are shifting, all impacting future changes in the magnitude and timing of water availability (Immerzeel et al., 2020; Bookhagen and Burbank, 2010). These changes can affect the mountain ecosystem and have far-fetching social, ecological and political consequences (Sorg et al., 2012).

Moreover, glacier retreat is associated with the rapid expansion and new formation of glacial lakes (Zheng et al., 2021; Zhang et al., 2015; Shugar et al., 2020). Allen et al. (2016) predicted the future appearance of glacial lakes in HMA and identified 5,000 overdeepenings that are at risk of turning into glacial lakes. These glacial lakes can cause glacial lake outburst floods when water is suddenly released from the lake. Floods of unstable lakes cause great risk up to a hundred kilometres downstream and have killed thousands of people, with the most frequent events observed in HMA (Dubey and Goyal, 2020).

Glaciers store and supply water for environmental and human demands downstream (Viviroli et al., 2011). These so-called 'water towers' provide a seasonal buffer as they provide higher water runoff in the warmer seasons and store water in the form of ice and glaciers in the colder seasons. The water runoff in downstream rivers is essential for irrigation, industrial water supplies, hydrodynamic power production and environmental needs (Viviroli et al., 2011). Shifts in seasonal runoff are already observed in some rivers, and this trend is expected to continue (Sorg et al., 2012). We find it essential to study regional changes in mountain water resources because of the social, ecological, and political consequences associated with water availability.

1.2 Spatial Variability in Glacier Mass Balance

It is widely recognized that glaciers are affected by climate change at a global level. However, different regions may behave remarkably dissimilar. Most glaciers in the Himalayas and on the Tibetan Plateau are losing mass (Bolch et al., 2012; Yao et al., 2012), but they show contrasting shrinkage patterns (Kääb et al., 2012; Maussion et al., 2014). Kaab et al. (2012) examine mass loss in five subregions of the Hindu Kush–Karakoram–Himalaya region (HKKH). On average, HKKH glaciers thinned 0.26 ± 0.06 m yr⁻¹, while glaciers in the northern and eastern parts of the subregion Karakoram thickened 0.14 ± 0.06 m yr⁻¹. These heterogeneous glacier responses can be partly explained by local morphological factors (such as slope, height and debris coverage) but also by the contrasting climatic setting in HMA (Brun et al., 2017; Maussion et al., 2014).

For example, the contrasting behaviour of Karakoram is associated with cooling summer temperatures over the Karakoram (Forsythe et al., 2017). Not only is the temperature highly variable in HMA, but the amount and timing of precipitation also differ across regions. Most glaciers in the western Himalayas and the Karakoram receive 60-70% of the annual accumulation from westerly extratropical cyclones (Shean et al., 2020), while glaciers in central and eastern Himalayas receive 80% of their annual accumulation from the summer monsoons (Bookhagen and Burbank, 2010).

Forsythe et al. (2017) identify heterogeneous climatology as a possible driving mechanism for changing glacial mass balance patterns in the Himalayas. However, heterogeneous climatology partly explains the regional patterns of glacier mass changes. Recent studies show that glacier morphology also explains the

variability of glacier mass balance (Brun et al., 2019). For example, a study by Huss (2012) investigates the mass balance and ice volume change for all glaciers in the European Alps with multiple regression. Glacier geometry variables perform best in reproducing observed spatial mass balance variability, with the most important morphological variables being: area, median elevation, lower 10 % of the slope, easting, northing and, aspect. Together these morphological variables explain 51% of the glacier mass balance variance.

A similar study by Brun et al. (2019) investigates 6470 glaciers in HMA, subdivided into 12 climatically homogeneous regions. They find that 8-48% of glacier mass balance variability can be explained by the slope of the glacier tongue, mean glacier elevation, percentage of supraglacial debris cover, and avalanche contributing area. They identified the slope of the glacier tongue and the mean glacier elevation as the best predictors of the mass balance. It is suggested that the significant influence of the slope is partially explained by the preferential development of glacial ponds on shallow slopes. This is in line with the findings by Zhang et al. (2015), who investigate the effect of glacial lakes and found that the presence of glacier-fed lakes negatively correspond to glacier mass balance patterns.

Kraaijenbrink et al. (2017) show that, in HMA, about 11% of glacier area is covered in debris. The amount and thickness of the debris layer are associated with the mass balance variability. A thin layer of debris accelerates glacial melt because debris has a lower albedo than debris-free ice. However, a debris layer thicker than a few centimetres has an insulating effect which inhibits glacial melt (Kraaijenbrink et al., 2017). Likewise, changes in ice velocity are associated with mass balance variability. Heid and Kääb (2012) analyzed ice velocity changes over recent decades. In regions with negative mass balance, they observed that ice flow slowed. However, they found no clear relation between mass balance and velocity change (Dehecq et al., 2019). In HMA, the slowdown of glaciers has been observed locally and is associated with a negative mass balance. In regions like Karakoram and West Kun Lun, where glaciers are stable or thickening, glacier flow accelerated. However, the relation between the flow response of glaciers to mass change is poorly understood at regional scales (Dehecq et al., 2019).

1.3 Research Question

In this thesis, we contribute to research on this topic by studying mass balance variability in HMA. We aim to discover which climatical and morphological predictors most contribute to explaining the specific mass balance (SMB) variability of glaciers in climatically heterogeneous region HMA and the following assumed climatically homogeneous subregions: Hissar Alay (HA), Pamir (P), West Tien Shan West (WTS), East Tien Shan (ETS), West Kun Lun (WKL), East Kun Lun East (EKL), Qilian Shan (QS), Inner Tibet (IT), South and East Tibet (T), Hindu Kush (HK), Karakoram (K), West Himalaya (WH), Central Himalaya (CH), East Himalaya (EH) and, Hengduan Shan (HS). We test this for 9098 individual glaciers ($<2\text{km}^2$) in HMA. Additionally, we test whether glaciers with a lake are associated with more negative SMB, debris coverage reduces mass loss and faster-flowing glaciers are associated with a more negative mass balance.

In Section 2, we will describe data sources and preprocessing. Next, in Section 3, we present the methods and motivation for selecting a multi-linear regression model. Finally in Section 4,5 and 6 we outline and discuss the results.

2 Data

In this thesis, we aim to investigate influence of climatical and morphological predictors on glacier mass balance. Section 2.1 describes all predictors that possibly influence SMB. Data is selected from multiple open source platforms. Table 1 provides an overview of all variables and their source. Next, in Section 2.2 we present data selection and outlier detection results. Finally, in Section 2.3 we present missing data analysis.

2.1 Documentation of Variables

Randolph Glacier Inventory Version 6.0

The Randolph Glacier Inventory (RGI) provides global estimates of all glacier outlines and is updated annually. Outlines are produced in coordination with the Global Land Ice Measurements from Space (GLIMS). Glacier outlines are obtained from satellite images acquired in 1990 or later using Landsat platforms like Landsat 5 TM and Landsat 7 ETM, ASTER, IKONOS and SPOT HRS sensors. Glaciers with areas less than 0.01 km^2 are removed from the data (Pfeffer et al., 2014). In this study, we examine 15 subregions (region 13_O1 - 15_O3) in HMA (region 13, 14, 15). Based upon the work of Brun et al. (2017), we assume that these regions are climatically homogeneous. From now on, we will refer to the climatically homogeneous subregions as 'subregions'. We will refer to all subregions combined as 'HMA'; this region is climatically heterogeneous.

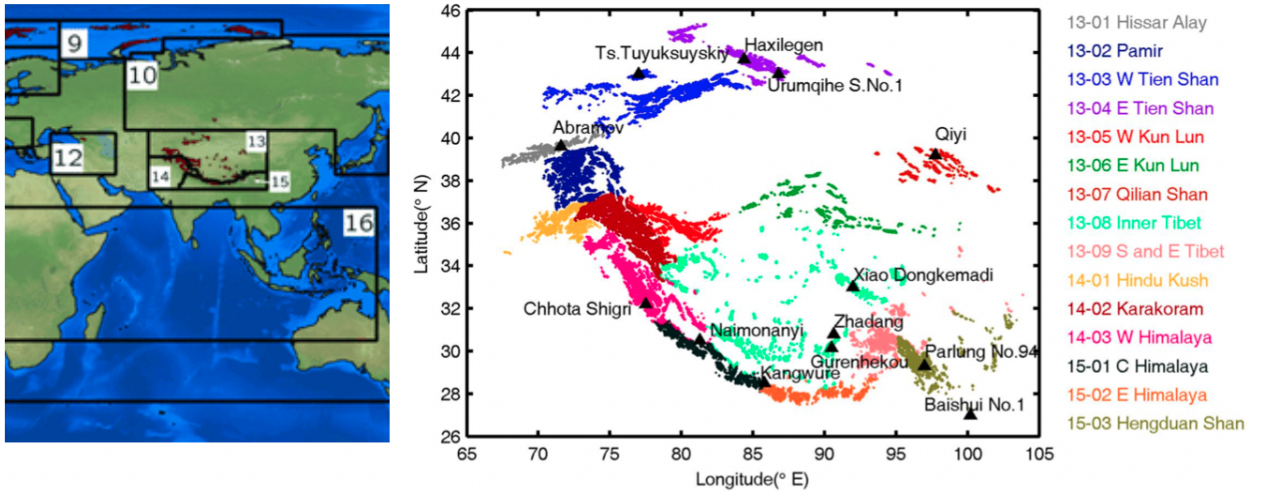


Figure 1: Left panel: Outlines of the RGI 6.0 regions. Region names: 9: Russian Arctic, 10: North Asia, 11: Central Europe, 12: Caucasus and Middle East, 13: Central Asia, 14: South Asia West, 15: South Asia East, 16: Low Latitudes. Image taken from (Wouters et al., 2019). Right panel: climatically homogeneous subregions. Image taken from (Zhao et al., 2016)

From the RGI (Consortium, 2017) we obtain i) Glacier area in km^2 , ii) The minimum, maximum elevation (m) above sea level iii) The median elevation (m) of the glacier iv) the mean slope of the glacier surface (in degree). v) The aspect (orientation) of the glacier surface (in degree) is defined as an integer azimuth relative to 0° at north. The mean aspect is calculated by the arctangent of the sum of the aspect sines and cosines of each of the glaciers. vi) Length (in m) of the longest surface flowline of the glacier. The length is measured with an algorithm developed by Frey et al. (2014).

ERA5

ERA5 is the fifth generation atmospheric reanalysis medium-range weather forecast of the global climate (ERA5, 2022). The Copernicus Climate Change Service produces ERA5 at the European Centre for Medium-Range Weather Forecasts. ERA5 covers the period from January 1950 to the present. It provides hourly estimates of climate data of many variables in the atmosphere, on land and in the ocean. The Earth is covered by 30km grid cells, and atmospheric data is provided up to a height of 80km. In addition, Era5 provides reanalysis data, which combines weather observations and past short-range weather forecasting models. This implies that reanalyses fill in the missing observational records, resulting in a globally complete map consistent in time.

Open Global Glacier Model

The Open Global Glacier Model (OGGM) is an open-source modelling framework for glaciers (OGGM, 2022). The model produces past and future mass balance, volume and geometry estimates of almost any glacier. The default directories are generated with the default parameters of the current stable OGGM version (May 2022). OGGM provides monthly temperature and precipitation data interpolated to the glacier location and height. Temperatures are corrected for altitude using a linear gradient (Maussion et al., 2019). In addition, OGGM provides estimates for the Equilibrium line Altitude (the altitude where accumulation is balanced by ablation over a one year period) and the Accumulation Area Ratio (the ratio between a glacier’s accumulation area and total area). These estimates are based on methods by Marzeion et al. (2012).

NASA MEaSURE

The NASA MEaSUREs project provides automated, low latency, global glacier flow and elevation change datasets (NASA, 2022). The ITSLIVE project provides a 120 m-resolution record of global ice velocity and elevation change from 1985 to the present. Data is obtained from multiple optical, radar, and laser satellite sensors. The U.S. Geological Survey provided Landsat 4,5,7,8 data. Regional Glacier and Ice Sheet Surface Velocity data are generated using auto-RIFT Dehecq et al. (2019) and provided by the NASA MEaSUREs ITSLIVE project. We obtain flow velocity fields on a 5 km grid. Each glacier’s average annual glacier velocity was extracted by overlaying the GeoTIFF file with the RGI polygons in QGIS and computing the zonal-statistics. 311 HMA glaciers larger than two km² are outside a grid and therefore missing.

Hugonnet et al.

Hugonnet et al. (2021) study glacier mass loss from 2000 to 2020. They use openly available satellite achieves to extract surface elevation changes at a 100m horizontal resolution covering 97.4% of all glaciers on Earth. They use a gap-filling method to extend the estimated elevation change up to 99.9% of all global glaciers. Estimates are validated against high-precision measurements. Using Gaussian process regression, Hugonnet et al. provide global estimates for glacier mass, volume and height variability per individual glacier along the period 2000-to 2020 (SEDOO, 2022).

Farinotti et al.

Farinotti et al. (2019) estimate the ice thickness distribution of all glaciers in HMA from surface characteristics, using an ensemble of five models. Glacier outlines are provided through RGI version 6.0 RGI. We use all ice thickness data in 25m spatial resolution.

Kraaijenbrink et al.

Kraaijenbrink et al. (2017) study the response of 5537 glacier melt in HMA to a global temperature rise of 1.5°C. Glacier outlines are provided from the Randolph Glacier Inventory (RGI) version 5.0. Additionally this study develops a debris classification method of HMA using Landsat 8 imagery. The Normalized Difference Snow Index (NDSI) is used to distinguish between debris-free and debris-covered ice (Hall et al., 1995). From this study, we obtain the volume- and area percentage of debris below the ELA.

Zheng et al.

Zheng et al. (2021) assess the changes in glacial lakes and the related glacial lake outburst flood in HMA. Glacial lakes are water bodies primarily formed and fed by glacier melt. They are located on the surface, below, adjacent or downstream of the glacier. Landsat 4,5,8 satellite images and US Geological Survey obtained lake outlines with a spatial resolution of 30m and 15m, respectively. Data is available from up to 1990. A 10 km buffer zone around the RGI establishes a preliminary distribution of possible glacier lakes. Images are taken from the warm season (June to November). An automatic water body classification algorithm locates all lakes from the Landsat images. We extracted all types of lakes from this study and selected the 2608 lakes that intersect with the RGI polygons using QGIS.

Below, Table 1 provides an overview of all variables and their source.

Variable	Documentation	Unit	Source	Period
RGI id	rgi_id	-	RGI 6.0	2022
O2 Region	O2Region	-	RGI 6.0	2022
SMB	dmdtda	w.e./yr	Huggonet	2000-2020
Area	Area	km ²	RGI 6.0	2022
Max hight	Zmax	km	RGI 6.0	2022
Mean hight	Zmed	km	RGI 6.0	2022
Min hight	Zmin	km	RGI 6.0	2022
Lenght	Lmax	km	RGI 6.0	2022
Slope	Slope	degree	RGI 6.0	2022
Ice Thickness	ice_thickness	m	Farinotti	2019
Mean ice flow velocity	velocity	m/year	Gardner	1985-2022
Apect	Aspect	0 = north (270°- 90°) 1 = south (90°- 270°)	RGI 6.0	2022
Glacial lake	Glacial_lake	0 = no lake 1 = lake	Zheng	2021
Equilibrium Line Altitude	TStar_ela_h	km	OGGM, ERA5 base	2022
Accumulation Area Ration	TStar_aar	-	OGGM, ERA5 base	2022
Volume percentage debris below ELA	debris_vol_ela_p	-	Kraaijenbrink	2017
Area percentage debris below ELA	debris_area_ela_p	-	Kraaijenbrink	2017
Mean temperature (aggregated by 10 year) Downscaled to glacier lvl.	temp_mean_2000_2020 temp_mean_1980_2000	degree	OGGM, ERA5 base	2000_2020 1980_2000
Temperature variability (aggregated by 10 year) Downscaled to glacier lvl.	temp_diff_2000_2020 temp_diff_1980_2000	degree	OGGM, ERA5 base	2000_2020 1980_2000
Mean precipitation (aggregated by 10 year)	prcp_mean_2000_2020 prep_mean_1980_2000	mm/year	OGGM, ERA5 base	2000_2020 1980_2000
Precipitation variability (aggregated by 10 year)	temp_diff_2000_2020 temp_diff_1980_2000	mm/year	OGGM, ERA5 base	2000_2020 1980_2000

Table 1: Documentation of variables used in this study. SMB stand for Specific Mass Balance in water equivalent per year

2.2 Data Preprocessing and Outlier Detection

Similar to Brun et al. (2017), we delete all glaciers with an area smaller than 2.0 km². As a result, we study 8098 glaciers in HMA. These glaciers contain 86.4% of the total ice volume. In addition, we have identified and deleted one outlier in the data, a glacier with a median elevation of 610 m (corresponding RGI id: RGI60-15.02228). Finally, we check whether the data is normally distributed and find that the area and the mean ice flow velocity are highly skewed. We log-transform the variables to reduce skewness and make patterns in the data more interpretable. The distribution of the variables are displayed in the supplementary material SC.1.

2.3 Missing Data Analyses

Thorough reanalysis and gap-filling methods in the source data minimize the presence of missing data. The mean ice flow velocity and ELA and AAR contain 3.8%, 0.2% and 0.2% of missing values, respectively. Missing data can cause bias and impact the overall statistical power of a study (Myers, 2000). In order to reduce bias, it is essential to apply the appropriate imputation method for the different types of missing data (missing completely at random, missing at random, or missing not at random). We apply an iterative imputation method, as this method can be applied to all types of missing data. We impute missing values with the IterativeImputer from scikit-learn [29]. This imputer uses the entire set of available data in the other features to estimate the missing value. At each step, the imputer treats a single feature column as the output y and treats the other feature columns as the inputs X . A regressor is fit on (X, y) for all known y . Then, missing values of y are imputed with the regressor predictions. This is iteratively done for each feature until the stopping criterium is met:

$$\frac{\max(|X_t - X_{t-1}|)}{\max(|X_v|)} < C_{\text{stop}},$$

where X_t is X at iteration t , X_v is the known value of X and C_{stop} is the stopping condition, in this case set to $C_{\text{stop}} = 10^{-3}$.

3 Methods

In this thesis, we aim to discover to what extent the variability of glaciers' specific mass balance is explained by the morphological and climate variables presented in Table 1. We do this by fitting a multivariate linear regression model as this model can quantify the relative impacts of the selected predictors on the outcome variable SMB. We select the best subset of predictors using LASSO regression and Pearson correlations.

3.1 Pearson Correlations

Multicollinearity occurs when several independent variables in a model are highly correlated. When this is the case, unnecessary noise is added to the model, making statistical results less reliable (Alin, 2010). We reduce multicollinearity in all following calculations by performing univariate linear analysis and disregarding all highly correlated independent variables ($> |0.7|$). We do this by calculating the Pearson's correlation coefficients ρ and their associated p values for all two-by-two combinations of available variables.

The Pearson correlation coefficient is a measure of linear correlation between two variables X and Y and is defined as:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}, \quad (1)$$

where cov is the covariance of X and Y, and σ is the standard deviation. All computed correlations are displayed in Figure 4.2.

3.2 Best Subset Selection with LASSO

Recent work showed that including unnecessary predictors that impact the exposure but not the outcome can increase standard errors without improving bias (Hawkins, 2004). Including predictors uncorrelated to the exposure but associated with the outcome can improve precision (Shortreed and Ertefaie, 2017). In order to include appropriate predictors for the mass balance variability, we use the Least Absolute Shrinkage and Selection Operator (LASSO) from Scikit Learn [31]. This operator favours simple models with fewer parameters. Variables with little impact on the outcome are entirely ignored by the model. Therefore, this type of regularization (known as L1 regularization), can be seen as a form of automatic feature selection. Selecting only relevant features reduces multicollinearity and makes model results easier to interpret.

LASSO is a linear model with an additional penalty term added to the loss function. Below we first introduce a simple multivariate linear model. A linear model aims to find the best set of coefficients $(\beta_1, \beta_2, \dots, \beta_n)$ which explain the maximum variability in \hat{y} (in this thesis the specific mass balance variability SMB) when multiplied with the predictors (x_1, x_2, \dots, x_n) (the selected predictors are displayed in Table 1).

$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n, \quad (2)$$

where we consider n instances in the data set. The maximum variability \hat{y} is explained if by optimizing the cost function J (known as the Residual Sum of Squares) with respect to β :

$$J(\beta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (3)$$

where y_i is the actual outcome and \hat{y}_i is the predicted outcome.

With LASSO regression an additional additional penalty λ is added to the loss function $J(\beta)$. Similarly, the aim is to optimize the new cost function J_{lasso} with respect to β :

$$J_{\text{lasso}}(\beta) = \sum_{i=1}^n \left(y_i - \sum_{j=1}^p X_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j|, \quad (4)$$

where the data set has n instances and p features and where $\hat{y}_i = \sum_{j=1}^p X_{ij} \beta_j$. The penalty term λ regularizes the coefficients. Meaning, the function is penalized if the coefficients take large values. So, LASSO regression shrinks the coefficients, and it helps to reduce the model complexity and multicollinearity. If the model complexity is high compared to the sample size, it is prone to overfitting. Choosing the optimal λ introduces a penalty and prevents the model from overfitting.

In order to fit the best LASSO regression (where the cost function J_{lasso} is minimized), it is essential to tune the hyper parameter λ . In this study, the best model is selected using 5-fold cross-validation. With cross-validation, the data is split into a test and train set. The single hyperparameter 'k' controls the number of test-train subsets into which the data is split. Predictions are made on test data that is not used during the model's training.

To allow for valid and direct comparison between the regression coefficients β_i we standardize all independent variables x_i by centring them around 0 and rescaling them to a standard deviation of 1:

$$x_{i,s} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}, \quad (5)$$

where $x_{i,s}$ is the standardize variable, \bar{x}_i is the mean of x_i and σ_{x_i} is the standard deviation of x_i

We perform LASSO regression on all glaciers in HMA and on all subregions. Here, we use LASSO to identify which variables are strongly associated with the SMB so we can perform feature selection. All predictors shrink towards zero, meaning the model expects no specific linear combination of any subset of the regressors to help predict the SMB. Therefore, in all further analyses, we select all features from Table 1 but exclude predictors that highly correlate ($> |0.7|$) with other predictors. When all variables shrink toward zero, the last term of equation 3 becomes increasingly small, i.e. LASSO regression reduces to OLS regression. Therefore, in the following Section, we will continue with OLS regression.

3.3 Ordinary Least Squares Multivariate Regression with LOOCV

We fit a simple Ordinary Least Squares (OLS) multivariate linear regression because this model allows us to determine the relative influence of one or more predictor variables on SMB. This may lead to a more accurate understanding of climate and morphological influences on glacier dynamics. We perform OLS regression on all glaciers in HMA and on all subregions. Here, we aim to find the best set of coefficients $(\beta_1, \beta_2, \dots, \beta_n)$ which explain the maximum variability in SMB when multiplied with the predictors (x_1, x_2, \dots, x_n) (the selected predictors are displayed in Table 1). For each analysis, we calculate Pearson's correlation coefficients ρ and their associated p values for all two-by-two combinations of all available variables and disregard all highly correlated predictor variables ($> |0.7|$) to reduce multicollinearity. We standardize all variables as described by equation 5.

We evaluate model performance with Leave One Out Cross Validation (LOOCV) by scikit-learn (Learn, 2022b)]. LOOCV is a special case of cross-validation. In LOOCV, k is set to the number of examples in the data set n . Meaning, LOOCV is trained on $n - 1$ samples of data. Cross-validation fits and evaluates k models and returns k estimates for model performance. In order to reduce bias and improve reliability, we report the mean performance of the k models.

4 Results

In this section we will present the results obtained from analyzing SMB variability in HMA and the selected climatically homogeneous subregions. First, we will present data exploration results in Section 4.1. Next, we present predictor-predictor (see Section 4.2) and predictor-outcome (SMB) Pearson correlation values (see Section 4.3). Finally, we continue by presenting the results obtained from LASSO and OLS regression in Section 4.4 and Section 4.5.

4.1 Data Exploration Results

Below in Table 2, we provide an overview of the mean of all selected variables in (HMA) and the selected subregions (HA-HS). In all regions, except for West Kun Lun, glacier mass is decreasing. The most negative SMB are observed in East Tien Shan, East Himalaya and Hengduan Shan.

	HMA	HA	P	WTS	ETS	WKL	EKL	QS	IT	T	HK	K	HW	CH	EH	HS
#glac.	8098	151	840	738	262	644	311	176	802	271	253	1481	734	518	467	449
SMB	-0.22	-0.09	-0.09	-0.27	-0.49	0.16	-0.06	-0.27	-0.31	-0.57	-0.18	-0.03	-0.34	-0.35	-0.46	-0.56
Area	8.2	5.5	7.9	8.4	5.1	9.6	6.8	4.5	6.3	8.2	6.3	12.6	6.4	7.0	7.2	6.3
Zmed	5.2	4.1	4.9	4.2	3.9	5.8	5.5	5.0	5.9	5.2	4.9	5.4	5.1	5.4	5.6	5.0
Slope	19.7	17.9	20.8	19.6	17.8	18.8	16.1	18.0	14.5	22.7	17.4	24.2	17.1	20.5	20.0	21.4
Vel	3.3	3.2	3.6	2.6	2.0	2.5	2.4	1.6	2.4	5.6	3.9	3.6	4.1	4.3	3.9	3.2
AAR	0.54	0.53	0.55	0.54	0.51	0.56	0.52	0.55	0.53	0.53	0.54	0.54	0.53	0.56	0.59	0.52
deb_a _ela_p	0.19	0.26	0.27	0.17	0.09	0.12	0.05	0.02	0.04	0.19	0.36	0.22	0.26	0.27	0.20	0.16
#Lake	814 (10.0)	6 (4.0)	55 (6.5)	79 (10.7)	30 (11.5)	26 (4.0)	47 (15.1)	26 (14.8)	111 (13.8)	51 (18.8)	17 (6.7)	83 (5.6)	44 (6.0)	62 (12.0)	129 (27.6)	48 (10.7)
#North facing	2235	125	672	543	201	531	227	148	582	166	187	1107	530	323	245	275
T mean '00-'20	-7.4	-3.2	-8.6	-6.4	-4.8	-10.2	-8.6	-7.3	-6.6	-4.5	-6.5	-12.1	-7.0	-3.7	-3.0	-2.3
dT '00-'20	0.36	0.48	0.46	0.18	0.18	0.47	0.63	0.53	0.64	0.59	0.47	0.28	0.07	0.18	0.28	0.39
dT '80-'00	0.38	0.16	0.18	0.36	0.56	0.62	0.46	0.88	0.20	0.25	-0.02	0.48	0.51	0.50	0.28	0.22
P mean '00-'20	8503	12919	6207	8597	7710	4491	4719	5194	5822	12677	7218	5445	9792	13763	16262	18133
dPrp '00-'20	603	-187.0	-176.3	-248.2	-205.6	47.7	182.0	304.7	-60.4	-106.3	-99.5	-45.7	372.3	833.7	418.0	67.2
dPrp '80-'00	93.1	250.6	-72.8	320.1	334.2	80.8	112.0	120.0	301.7	-270.3	-20.5	34.6	324.1	372.9	329.5	-1012.7

Table 2: Mean value of the selected variables in the climatically heterogeneous region (HMA) and the select climatically homogeneous subregions (HA-HS). glac are the number of glaciers. Lake is the number and percentage of glacial lakes. North facing are the number of north-facing glaciers. T mean and P mean are the mean temperature and precipitation, respectively. dT and dPrp are the changes in temperature and precipitation, respectively. All values are in standard units as displayed in Table 1.

Next, we investigate the influence of debris coverage below the ELA on the SMB. When considering the entire sample of 9098 glaciers in HMA, the results show that glaciers with more than 30% debris coverage below the ELA are associated with a lower specific mass balance (mean SMB = -0.24 w.e. yr^{-1}) than glaciers with less than 30% debris coverage below the ELA (mean SMB = -0.15 w.e. yr^{-1}). Additionally, results show that glaciers with more than 30% debris coverage below the ELA have a lower median elevation (5035 km) than glaciers with less than 30% debris coverage below the ELA (average median elevation = 5552 km). The left panel of Figure 2 displays the SMB of all glaciers with more than 30% debris coverage below the ELA in red and glaciers with less than 30% debris coverage below the ELA in black, as a function of their median elevation. The supplementary material provides similar results for each subregion (SC.2).

Furthermore, we investigate the influence of glacial lakes on the SMB. When considering the entire sample of 9098 glaciers in HMA, the results show that glaciers with lakes are associated with a lower specific mass balance (mean SMB = -0.40 w.e. yr^{-1}) than glaciers without a lake (mean SMB = -0.20 w.e. yr^{-1}). Additionally, results show that glaciers associated with lakes have lower slopes (average Slope = 17.0°) than glaciers that are not associated with a glacial lake (average Slope = 20.0°). The right panel of Figure 2 displays the SMB of all glaciers with a lake in red and without a lake in black, as a function of their slopes. The supplementary material provides similar results for each subregion (SC.3).

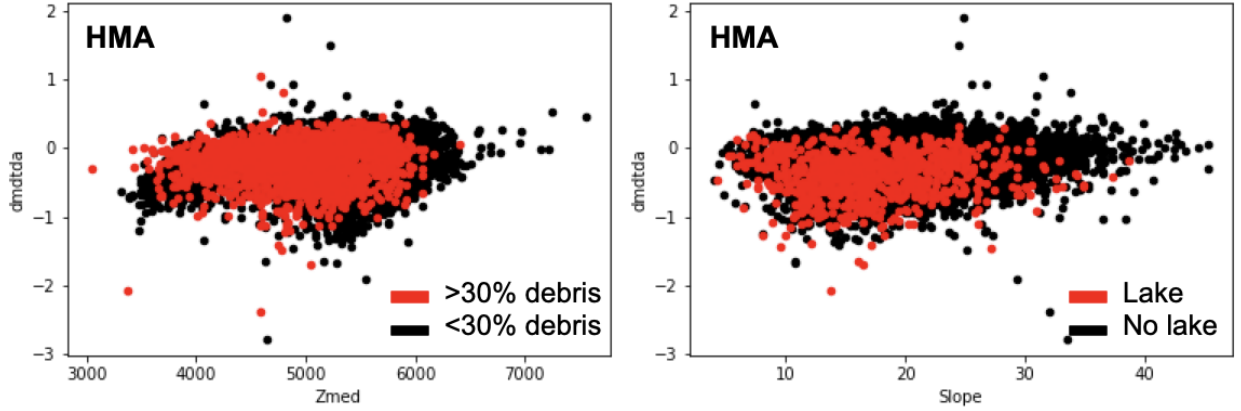


Figure 2: Left panel: SMB (dmdtda) as a function of median elevation (Zmed), grouped by more or less than 30% debris coverage below the ELA. The red dots represent glacier with more than 30% coverage. Right panel: SMB (dmdtda) as a function of the slope, grouped by the presence of a glacial Lake. The red dots represent glaciers associated with a lake.

Finally, we examine the influence of the glacier's aspect on the SMB change. In the sample, 72% of all glaciers are north-facing (5862 glaciers). On average, the north-facing glaciers have a lower temperature (mean temp = -7.6°) than south-facing glaciers (mean temp = -6.9°). In addition, north-facing glaciers are associated with a slightly less negative mass balance (SMB = -0.21 w.e. yr^{-1}) than south-facing glaciers (SMB = -0.23 w.e. yr^{-1}).

4.2 Predictor-Predictor Pearson Correlations

We compute Pearson correlations for all selected variables in HMA (see Figure 4.2). Similar calculations for all the subregions are displayed in the supplementary material SC.4.

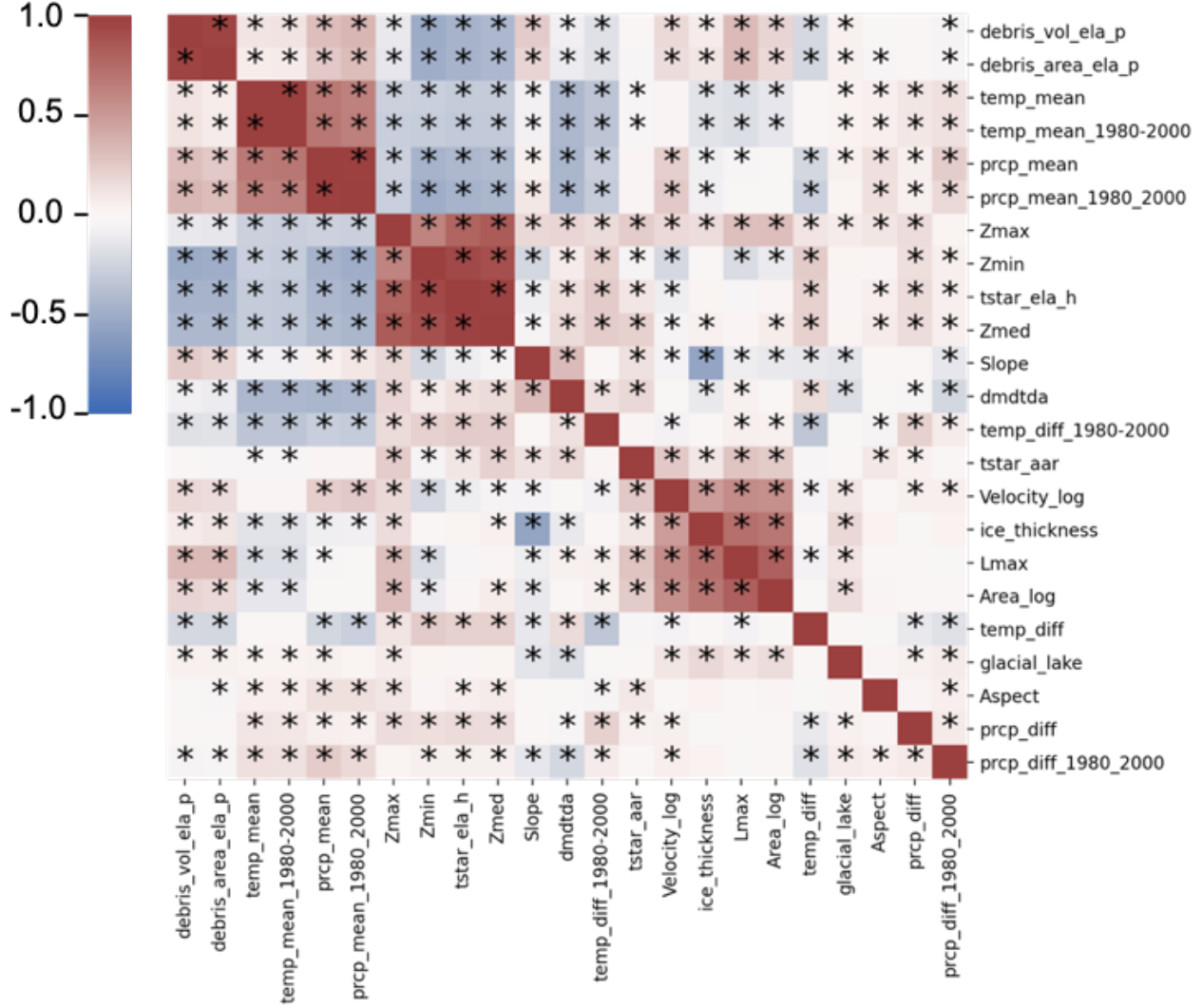


Figure 3: Pearson correlation matrix of all selected variables. The variable dmdtda stand for the Specific Mass Balance. The heat map represent the value of correlation, from red (+1.0) to blue (-1.0). The stars indicates a significant correlation for $p < 0.05$

The variables temp_mean_1980_2000 and prcp_mean_1980_2000 are excluded from all following analyses as they are highly ($> |0.7|$) correlated with temp_mean_2000_2020 and prcp_mean_2000_2020, respectively. Similarly, the variables ice_thickness, tstar_ela_h and, debris_vol_ela_p are excluded from the analyses as they are highly correlated with Area_log, tstar_aar and, debris_area_ela_p, respectively. Finally, Lmax, Zmin, and Zmax are excluded from the analyses as they are highly correlated with Zmed.

4.3 Predictor-SMB Pearson Correlations

We compute Pearson correlations between the variable SMB and all selected predictor variables, see Table 3. In this table we only display significant (< 0.05) correlations, missing values are non-significant.

	HMA	HA	P	WTS	ETS	WKL	EKL	QS	IT	T	HK	K	HW	CH	EH	HS
Area_log	-	-	-	-	-	-0.26	-0.13	-	-	-0.14	0.13	-0.16	-0.20	-	-	-
Zmed	0.17	-	0.26	0.52	0.57	-0.40	0.37	0.54	0.46	0.21	0.30	-0.16	-	0.24	0.33	-0.21
Slope	0.28	-	0.33	0.46	0.45	0.29	0.29	0.20	0.06	0.30	0.62	0.23	0.46	0.43	0.44	0.37
Velocity_log	-0.04	-	0.09	0.10	-	-	-	0.21	0.10	-	0.28	-0.08	0.11	0.14	-	-0.17
AAR	0.18	0.22	0.21	0.36	0.37	-	-	0.31	0.26	0.25	0.25	0.02	0.17	0.37	0.32	-
deb_a _ela_p	-	-	-	-	-	0.16	0.15	-	-0.25	-0.19	-	-	-0.10	-0.19	-0.19	0.11
T_mean 2000-2020	-0.38	-	-	-0.20	-0.31	0.47	-	0.25	-0.22	-	-0.25	0.21	0.17	0.26	-	0.14
T_diff 2000-2020	0.11	-0.22	-0.07	0.29	-	-	0.22	-	-	-	-	0.17	0.19	-	-0.12	-0.12
T_diff 1980-2000	0.14	-	0.31	-0.12	0.24	-	-	0.25	0.33	-	-0.47	-0.05	-0.30	-0.17	-	-0.22
P_mean 2000-2020	-0.33	0.36	-	-0.27	-	0.49	-0.37	-0.5	-0.57	0.15	-	0.12	0.13	0.35	-	0.2
P_diff 2000-2020	-0.05	-	-	0.40	-	0.37	-	0.28	-	-	-	-0.12	-0.12	0.32	-	0.17
P_diff 1980-2000	-0.03	-	0.27	-0.22	0.12	-	-0.2	-0.21	-0.32	-0.19	-	-	-0.14	-	-0.20	-0.20

Table 3: All significant correlations ($p < 0.05$) between all continuous variables and the SMB. All missing values are non-significant. T and P stand for temperature and precipitation, respectively.

First, we tested whether a shallower slope increases glacier mass losses. This hypothesis is accepted as a positive and significant correlation between SDM and slope is found for 14 of the 15 subregions. This means that the gentler the glacier slope, the more negative SMB.

In 9 out of 15 regions the median elevation is found to be associated with positive SMB. A negative association is found in 3 regions, with the biggest influence in West Kun Lun.

Next, we tested whether debris coverage increases glacier mass losses. This hypothesis is accepted in 4 out of 15 regions as a negative and significant correlation between SMB and percentage debris below the ELA is found. However, in all other regions, the hypothesis is rejected as a positive and significant correlation is observed in 3 out of 15 regions and, no significant correlation is observed in 8 out of 15 subregions.

In 7 out of 15 regions flow velocity is found to be associated with negative SMB. In 2 out of 15 regions the results show a negative and significant correlation (Karakoram and Hengduan Shan). In all other regions the correlation were not significant.

In 5 out of 15 subregions the area is found to be associated with negative SMB, with the most negative association found in West Kun Lun. Only in Hindu Kush a positive association is observed.

Generally, in the subregions, the climate predictors correlate less strong and less often with the SMB. Only in West Kun Lun a relative high correlation between the mean temperature (2000-2020), the mean precipitation (2000-2020) and the precipitation difference (2000-2020) is observed.

4.4 LASSO Regression

We performed L1 LASSO regression on all glaciers in HMA to find the best set of coefficients β_i , which explain the maximum variability in the SMB when multiplied by a combination of climate and morphological predictors variables as described in Section 4.2. The final LASSO model is produced by the optimal value: $\lambda = 0.59$. The results show a MAE training and test error of 0.20. Model performance is evaluated with with 5-fold cross-validation. This resulted in the following train en test performance: $R^2_{\text{train}} = 19.1$ and $R^2_{\text{test}} = 17.5$.

Additionally, we performed L1 LASSO regression on all glaciers in the 15 subregions to find the best set of coefficients β_i , which explain the maximum variability in the SMB when multiplied by the climate and morphological predictors (All highly correlated predictors are removed for the analysis, see supplementary SC.4). In 6 out of 15 regions, the model explained more than 40% of the SMB variability (in regions: WTS, ETS, EKL, QS, IT, HK). However, in 6 out of 15 regions the model explained less than 20% of the mass balance variability (in regions: HA, P, T, K, EH, HS).

In Table 4, we present an overview of all regression results. In this table, we also present the penalty term λ . We obtain relatively low values for λ . This means that no significant penalty is needed to prevent the model from overfitting.

	HMA	HA	P	WTS	ETS	WKL	EKL	QS	IT	T	HK	K	HW	CH	EH	HS
MAE train	0.20	0.09	0.11	0.14	0.12	0.10	0.11	0.09	0.15	0.20	0.12	0.11	0.12	0.17	0.17	0.30
MAE test	0.20	0.13	0.11	0.13	0.10	0.10	0.10	0.09	0.14	0.24	0.10	0.11	0.15	0.18	0.17	0.29
R^2 train	17.5	36.8	21.6	36.9	44.2	27.5	41.3	42.3	43.5	30.8	42.4	11.5	27.7	28.4	32.7	7.27
R^2 test	19.1	-5.9	16.2	40.8	47.6	24.6	45.0	53.2	43.6	-1.5	47.4	16.6	27.1	31.5	17.2	9.89
λ	0.59	0.22	0.03	0.05	0.24	0.02	0.04	0.35	0.25	0.03	0.05	0.03	0.09	0.06	0.05	16.6

Table 4: Summary of the LASSO regression results in the climatically heterogeneous region HMA and the homogeneous subregions HA-HS.

4.5 OLS Multivariate Regression

We performed OLS regression on all glaciers in HMA to find the best set of coefficients β_i , which explain the maximum variability in the SMB when multiplied by the predictor variables as described in Section 4.2. The results showed that 30.7% of the SMB variability can be explained by the selected variables (see Table 1). With the most important predictors variables: lake ($\beta_i = -0.13$), mean temperature ($\beta_i = -0.07$), mean precipitation ($\beta_i = -0.07$), and the slope ($\beta_i = 0.07$). The regression residuals showed no obvious signs of heteroscedasticity (see Figure 4)

Next, we performed OLS regression on all glaciers in the 15 climatically homogeneous subregions to find the best set coefficients β_i , which explain the maximum variability in the SMB when multiplied by the predictor variables as described in Section 4.2. Again, all highly correlated predictor variables were disregarded (correlation per subregion are provided in supplementary material SC.4). In the subregions, the results showed that the selected variables explain 18.1% to 50.0% of the SMB variability, with the most important predictors being the morphological variables: the presence of a lake, the slope, and the median elevation (Table 5 columns HA-HS). The presence of a lake had the most significant influence in 6 out of 15 regions (HA, QS, HW, CH, EH, HS), the slope had the most significant influence in 3 out of 15 regions (T, HK, HW) and median elevation had the most significant in 5 out of 15 regions (P, WTS, ETS, EKL, T). The mean temperature (in 2000-2020) and mean precipitation (in 2000-2020) were the most important predictors in only one region, West Kun Lun and Inner Tibet, respectively.

The results showed a larger influence of the climate predictors when examining a climatically heterogeneous region (Table 5, column HMA), than observed in the climatically homogeneous subregions (Table 5, column HA-HS). Generally, in the selected subregions, climate influences reduce (lower and more often not significant) and morphological influences increase (see Table 5 and Figure 4). Note that, in Figure 4, we only display the average of the standardized regression coefficients of all subregions. If a predictor is non-significant in a subregion, it is counted as a zero-contribution to the average. An overview of all significant ($p < 0.05$) regression results are displayed in Table 5. In all subregions, the regression residuals showed no obvious indication of heteroscedasticity (figures provided in supplementary material SC.5).

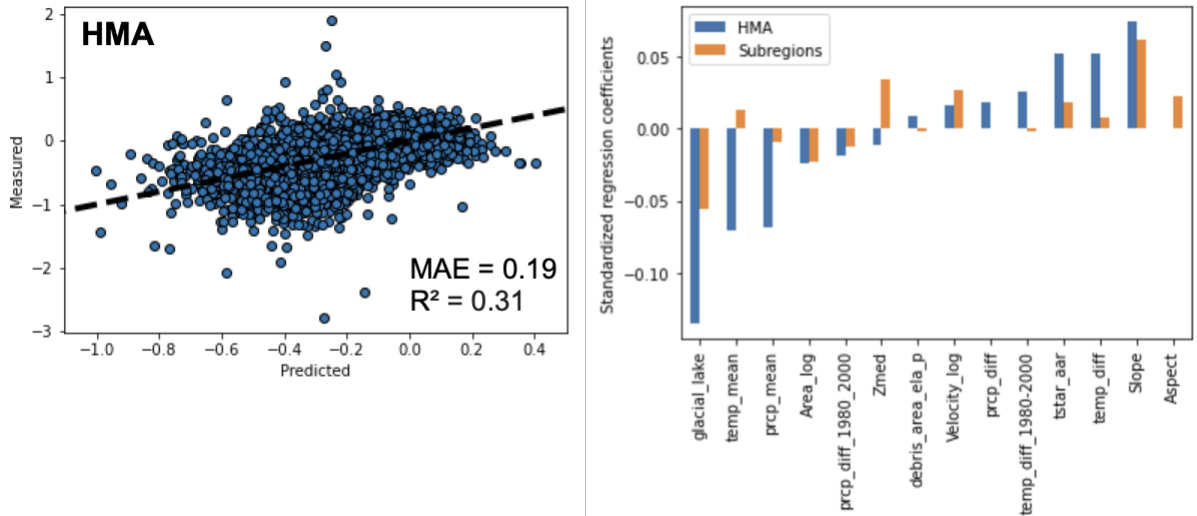


Figure 4: Left panel: Regression residuals of SMB predictions of all glaciers in HMA. Right panel: Standardized regression coefficients for each predictor. Coefficients from analyzing HMA are displayed in blue. The average of all coefficients from analyzing the subregions are displayed in orange.

Below in Table 5, we present a summary of the standardized regression results in all regions for significance level $p < 0.05$. Non-significant values are not displayed in the table.

	HMA	HA	P	WTS	ETS	WKL	EKL	QS	IT	T	HK	K	HW	CH	EH	HS
MAE	0.19	0.10	0.10	0.13	0.11	0.10	0.10	0.09	0.14	0.21	0.11	0.10	0.12	0.17	0.16	0.28
R²	30.7	28.4	23.7	40.8	44.6	27.1	45.7	48.0	50.0	22.0	42.7	20.2	38.1	33.6	38.0	18.1
Const	-0.21	-0.09	-0.09	-0.26	-0.48	0.16	-0.04	-0.27	-0.31	-0.59	-0.19	-0.04	-0.34	-0.37	-0.42	-0.60
Area	-0.02	-0.04	-0.04	-	-	-0.03	-0.02	-	-	-0.09	-	-0.01	-0.07	-0.05	-	-
Zmed	-0.01	0.04	0.06	0.09	0.13	-0.03	0.12	-	0.04	0.11	-	-0.04	0.04	0.05	-	-0.09
Slope	0.07	-	0.02	0.06	0.04	-	0.05	0.04	0.06	0.11	0.10	0.04	0.08	0.10	0.10	0.12
Velocity	0.02	-	0.04	-	-	0.02	-	0.04	0.04	0.10	-	-	0.06	0.08	0.03	-
Aspect	-	0.07	-	-	-	-	-0.09	-	-	-	0.06	0.04	-	0.09	-	0.16
Lake	-0.13	-0.17	-	-0.05	-	-	-	-0.07	-	-	-	-	-0.08	-0.14	-0.18	-0.16
AAR	0.05	-	-	0.02	0.03	-	-	0.02	0.02	-	-	0.02	0.02	-	0.07	0.08
DAEP	0.01	-	0.02	-	-	-0.02	0.02	-	-0.03	-	-	-0.02	-	-	-	-
temp mean	-0.07	0.07	0.03	0.02	-0.05	0.05	-	0.04	-	0.08	-0.04	0.02	-	-	-0.03	-
dT	0.05	-0.03	-	0.02	-	-	0.02	0.03	0.05	-0.06	0.04	0.04	0.02	-	-0.03	-
dT lag 20y	0.03	-	0.03	-0.03	-0.08	0.01	-	-	0.06	-	-0.04	0.02	-0.03	0.03	-	-
prcp mean	-0.07	0.07	-	-	0.03	-	-0.05	-0.06	-0.13	-	-	-	-	-	-	-
dP	0.02	-	-	-	-0.03	-	0.03	0.04	-	-	-0.03	-0.03	-0.02	0.05	-	-
dP lag 20y	-0.02	-0.08	0.02	-0.02	-	-	-	-	-	-	-0.03	-	-0.03	-	-0.05	-

Table 5: Summary of the standardized regression results when predicting the SMB in HMA and the subregions HA-HS. We only display significant ($p < 0.05$) regression results, missing values are non-significant.

The regression results showed a relatively large negative influence of the presence of a glacial lake on the SMB in 7 out of 15 subregions. The biggest influence is observed in Hissar Alay, where the presence of a lake results in a SMB decrease of -0.867 w.e. yr^{-1} . Similarly, the slope of the glacier showed a relatively large influence on the SMB. The biggest influence was observed in Hindu Kush (see Table 5), where a one-degree increase in slope results in an SMB increase of 0.022 w.e. yr^{-1} . The median elevation showed a significant contribution in 14 out of 15 subregions. The biggest influence was observed in East Kun Lun, where a one-kilometre increase results in an SMB increase of 0.0006 w.e. yr^{-1} .

The regression results showed a relatively low but significant velocity influence on the SMB. The most significant effect of flow velocity was observed in Central Himalaya, where one m/year increase in ice velocity results in an SMB increase of 0.0063 w.e. yr^{-1} (see Table 5). Similarly, the regression results showed a relatively small contribution of the area to mass balance ($-0.05 < \beta_i < -0.01$, except for West Himalaya $\beta_i = -0.09$). In West Himalaya, a one km increase in area results in an -0.0028 decrease in the SMB w.e. yr^{-1} .

5 Discussion

5.1 Contrasting Behaviour of West Kun Lun glaciers

In all subregions, except for West Kun Lun and Inner Tibet, we find that the morphological variables explain the variability in SMB best (see Section ??). We expect the relatively large climate influences in Inner Tibet to result from the possibly incorrect assumption that this region is climatically homogeneous (see Section 5.5). However, we expect the West Kun Lun abnormality to result from the unusual climate settings here. Generally, this region experiences a relatively dry and cold continual climate (see S2) because the wester Kun Lun and the Karakoram mountains block the transportation of moisture from the Indian Ocean (Bao et al., 2015). West Kun Lun displays more abnormal behaviours, as is it the only subregion where SMB is increasing. In West Kun Lun, the SMB correlates relatively strong with the area in a negative fashion (see Table 3). Meaning the bigger the area, the lower the mass loss. We expect this remarkable large correlation to result from the relatively large number of big glaciers located in this region (9 out of 33 biggest (>100 km²) glaciers in China are located in West Kun Lun (Bao et al., 2015). Additionally, in West Kun Lun, the SMB correlates relatively strong with the median Elevation in a negative fashion (see Table 3). Ke et al. (2015) suggest that the abnormal behaviour of West Kun Lun might be related to the heterogeneous spatial variability of elevation trends over different glacier parts.

5.2 Influence of Lakes and Slopes on Glacier Mass Balance

One of the most important predictors for the SMB variability is the presence of a glacial lake. That the presence of glacial-fed lakes negatively corresponds to glacier mass balance patterns (see Section 4.5) agrees with a study by Zhang et al. (2015). However, we expect that a part of this remarkably large influence of glacial lakes is explained by a spurious signal resulting from unobserved lake-slope correlations. Because of the preferential development of lakes on shallower slopes (Quincey et al., 2007), we expect a strong correlation between these variables. Our beliefs are supported by the fact that we find that glaciers associated with lakes have lower slopes than glaciers that are not associated with a glacial lake (see Section 4.1 and Figure 2).

5.3 Influence of Debris on Glacier Mass Balance

The percentage of debris coverage positively and negatively correlates with the SMB. This might be explained by the varying thickness of the debris layer, as a thin layer accelerates the melt and a layer thicker than a few centimetres inhibits the melt (Evatt et al., 2015; Kraaijenbrink et al., 2017). However, including debris thickness falls outside the scope of this thesis and is left for future study.

5.4 Non-linear Glacier Response to Climate Variables

When we compare the OLS and LASSO regression results from the climate heterogeneous region to the homogeneous subregions, the results showed:

- i) A more significant influence of the climate predictors in the climatically heterogeneous region (Table 5, column HMA) than in the climatically homogeneous sub regions (Table 5, column HA-HS).
- ii) Both LASSO and OLS performance have high Mean Absolute Errors in the climatical heterogeneous regions (OLS MAE = 0.22; LASSO MAE_{train} = 0.20, MAE_{test} = 0.20). Compared to other (non-linear) statistical methods, 0.20 is a relatively high MAE. For example, Hartmann (2022) studies SMB response to similar climate and morphological predictors. With three non-linear methods he finds a Means Absolute Errors of ~ 0.12 . Because our results showed a high test and train MAE, we expect the model to overfit, meaning the model is not complex enough to capture the true statistical nature of the data model.

iii) Model performance increased (meaning underfitting reduced) in the homogeneous climatical subregions. With OLS regression the results showed that the mean MAE of all regions is 0.13. With LASSO regression the results showed that mean MAE of all regions is 0.14. The difference between the test and train MAE was small (max difference = 0.04, see Table 4). This means the amount of underfitting was less in the climatical homogeneous subregions, and the linear model is better at capturing the true statistical response of the data.

The above observations show that model performance decreases in the region where SMB responses are dominated by temperature and precipitation influences. Here, in the climatically heterogeneous region, the proposed linear model is worse at capturing the true statistical response of the data. Because climate predictors predominantly influence this region, we expect they behave in a non-linear fashion and explain the reduced performance of the linear model. The non-linear response of glacier mass balance to climate variables agrees with studies by Bolibar et al. (2022) and Steiner et al. (2005). They develop deep learning models and find non-linear air temperature and precipitation contributions.

5.5 Limitation of the Analysis

In this thesis, we fit a linear model to non-linear data. This decreases model performance, especially in the climatically heterogeneous region HMA, which is predominantly influenced by climate predictors like temperature and precipitation. To improve performance it is essential to capture the non-linear behaviour of the data, for example by fitting a polynomial regression or using non-linear deep learning methods. Another possible way to improve the model is by including more explanatory variables, e.g. avalanche contributing area, debris thickness Kraaijenbrink et al. (2017) and velocity change Heid and Kääb (2012). A final way to improve model performance is by considering interaction between the predictor variables.

An other major limitation of this study is the inclusion of the non-independent variables Lakes and Aspect in the multiple linear regression model. Because of possible unobserved predictor-predictor correlations, regression results cannot be correctly interpreted. Because the variables Lake and Aspect are binary, we cannot compute their Pearson correlation. Therefore, we choose to include both variables in all the analyses. A more sophisticated approach would be to calculate a point-biserial correlation, which is used to measure the correlation between a continuous and dichotomous variable.

In this thesis, we investigate lagged (~ 40 y) climate effects on current changes in the SMB and do not find a significant contribution. However, we did not include lagged climate predictors at much larger (~ 100 kyr) time scales. We presume that these unobserved lagged effects impact current changes in SMB as the evolution of glaciation in HMA is dominated by a ~ 23 kyr periodicity during the last 425 kyr (Yan et al., 2021). Yan et al. (2021) finds that the ~ 23 kyr periodicity results from changes in the earth's precession, affecting summer temperatures and precipitation patterns induced by changes in the summer monsoons and the mid-latitude westerlies. The annual mean temperature shows a ~ 100 -kyr cycle dominated by CO_2 -induced effects. Including lagged climate effects at larger time scales is an interesting topic left for further study.

Another limitation of this study is the separation of glaciers into regions that we assumed were climatically homogeneous. This assumption might not hold, especially in regions that cover a larger spatial extent like Inner Tibet (covers more than 2000km) (Brun et al., 2017). A more sophisticated approach would be to spatially cluster glaciers based on climate data.

In this study, missing data is iteratively imputed, which can influence the predictor-predictor and predictor-outcome correlations. A study by Taylor et al. (2017) shows that, generally, the correlation's magnitude is reduced in imputed data sets, and this effect increases with the amount of missing data Taylor et al. (2017). Underestimation of correlation can lead to unobserved heterogeneity in regression results. However, we expect this effect to be minimized as the percentage of missing is low (see Section 2.3). Additionally, no significant sign of heterogeneity is observed in all regression residuals (see Section 4.5).

6 Conclusion and Outlook

In this thesis, we aim to discover which climatical and morphological predictors most contribute to explaining the SMB variability of glaciers in climatically heterogeneous region HMA and in the following climatically homogeneous subregions: Hissar Alay, Pamir, West Tien Shan West, East Tien Shan, West Kun Lun, East Kun Lun East, Qilian Shan, Inner Tibet, South and East Tibet, Hindu Kush, Karakoram, West Himalaya, Central Himalaya, East Himalaya and, Hengduan Shan. To investigate this, we performed OLS regression and excluded all significant predictor-predictor correlations $> |0.07|$.

We find that in HMA, 30.7% of the glaciers' SMB variability can be explained by the area, median elevation, slope, velocity, presence of a lake, accumulation area ratio, percentage of debris below the ELA, the mean temperature and temperature change in periods (1980-2000 and 2000-2020) and the mean precipitation and precipitation change in periods (1980-2000 and 2000-2020). Here, the most important predictors are the presence of a lake, the slope and the mean temperature and precipitation in 2000-2020. All predictors, except for the slope are found to be associated with negative SMB. We expect that a part of the remarkably large influence of glacial lakes on the SMB is explained by a spurious signal resulting from unobserved lake-slope correlations.

Generally, OLS regression showed that climate influences reduced (lower and more often not significant) and morphological influences increased in the subregions. Here, we found that the selected variables explain 18.1% to 50.0% of the SMB variability, with generally the most important predictors being the morphological variables: the presence of a lake, the slope, and the median elevation. The slope and the median elevation are found to be associated with positive SMB. The climatical variables mean temperature (in 2000-2020) and mean precipitation (in 2000-2020) were the most important predictors in only one region, West Kun Lun and Inner Tibet, respectively. We expect the relatively large climate influences in Inner Tibet to result from the possibly incorrect assumption that this region is climatically homogeneous. However, we expect the contrasting behaviour of West Kun Lun to result from the unusually dry and cold climate settings here.

Our analysis suggests that glacier mass balance responds in a non-linear fashion to climate influences like temperature and precipitation. We expect this because model performance decreases in the region where SMB responses are dominated by temperature and precipitation influences. Here, in the climatically heterogeneous region, the proposed linear model is worse at capturing the true statistical response of the data. Because climate predictors predominantly influence this region, we expect they behave in a non-linear fashion and explain the reduced performance of the linear model. This agrees with Bolibar et al. (2022) and Steiner et al. (2005), who developed deep learning models and found non-linear air temperature and precipitation contributions.

Additionally, results showed that a higher percentage of debris below the ELA does not significantly reduce glacier mass loss in 11 out of 15 subregions. In 7 out of 15 subregions, slower flowing glaciers are associated with more negative SMB.

A major limitation of this study is the uses of a linear model to non-linear data. This decreases model performance, especially in the climatically heterogeneous region HMA, which is predominantly influenced by climate predictors like temperature and precipitation. To improve performance it is essential to capture the non-linear behaviour of the data, for example by fitting a polynomial regression or using other non-linear deep learning methods. Another possible way to improve the model is by including more explanatory variables, e.g. avalanche contributing area, debris thickness and velocity change. A final way to improve model performance is by considering interaction terms between the predictor variables.

References

- Aylin Alin. Multicollinearity. *Wiley interdisciplinary reviews: computational statistics*, 2(3):370–374, 2010.
- Simon K Allen, Andreas Linsbauer, SS Randhawa, Christian Huggel, Pooja Rana, and A Kumari. Glacial lake outburst flood risk in himachal pradesh, india: an integrative and anticipatory approach considering current and future threats. *Natural Hazards*, 84(3):1741–1763, 2016.
- Wei-jia Bao, Shi-yin Liu, Jun-feng Wei, and Wan-qin Guo. Glacier changes during the past 40 years in the west kunlun shan. *Journal of Mountain Science*, 12(2):344–357, 2015.
- Stephan Bojinski, Michel Verstraete, Thomas C Peterson, Carolin Richter, Adrian Simmons, and Michael Zemp. The concept of essential climate variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society*, 95(9):1431–1443, 2014.
- Tobias Bolch, Anil Kulkarni, Andreas Kääb, Christian Huggel, Frank Paul, J Graham Cogley, Holger Frey, Jeffrey S Kargel, Koji Fujita, Marlene Scheel, et al. The state and fate of himalayan glaciers. *Science*, 336(6079):310–314, 2012.
- Jordi Bolibar, Antoine Rabatel, Isabelle Gouttevin, Harry Zekollari, and Clovis Galiez. Nonlinear sensitivity of glacier mass balance to future climate change unveiled by deep learning. *Nature communications*, 13(1):1–11, 2022.
- Bodo Bookhagen and Douglas W Burbank. Toward a complete himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *Journal of Geophysical Research: Earth Surface*, 115(F3), 2010.
- F Brun, Patrick Wagnon, Etienne Berthier, Vincent Jomelli, SB Maharjan, F Shrestha, and PDA Kraaijenbrink. Heterogeneous influence of glacier morphology on the mass balance variability in high mountain asia. *Journal of Geophysical Research: Earth Surface*, 124(6):1331–1345, 2019.
- Fanny Brun, Etienne Berthier, Patrick Wagnon, Andreas Kääb, and Désirée Treichler. A spatially resolved estimate of high mountain asia glacier mass balances from 2000 to 2016. *Nature geoscience*, 10(9):668–673, 2017.
- RGI Consortium. Randolph glacier inventory – a dataset of global glacier outlines: Version 6.0, 2017. URL https://nsidc.org/sites/nsidc.org/files/technical-references/RGI_Tech_Report_V6.0.pdf.
- Amaury Dehecq, Noel Gourmelen, Alex S Gardner, Fanny Brun, Daniel Goldberg, Peter W Nienow, Etienne Berthier, Christian Vincent, Patrick Wagnon, and Emmanuel Trouvé. Twenty-first century glacier slowdown driven by mass loss in high mountain asia. *Nature Geoscience*, 12(1):22–27, 2019.
- Saket Dubey and Manish Kumar Goyal. Glacial lake outburst flood hazard, downstream impact, and risk over the indian himalayas. *Water Resources Research*, 56(4):e2019WR026533, 2020.
- ERA5. ECMWF Atmospheric Reanalyses 5. <https://www.ecmwf.int>, 2022. [Online; accessed 2022-23-06].
- Geoffrey W Evatt, I David Abrahams, Matthias Heil, Christoph Mayer, Jonathan Kingslake, Sarah L Mitchell, Andrew C Fowler, and Christopher D Clark. Glacial melt under a porous debris layer. *Journal of Glaciology*, 61(229):825–836, 2015.
- Daniel Farinotti, Matthias Huss, Johannes J Fürst, Johannes Landmann, Horst Machguth, Fabien Maussion, and Ankur Pandit. A consensus estimate for the ice thickness distribution of all glaciers on earth. *Nature Geoscience*, 12(3):168–173, 2019.

- Nathan Forsythe, Hayley J Fowler, Xiao-Feng Li, Stephen Blenkinsop, and David Pritchard. Karakoram temperature and glacial melt driven by regional atmospheric circulation variability. *Nature Climate Change*, 7(9):664–670, 2017.
- Holger Frey, Horst Machguth, Matthias Huss, Christian Huggel, S Bajracharya, Tobias Bolch, A Kulkarni, Andreas Linsbauer, Nadine Salzmann, and Markus Stoffel. Estimating the volume of glaciers in the himalayan–karakoram region using different methods. *The Cryosphere*, 8(6):2313–2333, 2014.
- Dorothy K Hall, George A Riggs, and Vincent V Salomonson. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote sensing of Environment*, 54(2):127–140, 1995.
- David E. Hartmann. Impacts on glacier mass balance in high mountain asia assessed using machine learning. *Master Thesis, Applied Data Science, Universiteit Utrecht*, 2022.
- Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.
- T Heid and A Kääb. Repeat optical satellite images reveal widespread and long term decrease in land-terminating glacier speeds. *The Cryosphere*, 6(2):467–478, 2012.
- Romain Hugonnet, Robert McNabb, Etienne Berthier, Brian Menounos, Christopher Nuth, Luc Girod, Daniel Farinotti, Matthias Huss, Ines Dussaillant, Fanny Brun, et al. Accelerated global glacier mass loss in the early twenty-first century. *Nature*, 592(7856):726–731, 2021.
- Matthias Huss. Extrapolating glacier mass balance to the mountain-range scale: the european alps 1900–2100. *The Cryosphere*, 6(4):713–727, 2012.
- Walter W Immerzeel, AF Lutz, M Andrade, A Bahl, H Biemans, Tobias Bolch, S Hyde, S Brumby, BJ Davies, AC Elmore, et al. Importance and vulnerability of the world’s water towers. *Nature*, 577(7790):364–369, 2020.
- Andreas Kääb, Etienne Berthier, Christopher Nuth, Julie Gardelle, and Yves Arnaud. Contrasting patterns of early twenty-first-century glacier mass change in the himalayas. *Nature*, 488(7412):495–498, 2012.
- Linghong Ke, Xiaoli Ding, and Chunqiao Song. Heterogeneous changes of glaciers over the western kunlun mountains based on icesat and landsat-8 derived glacier inventory. *Remote Sensing of Environment*, 168: 13–23, 2015.
- Phillip DA Kraaijenbrink, MFP Bierkens, AF Lutz, and WW Immerzeel. Impact of a global temperature rise of 1.5 degrees celsius on asia’s glaciers. *Nature*, 549(7671):257–260, 2017.
- Scikit Learn. sklearn Impute IterativeImputer. <https://scikit-learn.org/stable/modules/generated/sklearn.impute.IterativeImputer.html#sklearn.impute.IterativeImputer>, 2022a. [Online; accessed 2022-23-06].
- Scikit Learn. **sklearn.linear_model.LOOCV**. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.LeaveOneOut.html#sklearn.model_selection.LeaveOneOut, 2022b. [Online; accessed 2022-23-06].
- Scikit Learn. **sklearn.linear_model.Lasso**. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html, 2022c. [Online; accessed 2022-23-06].
- Fabien Maussion, Dieter Scherer, Thomas Mölg, Emily Collier, Julia Curio, and Roman Finkelburg. Precipitation seasonality and variability over the tibetan plateau as resolved by the high asia reanalysis. *Journal of Climate*, 27(5):1910–1927, 2014.

- Fabien Maussion, Anton Butenko, Nicolas Champollion, Matthias Dusch, Julia Eis, Kévin Fourteau, Philipp Gregor, Alexander H Jarosch, Johannes Landmann, Felix Oesterle, et al. The open global glacier model (oggm) v1. 1. *Geoscientific Model Development*, 12(3):909–931, 2019.
- William R Myers. Handling missing data in clinical trials: an overview. *Drug information journal: DIJ/Drug Information Association*, 34(2):525–533, 2000.
- NASA. Its Live Jet Propulsion Laboratory, California Institute of Technology. <https://its-live.jpl.nasa.gov>, 2022. [Online; accessed 2022-20-05].
- OGGM. Open Global Glacier Model. <https://oggm.org>, 2022. [Online; accessed 2022-23-06].
- Nicholas Pepin, Raymond S Bradley, HF Diaz, Michel Baraër, EB Caceres, N Forsythe, H Fowler, Gregory Greenwood, MZ Hashmi, XD Liu, et al. Elevation-dependent warming in mountain regions of the world. *Nature climate change*, 5(5):424–430, 2015.
- W Tad Pfeffer, Anthony A Arendt, Andrew Bliss, Tobias Bolch, J Graham Cogley, Alex S Gardner, Jon-Ove Hagen, Regine Hock, Georg Kaser, Christian Kienholz, et al. The randolph glacier inventory: a globally complete inventory of glaciers. *Journal of glaciology*, 60(221):537–552, 2014.
- DJ Quincey, SD Richardson, A Luckman, Richard M Lucas, JM Reynolds, MJ Hambrey, and NF Glasser. Early recognition of glacial lake hazards in the himalaya using remote sensing datasets. *Global and Planetary Change*, 56(1-2):137–152, 2007.
- SEDOO. Accelerated global glacier mass loss in the early twenty-first century - Dataset. <https://www.sedoo.fr/theia-publication-products/?uuid=c428c5b9-df8f-4f86-9b75-e04c778e29b9>, 2022. [Online; accessed 2022-23-06].
- David E Shean, Shashank Bhushan, Paul Montesano, David R Rounce, Anthony Arendt, and Batuhan Osmanoglu. A systematic, regional assessment of high mountain asia glacier mass balance. *Frontiers in Earth Science*, 7:363, 2020.
- Susan M Shortreed and Ashkan Ertefaie. Outcome-adaptive lasso: variable selection for causal inference. *Biometrics*, 73(4):1111–1122, 2017.
- Dan H Shugar, Aaron Burr, Umesh K Haritashya, Jeffrey S Kargel, C Scott Watson, Maureen C Kennedy, Alexandre R Bevington, Richard A Betts, Stephan Harrison, and Katherine Strattman. Rapid worldwide growth of glacial lakes since 1990. *Nature Climate Change*, 10(10):939–945, 2020.
- Annina Sorg, Tobias Bolch, Markus Stoffel, Olga Solomina, and Martin Beniston. Climate change impacts on glaciers and runoff in tien shan (central asia). *Nature Climate Change*, 2(10):725–731, 2012.
- Daniel Steiner, Andreas Walter, and HJ Zumbühl. The application of a non-linear back-propagation neural network to study the mass balance of grosse aletschgletscher, switzerland. *Journal of Glaciology*, 51(173): 313–323, 2005.
- Sandra L Taylor, L Renee Ruhaak, Karen Kelly, Robert H Weiss, and Kyoungmi Kim. Effects of imputation on correlation: implications for analysis of mass spectrometry data from multiple biological matrices. *Briefings in bioinformatics*, 18(2):312–320, 2017.
- Daniel Viviroli, Hans H Dürr, Bruno Messerli, Michel Meybeck, and Rolf Weingartner. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water resources research*, 43(7), 2007.
- Daniel Viviroli, David R Archer, Wouter Buytaert, Hayley J Fowler, Gregory B Greenwood, Alain F Hamlet, Yan Huang, Gernot Koboltschnig, MI Litaor, Juan Ignacio López-Moreno, et al. Climate change and mountain water resources: overview and recommendations for research, management and policy. *Hydrology and Earth System Sciences*, 15(2):471–504, 2011.

- Bert Wouters, Alex S Gardner, and Geir Moholdt. Global glacier mass loss during the grace satellite mission (2002-2016). *Frontiers in earth science*, 7:96, 2019.
- Qing Yan, Lewis A Owen, Zhongshi Zhang, Huijun Wang, Ting Wei, Nanxuan Jiang, and Ran Zhang. Divergent evolution of glaciation across high-mountain asia during the last four glacial-interglacial cycles. *Geophysical Research Letters*, 48(11):e2021GL092411, 2021.
- Kun Yang, Hui Wu, Jun Qin, Changgui Lin, Wenjun Tang, and Yingying Chen. Recent climate changes over the tibetan plateau and their impacts on energy and water cycle: A review. *Global and Planetary Change*, 112:79–91, 2014.
- Tandong Yao, Lonnie Thompson, Wei Yang, Wusheng Yu, Yang Gao, Xuejun Guo, Xiaoxin Yang, Keqin Duan, Huabiao Zhao, Baiqing Xu, et al. Different glacier status with atmospheric circulations in tibetan plateau and surroundings. *Nature climate change*, 2(9):663–667, 2012.
- Guoqing Zhang, Tandong Yao, Hongjie Xie, Weicai Wang, and Wei Yang. An inventory of glacial lakes in the third pole region and their changes in response to global warming. *Global and Planetary Change*, 131: 148–157, 2015.
- Liyun Zhao, Yi Yang, Doying Ji, and John Moore. Glacier evolution in high mountain asia under stratospheric sulfate aerosol injection geoengineering. *Atmospheric Chemistry and Physics Discussions*, pages 1–21, 09 2016. doi: 10.5194/acp-2016-830.
- Guoxiong Zheng, Simon Keith Allen, Anming Bao, Juan Antonio Ballesteros-Cánovas, Matthias Huss, Guoqing Zhang, Junli Li, Ye Yuan, Liangliang Jiang, Tao Yu, et al. Increasing risk of glacial lake outburst floods from future third pole deglaciation. *Nature Climate Change*, 11(5):411–417, 2021.

A Acknowledgements

I wish to acknowledge the help provided by my primary supervisor Philip Kraaijenbrink, who guided me throughout this project. I would also like to thank my secondary supervisor Walter Immerzeel.

B Code Availability

Code is available at: <https://github.com/BovandenBosch/Glacial-Mass-Blance-HMA.git>

C Supplementary Tables and Figures

C.1 Distribution of predictors variables

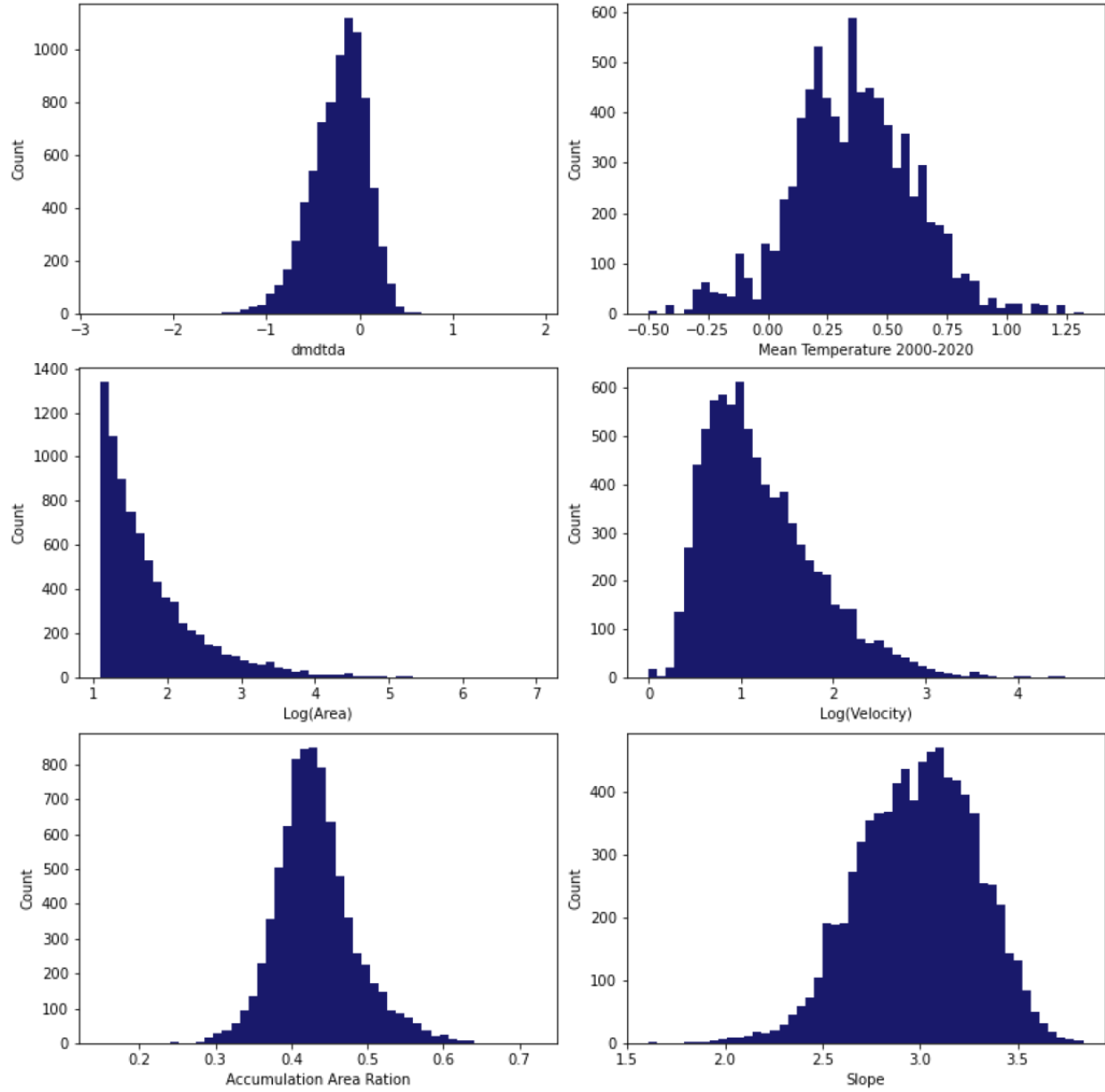


Figure 1: Count distributions of `dmdtda` (the SMB), the mean temperate 2000-2020, the log transform of the area, the log transform of the velocity, accumulation area ration and the slope. Bin size = 50

C.2 SMB as a function of glacier median elevation

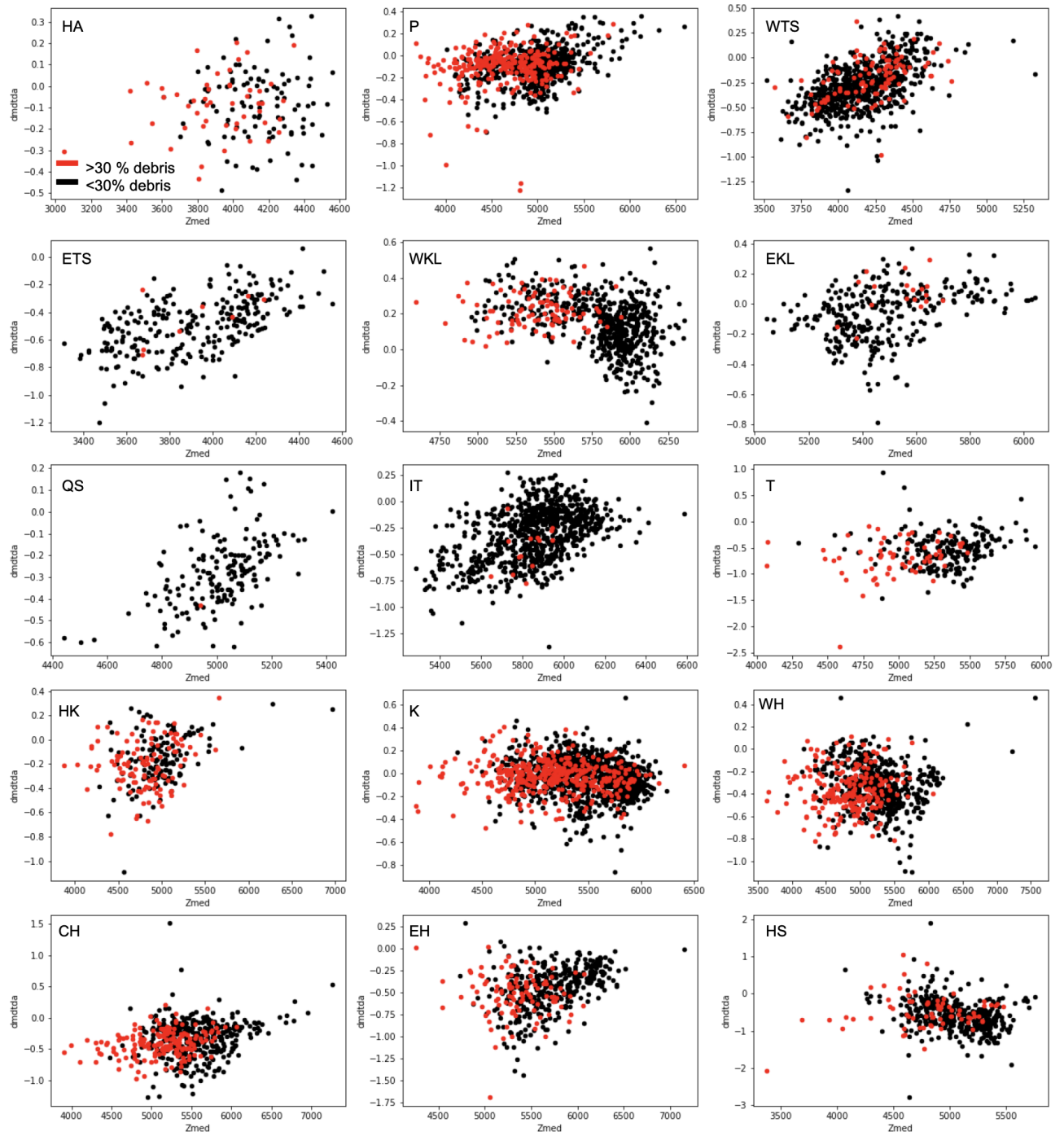


Figure 2: SMB (dmdtda) as a function of median elevation (Zmed), grouped by more or less than 30% debris coverage below the ELA. The red dots represent glacier with more than 30% coverage.

C.3 SMB as a function of glaciers slope

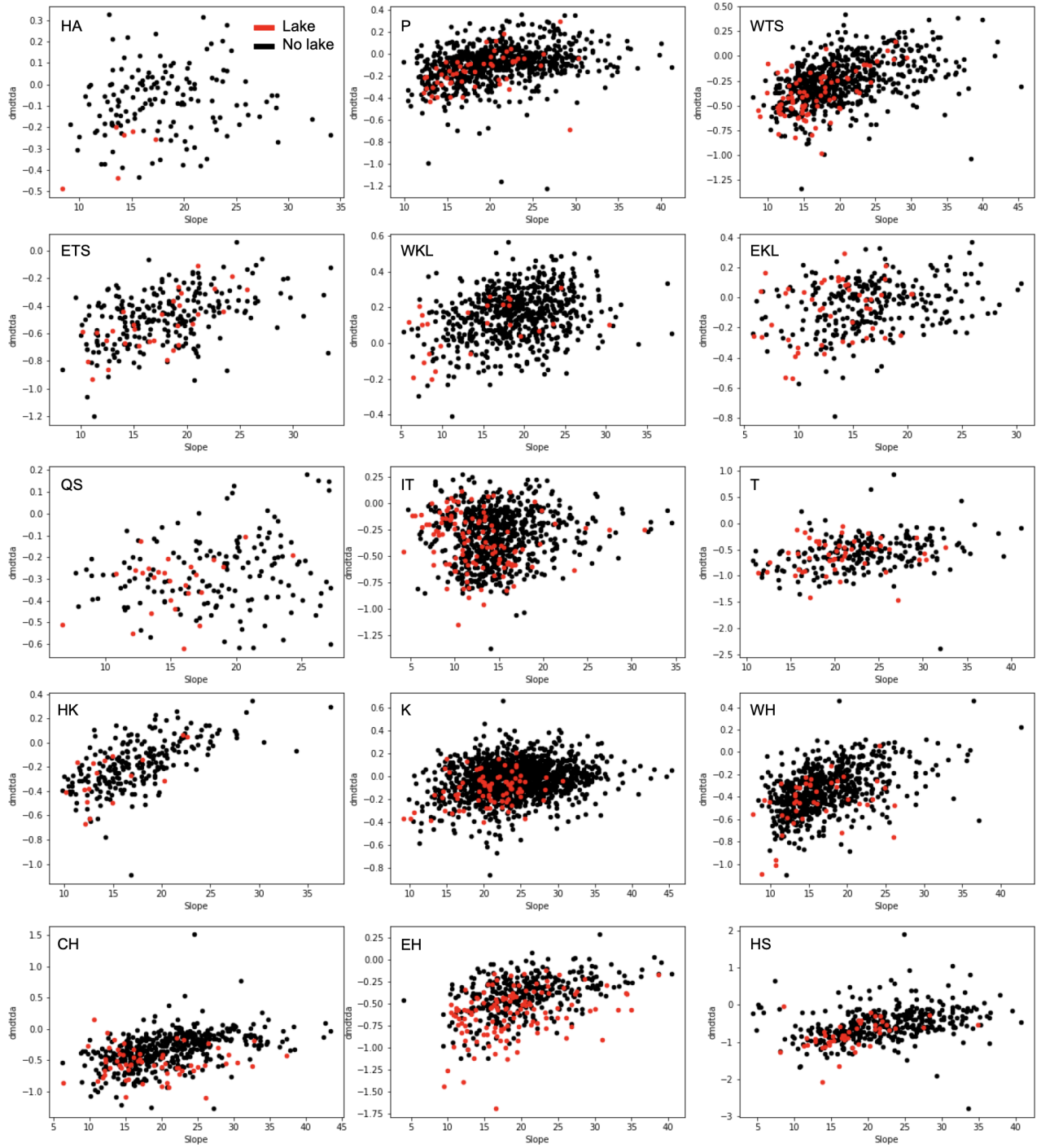


Figure 3: SMB (dmdtda) as a function of slope, grouped by the presence of a glacial Lake. The red dots represent glaciers associated with a lake

C.4 Pearson Correlations per Subregion

Pearson's Correlations Hissar Alay

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	**					**		***				
tstar_aar	0.22	-		***	***	*				***	*		
Slope	0.13	-0.02	-									**	***
Zmed	0.1	0.27	-0.15	-	***		**			*			**
debris_area_ela_p	-0.08	-0.33	0.11	-0.5	-	**				***		*	**
temp_mean	0.07	0.2	-0.05	-0.15	-0.21	-				***		*	
temp_diff	-0.22	0.1	-0.02	-0.23	-0.05	0.52	-		***	***			
prcp_diff	-0.09	-0.09	0.1	-0.1	0.02	-0.35	0.11	-	***	**	***	*	*
prcp_mean	0.36	0.16	0.12	-0.1	0.09	0.25	-0.36	-0.53	-		***		
temp_diff_1980-2000	-0.08	-0.28	0.01	0.21	0.27	-0.83	-0.37	0.23	-0.1	-	***		
prcp_diff_1980_2000	-0.05	0.18	-0.14	0.01	-0.08	0.63	0.08	-0.48	0.49	-0.48	-		
Area_log	-0.13	0.11	-0.26	-0.02	0.18	-0.17	0.03	0.19	-0.03	0.12	-0.11	-	***
Velocity_log	-0.05	-0.02	-0.32	-0.26	0.25	-0.14	0.09	0.17	-0.1	0.15	-0.09	0.55	-

Figure 4: Pearson correlation matrix of all selected variables in Hissar Alay. The variable dmdtda stand for the Specific Mass Balance. Significance levels: one star $p < 0.05$, two stars $p < 0.01$, three stars $p < 0.001$

Pearson's Correlations Pamir

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	***	***	***			*			***	***		*
tstar_aar	0.21	-	***	***	*		**	***		***	**	***	***
Slope	0.33	0.28	-	***			***		**	***	***		
Zmed	0.26	0.38	0.21	-	***	***	***	***	***	***		**	
debris_area_ela_p	-0.01	-0.09	0.03	-0.43	-	***		***	***		***	***	**
temp_mean	0.06	-0.06	0.02	-0.43	0.12	-	***	***	***	***	***	***	
temp_diff	-0.07	-0.11	-0.16	-0.35	0.03	0.69	-	***	***	***	*	***	**
prcp_diff	0.04	-0.13	0.05	0.55	-0.14	-0.38	-0.31	-	***				***
prcp_mean	-0.07	0.05	-0.11	-0.68	0.18	0.62	0.53	-0.82	-	***			***
temp_diff_1980-2000	0.31	0.23	0.24	0.33	-0.03	-0.2	-0.3	-0.04	-0.21	-	***	**	
prcp_diff_1980_2000	0.27	0.11	0.35	0.05	0.12	0.17	-0.08	0.06	0.01	0.24	-		
Area_log	-0.02	0.44	-0.04	0.1	0.21	-0.13	-0.19	-0.05	0.0	0.11	0.07	-	***
Velocity_log	0.09	0.41	0.02	-0.03	0.09	-0.02	-0.09	-0.2	0.18	0.02	0.05	0.68	-

Figure 5: Similar as figure 4

Pearson's Correlations Pamir

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	***	***	***			*			***	***		*
tstar_aar	0.21	-	***	***	*		**	***		***	**	***	***
Slope	0.33	0.28	-	***			***		**	***	***		
Zmed	0.26	0.38	0.21	-	***	***	***	***	***	***		**	
debris_area_ela_p	-0.01	-0.09	0.03	-0.43	-	***		***	***		***	***	**
temp_mean	0.06	-0.06	0.02	-0.43	0.12	-	***	***	***	***	***	***	
temp_diff	-0.07	-0.11	-0.16	-0.35	0.03	0.69	-	***	***	***	*	***	**
prcp_diff	0.04	-0.13	0.05	0.55	-0.14	-0.38	-0.31	-	***				***
prcp_mean	-0.07	0.05	-0.11	-0.68	0.18	0.62	0.53	-0.82	-	***			***
temp_diff_1980-2000	0.31	0.23	0.24	0.33	-0.03	-0.2	-0.3	-0.04	-0.21	-	***	**	
prcp_diff_1980_2000	0.27	0.11	0.35	0.05	0.12	0.17	-0.08	0.06	0.01	0.24	-		
Area_log	-0.02	0.44	-0.04	0.1	0.21	-0.13	-0.19	-0.05	0.0	0.11	0.07	-	***
Velocity_log	0.09	0.41	0.02	-0.03	0.09	-0.02	-0.09	-0.2	0.18	0.02	0.05	0.68	-

Figure 6: Similar as figure 4

Pearson's Correlations East Tien Shan

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	***	***	***		***				***	*		
tstar_aar	0.37	-		***		***	**	***	**	*	*	***	*
Slope	0.45	0.05	-	***					**	***	***	***	***
Zmed	0.57	0.21	0.54	-	**	***			***	***	***	**	
debris_area_ela_p	-0.05	-0.1	0.03	-0.18	-								
temp_mean	-0.31	-0.36	-0.03	-0.26	-0.03	-	*	***		***		***	
temp_diff	0.01	-0.16	0.01	-0.01	-0.08	0.13	-	***			**		
prcp_diff	-0.12	-0.22	0.09	-0.02	-0.03	0.41	0.22	-		***	***		*
prcp_mean	-0.09	0.18	-0.19	-0.54	0.08	-0.06	-0.04	-0.06	-	***	***		
temp_diff_1980-2000	0.24	0.15	0.26	0.54	-0.09	-0.64	-0.1	-0.48	-0.24	-		***	*
prcp_diff_1980_2000	0.12	-0.13	0.32	0.4	0.02	0.07	0.16	0.35	-0.49	0.05	-		*
Area_log	0.07	0.24	-0.29	0.16	0.02	-0.25	0.12	-0.1	-0.05	0.23	0.01	-	***
Velocity_log	-0.02	0.14	-0.26	0.09	-0.1	0.04	0.12	-0.13	-0.07	0.14	-0.16	0.67	-

Figure 7: Similar as figure 4

Pearson's Correlations West Kun Lun

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-		***	***	***	***		***	***			***	
tstar_aar	0.05	-		*	***	***		***	***	*	***	***	***
Slope	0.29	0.03	-	***	***	***	***	***	***	*	*	***	***
Zmed	-0.4	0.08	-0.35	-	***	***	***	***	***	***	***	***	***
debris_area_ela_p	0.16	0.2	0.37	-0.5	-	***	**	***	***	*	***		
temp_mean	0.47	0.19	0.39	-0.69	0.46	-	***	***	***	***	***	***	**
temp_diff	-0.06	0.03	-0.27	0.37	-0.13	-0.18	-	***		*	***		***
prcp_diff	0.37	0.16	0.24	-0.64	0.37	0.79	-0.16	-	***	***		*	
prcp_mean	0.49	0.15	0.34	-0.67	0.34	0.73	-0.02	0.79	-		***	**	*
temp_diff_1980-2000	-0.04	-0.09	-0.08	0.14	-0.08	-0.34	0.1	-0.25	-0.06	-	***		**
prcp_diff_1980_2000	-0.06	-0.13	-0.1	0.15	-0.22	-0.38	0.18	-0.04	0.16	0.46	-	***	***
Area_log	-0.26	0.35	-0.46	0.21	-0.02	-0.23	0.07	-0.09	-0.12	0.06	0.13	-	***
Velocity_log	-0.05	0.4	-0.19	0.13	0.03	-0.1	0.17	0.03	0.08	0.11	0.19	0.65	-

Figure 8: Similar as figure 4

Pearson's Correlations East Kun Lun

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-		***	***	**		***		***		***	*	
tstar_aar	0.03	-	**	***		**		***			**		***
Slope	0.29	0.15	-	***		***		*	**	***	***	***	***
Zmed	0.37	0.28	-0.22	-		***		***		***	***	***	***
debris_area_ela_p	0.15	0.11	0.05	0.03	-			***	**		**		***
temp_mean	-0.09	0.16	0.49	-0.48	0.03	-	***		***	***	***	***	***
temp_diff	0.22	-0.11	-0.01	0.03	-0.07	-0.33	-	***		**	**		
prcp_diff	-0.07	-0.23	-0.11	-0.34	-0.38	-0.06	0.35	-		**	***		**
prcp_mean	-0.37	0.05	-0.19	0.03	-0.16	-0.35	-0.03	0.04	-	***	***	*	
temp_diff_1980-2000	0.04	-0.03	-0.36	0.36	-0.06	-0.69	0.18	0.16	0.23	-	***	***	***
prcp_diff_1980_2000	-0.2	-0.16	-0.48	0.21	-0.18	-0.73	0.17	0.19	0.23	0.34	-	***	***
Area_log	-0.13	0.05	-0.5	0.32	-0.0	-0.32	0.03	0.0	0.11	0.26	0.32	-	***
Velocity_log	-0.08	0.21	-0.39	0.36	0.19	-0.25	0.04	-0.17	0.06	0.23	0.22	0.65	-

Figure 9: Similar as figure 4

Pearson's Correlations Qilian Shan

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	***	**	***		***		***	***	***	**		**
tstar_aar	0.31	-	**	**		***	***				***		*
Slope	0.2	0.23	-	*		***	***	***	**		***	***	**
Zmed	0.54	0.24	-0.16	-	***		***	***	***	***		*	***
debris_area_ela_p	-0.12	0.07	0.11	-0.31	-		**		***		*		
temp_mean	0.25	0.47	0.46	-0.08	0.1	-	***	***		***	***	***	
temp_diff	0.15	-0.27	-0.53	0.43	-0.19	-0.52	-	***	***	**	***	***	*
prcp_diff	0.28	-0.01	-0.25	0.42	-0.12	-0.49	0.47	-	***	***	***	**	*
prcp_mean	-0.5	0.12	0.2	-0.66	0.28	0.1	-0.43	-0.39	-	***			
temp_diff_1980-2000	0.25	-0.14	0.0	0.27	-0.04	-0.33	0.24	0.36	-0.43	-	*		**
prcp_diff_1980_2000	-0.21	-0.27	-0.32	0.12	-0.18	-0.67	0.26	0.28	0.04	0.19	-	***	**
Area_log	-0.03	0.08	-0.37	0.19	-0.1	-0.28	0.26	0.21	-0.05	0.12	0.26	-	***
Velocity_log	0.21	0.19	-0.2	0.33	-0.13	-0.13	0.16	0.16	-0.05	0.2	0.22	0.61	-

Figure 10: Similar as figure 4

Pearson's Correlations Inner Tibet

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	***		***	***	***			***	***	***		**
tstar_aar	0.26	-		***	***		*		**		*	***	***
Slope	0.06	-0.02	-	***	***	***			***	***	***	***	***
Zmed	0.46	0.44	0.23	-	***	**	*		***		**		
debris_area_ela_p	-0.25	-0.19	0.28	-0.13	-	***			***	***	***	**	**
temp_mean	-0.22	-0.05	0.45	0.1	0.17	-	***	***	***	***	***	***	
temp_diff	0.07	-0.08	-0.02	-0.09	0.01	0.13	-	***	**	***			**
prcp_diff	-0.02	-0.02	0.05	0.02	0.02	-0.28	-0.56	-	*	***			
prcp_mean	-0.57	-0.11	0.25	-0.39	0.24	0.55	0.1	-0.08	-	***	***	*	***
temp_diff_1980-2000	0.33	0.06	-0.25	-0.06	-0.13	-0.61	-0.27	0.31	-0.45	-	***	***	***
prcp_diff_1980_2000	-0.32	-0.09	0.41	0.11	0.24	0.69	-0.06	0.01	0.68	-0.52	-	***	
Area_log	0.06	0.21	-0.41	0.02	-0.09	-0.26	-0.05	0.02	-0.07	0.19	-0.17	-	***
Velocity_log	0.1	0.28	-0.16	0.04	-0.1	-0.06	-0.1	0.0	0.12	0.13	0.05	0.66	-

Figure 11: Similar as figure 4

Pearson's Correlations South and East Tibet

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	***	***	***	**				*		**	*	
tstar_aar	0.25	-		***	***		*					*	***
Slope	0.3	-0.07	-	**	**	***			***		**	***	***
Zmed	0.21	0.37	-0.17	-	***	***	***	***	***	**	***		
debris_area_ela_p	-0.19	-0.22	0.16	-0.58	-	***	**		***	*	***	***	
temp_mean	0.1	-0.02	0.22	-0.56	0.26	-	***	***	***	**	***		
temp_diff	-0.08	-0.15	-0.03	0.35	-0.17	-0.27	-	***	***	***	***	**	*
prcp_diff	0.1	0.02	0.05	-0.22	0.02	0.73	-0.21	-	***	***	***		
prcp_mean	0.15	0.04	0.23	-0.44	0.21	0.75	-0.48	0.78	-	***	***		
temp_diff_1980-2000	0.04	0.03	0.03	0.18	-0.14	-0.18	0.52	-0.25	-0.32	-	**		
prcp_diff_1980_2000	-0.19	-0.04	-0.17	0.5	-0.28	-0.83	0.43	-0.72	-0.88	0.19	-		
Area_log	-0.14	0.13	-0.26	-0.08	0.26	-0.01	-0.17	-0.03	0.04	0.05	-0.08	-	***
Velocity_log	0.04	0.21	-0.34	0.07	0.01	-0.11	-0.14	-0.04	-0.02	-0.03	0.07	0.68	-

Figure 12: Similar as figure 4

Pearson's Correlations Hindu Kush

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	***	***	***		***				***		*	***
tstar_aar	0.25	-	**	***						***		***	**
Slope	0.62	0.19	-	***	*	***				***			
Zmed	0.3	0.27	0.38	-	***	***		***	***	**			
debris_area_ela_p	-0.05	-0.11	-0.15	-0.37	-	**		**	***	*	*	***	*
temp_mean	-0.25	-0.06	-0.23	-0.43	-0.2	-	**	***	***	***	***	***	**
temp_diff	0.09	-0.01	-0.03	0.07	0.02	0.19	-	***	**		*		
prcp_diff	0.08	0.0	0.06	0.36	0.17	-0.47	0.35	-	***	*			
prcp_mean	-0.07	0.02	-0.03	-0.41	-0.22	0.74	-0.18	-0.66	-	**	***	*	
temp_diff_1980-2000	-0.47	-0.21	-0.46	-0.21	-0.16	0.39	-0.02	-0.13	0.16	-	***	***	***
prcp_diff_1980_2000	-0.05	-0.03	-0.01	0.09	0.15	-0.23	0.14	0.04	-0.47	-0.36	-		
Area_log	0.13	0.29	-0.07	0.09	0.25	-0.21	0.05	0.11	-0.16	-0.27	-0.03	-	***
Velocity_log	0.28	0.19	0.02	-0.07	0.15	-0.2	-0.11	-0.02	-0.03	-0.26	-0.12	0.7	-

Figure 13: Similar as figure 4

Pearson's Correlations Karakoram

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-		***	***		***	***	***	***	*		***	**
tstar_aar	0.02	-	***	***	***	**		***	***	***	***	***	***
Slope	0.23	0.12	-		***	**	*		***	*	*	***	***
Zmed	-0.16	0.09	0.05	-	***	***	***		***	***	***		***
debris_area_ela_p	0.03	0.12	0.21	-0.46	-	***			***	***	***	***	***
temp_mean	0.21	-0.09	-0.08	-0.56	0.19	-			***		***	***	*
temp_diff	0.17	0.03	-0.06	-0.12	0.02	0.04	-	***	***	***	***		
prcp_diff	-0.12	0.17	-0.03	-0.01	0.02	-0.04	0.4	-		***	***		
prcp_mean	0.12	0.21	0.23	-0.65	0.4	0.44	-0.21	-0.04	-	***	***	***	***
temp_diff_1980-2000	-0.05	-0.14	-0.06	0.57	-0.25	-0.05	-0.4	-0.26	-0.38	-	***	***	***
prcp_diff_1980_2000	-0.03	0.15	0.06	-0.19	0.1	-0.14	-0.15	-0.21	0.51	-0.3	-	***	***
Area_log	-0.16	0.42	-0.13	-0.02	0.3	-0.12	-0.02	0.03	0.15	-0.12	0.14	-	***
Velocity_log	-0.08	0.34	-0.15	-0.3	0.3	0.05	-0.02	0.02	0.37	-0.27	0.22	0.76	-

Figure 14: Similar as figure 4

Pearson's Correlations Karakoram

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-		***	***		***	***	***	***	*		***	**
tstar_aar	0.02	-	***	***	***	**		***	***	***	***	***	***
Slope	0.23	0.12	-		***	**	*		***	*	*	***	***
Zmed	-0.16	0.09	0.05	-	***	***	***		***	***	***		***
debris_area_ela_p	0.03	0.12	0.21	-0.46	-	***			***	***	***	***	***
temp_mean	0.21	-0.09	-0.08	-0.56	0.19	-			***		***	***	*
temp_diff	0.17	0.03	-0.06	-0.12	0.02	0.04	-	***	***	***	***		
prcp_diff	-0.12	0.17	-0.03	-0.01	0.02	-0.04	0.4	-		***	***		
prcp_mean	0.12	0.21	0.23	-0.65	0.4	0.44	-0.21	-0.04	-	***	***	***	***
temp_diff_1980-2000	-0.05	-0.14	-0.06	0.57	-0.25	-0.05	-0.4	-0.26	-0.38	-	***	***	***
prcp_diff_1980_2000	-0.03	0.15	0.06	-0.19	0.1	-0.14	-0.15	-0.21	0.51	-0.3	-	***	***
Area_log	-0.16	0.42	-0.13	-0.02	0.3	-0.12	-0.02	0.03	0.15	-0.12	0.14	-	***
Velocity_log	-0.08	0.34	-0.15	-0.3	0.3	0.05	-0.02	0.02	0.37	-0.27	0.22	0.76	-

Figure 15: Similar as figure 4

Pearson's Correlations Central Himalaya

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	***	***	***	***	***		***	***	***			**
tstar_aar	0.37	-	***	***	***	***	***	***	***	***		***	***
Slope	0.43	0.33	-	***	***	***	**	***	***	***		**	***
Zmed	0.24	0.61	0.22	-	***	**	*	*	***	***			
debris_area_ela_p	-0.19	-0.38	-0.23	-0.57	-			***		***		***	
temp_mean	0.26	0.19	0.34	-0.13	-0.06	-		***	***	***	***		
temp_diff	-0.06	-0.18	-0.14	-0.1	0.03	-0.0	-		**	***		**	
prcp_diff	0.32	0.26	0.32	0.09	-0.15	0.66	-0.02	-	***	***	***		*
prcp_mean	0.35	0.23	0.36	-0.19	0.0	0.74	-0.12	0.7	-	***	***		**
temp_diff_1980-2000	-0.17	-0.28	-0.29	-0.23	0.19	-0.53	0.41	-0.57	-0.33	-	***		
prcp_diff_1980_2000	-0.01	-0.01	0.03	-0.08	0.03	0.19	-0.08	0.31	0.25	-0.26	-		
Area_log	-0.08	0.18	-0.13	0.05	0.2	-0.05	-0.13	-0.01	-0.0	0.02	-0.03	-	***
Velocity_log	0.14	0.24	-0.17	0.02	0.03	0.05	0.0	0.11	0.14	0.0	0.02	0.52	-

Figure 16: Similar as figure 4

Pearson's Correlations East Himalaya

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	***	***	***	***		*				***		
tstar_aar	0.32	-	*	***	***							***	***
Slope	0.44	0.11	-			*		***		*			***
Zmed	0.33	0.72	0.05	-	***	***	***	***	***	***	***		
debris_area_ela_p	-0.19	-0.41	-0.09	-0.48	-		*	**		**	*	***	
temp_mean	-0.01	-0.01	0.11	-0.41	-0.01	-	***	***	***	***	***	***	*
temp_diff	-0.12	-0.04	-0.05	-0.24	-0.11	0.29	-		***	***	**	**	*
prcp_diff	-0.02	-0.08	0.17	-0.26	0.13	0.35	-0.06	-	***	***			*
prcp_mean	0.02	0.01	0.09	-0.36	-0.05	0.75	0.31	0.18	-	***	***		
temp_diff_1980-2000	0.08	0.0	0.11	0.18	0.14	-0.28	-0.79	0.32	-0.25	-		**	**
prcp_diff_1980_2000	-0.2	0.03	-0.09	0.16	-0.11	-0.29	-0.12	-0.04	-0.35	0.06	-		
Area_log	-0.03	0.17	-0.08	0.08	0.37	-0.22	-0.13	-0.04	-0.08	0.14	-0.09	-	***
Velocity_log	-0.02	0.22	-0.2	0.08	0.06	-0.09	0.12	-0.12	-0.03	-0.15	-0.02	0.52	-

Figure 17: Similar as figure 4

Pearson's Correlations East Himalaya

	dmdtda	tstar_aar	Slope	Zmed	debris_area_ela_p	temp_mean	temp_diff	prcp_diff	prcp_mean	temp_diff_1980-2000	prcp_diff_1980_2000	Area_log	Velocity_log
dmdtda	-	***	***	***	***		*				***		
tstar_aar	0.32	-	*	***	***							***	***
Slope	0.44	0.11	-			*		***		*			***
Zmed	0.33	0.72	0.05	-	***	***	***	***	***	***	***		
debris_area_ela_p	-0.19	-0.41	-0.09	-0.48	-		*	**		**	*	***	
temp_mean	-0.01	-0.01	0.11	-0.41	-0.01	-	***	***	***	***	***	***	*
temp_diff	-0.12	-0.04	-0.05	-0.24	-0.11	0.29	-		***	***	**	**	*
prcp_diff	-0.02	-0.08	0.17	-0.26	0.13	0.35	-0.06	-	***	***			*
prcp_mean	0.02	0.01	0.09	-0.36	-0.05	0.75	0.31	0.18	-	***	***		
temp_diff_1980-2000	0.08	0.0	0.11	0.18	0.14	-0.28	-0.79	0.32	-0.25	-		**	**
prcp_diff_1980_2000	-0.2	0.03	-0.09	0.16	-0.11	-0.29	-0.12	-0.04	-0.35	0.06	-		
Area_log	-0.03	0.17	-0.08	0.08	0.37	-0.22	-0.13	-0.04	-0.08	0.14	-0.09	-	***
Velocity_log	-0.02	0.22	-0.2	0.08	0.06	-0.09	0.12	-0.12	-0.03	-0.15	-0.02	0.52	-

Figure 18: Similar as figure 4

C.5 Regression Residuals per Subregion

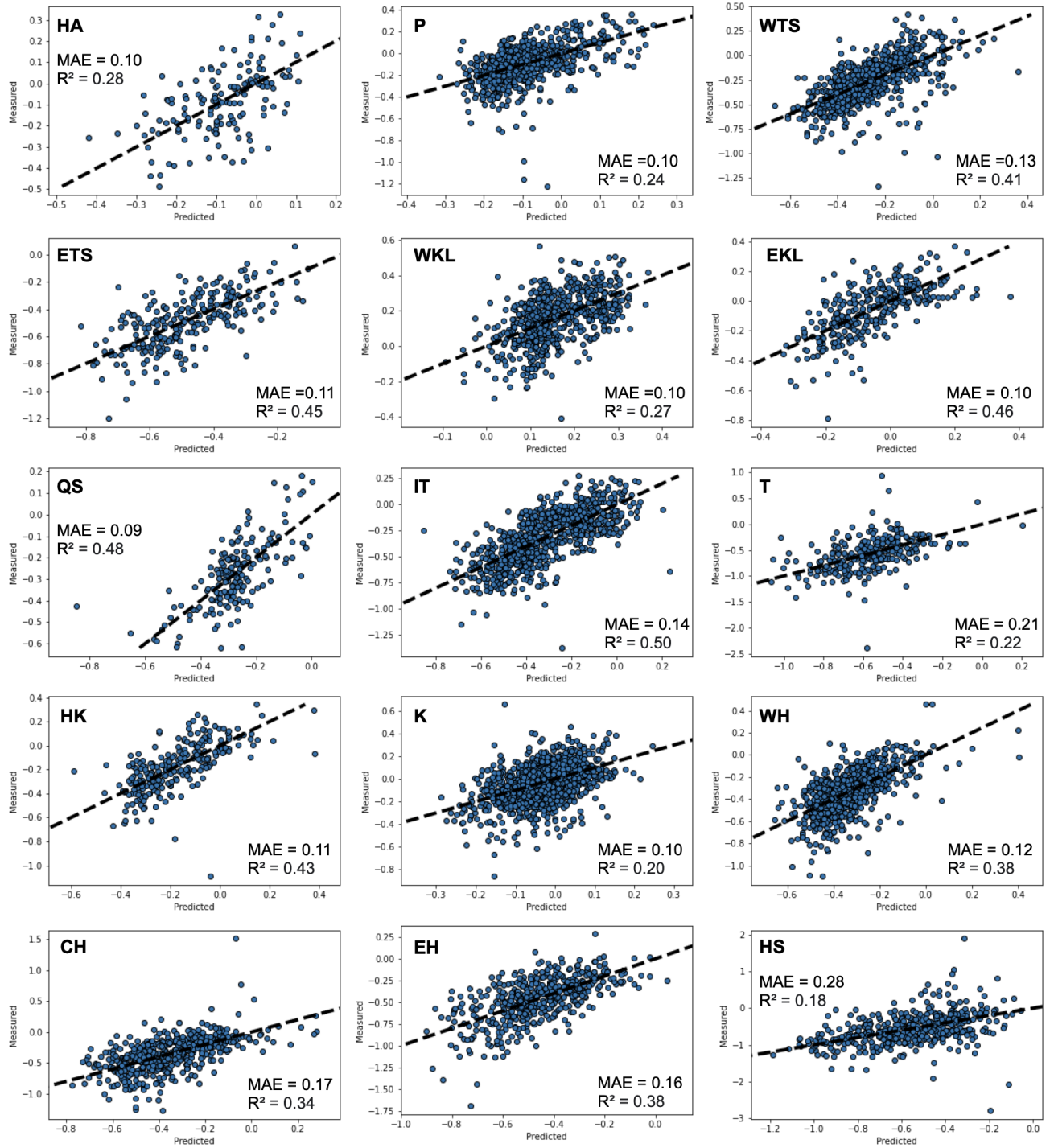


Figure 19: Regression residuals of SMB predictions of per subregion.