master thesis Applied Data Science

Impacts on glacier mass balance in High Mountain Asia assessed using machine learning

Universiteit Utrecht

Master candidate:

David Hartmann

Student no. 4780167

First supervisor:

Dr. Philip Kraaijenbrink

Second supervisor:

Prof. dr. Walter Immerzeel

July 1, 2022

Abstract

High Mountain Asia's glaciers are an important source of water for the major river systems in Asia and provide water to hundreds of millions of people. Understanding and predicting mass balance change in High Mountain Asia is essential to anticipate and mitigate the environmental and socioeconomic impacts of the melting glaciers. Nevertheless, large uncertainties exist in the contributions of climatic and morphological variables to the specific mass balance. In this study, the response of 10 morphological and 6 climatic variables to the specific mass balance of 8099 individual glaciers in High Mountain Asia is investigated with statistical models. These models are multiple linear regression, Random Forest, XGBoost and artificial neural network. The results show that the average precipitation between 2000 and 2020, and the slope of the glacier are the two most important variables. Other strong predictors are the average temperature between 2000 and 2020, the average glacier elevation, the precipitation change between 1980 and 2000, and the temperature difference between 2000 and 2000. The debris area under the ELA, accumulation area ratio and the mean velocity influence the specific mass balance but to a lesser degree. The presence of glacial lakes, aspect, average ice thickness and whether a glacier is surging are not important in explaining the total specific mass balance. The nonlinear models performed significantly better than the linear models, as the XGBoost model had 113.7% more explained variance ($R^2 = 0.685$) than the multiple linear regression model ($R^2 = 0.307$). Of the nonlinear models, the XGBoost had the highest performance. Substantial nonlinearities are captured in the nonlinear models, that indicate that temperature change, precipitation change between 1980 and 2000, debris under the ELA, and the median glacier elevation have nonlinear responses to the specific mass balance. This study has reiterated the importance of accounting for nonlinear responses when modelling the specific mass balance of High Mountain Asia glaciers. For further research, it would be interesting to perform the statistical analysis with the nonlinear models and the permutation and SHAP feature importance techniques for more climatically homogeneous regions.

Contents

1 Introduction

By the end of the century, it is expected that 200 million people that live on land will reside below the high tide lines of rising sea levels. Over the next three decades, more than one billion people could encounter water scarcity and food insecurity. Glaciers have a prominent impact on this, as they are after thermal expansion the largest predicted source of sea-level rise in the twenty-first century (Hugonnet et al., 2021). Besides, they are one of the most climate-sensitive components of the world's natural water towers (Immerzeel et al., 2020).

Nevertheless, glacier systems are one of the least understood elements of the global water cycle, identified as a critical research gap by the IPCC (Abram, Carolina, Bindoff, & Cheng, 2019). Recent studies showed the contrasted glacier mass changes within large glacier regions (e.g. Brun et al., 2019; Wang, Liu, Shangguan, Radić, & Zhang, 2019; Sakai & Fujita, 2017; (Hugonnet et al., 2021)). The causes for these spatiotemporal differences are still not fully understood, because there is uncertainty regarding the mass balance response to different climatic and morphological conditions (Brun et al., 2019). Previous studies showed that important predictors for the specific mass balance (glacier mass change rate per unit area in meters water-equivalent per year, SMB) are the slope of the tongue and the mean elevation of the glaciers (Brun et al., 2019). Gentler slopes are associated with more negative SMB values and the higher the glacier elevation, the more positive the SMB (Brun et al., 2019). Uncertainties exist about the degree of influence of the debris coverage, the glacier's flow velocity, temperature rise, presence of glacial lakes and surging behaviour among other variables (Brun et al., 2019; Dehecq et al., 2019; Huo, Bishop, & Bush, 2021).

Modelling the SMB at a large scale can be done in several ways (Bolibar et al., 2020). Firstly, the SMB can be modelled with empirical models that are based on empirical relationships between temperature and melt and accumulation, such as the temperature-index model from Hock (2003). Secondly, statistical or machine learning models can be used. They predict the SMB with statistical relationships of topographical and climate predictors (e.g. Brun et al., 2019; Bolibar et al., 2020). Lastly, the SMB can be modelled with physical and surface energy balance models (e.g. Gerbaux, Genthon, Etchevers, Vincent, & Dedieu, 2005), that take all energy changes between the glacier and the atmosphere into account (Bolibar et al., 2020).

Statistical models have been used in glaciology for more than 50 years, beginning with linear regressions on a few climatic variables (e.g Hoinkes, 1968). In the last decades, statistical modelling has made enormous progress, mainly due to advances in machine learning. Steiner, Walter, and Zumbühl (2005) were the first to use artificial neural networks (ANNs) in glaciology to model glacier mass balances. They showed that a nonlinear model is better capable of modelling glacier mass balances of the Great Aletsch Glacier in Switzerland than a multiple linear regression model. Bolibar et al. (2020) used a deep learning model to simulate the mass balances of 32 French Alpine glaciers and showed that it explained 64% more of the

variance in space than a linear method. As glacier and climate systems are nonlinear, nonlinear models allow for exploration of the nonlinear structures of the systems (Bolibar et al., 2020; Steiner et al., 2005).

In this study, statistical models are used to model the SMB between 2000 and 2020 of glaciers in High Mountain Asia (HMA). HMA contains the largest concentration of glacier ice outside of the polar regions and is therefore referred to as "The Third Pole". The water demands of hundreds of millions of people are supplied by the glaciers of HMA (Biemans et al., 2019). The current glacier retreat in this region temporarily relieves water stress, but this short-lived effect will eventually decline. To model the SMB in this region, 8 climatic and 16 glacier morphological variables are considered. Modelling the SMB with different climatic and topographic variables as input can help to understand how glaciers respond to different processes, which is vital for mass balance models to improve. As a consequence, better predictions of glacier melt in HMA will be possible. This is essential to avoid water scarcity and geopolitical instability in this region, and to anticipate and mitigate the environmental impacts that are associated with glacier melt (Immerzeel et al., 2020; Hugonnet et al., 2021). The research question that is posed in this study is:

- How does the specific mass balance of High Mountain Asia glaciers respond to different climatic and morphological variables?

To answer the main research question, the focus lies on the subquestions:

- Which machine learning algorithm performs best in predicting the mass balance change?
- To which variables does the specific mass balance respond nonlinearly?

To assess how the SMB responds to every variable, a multiple linear regression model, Random Forest model, XGBoost model and an ANN model are made. Also, the model performances are compared to evaluate which statistical model predicts the SMB best. Finally, there is examined which variables have a nonlinear response on the SMB.

2 Study area

High Mountain Asia is divided into three main regions in terms of glaciers (RGI Consortium, 2017): Central Asia, South Asia West and South Asia East. It consists of the Himalayas, other mountain ranges such as Pamir, Hundu Kush, Karakoram and the Tien Shan, and the vast Tibetan Plateau (Shean et al., 2020). Each of these regions hosts a range of topographic and climatic regimes (Treichler, Kääb, Salzmann, & Xu, 2018). A map of HMA with its second-order regions is shown in Figure 1.

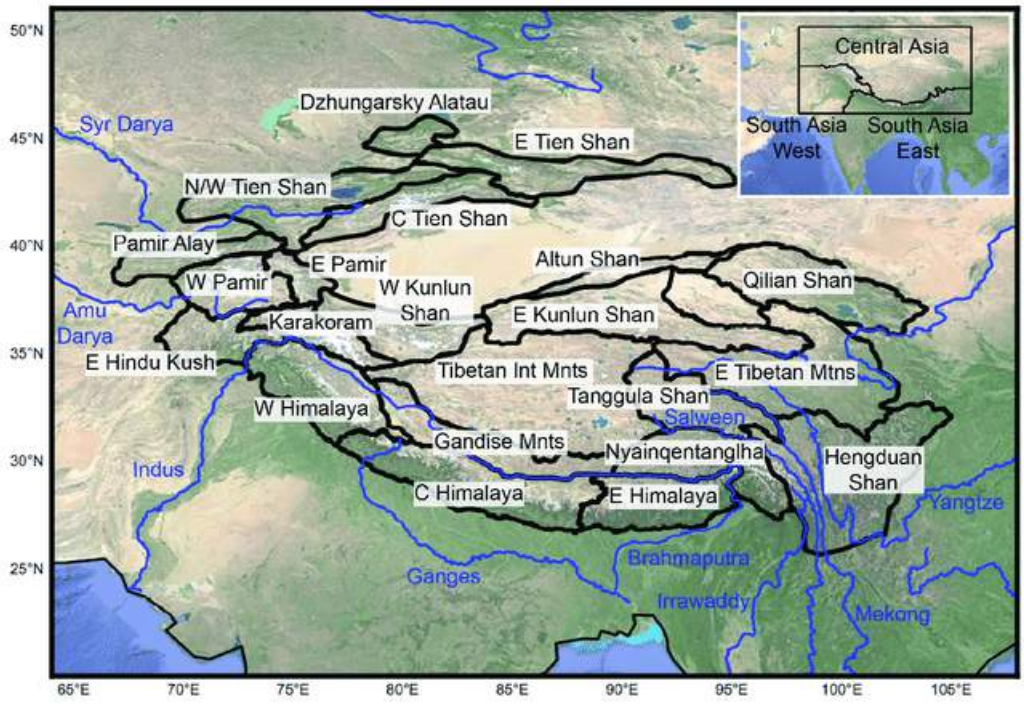

Figure 1: Map of High Mountain Asia showing the three primary RGI regions and its subregions. Replicated from Rounce et al. (2020).

The precipitation in HMA is characterized by different climate systems, that are shown in Figure 2. The precipitation of the central and eastern part of the Himalayas is brought for circa 80% by the Indian and South-East Asian monsoons, which occur in the summer months (Shean et al., 2020). The influence of the monsoon systems decreases from southeast to northwest. The Pamir, Hindu Kush and the Karakoram receive 60 - 70% of the precipitation from the westerlies, of which most of its precipitation occurs in the winter months (Shean et al., 2020). The influence of the westerlies decreases from southeast to northwest (Lutz, 2018). A dry and cold continental climate is present in the Tibetan Plateau (Lalande, Ménégoz, Krinner, Naegeli, & Wunderle, 2021). The climate in HMA has been warming during recent decades, with

steeper trends than in non-mountainous regions (Kraaijenbrink, Bierkens, Lutz, & Immerzeel, 2017). The warming rate of HMA is circa 11% higher than other continental areas south of 60°north latitude. This is likely caused by changes in albedo and other processes specific to mountainous areas (Lalande et al., 2021; Kraaijenbrink et al., 2017).

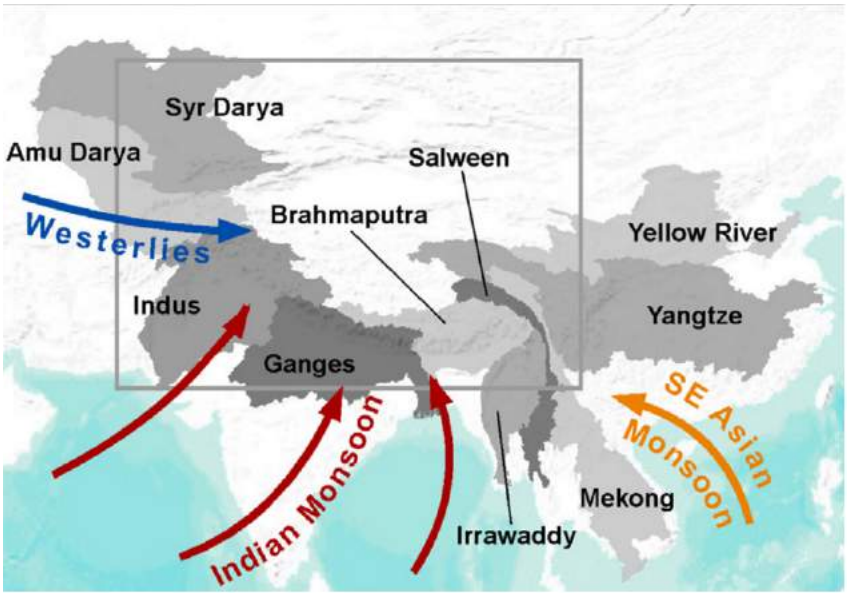

Figure 2: Map of most important climate systems in High Mountain Asia (grey outlined) and its river basins. Replicated from Lutz (2018).

HMA consists of 95,536 glaciers that cover an area of 97,606 km^2. Between 2000 and 2018, the total mass balance of HMA glaciers was -19.0 ± 2.5 Gt/a, equivalent to 0.19 ± 0.03 m.w.e./a. This contributed to circa 0.7 mm sea-level rise (Shean et al., 2020). By the end of 2100, it is estimated that the glaciers in HMA could lose 45 ± 8% for the RCP 2.6 scenario and up to 69 ± 14% for the RCP 8.5 scenario (Hock et al., 2019). Thus, the general tendency is that HMA glaciers are retreating and losing mass. However, due to the complex topography and large-scale climate systems that affect the glaciers, the evolution of glaciers in HMA is not homogeneous. For example, Hissar Alay and the Qilian Shan are regions that show the most negative mass balances (-0.36 ± 0.31 and -0.49 ± m.w.e/a) (Kraaijenbrink et al., 2017). In the Karakoram and Kunlun Shan regions, glacier mass budgets are neutral or even positive (Rounce et al., 2020).

3 Data

In this study, the influence of 16 glacier morphological and 8 climatic variables on the SMB were explored. The variables were selected based on findings of previous studies (e.g. Brun et al., 2019; Bolibar et al., 2020; Hock, 2010. An overview of all variables used in this study is shown in Table A.1.

3.1 Mass balance

The mass balance of a glacier is the difference between accumulation and ablation. For this study, the mass balance dataset of Hugonnet et al. (2021) is used. This dataset provides mass balance change estimates between 2000 and 2019 at 100 meters resolution for 99.8% of the glaciers in HMA (Hugonnet et al., 2021). These changes were acquired from surface elevation changes that were determined by using stereo images from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) satellite sensor (NASA, 2022). From the dataset of Hugonnet et al. (2021), the following variable is used as the dependent variable in this study:

dmdtda
The mass change rate per unit area (specific mass balance) in meters water-equivalent per year between 2000 and 2020.

3.2 Glacier geometry

Randolph Glacier Inventory
The glacier outlines from the Randolph Glacier Inventory version 6.0 (RGI 6.0) are used, which is a global inventory of glacier outlines (RGI Consortium, 2017). In this study, the glacier regions of High-Mountain Asia are used, respectively primary regions Central Asia (Region 13), South Asia West (Region 14), and South Asia East (Region 15). For the HMA glaciers, the RGI outlines are based on remote sensing data that are gathered between 1998 and 2013. The RGI is provided as shapefiles per primary region, containing the outlines of glaciers in geographic coordinates (longitude and latitude, in degrees) (Arendt et al., 2017).
In this study, the following variables from the RGI are used:

RGIId
A 14-character unique identifier of the form RGIvv-rr.nnnnn, where *vv* is the version number, *rr* is the primary region and *nnnnn* is an arbitrary identifying code that is unique within the primary region (Arendt et al., 2017). The RGIId is used as the primary key to link multiple datasets.

Area

The area of the glacier in km^2, calculated in cartesian coordinates on a cylindrical equal-area projection of the authalic sphere of the WGS84 ellipsoid (Arendt et al., 2017). Small glaciers (<0.1 km^2) have a small contribution to the total HMA mass balance (Shean et al., 2020). 46 glaciers with an area larger than 100 km^2 lost more mass than > 55,000 glaciers with an area smaller than 0.3 km^2. The error of mass balance estimations is higher for small glaciers, due to low-resolution issues and static RGI polygon outlines for the 2000 - 2018 period (Shean et al., 2020; Brun et al., 2019). Therefore, in this study, only the glaciers larger than 2 km^2 are considered, comprising 8099 glaciers.

Zmin, Zmax

Minimum and maximum elevation of the glacier in meters above sea level. For most glaciers, it was obtained directly from a DEM (Arendt et al., 2017).

Zmed

Median elevation (m) of the glacier. It is acquired by sorting the elevation of the DEM cells of the glacier and selecting the 50th percentile of the cumulative frequency distribution (Arendt et al., 2017).

Slope

Mean slope of the glacier surface in degrees. It is acquired by averaging the slopes of every cell from the DEM (Arendt et al., 2017).

Aspect

Mean aspect (orientation) of the glacier, expressed as an integer azimuth relative to 0° at due north. For each grid cell of the glacier's DEM, the aspect sines and cosines are calculated. The mean aspect is calculated as the arctangent of the quotient of the sum of the aspect sines and cosines (Arendt et al., 2017).

Lmax

Length of the longest surface flowline of the glacier in meters. The length is measured with the algorithm of Machguth and Huss (2014). This algorithm selects points on the glacier outline above *Zmed* as candidate starting points. The flowline is computed by choosing successive DEM cells according to a weighted blend of steepest descent and greatest distance from the glacier outline. The longest of the resulting lines is chosen as the glacier's centreline (Arendt et al., 2017).

Open Global Glacier Model (OGGM)

For the climate data, the data from the OGGM is used. The OGGM is an open-source modelling framework for glaciers. It is a 'glacier-centric model', which means that it runs for each glacier independently of the others. The main advantage of using the OGGM framework is its pre-processing capabilities. It is designed for large-scale applications, where a large number of datasets are ready to use from within the OGGM framework. There are different available levels of pre-processing. Level 0 is the lowest level, with directories containing the glacier outlines only. At level 3, CRU and ERA5 climate time series are added to the directories. The level 3 directories contain the most data and are thus the largest. With levels 4 and 5, intermediate output files are removed (Maussion, Butenko, et al., 2018). From the OGGM level 3 directories (Maussion, Rothenpieler, et al., 2018), the following glacier variables are included in this study:

tstar_ela_h

The t* ELA in meters. The ELA stands for the equilibrium-line altitude, which is the altitude where the ablation is equal to accumulation. tstar stands for t*, which is the year that the glaciers on average are in equilibrium with the climate, if the glaciers had the geometry during the measurement of mass balances (Marzeion, Jarosch, & Hofer, 2012).

tstar_aar

The Accumulation Area Ratio (AAR) is the ratio of the area of the accumulation zone to the area of the glacier, ranging between 0 and 1 (Cogley et al., 2010). The 'Balanced Budget AAR' is the AAR where the mass balance of a glacier is equal to 0. The balanced budget AAR of non-calving glaciers is on average between 0.5 and 0.6, but the variance is substantial and depends on climatic and topographic factors.

Ice thickness

For the ice thickness for every glacier, the consensus estimate by Farinotti et al. (2019) is used. This dataset consists of ice thickness estimates of all glaciers of the RGI. The estimate is based on an ensemble of models that use characteristics of the glacier surface, for example, the slope and surface velocities. Also, principles of ice-flow dynamics were used for ice thickness inversion (Farinotti et al., 2019). The ice thickness estimates are provided in .TIF files per glacier, with a resolution of 25 m for the HMA glaciers. The following variables regarding ice thickness are included in this study:

avg_ice_thickness

Average ice thickness of the glacier in meters. Acquired by computing the average of every grid.

ice_volume

Ice volume of the glacier in km^3. Computed by multiplying the *avg_ice_thickness* with the glacier area from the RGI.

3.3 Climate data

Climate data between 1980 and 2000, and 2000 and 2020 is used from the OGGM framework. The OGGM offers monthly precipitation and temperature time-series for every RGI glacier stored in a netCDF file per glacier, named *climate_historical.nc*. The climate data comes from the ERA5 dataset, which is a reanalysis of high quantities of historical measurements into global estimates using advanced modelling systems. ERA5 contains hourly weather data ranging from 1979 until the present in 31 km spatial resolution (Hersbach et al., 2020). In the OGGM framework, the temperature data are scaled to the glacier height (Maussion, Rothenpieler, et al., 2018). From the *climate_historical.nc* files, eight variables are constructed in this study:

mean_prcp_1980_2000 & mean_prcp_1980_2000

The average annual precipitation of the glacier in mm over 1980 to 2000 and 2000 to 2020.

mean_temp_1980_2000 & mean_temp_2000_2020

The average annual temperature of the glacier in degrees Celcius over 1980 to 2000 and 2000 to 2020.

prcp_diff_1980_2000 & prcp_diff_2000_2020

The difference in the average annual precipitation of the glacier in mm between 1980 and 2000, and between 2000 and 2020.

temp_diff_1980_2000 & temp_diff_2000_2020

The difference in the average annual temperature of the glacier in degrees Celcius between 1980 and 2000, and between 2000 and 2020.

3.4 Debris cover

In this study, the debris cover estimates of Kraaijenbrink et al. (2017) are used. They determined the spatial distribution of debris by using remote sensing data. Furthermore, they estimated the debris thickness and its relation to ice melt with debris surface-temperature data. The debris-cover extent and thickness were considered to be static over time (Compagno et al., 2022). The following variables from the dataset from Kraaijenbrink et al. (2017) are used in this study:

debris_area_ela_p

The percentage of the glacier area below the ELA that is covered with debris.

debris_vol_ela_p

The percentage of debris volume below the ELA of the total glacier volume below the ELA.

3.5 Glacial lakes

Glaciers and lakes are intrinsically linked. Supraglacial lakes form when meltwater flows in depressions on the ice surface of a glacier (Stokes, Sanderson, Miles, Jamieson, & Leeson, 2019). When a lake is formed in front of the ice, it is called a proglacial lake. Lakes can also manifest underneath the ice (subglacial lakes). In this study, the glacial lake dataset of Zheng et al. (2021) is used. They mapped all glacial lakes of HMA larger than 0.01 km^2 with a spatial resolution of 15 m, based on Landsat imagery between 1990 and 2015. They used an automated waterbody classification algorithm, accompanied by a slope threshold of $<20°$ and a shaded relief threshold of >0.25, to account for disturbances of mountain shadows (Zheng et al., 2021). In this study, lakes that intersect with the RGI polygons were selected, which amounted to 2608 lakes. From this data, the following variable is made:

glacial_lake

Binary variable of the presence of a glacial lake intersecting the glacier. 0 stands for no glacial lake present at the glacier and 1 stands for the presence of a glacial lake.

3.6 Glacier velocity

The glacier flow is expected to change in response to mass changes. However, the link between these components is poorly understood (Dehecq et al., 2019). Therefore, the glacier's flow velocity is an interesting variable to consider when modelling mass balance. In this study, the mean annual velocity measurements of Dehecq et al. (2019) are used. They used optical satellite imagery to measure the flow velocity between 2000 and 2017 for each HMA glacier. Furthermore, they used inventories from previous studies to analyze which glaciers are surging. The surging of glaciers is related to the glacier velocity, as a surging glacier moves rapidly in short semi-periodical periods with little forward movement in between the surging events. The mean annual velocity fields are downloaded from the NASA MEaSUREs - ITS_LIVE project in GeoTIFF format. The flow velocity is measured in 5 km grids, with a spatial resolution of 120 meters (Gardner, Fahnestock, & Scambos, 2021). In this study, the average annual glacier velocity of each glacier was extracted by overlaying the GeoTIFF file with the RGI polygons in QGIS and computing the zonal_statistics. 311 HMA glaciers larger than 2 km^2 are outside a grid and thus missing. The following variables from this dataset are

used:

velocity_mean

The average annual glacier's flow velocity in meters per year.

surging

Binary variable of whether a glacier is a surging glacier or not. 0 stands for no surging glacier and 1 stands for surging glacier.

4 Methods

4.1 Preprocessing

In total, the dataset consists of 8099 glaciers with a minimum area of 2 km^2. The first preprocessing step was to impute missing data. 18 *tstar_aar*, *tstar_ela_h* and 311 *velocity_mean* values are missing. To impute the missing data, the IterativeImputer class from the machine learning library *scikit-learn* was used. The IterativeImputer is a multivariate feature imputer and models each variable with missing values as a function of other variables. The IterativeImputer designates a variable as the input feature and the other features columns are used as inputs. Next, a regressor is fit on the inputs and output for known output, for which the RandomForestRegressor was used in this study. Then, the regressor predicts the missing values of the input feature and this estimate is then used for imputation. These steps are conducted for each feature iteratively (scikit learn, 2022b).

Next, data normalization for the independent variables was conducted. Data normalization is the technique to transform features to be on a similar scale, which can improve the performance and training stability of the model (Nawi, Atomi, & Rehman, 2013). In this study, the Z-Score normalization technique is used. With Z-Score normalization, the input attribute values are normalized by using the mean and standard deviation. The transformation is given by:

$$z = \frac{v - \mu}{\sigma} \tag{1}$$

Where μ is the mean value and σ represents the standard deviation of the data.

In this study, only the independent variables were normalized and not the dependent variable, due to easier interpretation. The Z-score transformation did not impact the results, as the Z-scores can be transformed back to the original values by dividing them by the variable's standard deviation.

4.2 Machine learning models

Four types of machine learning models were constructed and compared in their performance in modelling the SMB: multiple linear regression, Random Forest, XGBoost and an artificial neural network.

4.2.1 Multiple linear regression

First, feature selection was conducted, which is the process of reducing the dataset by choosing the relevant features from the entire dataset. One of the measures for feature selection is dependency measures (Blessie & Karthikeyan, 2012). In this study, Pearson's correlation coefficients are computed to find multicollinearity in

the data. With the Pearson's correlation, linear correlations between features are calculated and expressed as r, ranging between -1 and 1. A perfect negative correlation between features is expressed by an r of -1. An r of 0 indicates no relationship between the features, and 1 stands for a perfect positive linear relationship. Highly correlated features have got a similar effect on the dependent variable and therefore one of the features can be dropped out. The formula of the Pearson's correlation coefficient is:

$$r = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \tag{2}$$

Where *cov* is the covariance, σ_X is the standard deviation of X and σ_Y is the standard deviation of Y.

After removing one feature for each highly correlated feature pair, a multiple linear regression model was constructed. The multiple linear regression model is a linear regression model with more than one explanatory variable. The model shows how a set of explanatory variables is associated with a response variable of interest but does not allow to make causal inferences (Constantine, 2012). The equation for multiple linear regression is:

$$y_i = \beta_0 + \beta_i x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} + e_i \tag{3}$$

Where β_0 is the intercept, β_i are the partial regression coefficients of the explanatory variables and e_i is the error term, which is the amount by which the predicted value is different to the actual value, also referred to as the residual (Tranmer & Elliot, 2008).

Estimation of a multiple linear regression model is done with the least squares criterion. With this criterion, the β_i coefficients are chosen that minimize the sum of squared vertical distances between observed y_i and the fitted model, by:

$$\widehat{\beta} = min \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_i x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}))^2 \tag{4}$$

Where $\widehat{\beta}$ are the regression coefficients that minimize the sum of the squares.

Next, the optimal combination of variables was found that minimize the Aikaike Information Criterion (AIC) and maximize the explained variance (R^2). The AIC is a metric that rewards models with a lower number of parameters (Bozdogan, 1987). The model's performance was tested by running the model ten times with different *train_test_split* seeds that were chosen at random and not cherry-picked. The multiple linear regression model is evaluated on the (R^2), the Root Mean Squared Error (RMSE), and the Mean Absolute Error (MAE). These scores were calculated by averaging the values from the ten model runs. There

was chosen not to conduct k-fold cross-validation due to the issue of data leakage, as with this method parts of test data are also used to train the model. This results in overfitting (Cawley & Talbot, 2010).

4.2.2 Random Forest regression

The Random Forest algorithm has been regarded as one of the most precise methods for classification and regression, as it can model complex interactions among input variables and is robust against outliers (X. Zhou, Zhu, Dong, Guo, et al., 2016). The Random Forest algorithm has got certain advantages compared to other machine learning methods. Firstly, it runs efficiently on large datasets. Secondly, it is not sensitive to noise or overfitting. Furthermore, it can handle thousands of variables without variable deletion and lastly, it has fewer parameters compared to other machine learning algorithms, like the ANN and XGBoost algorithms.

The Random Forest regression algorithm is an ensemble-learning algorithm that combines the performance of a large number of decision trees to predict the value of a variable. The decision trees are created by drawing a subset of training samples with replacement, called bagging. A decision tree represents a set of binary conditions that are hierarchically organized and used from the root to a leaf of a tree. A decision tree is comprised of three main parts: root, internal and leaf nodes. The root node is the training dataset and is followed by the internal node, which acts as a decision-making node. The leaf node, or termination node, is the final node and holds the decision. At each node, a subset of features will be drawn at random. For each of these features, different splitting thresholds are tested to find the best split to subset the data according to the mean absolute error at the node. The MSE of a node gives a measure of goodness of fit, with low values representing good fit (Breiman, 2001).

Once a split has been selected, the node is partitioned into two descendant nodes. These split nodes are treated the same until a stopping criterion is met. For example, the training may stop when all unsplit nodes contain less than a certain amount of samples, or when a maximum depth for the decision tree is reached. Once a tree has been built, simple regression is executed for each of the termination nodes to predict the dependent variable based on the observed values. A downside of decision trees is that they can create over-complex trees that predict well for a specific combination of input variables and thus have low bias but do not generalize well on the data. This is a phenomenon called overfitting. The Random Forest algorithm uses a bagging approach, based on the bias-variance tradeoff principle. It predicts the dependent variable by averaging the estimations of every decision tree, resulting in estimations with the low-bias property of the decision trees (Qi, 2012). Furthermore, the variance of the predictions is low, according to the Central Limit Theorem:

$$Standarderror = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{5}$$

Where $\sigma_{\bar{x}}$ is the standard deviation of the sample, σ the standard deviation of the population, and n is the

number of data points.

In this study, scikit-learn's adaptation of the Random Forest regression algorithm was used to explain the SMB between 2000 and 2020. The hyperparameters were tuned with GridSearchCV of scikit-learn. GridSearchCV calculates the model performance of all combinations of the specified hyperparameters and their values (scikit learn, 2022a). The hyperparameters and its values of the Random Forest model are shown in Table 1. After tuning the parameters, predictions were made. The model's performance was tested by running the model ten times with the same *train_test_split* seeds as used for evaluating the multiple linear regression model. From these runs, the average R^2, RMSE and MAE are calculated.

Table 1: Hyperparameters of the Random Forest model that is used in this study.

Parameter	Description	Value used
max_depth	The maximum depth of the tree.	30
max_features	The number of features to consider when searching for the best split.	11
min_samples_leaf	The minimum number of samples required that a node consists of.	2
min_samples_split	The minimum number of samples required for a node to split.	4
n_estimators	The number of decision trees in the forest.	200

At last, the feature importance of the Random Forest model was calculated, for which the permutation and SHAP feature importances techniques were used. The permutation feature importance is defined to be the decrease in the model score when the value of a feature is randomly shuffled. This random shuffling removes the relationship between the feature and the dependent variable. The permutation score is interpreted as the decrease of the R^2 after the removal of that variable and is thus an indicator of how much the model is depending on it. This feature importance technique benefits from being model agnostic and is useful when non-linear features are present. Another advantage of this feature importance measure is that it takes all interactions with other features into account. By permutating the feature, the interaction effects with other features are destroyed, which is not the case when removing variables (scikit learn, 2022c). A downside of this technique is that it tends to decrease the importance of correlated predictors (Molnar, 2020). The SHAP (SHapley Additive exPlanations) feature importance technique is based on the Shapley values from cooperative game theory. In game theory, Shapley values are used to determine the contribution of each player in a coalition or a cooperative game. The SHAP method was introduced by Lundberg and Lee (2017) to explain the feature importance of tree-based models. For every variable, the SHAP value is computed

by calculating the marginal contribution of the variable to individual glacier SMB predictions. The feature importance is then calculated by averaging all SHAP values (Lundberg et al., 2020). A high SHAP value is interpreted as a high contribution of the feature to the prediction. A benefit of using the SHAP method is that it suffers less from underestimating the importance of correlated features. A disadvantage is that calculating the Shapley values requires a lot of computing time (Molnar, 2020).

4.2.3 XGBoost

XGBoost stands for eXtreme Gradient Boosting and is a gradient boosted machine learning library. It provides parallel tree boosting and is one of the leading machine learning libraries for regression and classification. The XGBoost algorithm is similar to the Random Forest algorithm, as they both are decision tree algorithms. The difference lies in the training methods, where the XGBoost algorithm learns by boosting and the Random Forest algorithm by bagging. Boosting is an ensemble learning method to build a strong model from several weak models:

$$F(x) = \sum_{i=1}^{M} f_i(x) \tag{6}$$

Where $f_i(x)$ is the weak learner.

A weak learner is an underpowered linear regression. The goal of the weak learner is not to capture all the dynamics in the data in one of the weak learners, but to capture a small part of the dynamics. The idea of boosting is that the next weak learner learns from the previous weak learners. At the end of the training process, although each of the weak learners predicts poorly by itself, when the weak learners are combined the model is powerful. With gradient boosting, extra steps are taken. Firstly, the loss function is defined. The role of the loss function is to estimate how good the model makes predictions with the given data. The loss function needs to be differentiable in an efficient way, as gradients of the loss function need to be calculated. In this study, the MSE was used. Next, the learning process is started by initiating a weak learner as the base learner whose mistakes all subsequent trees learn from, as is shown in Equation 7. The base learner in the XGBoost ensemble is the first decision tree.

$$F_1(x) = f_1(x) \tag{7}$$

Where $F_1(x)$ is the model in iteration 1 and $f_1(x)$ is the first decision tree.

Next, the derivative of the loss function is calculated for each data point with respect to the prediction of the first weak learner, using:

$$\widehat{r}_{1i} = -\frac{\delta L(y_i, \widehat{y_i})}{\delta \widehat{y_i}} \quad \Big| \quad \widehat{y_1} = F_1(x_i) \tag{8}$$

Where $\widehat{r}_{1i}$ is the gradient of the loss function for data point i and model number 1.

The gradient informs the model in which way the model has to learn in the next iteration to minimize the loss. The following step is to fit the next weak learner, where it learns from the mistakes of the previous weak learner. The target variable of the weak learner are residuals of the previous model and the input are the input variables:

$$f_2(x) = [\widehat{r_1} \sim X] \tag{9}$$

Where $f_2(x)$ is the weak learner decision tree in iteration 2.

Then, there will be determined how much of $f_2(x)$ is to be added to the current model $F_1(x)$. This is done by summing the loss for every data point of the true value and the new model, that is composed of the old weak learner and certain quantity of the new weak learner:

$$\widehat{\gamma_2} = \underset{\gamma}{argmin} \left[\sum_{i=1}^{N} L(y_{i1} f_1(x_i) + \gamma f_2(x_i)) \right] \tag{10}$$

Where $\widehat{\gamma_2}$ is the relative amount of the new weak learner to add to the model where the sum of the loss functions is minimized for all N observations.

Next, the new model is the old weak learner added with the optimal amount of the new weak learner:

$$F_2(x) = f_1(x) + \widehat{\gamma_2} f_2(x) \tag{11}$$

These steps are iterated until the loss on the validation set starts to increase. The final prediction is a weighted sum of all of the weak learners. The predictive quality of the XGBoost algorithm is generally high. Each decision tree contains low variance, but high bias. Due to the boosting of each decision tree, bias and underfitting are minimalized, resulting in a model with low variance and low bias (Chen et al., 2015).

The parameters were tuned with GridSearchCV, to find the combination that predicts the SMB with the lowest mean squared error. The parameters of the XGBoost model used in this study are shown in Table 2.

Table 2: Hyperparameters of the XGBoost model that is used in this study.

Parameter	Description	Value used
max_depth	The maximum depth of the tree.	9
min_child_weight	The minimum weight required to create a new node in the decision tree. A smaller value allows the algorithm to create nodes that correspond to fewer samples.	1
gamma	Specifies the minimum loss reduction required for a node to make a split.	0
subsample	The fraction of observations to subsample at each iteration.	0.6
colsample_bytree	The fraction of features to be used.	0.7
reg_alpha	L1 regularization term on weights, used in case of high dimensionality to reduce features.	0.00005
reg_lambda	L2 regularization term on weights.	1
learning_rate	Determines the step size at each iteration while moving towards a minimum of the loss function.	0.01
n_estimators	The number of decision trees in the forest.	1000

4.3 Deep learning

Artificial neural networks are nonlinear statistical models. The name and structure are inspired by the human brain, as they mimic the way that biological neurons signal to each other. ANNs are comprised of node layers, that contain an input layer, one or more hidden layers and an output layer. The input layer is fed with a multidimensional vector, which will distribute it to the hidden layers. Every node in one layer is connected to every other node in the next layer. A node takes the weighted sum of its inputs and passes it through a nonlinear activation function. If the input value is higher than a certain threshold value, the nonlinear activation function activates. The function transforms the features in a nonlinear manner and may extract complex features from simpler features. This results in the output of the node, which becomes the input of another node in the following layer. The model learns by a certain optimizer algorithm, which determines the weights of the connections between the nodes (O'Shea & Nash, 2015). During the training process, the weights and thresholds are continuously modified to minimize a particular loss function and generate output that is increasingly similar to the target output. After a number of iterations, the training can be stopped based on certain criteria. Recently, ANNs have gained interest due to improved optimization algorithms, which made the training of deep learning networks possible, leading to improved modelling of complex data patterns. The ability of ANNs to make complex representations of input parameters makes them suitable to model nonlinear systems such as glacier systems (Bolibar et al., 2020).

In this study, a six-layered ANN was constructed to model the SMB for the 8099 glaciers in HMA. To construct an ANN with high performance, the number of layers and hyperparameters need to be tuned. At the current stage, the neural network architecture selection is driven more by empirical results rather than mathematical theory (Crone, 2005). Various combinations of different amounts of dense layers *[1..10]* and neurons per layer *[10, 30, 50, 100, 250, 500, 1000]* were evaluated. An ANN with six dense layers performed best and was therefore chosen, consisting of respectively 150, 100, 50 and 50 neurons from top to bottom. The final layer of the model consists of a dense layer of size 1 that predicts one output. The ANN architecture is shown in Figure 3.

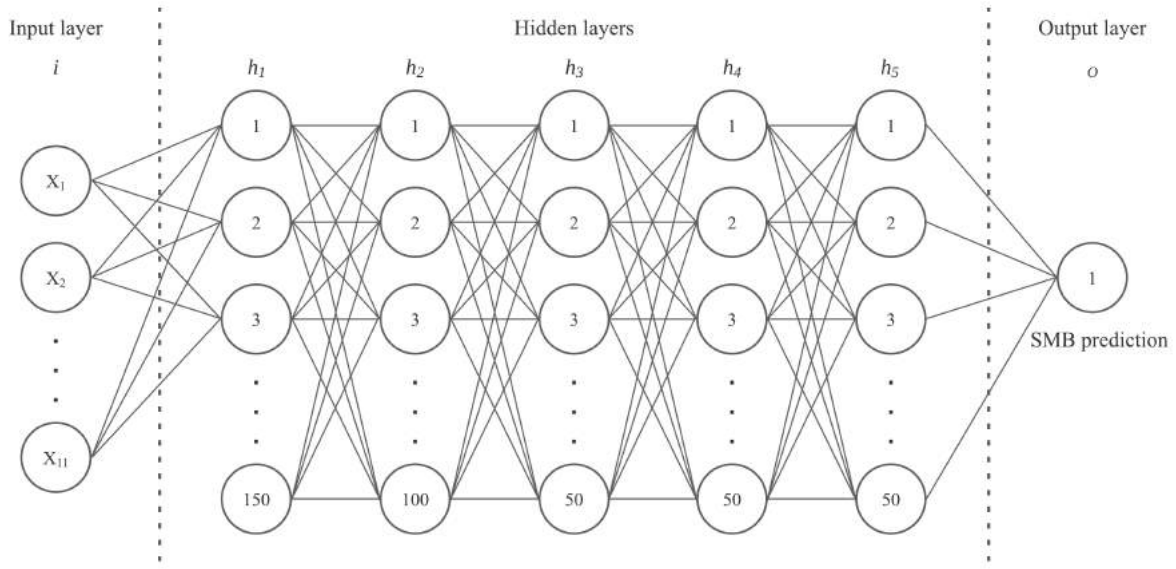

Figure 3: Architecture diagram of the Artificial Neural Network used in this study.

Also, the use of batch normalization was investigated. Batch normalization applies a transformation that maintains an intermediate output with a mean of 0 and a standard deviation of 1. Not using batch normalization in the network showed the highest model performances. Also, the use of different dropout values *[0..0.9]* was assessed. Dropout is a regularization method that randomly disconnects a certain number of connections between nodes. Large connection weights in a neural network are indicative of a more complex network that has overfitted the training data. Dropout is effective to reduce overfitting and improving generalization, as the model cannot rely on connections with very high weights and has to continue learning by finding other patterns in the data (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). Dropout values of 0.3 in the first layer, and 0.1 in the consecutive layers resulted in the lowest test error. With higher dropout values the model had more bias and with lower values the model overfitted quickly.

As activation function, the LeakyReLu function was used. ReLU stands for Rectified Linear Unit and is

a piecewise linear activation function that outputs the input if the value is higher than 0, which introduces nonlinearity in the data. LeakyReLU is similar to the ReLu, but it allows a small gradient when the unit is negative (Agarap, 2018). For this dataset, it performed better than the ReLU, tanh and sigmoid activation functions. The ReLu activation function does not have the vanishing gradient problem that tanh and sigmoid activations suffer with, as no derivative is involved in the ReLu algorithm. However, it can suffer from another problem, where a neuron dies if all inputs are below 0. With ReLU the neuron has no chance to be activated again, which is called the dying ReLU problem. With the LeakyReLU activation function, nodes that receive negative weights still have a small positive gradient, which gives the node a possibility to recover from being inactive (Xu, Wang, Chen, & Li, 2015).

The Adam (Adaptive Moment Estimation) optimizer technique for gradient descent was used as the optimization algorithm. It performed better than the stochastic gradient descent, which is considered to generalize better (P. Zhou et al., 2020). Optimizers are algorithms that change the weights of the network to reduce the loss function. The Adam optimizer accelerates the gradient descent by taking the exponentially weighted average and moving average of the gradients into account. This results in minimum oscillation in reaching the global minimum while taking big enough steps to not get stuck in local minima (Kingma & Ba, 2014).

The learning rate was also tuned, which is a hyperparameter that determines how much the weights are changed in each epoch. When the learning rate is too small, the training process will be long and could get stuck in a local minimum, while a value too large may lead to an unstable training process and cause the model not to achieve convergence. Values of 0.0001, 0.0005, 0.001, 0.01 and 0.1 were tested. A value of 0.0005 lead to the lowest test error.

Lastly, different combinations of features were selected as input, based on the permutation and SHAP feature importances from the XGBoost and Random Forest models. The model performed best when features with very low importance were not used as input. This resulted in the following input of the model: *prcp_mean_2000_2020, Slope, Zmed, temp_mean_2000_2020, temp_diff_1980_2000, prcp_diff_1980_2000, debris_area_ela_p, tstar_aar, velocity_mean, glacial_lake* and *surging*. The model's performance was tested by running the model ten times with the same *train_test_split* seeds as used before. The R^2, RMSE and MAE were calculated by averaging the values from the model runs. In Table A.2, the exact parameters of the constructed ANN are presented.

5 Results

5.1 Pearson's correlations

The correlations between the variables were calculated to remove a variable for each correlated variable pair. In Figure 4, the Pearson correlation coefficients between the selected variables is shown. Certain variables correlate strongly with each other. The area of the glacier correlates strongly with the ice volume and *Lmax*. *Lmax* also correlates with the average ice thickness and *velocity_mean*. *debris_area_ela_p* correlates very strong with *debris_vol_ela_p*. *Zmax, Zmin and Zmed* correlate very strongly with each other and also with *tstar_ela_h*. The slope correlates moderately high with the average ice thickness. Lastly, many climatic features correlate strongly but are not shown in the graph for visibility reasons. The *temp_mean_2000_2020* has a 1.0 correlation with the *temp_mean_1980_2000*, while the *temp_diff_2000_2020* and *temp_diff_1980_2000* have weak inverse correlation (-0.32). The *prcp_mean_2000_2020* and *prcp_mean_1980_2000* also have nearly perfect correlation (0.99). The *prcp_diff_2000_2020* and *prcp_diff_1980_2000* have 0 correlation, however. The *prcp_mean_2000_2020* and *prcp_mean_1980_2000* have moderately strong correlation with the average temperature in the respective time periods, with values of 0.66 and 0.64. All correlations mentioned above are significant ($P \leq 0.05$).

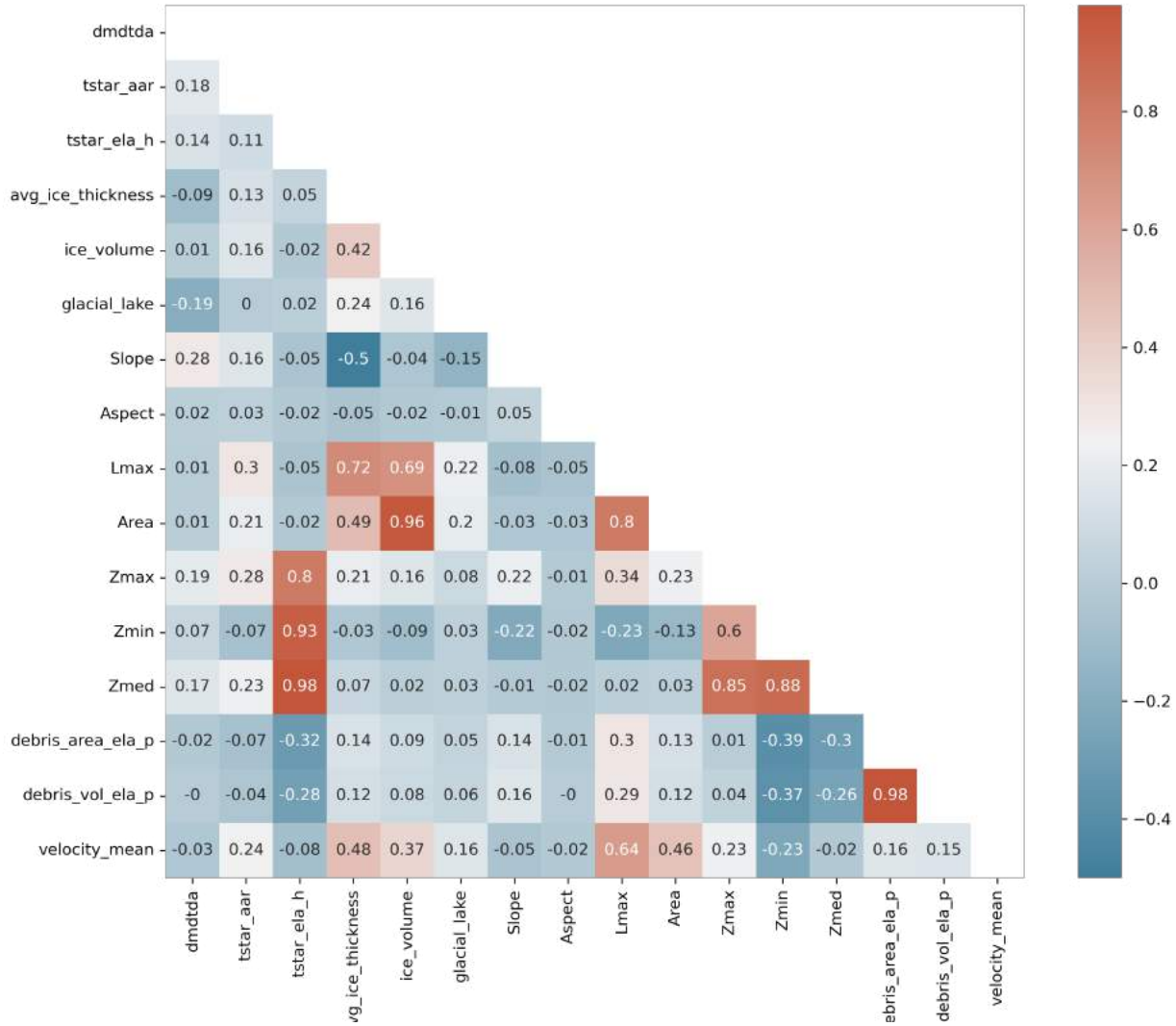

Figure 4: Correlation matrix showing the Pearson's correlation between each combination of variables used in this study (Table A.1).

5.2 Features importances

5.2.1 Multiple linear regression coefficients

The multiple linear regression was conducted to investigate linear responses of the variables on the SMB. The multiple linear regression coefficients are shown in Figure 5. The regression table is shown in Table A.3. The coefficients represent coefficients of the Z-scored features. This makes it possible to interpret which features explain most of the variance. The regression model without Z-scored features is presented in Table A.4, which is useful to interpret the inference of the features on the SMB in terms of the predictors' units.

The model has an R^2 of 0.309, meaning that the model explains 30.9% of the variability of the specific mass balance. The glacial lake variable has the highest β-coefficient, with a value of -0.1401 (P = 0.00001). This means the specific mass balance is on average 0.14 m.w.e/a. lower if a glacial lake is present. The second most important variable is the slope, with a β-coefficient of 0.0738 (P = 0.00001), meaning that if the slope increases by 1 standard deviation, the SMB increases by 0.0738 m.w.e./a on average. In more interpretable terms, the SMB increases by 0.012 m.w.e/a for every degree increase of the slope. The third most important variable is the average temperature increase between 2000 and 2020, signifying a 0.017 m.w.e/a SMB decrease for every degree Celcius increase (P = 0.00001). The average precipitation between 2000 and 2020 has got the next highest coefficient, suggesting a 0.01 m.w.e/a. SMB decrease for every 100 mm increase in average precipitation (P = 0.00001). The variable with the fourth-highest coefficient is the temperature difference between 2000 and 2020, with a value of 0.0512 (P = 0.00001) which implies a 0.199 m.w.e/a. increase for every degree Celcius increase in temperature. The *tstar_aar* is the last coefficient with a high coefficient, with a value of 0.05 (P = 0.00001). The model suggests that for every 0.01 increase in the Accumulation Area Ratio, the SMB increases by 0.006 m.w.e/a. Variables that are significant ($P \leq 0.05$), but not highly explanatory are the temperature difference between 1980 and 2000 (β = 0.024 and 0.10 m.w.e./a increase per degree Celcius, P = 0.0001), precipitation difference between 2000 and 2020, and 1980 and 2000 (β = 0.0184 and -0.0168, P = 0.0001), the median elevation (β = -0.011, P = 0.002), the aspect (β = 0.010, P = 0.0001) and the percentage of the glacier area under the ELA that is covered by debris (β = 0.007, P = 0.027). The area, average ice thickness, the average flow velocity of the glacier and whether a glacier is surging are variables that are not significant ($P \geq 0.05$) in predicting SMB in the multiple linear regression model. As homoscedasticity of the residuals was not met, according to the Breusch–Pagan test and White test, the p-values are to be interpreted carefully. For instance, the residuals of the *debris_area_ela_p* have a higher spread in low debris area than in high debris area. The residual plots are shown in Figure A.1. Heteroscedasticity tends to result in smaller p-values than it should be because it increases the variance of the coefficient estimates that the OLS procedure does not detect. Heteroscedasticity does not cause bias in the β-coefficients (White, 1980). Furthermore, the distribution of the residuals is heavy-tailed at each end of the Q-Q Plot, indicating that the residuals are not normally distributed.

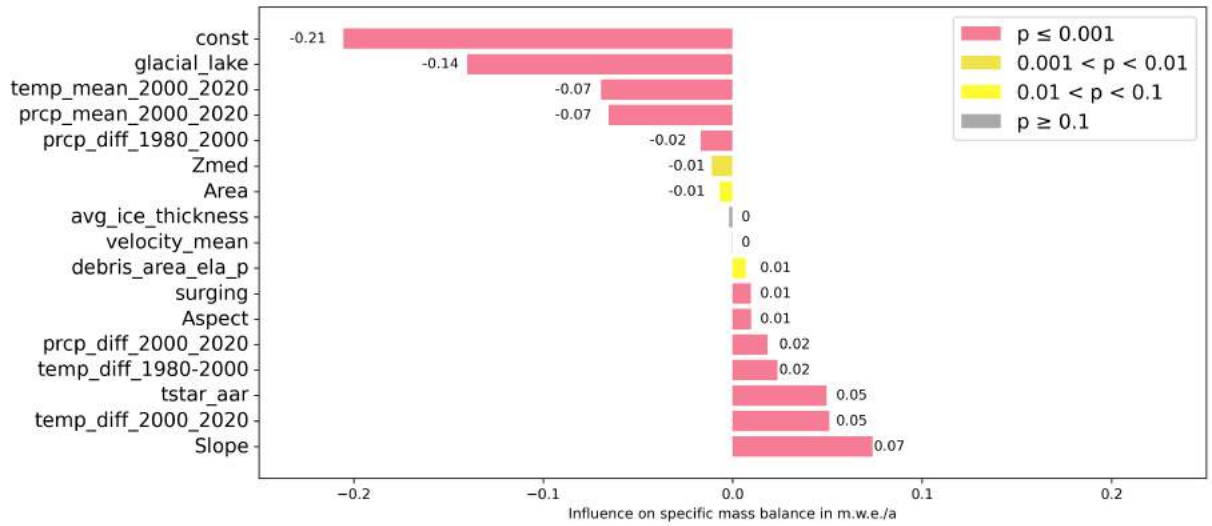

Figure 5: β-coefficients of the multiple linear regression model to explain the specific mass balance. The coefficients represent the SMB change per 1 standard deviation increase of the independent variable.

5.2.2 Feature importance of Random Forest and XGBoost models

The feature importances from the Random Forest and XGBoost models, calculated with the permutation and SHAP feature importance techniques, are shown in Figure 7. For both techniques and models, the feature importances are very similar. The most important feature for both models and feature importance techniques is the mean precipitation between 2000 and 2020. According to the permutation technique for the Random Forest model, the R^2 will drop by 45.8% when the mean precipitation between 2000 and 2020 is not included in the model. For the XGBoost model, this value is lower at 25.1%. Figure 8 shows beeswarm plots of the SHAP values, coloured by feature values. It shows that high mean precipitation 2000 - 2020 values correspond with negative SHAP values. Glaciers with lower average precipitation between 2000 and 2020 mostly have got positive SHAP values, between circa 0 and 0.25. In other words, it shows that the mean precipitation is the strongest positive predictor in modelling the SMB, where it contributes up to 0.25 m.w.e./a of the SMB prediction.

The next most important variable in the nonlinear models is the slope. The SHAP measure indicates that glaciers with higher slopes are related to higher SMB values. A high slope contributes on average 0.05 m.w.e./a towards a more positive SMB prediction. The mean temperature between 2000 and 2020, and the median glacier elevation have the third and fourth highest permutation scores for the nonlinear models. According to the SHAP values, higher mean temperatures and glacier elevation contribute to lower SMB values. In the linear model, the median glacier elevation was one of the least explanatory variables as it explained only 1.4% of the variance. Another difference between the nonlinear models and the multiple linear

regression model is regarding the precipitation difference between 1980 and 2000. For the XGBoost model, it is the fifth most important feature where the model performs 10.3% worse without this variable, while it only explained 2.4% of the variance in the linear model. A noteworthy pattern is visible in the SHAP values for this variable. In general, increases in precipitation between 1980 and 2000 contribute slightly negative to the SMB and a decrease in precipitation is associated with higher SMB values. However, the strongest negative contributions of this variable to the SMB is caused by decreases in precipitation.

The temperature difference between 2000 and 2020 is an important feature in the nonlinear models. The SHAP values indicate that more positive temperature changes contribute to higher SMB values. The temperature difference between 1980 and 2000 is an important feature with the permutation measure, but less important than the temperature change between 2000 and 2020. For the SHAP measure, this variable is the ninth most contributing variable to the SMB and thus scores moderately low. The *debris_area_ela_p* is moderately important for the nonlinear models, as it has the seventh-highest mean SHAP value and the nonlinear models perform circa 5% worse without this variable. A dependence plot of the SHAP values of the *debris_area_ela_p* is shown in Figure 6. The figure shows that clean ice under the ELA contributes positively to the SMB. When the debris area increases, the SHAP values initially decrease to negative values of circa -0.05 m.w.e./a. However, when the ELA is mostly covered with debris, the SHAP measure indicates that the debris contributes positively to the SMB, up to 0.05 m.w.e./a. For the linear model, the debris area is one of the least explanatory variables as it explains less than 1% of the variance.

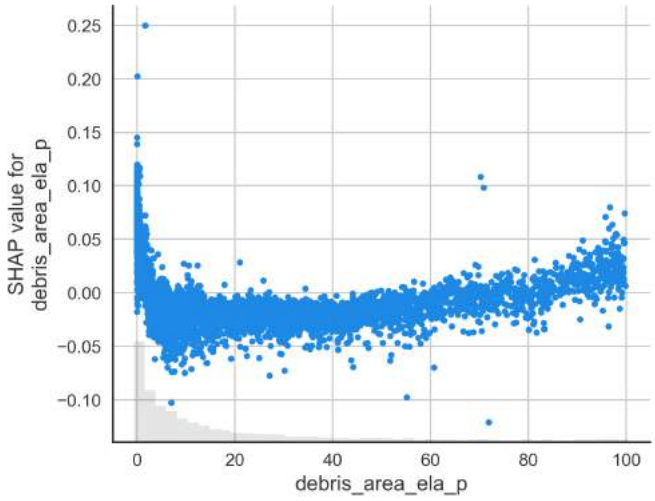

Figure 6: Dependence plot of the percentage of area under the ELA that is covered with debris, versus its contribution on the specific mass balance.

The mean flow velocity, aspect, average ice thickness, area, whether a glacier is surging, and presence of glacial lakes are the least important variables for the nonlinear models. Non the less, the SHAP measure shows that higher average ice thickness, area values and surging behaviour contribute slightly to more negative SMB values. Also, it shows that when a glacier lake is present, it could contribute up to -0.18 m.w.e/a to the SMB, which corresponds to the high absolute β-coefficient of the linear model.

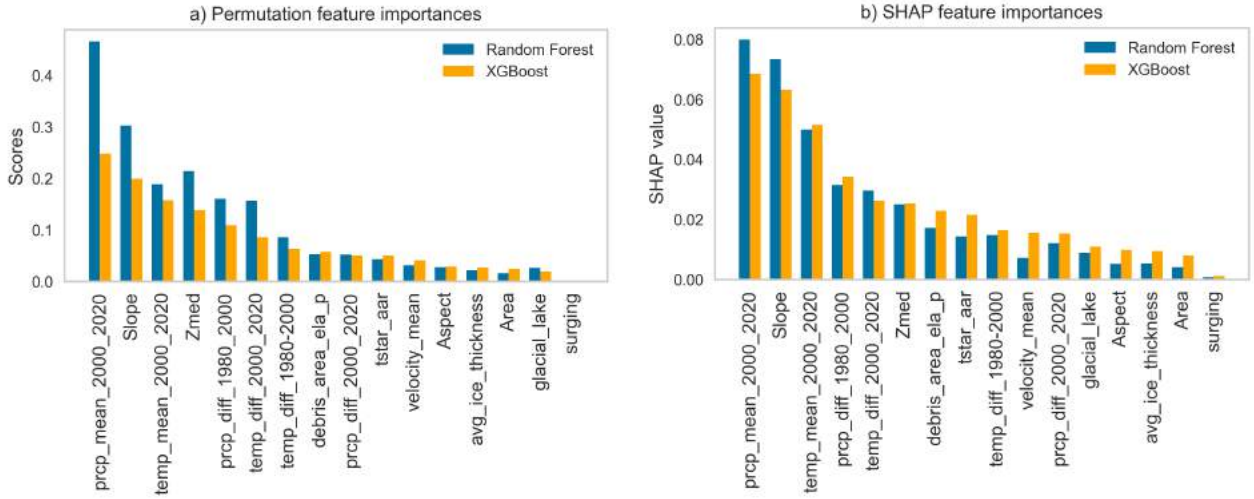

Figure 7: Feature importances using permutation and SHAP methods on Random Forest and XGBoost models.

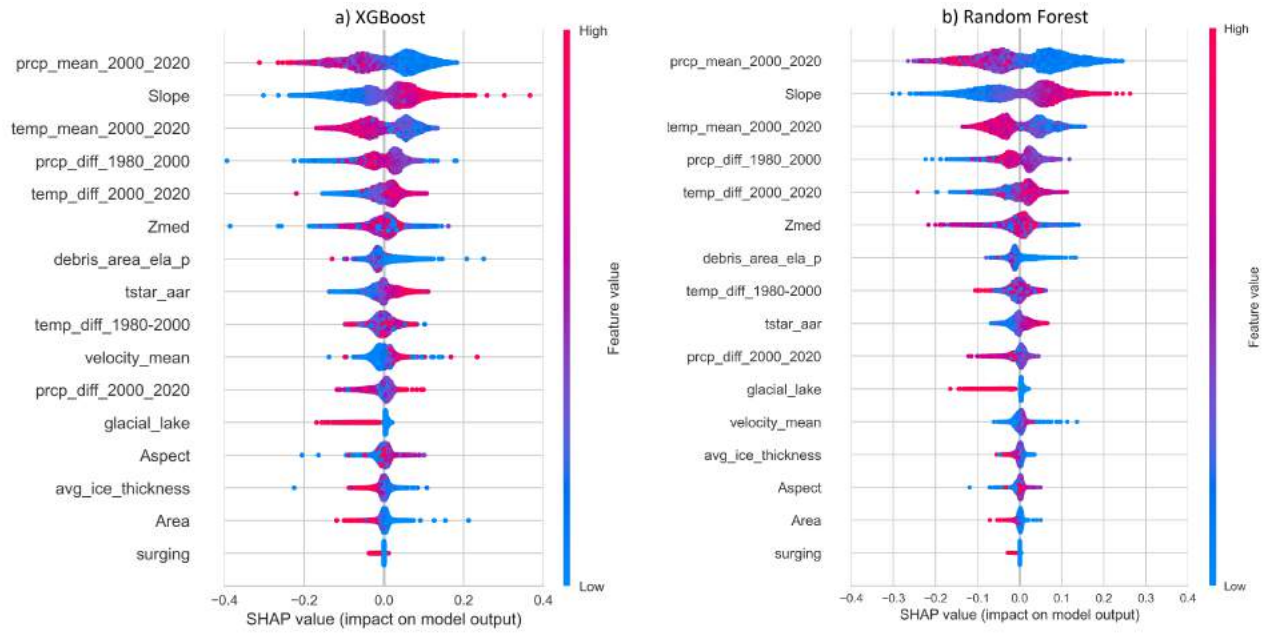

Figure 8: SHAP feature importances for every SMB prediction with Random Forest and XGBoost models.

5.3 Model performances

The nonlinear models perform better than the multiple linear regression model, regarding the R^2, RMSE and MAE. The exact model results for each of the ten runs are shown in Table A.5 and Table A.6. The multiple linear regression model has got on average an R^2 of 0.307, RMSE of 0.251 and MAE of 0.189. The Random Forest model has a mean R^2 of 0.656, RMSE of 0.177 and MAE of 0.119. For the XGBoost model, the average R^2 is 0.685, the RMSE is 0.172 and the MAE is 0.117. The ANN model has got a mean R^2 of 0.659, RMSE of 0.176 and MAE of 0.119. Therefore, the XGBoost model performs the best regarding all these scoring metrics. The explained variance of the XGBoost is 113.7% higher than of the multiple linear regression model and 4.4% and 3.9% higher than the Random Forest and Artificial Neural Network model. The ANN and Random Forest models have got very similar values for the scoring metrics. However, the ANN had the highest maximum R^2 in the ten runs of all models, with 0.705, compared to 0.701 from the XGBoost and 0.688 from the Random Forest model. The ANN had a worse minimum R^2 than the XGBoost model though, with 0.636 compared to 0.650. The R^2 of the XGBoost was more stable between the different runs than the other nonlinear methods. The multiple linear regression model performs worst of all models but is the most stable model.

Figure 9 shows the Predicted vs Actual plots of the different models for a seed where the models performed average. The multiple linear regression model shows a bad fit to the $y = x$ line. The spread in the predicted values is large, ranging from circa -1 to 0.5 m.w.e./a, whereas measured values range from circa

-0.75 to 0.2 m.w.e./a. A slight trend is visible nonetheless, as the cluster of predictions tilts slightly to the right. The Predicted vs Actual plots of the XGBoost, Random Forest and ANN models look similar, in that they follow the $x = y$ line well. However, the models appear to underpredict the extreme negative SMB values and overpredict the extreme positive SMB values. Also, certain glaciers are predicted with a high error. No heteroscedasticity was observed in the residuals of the nonlinear models.

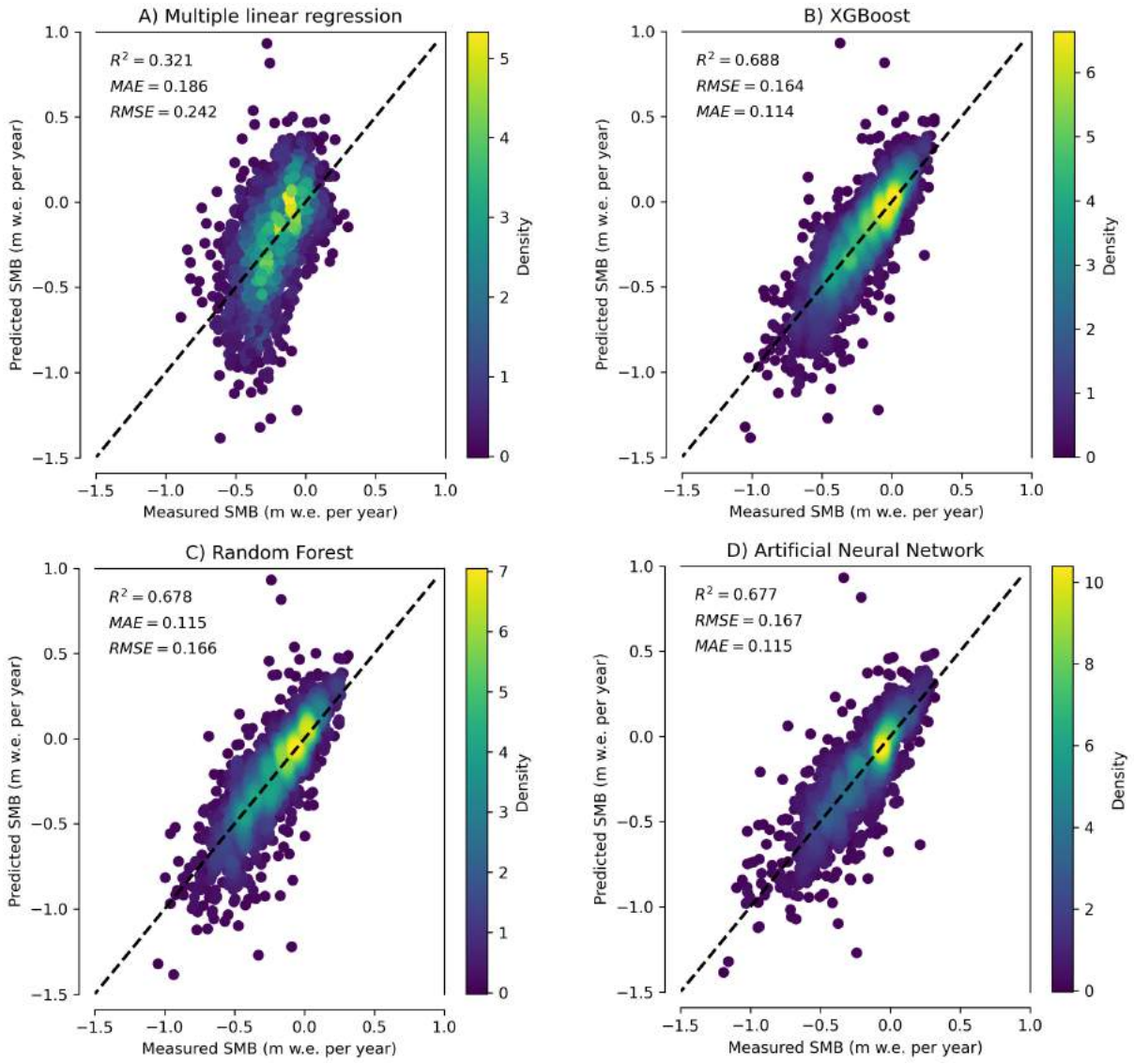

Figure 9: Plots of measured vs predicted specific mass balance for A) multiple linear regression, B) XGBoost, C) Random Forest and D) Artificial Neural Network model.

6 Discussion

6.1 Influence of the variables on the specific mass balance

6.1.1 Influence of the mean precipitation between 2000 and 2020

The results indicate that the average precipitation between 2000 and 2020 is the most important predictor in explaining mass balance. This finding corresponds with the results from Bolibar et al. (2020). They found that accumulation-related predictors (winter snowfall, summer snowfall and March snowfall) are the most important mass balance predictors for the French Alps, although this region has got different topographical and climate conditions than HMA. For glaciers in the French Alps, the SMB decreases when there is less snowfall in winter and summer (Bolibar et al., 2020). However, the SHAP values of the nonlinear models and the coefficients of the multiple linear regression model used in this study suggest that the SMB increases with decreased precipitation. The linear model suggests that for every 100 mm increase in precipitation, the SMB decreases by 0.01 m.w.e./a.

Exploratory data analysis of the mass balance and climate data showed that in 9 out of 14 second-order regions in HMA, melting glaciers have got lower mean precipitation than growing glaciers. In 5 regions, the mean precipitation for melting glaciers is higher than for glaciers with positive SMB values. Furthermore, in the regions where the glaciers are growing such as the Karakoram, the glaciers grow mainly due to an increase in snowfall (De Kok, Kraaijenbrink, Tuinenburg, Bonekamp, & Immerzeel, 2020). Therefore, these findings do not correspond with the negative β-coefficient and the negative SHAP values of the mean precipitation between 2000 and 2020 variable.

An explanation for this inverse relationship is that there is strong spatial heterogeneity present regarding relationship between the precipitation and the SMB. Exploratory data analysis showed that higher mean precipitation is an important predictor for higher SMB in the West Kunlun and Central Himalayas regions. On the other hand, for the East Kunlun, Qilian Shan and Inner Tibet regions, strong negative relationships exist between the mean precipitation and SMB. A cause for the negative relationship between the mean precipitation and SMB for the entire HMA region is that glaciers with high precipitation are usually more maritime than continental. Maritime glaciers have been shown to have a higher sensitivity to the climate. In other words, maritime glaciers have a shorter response time than continental glaciers, because they have a large ice turnover (Wang et al., 2019). Therefore, higher precipitation values are associated with more glacier melt for maritime glaciers. Thus it appears that the negative association of the mean precipitation with the SMB for the entire HMA is caused by heterogeneous climate conditions, rather than the effect of the underlying process of precipitation.

What may also play a role in a more negative relationship between higher precipitation and more negative

SMB, is that since 1975 relatively more precipitation fell in the form of rain instead of snow in most of HMA Kapnick (2014). Therefore, the fraction of precipitation falling as rain has increased. Precipitation in the form of rainfall is considered to be lost to the system and thus explains less accumulation of the glacier (Cogley et al., 2010). However, rainfall has negligible influence on ablation, as even in high-precipitation environments only a few per cent of the energy available for melt is supplied by rainfall (Anderson et al., 2010).

6.1.2 Influence of the glacier tongue slope

The models in this study show that the tongue slope is the second most important variable in explaining the SMB, with steeper slopes contributing positively to the SMB. This is in line with other research (e.g. Brun et al., 2019; Huss, 2012; Mannerfelt et al., 2022; Huss & Fischer, 2016). An explanation of the positive relationship between steeper slopes and higher SMB values is that a glacier with a steep terminus has a shorter response time. This results in that after climatic forcing the glacier with a steep terminus retreats quicker to a higher elevation. Here it will experience lower temperatures and reach a new equilibrium with the climate. On the other hand, a glacier with a more flat terminus will lose most of its tongue before it can reach a new equilibrium (Zekollari, Huss, & Farinotti, 2020).

6.1.3 Influence of the median glacier elevation

According to the nonlinear XGBoost and Random Forest models from this study, the median glacier elevation is an important variable in explaining SMB. This is in line with previous studies (e.g. Rabatel, Letréguilly, Dedieu, & Eckert, 2013; Brun et al., 2019; Huss, 2012). The nonlinear models showed that lower median elevation values contribute to more positive SMB values, which is not in line with the research of Brun et al. (2019) and Rabatel et al. (2013), who found more positive SMB values for glaciers with higher altitudes. However, when running the nonlinear models separately for each second-order region, positive relationships between the median altitude and SMB was observed for every region apart from West Kunlun, the Karakoram and Hengduan Shan regions. Huss and Fischer (2016) explain that glaciers with a low slope and elevation are more sensitive to temperature change and that thus a positive relationship exists between the median glacier elevation and the SMB.

Exploratory data analysis showed that glaciers in Central Asia lie on average more than 200 m lower than glaciers in South Asia West and South Asia East. Glaciers of South Asia East on average have almost 3 times more negative SMB values than the glaciers of Central Asia. However, within the South Asia East region, there is still a positive relationship present between the median elevation and the SMB. Thus an explanation for the negative relationship between the median glacier elevation and SMB for the HMA in its entirety, is that the regional and process effects of the median elevation on the SMB differ.

6.1.4 Influence of temperature and precipitation

The mean temperature between 2000 and 2020 was in both the linear and nonlinear models found to be one of the most explanatory variables, which complies with the findings of Bolibar et al. (2020). For glaciers in the French Alps, they found that the October temperature is the most important temperature variable. The reason is that temperatures higher than 0 degrees Celsius favour lengthening of the ablation season, while negative temperatures allow snowfall to protect the ice and contribute to accumulation. Furthermore, the cumulative positive degree days (CPDD) and summer- and shoulder season months are important ablation predictors (Bolibar et al., 2020).

The difference in temperature between 2000 and 2020 and 1980 and 2000 are important variables for the nonlinear models. In the multiple linear regression model, these variables have got low coefficients and thus do not explain a high amount of the variance. These results imply that the temperature change is a nonlinear variable for mass balance change, which is in line with the research of Hock (2010). The SHAP measure showed that an increase in temperature between 2000 and 2020 contributes to more positive SMB values. The coefficients of the linear model showed a similar pattern. The same trend is present for the second-order regions separately. These findings are inexplicable, as glaciers in the southern and western HMA mostly melt due to temperature rise (De Kok et al., 2020). Also, other research has shown that the altitude of the equilibrium line can shift dramatically if the temperature rises even a little, resulting in a high negative mass balance change (Salinger, Chinn, Willsman, & Fitzharris, 2008). It is possible that the temperature estimations from the *climate_historical.nc* dataset have large discrepancies with the actual temperature values of the glaciers.

The high importance of the precipitation between 1980 and 2000 implicates that a time lag effect appears to be present for the climatic variables. The presence of a time lag effect for glaciers is confirmed by Raper and Braithwaite (2009). The glacier mass balance has a certain response time to a climate forcing, which can be approximated by the ratio of the glacier thickness to ablation at the glacier terminus (Raper & Braithwaite, 2009). Under the same climate conditions, larger glaciers have got longer response times than smaller, thinner glaciers. The individual glaciers' response times to climatic changes vary from 5 to more than 100 years, partly determined by the size of the glacier and whether it is a continental or maritime glacier (Salinger et al., 2008; Wang et al., 2019).

6.1.5 Influence of debris area under the ELA

The results showed that the area under the ELA that is covered with the debris is a moderately important variable for the nonlinear models. In the multiple linear regression model, the variable was barely significant. This is in line with findings of (Brun et al., 2019) who observed no significant differences between glaciers with and without debris cover for the entire HMA region. This indicates that the SMB responds nonlinearly

to the presence of debris cover on the glacier. The SHAP values have shown that a low debris area contributes to higher SMB values, but that a high debris area is related to higher SMB. Compagno et al. (2022) have shown that if the debris is thinner than a few centimetres, ice melting is enhanced due to lowering the albedo of the glacier surface. For thicker, continuous debris layers, ice melt is reduced due to the insulating properties of the debris. As the same pattern is present for the debris area under the ELA, I hypothesize that this variable is correlated to the debris thickness.

6.1.6 Influence of glacier lakes

The multiple linear regression model suggests that the SMB decreases with an average of 0.14 m.w.e./a when a glacial lake is present. The nonlinear models showed that the variable is not an important variable in explaining the mass balance change for the HMA in its entirety, presumably because only 814 out of 8099 glaciers intersect with a lake. Likely, the variable scored low in the nonlinear models due to this skewness. Also, glacier lakes were expressed in binary terms in this study, which induces bias in the tree algorithms as they prioritize variables with higher amounts of values (Strobl, Boulesteix, Zeileis, & Hothorn, 2007). For glaciers that do intersect with a lake, it was shown that the presence of a glacial lake can contribute negatively to the SMB with values ranging from -0.025 to -0.18 m.w.e/a. These values are comparable with findings from Maurer, Schaefer, Rupper, and Corley (2019), who found that for the 650 largest Himalayan glaciers, lake-terminating glaciers have got on average -0.18 $\pm$ 0.08 m.w.e./a lower SMB values than glaciers that do not terminate into a lake. There are several reasons for higher mass loss rates when a glacial lake is present. Glacier lakes can affect the ice flow by reducing friction at the ice-bed interface, which encourages basal sliding (Pronk, Bolch, King, Wouters, & Benn, 2021). Also, they can have an impact on the albedo of the ice surface, resulting in more surface melt. Proglacial lakes cause ice calving, which can decouple glaciers from the climate due to heat absorption by the lake, resulting in reduced mass balance (Davies, 2021; Watson et al., 2020). The variability of mass loss rates within glaciers with glacial lakes is caused by the glacier's size and development stage (Brun et al., 2019; Sakai, Nishimura, Kadota, & Takeuchi, 2009; Benn et al., 2012).

6.1.7 Variables with low influence

The results showed that the mean glacier flow velocity, surging behaviour, the glacier area, aspect, and the average ice thickness are not strong predictors for the SMB.

The mean velocity was a variable with low or no explanation for the SMB in the models from this study. This is indicative of the mean velocity being not a good proxy for the change in flow velocity. The change in flow velocity appears to be a strong predictor for the SMB (Dehecq et al., 2019). A decrease in the flow velocity was found to be accompanied by mass balance loss. In contrast, the stable or growing glaciers of

the Karakoram and West Kunlun regions undergo accelerated glacier flow (Dehecq et al., 2019).

Regarding the interpretation of the influence of surging on the SMB, caution ought to be taken as only 382 glaciers out of 8099 glaciers were classified as surging. Also, not all glaciers that were labelled as non-surging are likely classified correctly (Dehecq et al., 2019). In the nonlinear models, whether a glacier is surging is of the lowest permutation importance and the variable is not significant in the multiple linear regression model. The SHAP values suggest a very small contribution of surging behaviour to negative SMB, although the significance of this is debatable. This is in line with findings from Bhattacharya et al. (2021), who found that surging glaciers have got slightly but insignificantly lower mass balance loss than non-surge type glaciers. However, they observed that after surge termination the mass loss rates of surge-type glaciers in HMA increased substantially by -0.62 $\pm$ 0.10 m.w.e./a bwteen 1980 and 2019. A hypothesis is that glaciers are more vulnerable to melting after the ice is transferred from high to low during a surge event (Bhattacharya et al., 2021). This effect is not considered in this study, as ongoing surging and not historical surging were examined.

The results show that there is a negative relationship between the glacier area and the SMB, but that it is of low importance or no significance which is in line with research of Brun et al. (2019). In past research, only a negative correlation between the area and SMB for glaciers smaller than 0.1 km^2 was found in the Alps (Fischer, Huss, & Hoelzle, 2015). The aspect is a significant but very low explanatory variable in the multiple linear regression model. It is also not an important variable in the nonlinear models. These results are in line with Brun et al. (2019), who found that the aspect is not a significant variable in explaining the SMB for most HMA regions. Furthermore, the aspect was the least important variable for glaciers in the French Alps (Bolibar et al., 2020). The average ice thickness was not an important variable in the models of this study, but a relationship between higher ice thickness and lower SMB values appears to be present. No comparative research was found that included this variable.

6.2 Model comparison

The results suggest that the nonlinear XGBoost, Random Forest, and ANN models perform better than the multiple linear regression model, as more than twice the variance is explained in the nonlinear models. This is in line with previous studies (Steiner et al., 2005; Bolibar et al. 2020, 2022). As discussed before, certain variables appear to have nonlinear responses to the SMB, which are the climatic variables, median glacier elevation and the presence of debris cover under the ELA. A shortcoming of linear models for modelling mass balance is that they do not capture these nonlinearities. For instance, linear models appear to be oversensitive to extreme positive and negative snowfall anomalies (Bolibar, Rabatel, Gouttevin, Zekollari, & Galiez, 2022). The multiple linear model thus does not meet its assumptions, as one of the assumptions is that the relationship between the dependent and independent variables is linear. Furthermore, the multiple

linear regression model violates the assumption of homoscedasticity. Lastly, the assumption of normality is not met as the Q-Q Plot shows a heavy-tailed distribution, indicating that the residuals are not distributed normally. Therefore, the use of the multiple linear regression model in this study to model the SMB is debatable, as the confidence intervals of the coefficients are unreliable (Duke University, 2022). However, they have low computation costs and show the response of predictors in their unit which allows for easier interpretation. Consequently, when used in parallel with nonlinear models, linear models still have a place in SMB modelling (Bolibar et al., 2020).

From the nonlinear models in this study, the XGBoost model performed optimally. The model scores highest regarding the R^2, RMSE and MAE metrics and is computationally less expensive than the ANN model. No studies were found that model the SMB with the Random Forest or XGBoost models. For most regression tasks, XGBoost tends to perform slightly better than Random Forest (MLJAR, 2022). Compared to Random Forest, XGBoost has a more complex relationship between the hyperparameters and accuracy. This results in more time-consuming hyperparameter tuning and higher computational power, but possibly also higher scoring metrics than Random Forest models (Pafka, 2022). It is possible that the ANN performs better in modelling the SMB after further tuning the hyperparameters. A downside of the ANN models is that they require very intensive model tuning based on trial and error. Using GridSearchCV, which was used to tune the hyperparameters of the Random Forest and XGBoost models, was not feasible for the ANN in this study due to too high computational costs.

According to the 'no free lunch theorem', any two algorithms are equivalent when their performance is averaged across all possible problem. Therefore, it is recommended to test as many machine learning models as possible for modelling the SMB, because the best performing model depends on the regression task and data characteristics (Wolpert & Macready, 1997).

6.3 Limitations and recommendations

First of all, the statistical analysis in this study is sensitive to outliers, biases and uncertainties in the data. The mass balance data of Hugonnet et al. (2021) suffer from relatively high uncertainties, with a median uncertainty of 0.22 m.w.e./a (Brun et al., 2019). Other data sources are also affected by high levels of uncertainty, such as the climate data due to a sparse network of weather stations that are typically placed at lower elevations. Additionally, the ERA5 reanalysis dataset tends to greatly underestimate the precipitation in mountainous areas (De Kok et al., 2020). Lastly, there are large uncertainties present regarding the estimations of the debris area under the ELA (Rounce et al., 2021).

The second limitation of this study is that the statistical analysis is conducted for the entire HMA region, instead of analyzing second-order HMA regions separately. The factors that drive glacier melt are highly variable between different HMA regions (Bonekamp, De Kok, Collier, & Immerzeel, 2019). The results of

Brun et al. (2019) also showed that many variables were only statistically significant for certain regions and not for the entire HMA region. Therefore, a recommendation for future research is to execute the nonlinear models and use the feature importance techniques for separate regions that are more homogeneous climate and topography-wise.

Thirdly, it is interesting to analyze the SMB response of certain variables that were not included in this study. The mean flow velocity was used in this study to investigate the effect of flow velocity on the SMB. The change in flow velocity would have been a more interesting variable to include in the model as this appears to be an important predictive variable for the SMB (Dehecq et al., 2019). Regarding the climatic variables, it would have been interesting to split the annual temperature up into months, because the summer- and shoulder season months are the most significant predictors for SMB in the French Alps (Bolibar et al., 2020). Furthermore, the interannual variability in precipitation would be an interesting variable to investigate, as this potentially has got a large effect on the SMB (Bonekamp et al., 2019). It would also be of interest to discriminate the precipitation into rainfall and snowfall to get a better insight into the SMB response to rainfall. This could be executed with a temperature threshold (De Woul, 2008).

Furthermore, spatial autocorrelation was not considered, while it is shown that it is present for the SMB of glaciers in HMA (De Kok et al., 2020; Wang et al., 2019). Glaciers close to each other undergo more similar climate conditions and have on average more related morphological characteristics, such as elevation, than glaciers very far apart (Wang et al., 2019). Accounting for spatial autocorrelation could improve the predictive power of the nonlinear models and allow for easier separation of spatial effects and process effects of the variables on the SMB (Li, 2022). For further research, it would be interesting to add spatial lag and eigenvector spatial filtering features to the XGBoost and Random Forest models, as research has indicated that this can improve the nonlinear models' predictions for geospatial inference tasks (Liu, 2020). Also, combining the predictions of a Geospatial Weighted Regression model and Random Forest model has shown promising results for geospatial inference (Shahneh, Oymak, & Magdy, 2021). For the imputation of the *velocity_mean* and *tstar_aar* variables, spatial autocorrelation was also not taken account of. For further research, Kriging interpolation could be considered for the imputation technique. With this method, spatial autocorrelation of the variables is considered as it uses nearby observations to impute the missing values (Cressie, 1988).

Lastly, the validations of the models could be more robust when nested k-fold cross-validation is used. In this study, the models were executed on ten different seeds than the models were trained for. This procedure resulted in an objective model comparison. Nonetheless, the scoring metrics of the models could increase when the models are trained multiple times with different training sets with optimized hyperparameters for each training set. Also, the generalizability of the models would improve as all glaciers will be used without inducing bias, which is the issue with regular k-fold cross-validation (Cawley & Talbot, 2010).

7 Conclusion

In this study, the SMB of 8099 HMA glaciers for the period 2000 - 2020 was simulated with 16 independent variables as input. The SMB was modelled with a linear model and three nonlinear models. The responses of the variables on the SMB were assessed with the coefficients of the linear model and with the SHAP and permutation feature techniques of the XGBoost and Random Forest model. Also, the predictive power of the models was compared. The two most important variables appear to be the average precipitation between 2000 and 2020 and the slope. The SHAP measure indicates a negative relationship between the average precipitation and SMB, which is not in line with other research. An explanation for this inverse relationship is the difference in regional effects and the effects of the underlying process. Other strong predictors are the average temperature between 2000 and 2020, the median glacier elevation, the precipitation change between 1980 and 2000 and the temperature difference between 2000 and 2020. The higher the average temperature, the more negative the SMB. However, more positive SMB values were associated with temperature rise between 2000 and 2020, which is inexplicable and could be caused by inaccuracies in the temperature change estimations. Certain variables appear to have nonlinear responses to the SMB, which are the climatic variables, median glacier elevation and the presence of debris cover under the ELA. The nonlinear models have performed significantly better than the linear models, as they capture these nonlinearities of the glacier systems. Of the nonlinear models, the XGBoost performed best. This model scored highest regarding the R^2, RMSE and MAE.

8 Appendix

Table A.1: Overview of variables used in study.

Variable	Unit	Description	Source
dmdtda	m w.e./a	Mass change rate per unit area in meters water-equivalent per year between 2000 and 2020.	Hugonnet et al. (2021)
rgi_id		Primary key of the dataset. Unique identifier for each RGI glacier.	RGI 6.0 (2017)
Area	km^2	Area of the glacier.	RGI 6.0 (2017)
Zmin, Zmax, Zmed	m	Minimum, maximum and median elevation of the glacier in meters above sea level.	RGI 6.0 (2017)
Slope	degree	Mean slope of the glacier surface	RGI 6.0 (2017)
Aspect	degree	Mean aspect of the glacier, expressed as integer azimuth relative to 0° at due north.	RGI 6.0 (2017)
Lmax	m	Length of the longest surface flowline of the glacier.	RGI 6.0 (2017)
tstar_ela_h	m	Equilibrium-line altitude for year t*.	Maussion et al. (2018)
tstar_aar		Accumulation Area Ratio, which is the ratio of the area of the accumulation zone to the area of the glacier.	Maussion et al. (2018)
avg_ice_thickness	m	Average ice thickness of the glacier.	Farinotti et al. (2019)
ice_volume	km^3	Ice volume of the glacier.	Farinotti et al. (2019)
mean_prcp_*	mm	The average annual precipitation of the glacier over 1980 to 2000 and 2000 to 2020.	Maussion et al. (2018)
mean_temp_*	C°	The average annual temperature of the glacier over 1980 to 2000 and 2000 to 2020.	Maussion et al. (2018)
prcp_diff_*	mm	The difference in the average annual precipitation of the glacier between 1980 and 2000, and between 2000 and 2020.	Maussion et al. (2018)
temp_diff_*	C°	The difference in the average annual temperature of the glacier between 1980 and 2000, and between 2000 and 2020.	Maussion et al. (2018)
debris_area_ela_p	%	The percentage of the glacier area that is covered with debris.	Kraaijenbrink et al. (2017)
debris_vol_ela_p	%	The percentage of debris volume of the total glacier volume.	Kraaijenbrink et al. (2017)
glacial_lake		Binary variable of the presence of a glacial lake intersecting the glacier. 0 stands for no presence and 1 stands for the presence of a glacial lake.	Zheng et al. (2021)
velocity mean	m/a	The average annual glacier flow velocity.	Dehecq et al. (2019)
surging		Binary variable of whether a glacier is a surging glacier or not. 0 stands for no surging glacier and 1 stands for surging glacier.	Dehecq et al. (2019)

Table A.2: Architecture of the ANN used in this study.

Layer	Number of neurons	Activation function	Number of parameters
Input layer	-	-	0
Dense	150	LeakyReLU	1800
Dropout (0.3)	-	-	0
Dense	100	LeakyReLU	15,100
Dropout (0.1)	-	-	0
Dense	50	LeakyReLU	5050
Dropout (0.1)	-	-	0
Dense	50	LeakyReLU	2550
Dropout (0.1)	-	-	0
Dense	50	LeakyReLU	255
Dropout (0.1)	-	-	0
Dense	1	-	51
Total parameters: 27,101; trainable parameters: 27,100; non-trainable parameters: 0.			

Table A.3: Regression table of multiple linear regression with Z-scored coefficients.

Variable	\textbeta	95% CI	t	P>t
const	-0.205	[-0.211, 0.066]	4.769	0.000
Area	-0.022	[-0.044, 0]	-1.939	0.052
avg_ice_thickness	-0.011	[-0.038, 0.017]	-0.758	0.448
debris_vol_ela_p	0.038	[0.018, 0.058]	3.693	0.000
Slope	0.239	[0.214, 0.264]	18.785	0.000
Aspect	0.033	[0.015, 0.051]	3.534	0.000
Zmed	-0.034	[-0.056, -0.012]	-3.006	0.003
tstar_aar	0.166	[0.146, 0.186]	16.228	0.000
temp_diff_2000_2020	0.169	[0.147, 0.190]	15.647	0.000
temp_mean_2000_2020	-0.232	[-0.260, -205]	-16.533	0.000
prcp_diff_2000_2020	0.059	[0.038, 0.080]	5.519	0.000
prcp_mean_2000_2020	-0.217	[-0.244, -0.19]	-15.787	0.000
glacial_lake	-0.464	[-0.528, -0.402]	-14.476	0.000
temp_diff_1980-2000	0.078	[0.056, 0.099]	7.172	0.000
prcp_diff_1980_2000	-0.056	[-0.075, -0.037]	-5.650	0.000
Note: $R^2 = 0.309$ (N = 8099, P = 0.001). CI = confidence interval for β.				

Table A.4: Regression table of multiple linear regression.

Variable	β	95% CI	t	P>t
const	-0.8301	[-0.898, -0.762]	-23.848	0
temp_mean_2000_2020	-0.0166	[-0.019, -0.015]	-16.223	0
temp_diff_2000_2020	0.1988	[0.174, 0.224]	15.647	0
temp_diff_1980-2000	0.0974	[0.071, 0.124]	7.146	0
prcp_mean_2000_2020	-0.0001	[0, -9.83E-05]	-15.683	0
prcp_diff_2000_2020	0.0004	[0, 0.001]	5.719	0
prcp_diff_1980_2000	-3.46E-05	[-4.68E-05, -2.25E-05]	-5.577	0
Area	-0.0003	[-0.001, 4.19E-06]	-1.93	0.054
Slope	0.0121	[0.011, 0.013]	19.053	0
Aspect	7.73E-05	[3.44E-05, 0]	3.532	0
Zmed	-1.71E-05	[-2.78E-05, -6.37E-06]	-3.124	0.002
avg_ice_thickness	-8.84E-05	[-0.001, 0]	-0.406	0.685
debris_area_ela_p	0.0003	[3.32E-05, 0.001]	2.215	0.027
tstar_aar	0.6082	[0.533, 0.684]	15.783	0
glacial_lake	-0.1401	[-0.159, -1.121]	-14.402	0
velocity_mean	-3.93E-05	[-0.002, 0.001]	-0.052	0.958
surging	0.0097	[-0.018, 0.038]	0.686	0.493

Note: R = 0.309 (N = 8099, P = 0.001). CI = confidence interval for β

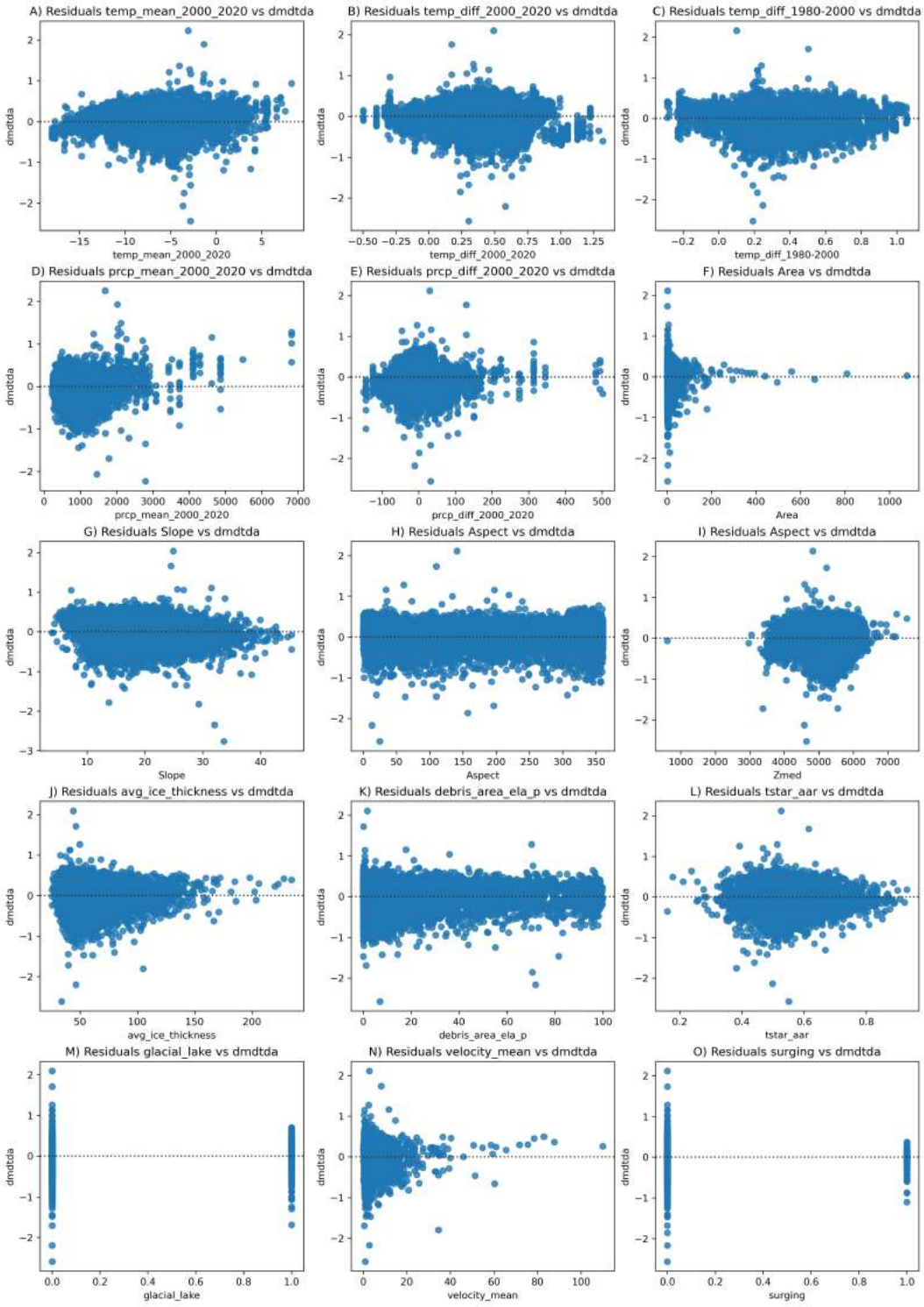

Figure A.1: Residuals of each variable used in this study (Table A.1) versus the dmdtda, based upon the selection of 8099 glaciers in HMA.

Table A.5: Performance of Multiple Linear Regression and Random Forest models for ten random seeds.

	Multiple linear regression			Random Forest		
Seed train_test_split	MAE	RMSE	R^2	MAE	RMSE	R^2
94869	0.197	0.265	0.293	0.126	0.193	0.626
32389	0.186	0.242	0.321	0.115	0.167	0.676
58311	0.187	0.247	0.311	0.119	0.174	0.660
43783	0.192	0.252	0.311	0.118	0.170	0.688
73727	0.182	0.243	0.312	0.118	0.175	0.643
24894	0.190	0.254	0.292	0.118	0.174	0.668
35349	0.195	0.265	0.287	0.122	0.189	0.638
85399	0.186	0.249	0.309	0.121	0.179	0.644
45689	0.186	0.251	0.309	0.118	0.177	0.656
19359	0.186	0.243	0.328	0.118	0.172	0.664
Average	0.189	0.251	0.307	0.119	0.177	0.656
Standard deviation	0.005	0.008	0.013	0.003	0.008	0.019

Table A.6: Performance of XGBoost and Artificial Neural Network for ten random seeds.

	XGBoost			ANN		
Seed train_test_split	MAE	RMSE	R^2	MAE	RMSE	R^2
94869	0.123	0.187	0.650	0.126	0.191	0.641
32389	0.114	0.164	0.688	0.115	0.167	0.677
58311	0.116	0.169	0.680	0.117	0.169	0.677
43783	0.117	0.166	0.701	0.115	0.165	0.705
73727	0.116	0.170	0.664	0.118	0.175	0.639
24894	0.116	0.171	0.682	0.120	0.182	0.636
35349	0.119	0.184	0.656	0.117	0.181	0.649
85399	0.119	0.173	0.666	0.120	0.179	0.654
45689	0.116	0.173	0.673	0.117	0.180	0.650
19359	0.115	0.166	0.685	0.119	0.170	0.665
Average	0.117	0.172	0.675	0.119	0.176	0.659
Standard deviation	0.003	0.008	0.016	0.003	0.008	0.022

9 Code and data availability

The source code of the SMB models is available at https://github.com/DEHartmann/SMBmodels. All scripts used to generate plots and results are included in the repository.

References

Abram, N., Carolina, A., Bindoff, N., & Cheng, L. (2019). Special report on the ocean and cryosphere in a changing climate. *Intergov. Panel Clim. Chang*, *1*, 1–36.

Agarap, A. F. (2018). Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*.

Anderson, B., Mackintosh, A., Stumm, D., George, L., Kerr, T., Winter-Billington, A., & Fitzsimons, S. (2010). Climate sensitivity of a high-precipitation glacier in new zealand. *Journal of Glaciology*, *56*(195), 114–128.

Arendt, A., Bliss, A., Bolch, T., Cogley, J., Gardner, A., Hagen, J.-O., ... others (2017). Randolph glacier inventory–a dataset of global glacier outlines: Version 6.0: Technical report, global land ice measurements from space.

Benn, D., Bolch, T., Hands, K., Gulley, J., Luckman, A., Nicholson, L., ... Wiseman, S. (2012). Response of debris-covered glaciers in the mount everest region to recent warming, and implications for outburst flood hazards. *Earth-Science Reviews*, *114*(1-2), 156–174.

Bhattacharya, A., Bolch, T., Mukherjee, K., King, O., Menounos, B., Kapitsa, V., ... Yao, T. (2021). High mountain asian glacier response to climate revealed by multi-temporal satellite observations since the 1960s. *Nature communications*, *12*(1), 1–13.

Biemans, H., Siderius, C., Lutz, A., Nepal, S., Ahmad, B., Hassan, T., ... others (2019). Importance of snow and glacier meltwater for agriculture on the indo-gangetic plain. *Nature Sustainability*, *2*(7), 594–601.

Blessie, E. C., & Karthikeyan, E. (2012). Sigmis: a feature selection algorithm using correlation based method. *Journal of Algorithms & Computational Technology*, *6*(3), 385–394.

Bolibar, J., Rabatel, A., Gouttevin, I., Galiez, C., Condom, T., & Sauquet, E. (2020). Deep learning applied to glacier evolution modelling. *The Cryosphere*, *14*(2), 565–584.

Bolibar, J., Rabatel, A., Gouttevin, I., Zekollari, H., & Galiez, C. (2022). Nonlinear sensitivity of glacier mass balance to future climate change unveiled by deep learning. *Nature communications*, *13*(1), 1–11.

Bonekamp, P. N., De Kok, R. J., Collier, E., & Immerzeel, W. W. (2019). Contrasting meteorological drivers of the glacier mass balance between the karakoram and central himalaya. *Frontiers in Earth Science*, *7*, 107.

Bozdogan, H. (1987). Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345–370.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Brun, F., Wagnon, P., Berthier, E., Jomelli, V., Maharjan, S., Shrestha, F., & Kraaijenbrink, P. (2019). Heterogeneous influence of glacier morphology on the mass balance variability in high mountain asia.

Journal of Geophysical Research: Earth Surface, *124*(6), 1331–1345.

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, *11*, 2079–2107.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... others (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1–4.

Cogley, J. G., Arendt, A., Bauder, A., Braithwaite, R., Hock, R., Jansson, P., ... others (2010). Glossary of glacier mass balance and related terms.

Compagno, L., Huss, M., Miles, E. S., McCarthy, M. J., Zekollari, H., Dehecq, A., ... Farinotti, D. (2022). Modelling supraglacial debris-cover evolution from the single-glacier to the regional scale: an application to high mountain asia. *The Cryosphere*, *16*(5), 1697–1718.

Constantine, N. A. (2012). Regression analysis and causal inference: cause for concern? *Perspectives on sexual and reproductive health*, 134–137.

Cressie, N. (1988). Spatial prediction and ordinary kriging. *Mathematical geology*, *20*(4), 405–421.

Crone, S. F. (2005). Stepwise selection of artificial neural network models for time series prediction. *Journal of Intelligent Systems*, *14*(2-3), 99–122.

Davies, B. (2021). *Glacial lakes.* Retrieved from `https://www.antarcticglaciers.org/glacier-processes/glacial-lakes/`

Dehecq, A., Gourmelen, N., Gardner, A. S., Brun, F., Goldberg, D., Nienow, P. W., ... Trouvé, E. (2019). Twenty-first century glacier slowdown driven by mass loss in high mountain asia. *Nature Geoscience*, *12*(1), 22–27.

De Kok, R. J., Kraaijenbrink, P. D., Tuinenburg, O. A., Bonekamp, P. N., & Immerzeel, W. W. (2020). Towards understanding the pattern of glacier mass balances in high mountain asia using regional climatic modellingrounce2021distributed. *The Cryosphere*, *14*(9), 3215–3234.

De Woul, M. (2008). *Response of glaciers to climate change: mass balance sensitivity, sea level rise and runoff* (Unpublished doctoral dissertation). Institutionen för naturgeografi och kvartärgeologi.

Duke University. (2022). *Testing the assumptions of linear regression.* `https://people.duke.edu/~rnau/testing.htm#normality`. (Accessed on 06/30/2022)

Farinotti, D., Huss, M., Fürst, J. J., Landmann, J., Machguth, H., Maussion, F., & Pandit, A. (2019). A consensus estimate for the ice thickness distribution of all glaciers on earth. *Nature Geoscience*, *12*(3), 168–173.

Fischer, M., Huss, M., & Hoelzle, M. (2015). Surface elevation and mass changes of all swiss glaciers 1980–2010. *The Cryosphere*, *9*(2), 525–540.

Gardner, A., Fahnestock, M., & Scambos, T. (2021). *Measures its_live landsat imagepair glacier and ice sheet surface velocities, version 1.* NASA National Snow and Ice Data Center DAAC. Retrieved from

https://nsidc.org/apps/itslive/ doi: 10.5067/IMR9D3PEI28U

Gerbaux, M., Genthon, C., Etchevers, P., Vincent, C., & Dedieu, J. (2005). Surface mass balance of glaciers in the french alps: distributed modeling and sensitivity to climate change. *Journal of Glaciology*, *51*(175), 561–572.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... others (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of hydrology*, *282*(1-4), 104–115.

Hock, R. (2010). Glacier mass balance. *Encyclopedia of Snow, Ice and Glaciers; Springer: Dordrecht, The Netherlands*, 399–408.

Hock, R., Bliss, A., Marzeion, B., Giesen, R. H., Hirabayashi, Y., Huss, M., ... Slangen, A. B. (2019). Glaciermip–a model intercomparison of global-scale glacier mass-balance models and projections. *Journal of Glaciology*, *65*(251), 453–467.

Hoinkes, H. C. (1968). Glacier variation and weather. *Journal of Glaciology*, *7*(49), 3–18.

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., ... others (2021). Accelerated global glacier mass loss in the early twenty-first century. *Nature*, *592*(7856), 726–731.

Huo, D., Bishop, M. P., & Bush, A. B. (2021). Understanding complex debris-covered glaciers: Concepts, issues, and research directions. *Frontiers in Earth Science*, 358.

Huss, M. (2012). Extrapolating glacier mass balance to the mountain-range scale: the european alps 1900–2100. *The Cryosphere*, *6*(4), 713–727.

Huss, M., & Fischer, M. (2016). Sensitivity of very small glaciers in the swiss alps to future climate change. *Frontiers in Earth Science*, *4*, 34.

Immerzeel, W. W., Lutz, A., Andrade, M., Bahl, A., Biemans, H., Bolch, T., ... others (2020). Importance and vulnerability of the world's water towers. *Nature*, *577*(7790), 364–369.

Kapnick, S. B. (2014). *Powerpoint presentation.* https://www.wilsoncenter.org/sites/default/files/media/documents/event/sarah_kapnick_climate_precipitation_trends_over_high_mountain_asia.pdf. (Accessed on 06/20/2022)

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kraaijenbrink, P. D., Bierkens, M., Lutz, A., & Immerzeel, W. (2017). Impact of a global temperature rise of 1.5 degrees celsius on asia's glaciers. *Nature*, *549*(7671), 257–260.

Lalande, M., Ménégoz, M., Krinner, G., Naegeli, K., & Wunderle, S. (2021). Climate change in the high mountain asia in cmip6. *Earth system dynamics*, *12*(4), 1061–1098.

Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An

example of shap and xgboost. *Computers, Environment and Urban Systems*, *96*, 101845.

Liu, X. (2020). *Incorporating spatial autocorrelation in machine learning* (Unpublished master's thesis). University of Twente.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, *2*(1), 56–67.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Lutz, A. F. (2018). *Impact of climate change on the hydrology of high mountain asia.* Faculty of Geosciences, Universiteir Utrecht, The Netherlands.

Machguth, H., & Huss, M. (2014). The length of the world's glaciers ndash; a new approach for the global calculation of center lines. *The Cryosphere*, *8*(5), 1741–1755. Retrieved from `https://tc.copernicus.org/articles/8/1741/2014/` doi: 10.5194/tc-8-1741-2014

Mannerfelt, E. S., Dehecq, A., Hugonnet, R., Hodel, E., Huss, M., Bauder, A., & Farinotti, D. (2022). Halving of swiss glacier volume since 1931 observed from terrestrial image photogrammetry. *The Cryosphere Discussions*, 1–32.

Marzeion, B., Jarosch, A., & Hofer, M. (2012). Past and future sea-level change from the surface mass balance of glaciers. *The Cryosphere*, *6*(6), 1295–1322.

Maurer, J., Schaefer, J., Rupper, S., & Corley, A. (2019). *Acceleration of ice loss across the himalayas over the past 40 years, sci. adv., 5, eaav7266.*

Maussion, F., Butenko, A., Eis, J., Fourteau, K., Jarosch, A. H., Landmann, J., ... Marzeion, B. (2018). The open global glacier model (oggm) v1.0. *Geoscientific Model Development Discussions*, *2018*, 1–33. Retrieved from `https://www.geosci-model-dev-discuss.net/gmd-2018-9/` doi: 10.5194/gmd-2018-9

Maussion, F., Rothenpieler, T., Recinos, B., Vlug, A., Marzeion, B., Oesterle, F., ... Smith, S. (2018, January). *Oggm/oggm: v1.4.* Retrieved from `https://doi.org/10.5281/zenodo.1149701` doi: 10.5281/zenodo.1149701

MLJAR. (2022). *Random forest vs xgboost.* `https://mljar.com/machine-learning/random-forest-vs-xgboost/`. (Accessed on 06/29/2022)

Molnar, C. (2020). *Interpretable machine learning.* Lulu. com.

NASA. (2022). *Aster.* `https://terra.nasa.gov/about/terra-instruments/aster`. (Accessed on 06/27/2022)

Nawi, N. M., Atomi, W. H., & Rehman, M. Z. (2013). The effect of data pre-processing on optimized training of artificial neural networks. *Procedia Technology*, *11*, 32–39.

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Pafka, S. (2022). *Simple/limited/incomplete benchmark for scalability, speed and accuracy of machine learning libraries for classification.* `https://github.com/szilard/benchm-ml`. (Accessed on 06/30/2022)

Pronk, J. B., Bolch, T., King, O., Wouters, B., & Benn, D. I. (2021). Contrasting surface velocities between lake-and land-terminating glaciers in the himalayan region. *The Cryosphere*, *15*(12), 5577–5599.

Qi, Y. (2012). Random forest for bioinformatics. In *Ensemble machine learning* (pp. 307–323). Springer.

Rabatel, A., Letréguilly, A., Dedieu, J.-P., & Eckert, N. (2013). Changes in glacier equilibrium-line altitude in the western alps from 1984 to 2010: evaluation by remote sensing and modeling of the morpho-topographic and climate controls. *The Cryosphere*, *7*(5), 1455–1471.

Raper, S. C., & Braithwaite, R. J. (2009). Glacier volume response time and its links to climate and topography based on a conceptual model of glacier hypsometry. *The Cryosphere*, *3*(2), 183–194.

RGI Consortium. (2017). *Randolph glacier inventory 6.0.* NSIDC. Retrieved from `http://www.glims.org/RGI/randolph60.html` doi: 10.7265/N5-RGI-60

Rounce, D. R., Hock, R., McNabb, R., Millan, R., Sommer, C., Braun, M., ... others (2021). Distributed global debris thickness estimates reveal debris significantly impacts glacier mass balance. *Geophysical Research Letters*, *48*(8), e2020GL091311.

Rounce, D. R., Hock, R., & Shean, D. E. (2020). Glacier mass change in high mountain asia through 2100 using the open-source python glacier evolution model (pygem). *Frontiers in Earth Science*, *7*, 331.

Sakai, A., & Fujita, K. (2017). Contrasting glacier responses to recent climate change in high-mountain asia. *Scientific reports*, *7*(1), 1–8.

Sakai, A., Nishimura, K., Kadota, T., & Takeuchi, N. (2009). Onset of calving at supraglacial lakes on debris-covered glaciers of the nepal himalaya. *Journal of Glaciology*, *55*(193), 909–917.

Salinger, J., Chinn, T., Willsman, A., & Fitzharris, B. (2008). Glacier response to climate change. *Water & Atmosphere*, *16*(3), 16–17.

scikit learn. (2022a). *Gridsearchcv.* `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html`. (Accessed on 06/11/2022)

scikit learn. (2022b). *Imputation of missing values.* `https://scikit-learn.org/stable/modules/impute.html#iterative-imputer`. (Accessed on 06/10/2022)

scikit learn. (2022c). *Permutation feature importance.* `https://scikit-learn.org/stable/modules/permutation_importance.html#:~:text=The%20permutation%20feature%20importance%20is,model%20depends%20on%20the%20feature.` (Accessed on 06/13/2022)

Shahneh, M. R., Oymak, S., & Magdy, A. (2021). A-gwr: Fast and accurate geospatial inference via augmented geographically weighted regression. In *Proceedings of the 29th international conference on*

advances in geographic information systems (pp. 564–575).

Shean, D. E., Bhushan, S., Montesano, P., Rounce, D. R., Arendt, A., & Osmanoglu, B. (2020). A systematic, regional assessment of high mountain asia glacier mass balance. *Frontiers in Earth Science*, *7*, 363.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929–1958.

Steiner, D., Walter, A., & Zumbühl, H. (2005). 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.

Stokes, C. R., Sanderson, J. E., Miles, B. W., Jamieson, S. S., & Leeson, A. A. (2019). Widespread distribution of supraglacial lakes around the margin of the east antarctic ice sheet. *Scientific reports*, *9*(1), 1–14.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, *8*(1), 1–21.

Tranmer, M., & Elliot, M. (2008). Multiple linear regression. *The Cathie Marsh Centre for Census and Survey Research (CCSR)*, *5*(5), 1–5.

Treichler, D., Kääb, A., Salzmann, N., & Xu, C.-Y. (2018). High mountain asia glacier elevation trends 2003–2008, lake volume changes 1990–2015, and their relation to precipitation changes. *The Cryosphere Discuss., https://doi. org/10.5194/tc-2018-238, in review*.

Wang, R., Liu, S., Shangguan, D., Radić, V., & Zhang, Y. (2019). Spatial heterogeneity in glacier mass-balance sensitivity across high mountain asia. *Water*, *11*(4), 776.

Watson, C. S., Kargel, J. S., Shugar, D. H., Haritashya, U. K., Schiassi, E., & Furfaro, R. (2020). Mass loss from calving in himalayan proglacial lakes. *Frontiers in Earth Science*, *7*, 342.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, 817–838.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, *1*(1), 67–82.

Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*.

Zekollari, H., Huss, M., & Farinotti, D. (2020). On the imbalance and response time of glaciers in the european alps. *Geophysical Research Letters*, *47*(2), e2019GL085578.

Zheng, G., Allen, S. K., Bao, A., Ballesteros-Cánovas, J. A., Huss, M., Zhang, G., ... others (2021). Increasing risk of glacial lake outburst floods from future third pole deglaciation. *Nature Climate Change*, *11*(5), 411–417.

Zhou, P., Feng, J., Ma, C., Xiong, C., Hoi, S. C. H., et al. (2020). Towards theoretically understanding why sgd generalizes better than adam in deep learning. *Advances in Neural Information Processing Systems*, *33*, 21285–21296.

Zhou, X., Zhu, X., Dong, Z., Guo, W., et al. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal*, *4*(3), 212–219.