JGR Earth Surface

RESEARCH ARTICLE
10.1029/2022JF006597

A Machine Learning Framework to Automate the Classification of Surge-Type Glaciers in Svalbard

C. Bouchayer[1], J. M. Aiken[1], K. Thøgersen[1], F. Renard[1,2], and T. V. Schuler[1]

[1]The Njord Centre, Departments of Geosciences and Physics, University of Oslo, Oslo, Norway, [2]ISTerre, University Grenoble Alpes, Grenoble INP, University Savoie Mont Blanc, CNRS, IRD, University Gustave Eiffel, Grenoble, France

Key Points:

- We establish a machine learning framework to evaluate the probability of glacier surge
- We build a combined database of glaciers in Svalbard that contains thirteen features
- We compute the first map of glacier surge probability in Svalbard and we quantify the relative importance of relevant features

Correspondence to:

C. Bouchayer,
colili@uio.no

Citation:

Bouchayer, C., Aiken, J. M., Thøgersen, K., Renard, F., & Schuler, T. V. (2022). A Machine learning framework to automate the classification of surge-type glaciers in Svalbard. *Journal of Geophysical Research: Earth Surface*, *127*, e2022JF006597. https://doi.org/10.1029/2022JF006597

Received 12 JAN 2022
Accepted 27 JUN 2022

Abstract Surge-type glaciers are present in many cold environments in the world. These glaciers experience a dramatic increase in velocity over short time periods, the surge, followed by an extended period of slow movement, the quiescence. This study aims at understanding why only few glaciers exhibit a transient behavior. We develop a machine learning framework to classify surge-type glaciers, based on their location, exposure, geometry, climatic mass balance and runoff. We apply this approach to the Svalbard archipelago, a region with a relatively homogeneous climate. We compare the performance of logistic regression, random forest, and extreme gradient boosting (XGBoost) machine learning models that we apply to a newly combined database of glaciers in Svalbard. Based on the most accurate model, XGBoost, we compute surge probabilities along glacier centerlines and quantify the relative importance of several controlling features. Results show that the surface and bed slopes, ice thickness, glacier width, climatic mass balance, and runoff along glacier centerlines are the most significant features explaining surge probability for glaciers in Svalbard. A thicker and wider glacier with a low surface slope has a higher probability to be classified as surge-type, which is in good agreement with the existing theories of surging. Finally, we build a probability map of surge-type glaciers in Svalbard. The framework shows robustness on classifying surge-type glaciers that were not previously classified as such in existing inventories but have been observed surging. Our methodology could be extended to classify surge-type glaciers in other areas of the world.

Plain Language Summary 1% of the glaciers in the world exhibit intermittent phases of accelerated motion, called surge. These accelerations are not fully understood and we do not know why only few glaciers experience such behavior. Surging glaciers may lead to dramatic advances over rivers and damming up lakes that are then prone to a sudden and possibly catastrophic drainage. The Svalbard archipelago, located in the high Arctic, hosts more than one hundred surging glaciers. By analyzing statistically several data-sets, we calculate the probability for every glacier to experience surge events. Our results show that specific combinations of surface and bed slopes, glacier width and ice thickness control glacier surge probability. To a smaller extent climatic parameters such as the mass a glacier may lose or gain during the year and the amount of melt water available also contribute to the surge probability. These findings are in good agreement with existing theories explaining surge dynamics. We produce the first probabilistic map of surging for all the glaciers in Svalbard. Our classification highlights some glaciers that were not classified as surging glaciers in glacier inventories but have been observed surging, confirming the robustness of our framework. Our method is applicable to other world regions.

1. Introduction

Glacier instabilities, such as surges, are primary contributors to uncertainties of future sea-level rise projections (Ritz et al., 2015). Surge-type glaciers exhibit long periods of quiescence and short periods of accelerated motion, transferring substantial ice mass to lower elevation and thus often leading to rapid ice loss (Cuffey & Paterson, 2010; Meier & Post, 1969). They represent approximately 1% of the glaciers in the world (Sevestre & Benn, 2015) and pose a considerable hazard potential (Truffer et al., 2021). Surges can occur at quasi-regular time intervals and a huge spatial variability has been observed, with surging and non-surging glaciers located next to each other (Bhambri et al., 2017; Cuffey & Paterson, 2010; Meier & Post, 1969). Thus, identifying surge-type glaciers may contribute to a reduction in the uncertainties of future sea-level rise and may provide better hazard mitigation (e.g., surges related to glacier lake outburst floods (Bazai et al., 2021)). In the present study, we use the term surge for quasi-cyclic increases of ice flow velocity that "result from oscillations in conditions at the bed of the glacier" (Benn & Evans, 2014).

Surge dynamics is considered to be governed by changes in the sub-glacial drainage system configuration, switch in the thermal basal conditions and interactions between the hydrology and the glacier substrate (Benn et al., 2019; Cuffey & Paterson, 2010; Minchew & Meyer, 2020; Thøgersen et al., 2019, 2021). Due to the limited accessibility of subglacial environments, the physical processes at the ice-bed interface are difficult to measure. Recently, three approaches have been proposed to unify the theories of glacier instabilities. On the one hand, Benn et al. (2019) proposed that a glacier remains stable when the variations of enthalpy at the glacier bed, which impact the ice flow, are in equilibrium with the variations of ice mass. Enthalpy increases through geothermal and frictional heating and decreases by heat conduction and melt water exiting the system. If the ice mass and enthalpy budget are out of equilibrium, the glacier dynamic will alternate between periods of quiescence and surge phases. On the other hand, Thøgersen et al. (2019); Thøgersen et al. (2021) developed an evolution model for subglacial friction based on the rate-and-state friction law (Dieterich, 1992), suggesting that large enough perturbations can propagate and cause a glacier surge. They concluded that a better understanding of the feedback between the subglacial drainage and basal friction is critical to describe such perturbations. Other studies have examined the rate-and-state friction law to describe mechanical processes at the ice-bed interface (Zoet et al., 2020). Finally, Minchew and Meyer (2020) proposed a model describing the mechanical evolution of the till (internal friction, porosity and pore water pressure) while the glacier is flowing. They suggested that changes on both hydromechanical properties of the sediment layer and the thickness of the glacier may control surge behavior. Based on these three approaches, we select a series of features detailed below, which have been proposed to control the process of glacier surge. In the following, we use the term features to denominate physical parameters that may have an effect on glacier surging.

Previous studies have established that surge-type glaciers have the following properties: (a) they are more likely to be longer and/or wider (Barrand & Murray, 2006; Clarke, 1991; Clarke et al., 1986; Jiskoot et al., 1998) than non-surging glaciers. These variables are highly correlated to the bed and surface slopes of the glacier and so the relative importance of each individual feature is hard to assess (Clarke, 1991; Clarke et al., 1986); (b) their bed contains more likely younger and mechanically weaker lithologies than hard beds (Jiskoot et al., 1998, 2000); (c) they are clustered in climatic envelopes between cold-dry and warm-humid environments (Sevestre & Benn, 2015); and (d) they are more likely polythermal in Svalbard (only region where the thermal regime has been integrated into a statistical study) (Jiskoot et al., 2000). Based upon these studies, Sevestre and Benn (2015) built an entropy maximization model to qualitatively classify the glaciers in the Randolph Glacier Inventory database, RGI (2017), into five surging categories, from no surge to surge-type (Figure 1). However, statistical studies of glacier surges have two limitations: (a) they use integrated features for entire glaciers, and (b) none of them are comparing different type of models. Although Barrand and Murray (2006) explored differences between generalized linear models and the features that are included in each model, their study does not compare different types of models.

Here we address the question of why some glaciers exhibit such intermittent behavior while others do not, and whether these differences can be explained based on geometric and climatic characteristics. By limiting the geographical extent of our study area to a climatically relative homogeneous setting, we exclude overall climatic controls (Sevestre & Benn, 2015) and aim to isolate the non-climatic influences. The climate in Svalbard is assumed to be relatively homogeneous compare to other regions. Around 22% of Svalbard's glaciers are surge-type, which represents a relatively large proportion of the 1,615 glaciers of this region reported in the Randolph Glacier Inventory (RGI, 2017). We propose a framework to regularize the evaluation of several machine learning models for determining glacier surge probability. The machine learning framework aims at identifying glacier areas that might enter an unstable regime due to their geometrical and climatic configuration (Figure 2). Selecting the best performing model, the Extreme Gradient Boosting (XGBoost) (Chen & Guestrin, 2016), we identify the features that control the classification of surge-type glaciers. By applying this framework on a custom-built database, we produce a map of surge probability for Svalbard glaciers. Using this model, we demonstrate that geometrical features have a high impact on the classification, and these findings are discussed in the context of the existing glacier surge theories. The machine-learning framework can be applied for assessing the surge probability of glaciers in other regions of the world, expanded when new data are available and/or adapted to other fields (e.g., landslides, earthquakes dynamics).

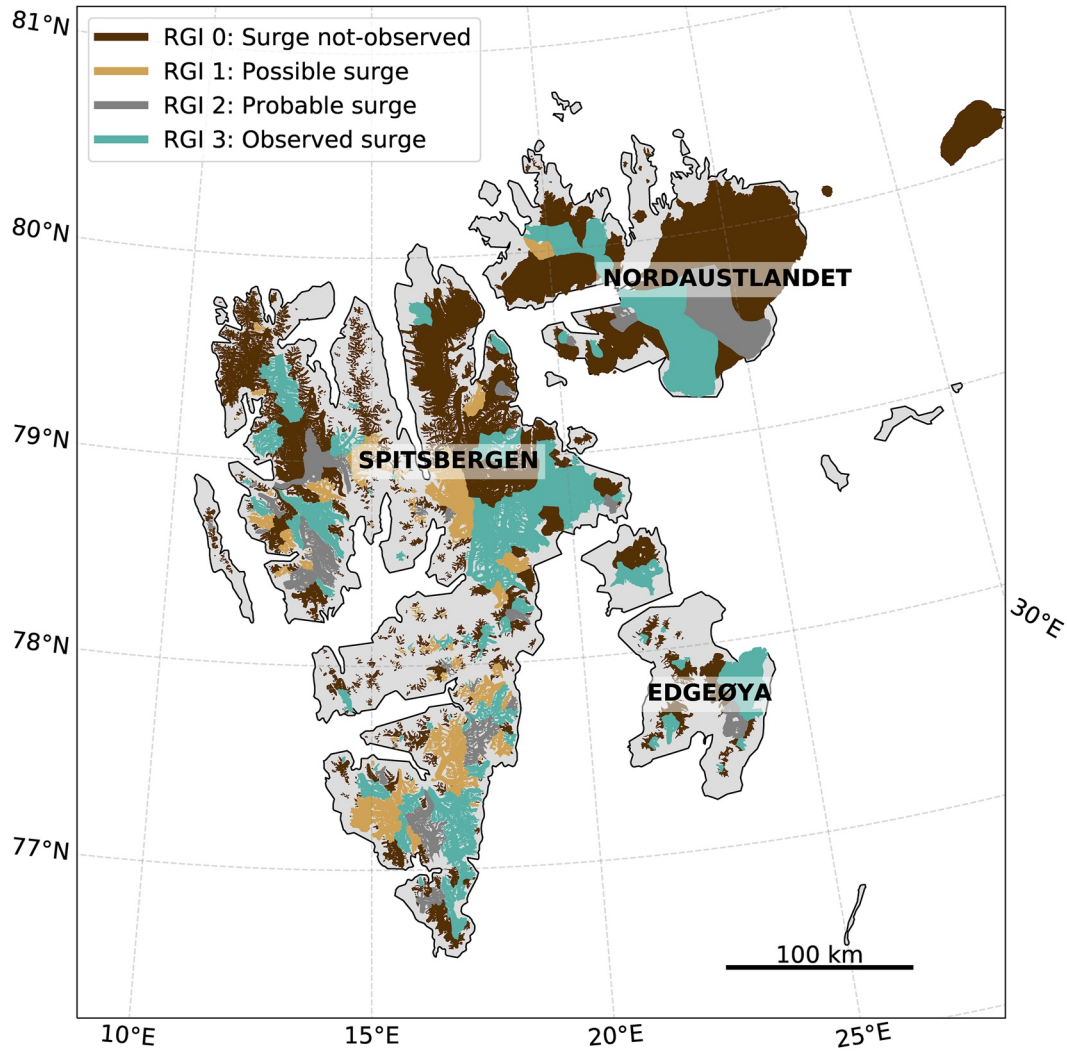

Figure 1. Classification of glaciers in Svalbard in the Randolph Glaciological Inventory database (RGI, 2017). This database contains five classes that characterize the surge potential of glaciers (Sevestre & Benn, 2015): Not observed (0), Possible (1), Probable (2), Observed (3), Not assigned (9). The class 9 is not represented in the Svalbard region.

2. Data and Methods to Assess the Surge Probability of Glaciers in Svalbard

We develop a machine learning framework for classifying surge-type glaciers. This framework includes the development of a custom-built database, a method for training machine learning models consistent with best machine learning practices, methods for evaluating the model outputs (i.e., the probability for a glacier to be classified as surge-type), and finally a method for mapping the surge probability of Svalbard glaciers. Additionally, we identify the key features that control the predictions of the models.

We build a glacier database by combining the Randolph Glacier Inventory (RGI, 2017), geometrical features (Fürst et al., 2018; Maussion et al., 2019), and climatic data (Pelt et al., 2019). These data are discretized along the glacier centerlines. After discretizing and post-processing the data, the custom-build database combines 981 glaciers which are discretized along 97,140 points over Svalbard.

The database is used in three different supervised machine learning models: logistic regression, random forest, and XGBoost. Data are split between training and testing data-sets. Training data are used to teach the machine learning models whether a glacier is surge-type. Testing data are used to evaluate the ability of the models to classify surge-type glaciers.

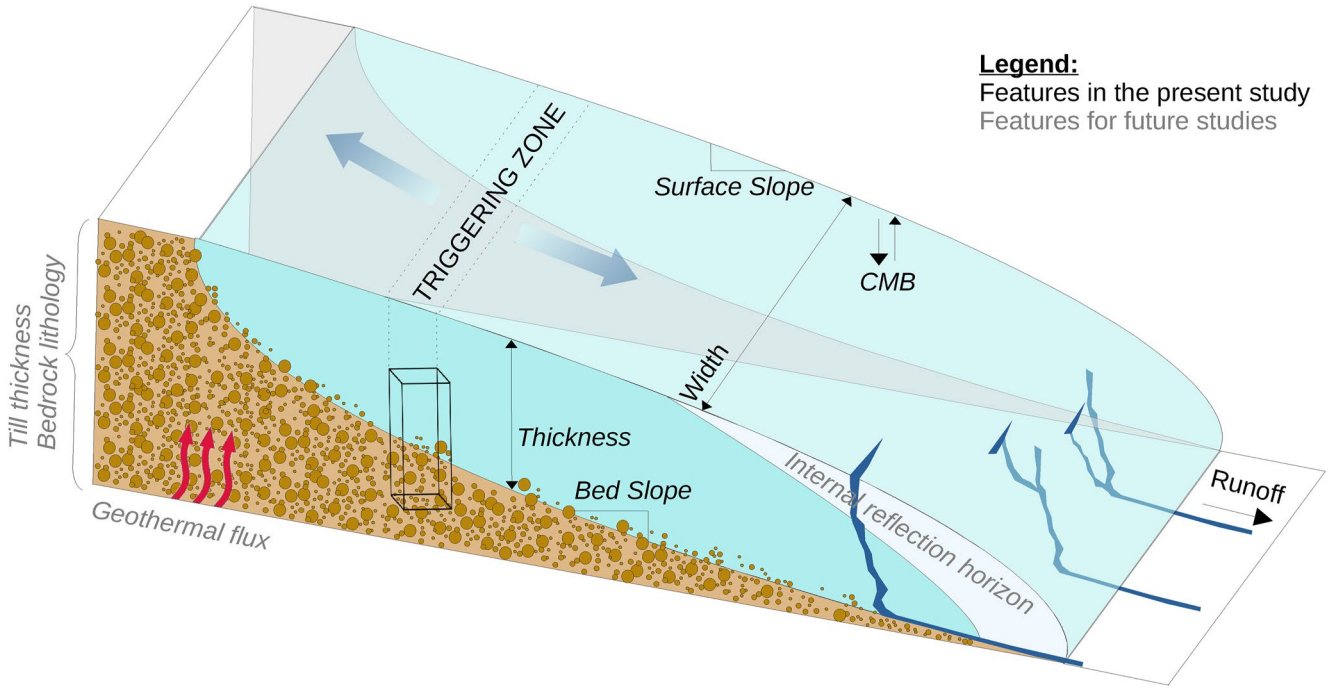

Figure 2. Sketch of the features that have been implemented in our model and new features that could be implemented for evaluating the surging potential of glaciers.

These models are evaluated using multiple statistic metrics, such as the area under the Receiver Operator Characteristic curve (Hanley & McNeil, 1982). After the models are evaluated, the best model, in our case XGBoost, is used to calculate the surge probability of each centerline point in each glacier. These values are then used to build a probability map of surge glaciers in Svalbard.

In addition to generating the probability map, we identify the features in the training data set that most strongly control the classification. We calculate the feature importance scores for each model. We also perform a recursive feature elimination to quantify the contribution of each features in the model performance (Chen & Jeong, 2007), and finally we use the Shapley Additive values (Lundberg & Lee, 2017). The sketch in Figure 3 illustrates our framework.

2.1. Data

2.1.1. Randolph Glacier Inventory Features

The Randolph Glacier Inventory (RGI, 2017) is a globally complete database of digital outlines of glaciers worldwide, excluding the ice sheets. This database was developed to provide better estimates of past and future surface mass balance of glaciers (Pfeffer et al., 2014). It includes integrated features such as glacier surface area and length. Glaciers are classified into five different surging categories: 0 - Surge not observed, 1 - Possible surge-type, 2 - Probable surge-type, 3 - Surge Observed, 9 - Not surging (Figure 1). This classification has been established following the work of Sevestre and Benn (2015). While the classes 0 and 3 are based on field observations, the classes 1 and 2 are based on statistical modeling (Sevestre & Benn, 2015). However, no quantitative predictions of surge probability are assessed.

The Randolph Glacier Inventory is distributed through the Global Land Ice Measurements from Space, and the National Snow and Ice Data Center (GLIMS/NSIDC) website (RGI, 2017). It is continuously developed and new versions are released regularly. In the present study, we use the most recent version (v6.0) for the Svalbard region, which is the region 7 in this database. From the Randolph Glacier Inventory, we use only the unique identifier allocated for each glacier in Svalbard (RGIId), the corresponding glacier name and the surging class (Figure 3). Other features present in the Randolph Glacier Inventory, such as the surface area and the length of glaciers, are

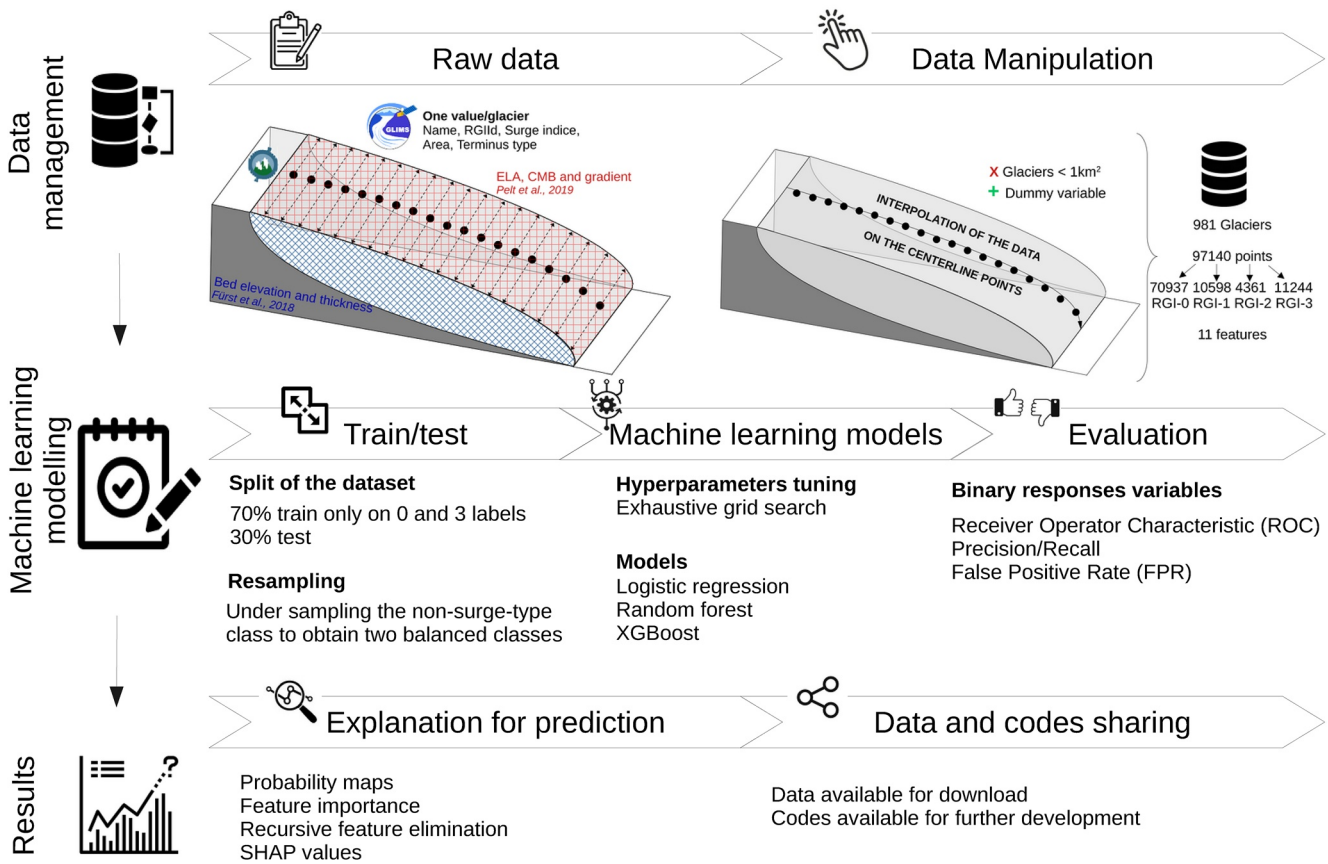

Figure 3. Workflow of the machine learning methods used to classify surge-type glaciers. Once the raw data are collected, the features are interpolated along the centerlines points. The database is then filtered and separated into a train data set and a test data set. Data are re-sampled to obtain balanced classes between surge-type and non-surge-type glaciers. The machine learning models are run and evaluated. The best model is XGBoost after evaluation. By looking at the contribution of each feature contribution in the model, the surge probability map of glaciers in Svalbard is produced.

not used because these are integrated features across each glaciers while we focus in this study on discretized variables along glacier centerlines.

2.1.2. Geometric Features

Many studies investigating glacier surges have highlighted the importance of geometrical features (Barrand & Murray, 2006; Björnsson et al., 2003; Clarke, 1991; Clarke et al., 1986; Hamilton & Dowdeswell, 1996; Jiskoot et al., 1998, 2000; Sevestre & Benn, 2015). In the present study, we include the width, the thickness, the bed elevation and the surface elevation of each glacier, and the associated bed and surface slopes. The geometrical widths have been computed using the Open Global Glacier Model (Maussion et al., 2019). This model is open-source and is partly used to simulate past and future changes of any glacier in the world. Glacier outlines are extracted from the Randolph Glacier Inventory and projected onto a local glacier grid. The spatial resolution depends on the size of the glacier (Maussion et al., 2019). The geometrical widths are computed by intersecting lines perpendicular to the flow lines at each grid point with the glacier outlines and the tributary catchment areas. The detailed workflow is described in Maussion et al. (2019).

The bed elevation and glacier thickness are retrieved from Fürst et al. (2018). These authors presented a first version of the ice-free topography (SVIFT1.0), which was computed using a mass conservation approach for mapping glacier ice thickness. This database is built from more than one million point measurements coming from radio-echo soundings. In total, it corresponds to an accumulated length of 700 km of measured thickness profiles. The ice thickness was also computed where measurements were not available and an error estimation map is available by comparing the model calculation and the observations. The reconstructed ice thickness corresponds to the status of the glaciers in year 2010 (Fürst et al., 2018). We also estimated the surface and bed slopes

by calculating the gradient between two successive points along the centerlines of the surface and bed elevation data.

2.1.3. Climatic Features

We added climatic features to the database, that is, runoff and climatic mass balance (CMB). Pelt et al. (2019) created a long-term (1957–2018) data set of CMB for the glaciers, snow conditions, and runoff with a 1 × 1 km spatial resolution and 3-hr temporal resolution over Svalbard. These authors used a coupled energy balance–subsurface model, forced with down-scaled regional climate model fields, and apply it to both glacier-covered and land areas in Svalbard. In our study, we characterize CMB by spatially distributed values of the Equilibrium Line Altitude (ELA) and mass balance gradient. The runoff is the local discharge corresponding to the available water coming from rainfall and melt at the bed after accounting for retention by refreezing and liquid water storage (Pelt et al., 2019). We use the latest computed data corresponding to the year 2018.

2.2. Data Management

2.2.1. Discretization

Using the Open Global Glacier Model, we computed the centerline coordinates for each glacier in Svalbard with the algorithm developed by Kienholz et al. (2014) and modified by Maussion et al. (2019). Once the termini and the heads of each glacier are identified, the least-cost route is calculated to derive the centerlines. The centerline points are not equidistant after this calculation. Then, the centerlines points are interpolated to be equidistant from each other. Depending on the size of each glacier, the distance between successive points varies between 20 and 400m for different glaciers. Some glacier catchments contain a main glacier accompanied by its tributary glaciers and so several centerlines are computed for the same catchment. In our study, we use the longest centerline as the main centerline of the principal glacier. Once the centerlines have been extracted, we interpolate or extrapolate all other data along the centerlines coordinates.

2.2.2. Custom-Built Database of Svalbard Glaciers

The database is the combination of all the features discretized along the centerlines. Since the climatic data have a spatial resolution of 1 × 1 km, we exclude all the glaciers with a surface area less than 1 km^2 and a length less than one km.

As a consequence, our custom-built database contains 981 glaciers which are discretized along 97,140 points: 70,937 points belong to the class "Not Observed Surging," 10,598 belong to the class "Possible Surge," 4,361 belong to the class"Probable Surge-type," and 11,244 belong to the class "Observed Surging." The surge class of an entire glacier is assigned to every points discretized along the centerline of this glacier. Eleven features are used to train the statistical models: the surging class (1), the bed elevation and slope (2, 3), the surface elevation and slope (4, 5), the thickness (6), the CMB (7), the glacier width (8), the width divided by the thickness (9), and the driving stress (10). A random number is also added as a dummy feature (11) that does not have a physical interpretation and is used here to compare the importance of the other features to a random value (McBeck et al., 2020). The final database contains as well the Randolph Glacier Inventory identifier and the corresponding glacier name.

Figure 4 displays the correlations between these features. The features clustered in the upper left corner of the correlation matrix show high positive or negative correlations. Following the diagonal of the matrix toward the lower right corner, the correlations are decreasing. The bed elevation, thickness, width, runoff, bed and surface slope are highly correlated with each other. The driving stress, width times thickness (W × H), and the dummy features show correlation values close to 0, indicating that they are not correlated to other features.

2.3. Machine Learning Modeling

2.3.1. Training and Testing Data Sets

The training data set is organized in the following ways: (a) only glaciers classified as Not-Observed surge (class 0) or Observed surge (class 3) in the Randolph Glacier Inventory are used; (b) the training data set is resampled such that it contains an equal number of surge-type and non-surge-type glaciers; and (c) the training and testing

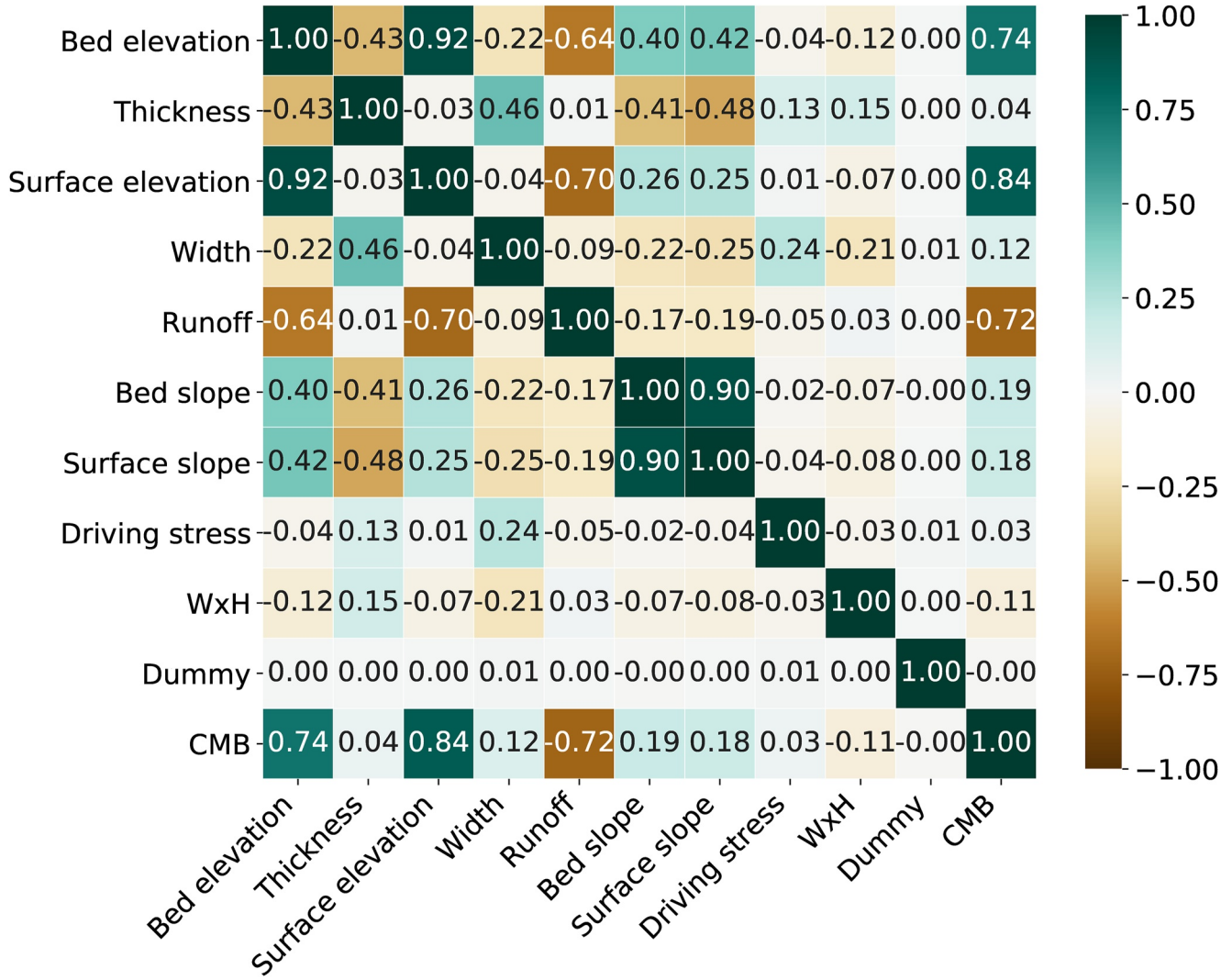

Figure 4. Correlation matrix between the most important features using only the training data. The colors shows the value of the coefficient of correlation.

data-sets are split such that all the data of a given glacier belong either to the training data set or to the testing data set, but not to both. We only use glaciers from the classes Not-Observed (0) and Observed (3) surge to avoid systematic errors that may be associated with glaciers labeled in the Randolph Glacier Inventory as having some likelihood to be surge-type but no direct evidence of surging behavior has been observed (i.e., Possible surge (1) and Probable surge (2) classes). We train the data considering only two classes, surge-type or not surge-type (class 0 and 3 of the RGI classification). Every point along the centerline will then be assigned to a 100% probability to be classified as surge-type or 100% probability to be classified as not-surge type. The results of the statistical modeling will however assess the probability for each point in the centerline of a glacier to be classified as surge-type or not rather than a binary class allocation for the entire glacier.

The glaciers classes are highly unbalanced with almost seven times more glaciers of the Not-Observed surge class than Observed surge class. An unbalanced training data set can lead to erroneous results in classification problems (Ganganwar, 2012). Therefore, we under-sampled the Not-Observed surge glaciers such that the data set contains a 50%–50% distribution of non-surging and surging glaciers. This data set is then split into a training and a testing set, respectively 70% and 30% of the database. We justify this split by running ten different models trained with a split between 10% and 90% for the training data set. The values of the Area Under The Curve (AUC) is acceptable when 70% of the database is used for training, while higher proportion might lead to overfitting model. A precise description of the analysis is presented in the appendix (Appendix A, Figure A1).

2.3.2. Machine Learning Models

We use three different supervised machine learning models: logistic regression, random forest and Extreme Gradient Boosting (hereafter, XGBoost). Logistic regression is based on the maximum likelihood and was used in previous glaciological statistical studies (Jiskoot et al., 1998, 2000), while random forest is a commonly used tree-based ensemble model. XGBoost is a new tree ensemble model that is leading machine learning competitions (e.g., Kaggle). We choose to compare traditionally used models in the glaciology community, with commonly used models in other scientific communities and a new leading model. Using a data set with a known outcome (i.e., whether a glacier is surge-type or not), we train models to identify this outcome. Each model requires selecting at least one hyperparameter (e.g., the depth of decision trees used in a random forest). We selected the values of the hyperparameters after an exhaustive grid search (Appendix C, Table C1).

2.3.2.1. *Logistic Regression*

Logistic regression is commonly used in machine learning for classification tasks. This algorithm produces a probabilistic estimate of whether a particular set of input features belongs to a class or not. Logistic regression has been used in several studies in glaciology to better understand glacier surges (e.g., Jiskoot et al. (2000); Barrand and Murray (2006)). We used the logistic regression equation (Nelder & Wedderburn, 1972):

$$\ell = \log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_M x_M = \boldsymbol{\beta} \cdot \mathbf{X} \tag{1}$$

where $\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_M\}$ is the vector containing the model parameters that weight the impact of the input features containing in the vector $\mathbf{X} = \{x_0, x_1, \ldots, x_M\}$. $p = P(Y = 1)$ is the response of one binary feature Y. We implemented this method using the scikit-learn library in Python (Pedregosa et al., 2011). The inverse regularization length C is set to 1×10^{-5} and the penalty to L2 (Appendix C, Table C1).

2.3.2.2. *Random Forest*

Random forest is a tree-based ensemble machine learning technique that is constructed by a multitude of decision trees. Each tree in the random forests is producing a class prediction and the class with the most votes becomes the model prediction (Breiman, 1999). We implemented the random forest models with the scikit-learn library of Python (Pedregosa et al., 2011), and using the Gini impurity:

$$\text{Gini} = \sum_{i=1}^{C} f_i \, (1 - f_i) \tag{2}$$

where f_i is the frequency of the label i at a node and C is the number of unique labels. We used 1000 trees in the forests with a maximum depth of 2 and the number of features to consider when looking at the best split is the square root of the number of features (Appendix C, Table C1).

2.3.2.3. *Extreme Gradient Boosting - XGBoost*

Boosting is an ensemble technique where new models are added to correct the errors made by pre-existing models. The models are added sequentially until no further improvements is made. The algorithm attributes more weights to the misclassified data to improve the predictions. To minimize the loss function, the algorithm uses gradient descent (Hastie et al., 2009). We use a specific implementation of gradient boosting called Extreme Gradient Boosting (XGBoost) (Chen & Guestrin, 2016). XGBoost is an implementation of a stochastic gradient boosting machine (Friedman, 2001; Friedman, 2002; T. Chen et al., 2015; Chen & Guestrin, 2016). XGBoost can use a variety of learners as its base learners such as linear models or decision trees (Chen & Guestrin, 2016). We use decision trees as the base learners. The gradient boosted equation is formulated as follows:

$$\log \frac{p}{1-p} = F_0 + \sum_{m=1}^{M} r_m(\mathbf{X}) F_m(\mathbf{X}) \tag{3}$$

where m is the iteration index over M maximum iterations. $F_m(\mathbf{X})$ is the current iteration fitted to the previous iterations residuals $r_m(\mathbf{X})$. F_0 is the base estimate.

We implemented XGBoost using the xgboost library in Python (Chen & Guestrin, 2016). The objective is the logistic regression, we define 20,000 boosting learners, trees have a maximum depth of 2, and the minimum child weight is 1 (Appendix C, Table C1).

2.3.3. Evaluation of the Models

We use evaluation metrics based on comparison to random chance. These evaluation metrics include the Area Under The Curve (AUC) (Hanley & McNeil, 1982), the precision and recall, and the F1-score. Each of these metrics is used widely in machine learning studies (Hastie et al., 2009).

The AUC value is within the range [0.5–1.0], where the minimum value represents the performance of a random classifier and the maximum value would correspond to a perfect classifier. A value of 0.5 would suggest no discrimination between surge-type and no surge-type glaciers. AUC values between 0.70 and 0.80 are considered acceptable for classification (Hosmer Jr et al., 2013).

The receiver operating characteristic curve is the true positive rate TP_{rate} against the false positive rate FP_{rate}:

$$TP_{rate} = \frac{TP}{TP + TN} \tag{4}$$

$$FP_{rate} = \frac{FP}{FP + FN} \tag{5}$$

where TP stands for true positive, TN for true negative, FP for false positive and FN for false negative. False positive indicates predictions that have been labeled as surge-type while the true label should have been non-surge-type. The true positives correspond to surge-type glaciers that have been labeled correctly. The same logic applies for false negative and true negative rates.

The performance of a classifier with respect to test data can be assessed by the value of the precision, which is the ratio of correctly predicted positive observations to the total predicted positive observations:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

and the value of the recall, which is the ratio of correctly predicted positive observations to the all observations in an actual class:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

The models have been trained fifty times with different training sets randomly picked and then tested in fifty different testing sets. The corresponding AUC has been calculated for each of these models. The method and results are presented in Appendix B. In addition, a K-Fold cross-validation has been performed to evaluate the model performance stability. A detailed description of the method and the results is discussed in the appendix Appendix B.

2.4. Explanation for Prediction

We examine how the eleven features impact model decision in several ways: (a) we compute the relative feature importances across all models and compare them, (b) we examine two feature importances relevant to XGBoost (gain and weight), the model that has the highest performance, (c) we calculate the Shapely Additive values for the XGBoost model.

We compare the feature importances of three models in a stacked diagram (Figure 8). For each model, the features importance score is calculated and the scores are summed together for the three models. The feature importance score informs on the gain of information a feature gives to the model for classification (a detailed description of the feature importances can be found in Appendix D). For comparison purpose, we normalized all the scores using a min-max normalization. To add more weight on best performing models, the feature importance scores are multiplied by the AUC of each corresponding models. For XGBoost, only the gain scores are taken into account.

Another way to evaluate the feature contributions to the model predictions is to compute the Shapley Additive exPlanations (SHAP) values (Lundberg & Lee, 2017). SHAP values quantify the impact of having a certain value for a given feature in comparison to the prediction the model would have made if that feature had some baseline value. SHAP values allow for (a) a global interpretation of the predictions by analyzing how much each predictor contributes positively or negatively to the target feature; (b) a local interpretation because each observation gets its own SHAP value while most of the traditional feature importance algorithms only show results across an entire class. Based on the value of the features, SHAP analysis allocates a positive or a negative impact on the model output, for example, a high value of a certain feature has a positive impact on the model output meaning that a high value will influence the model toward a high potential of surging.

2.5. Interpolation of a Surge Probability Map

The surge probability is assessed for each discrete centerline point of a glacier using the XGBoost model. Only the XGBoost model is used to produce the map because results show it is the best-fit model (see Section 3 for more details). We average the centerline points probabilities to produce a single probability per glacier centerline. If the average probability along the centerline is under 0.5, the glacier is not considered to be surge-type. If the average probability along the centerline is equal or larger than 0.5 the glacier is considered to be surge-type. The Randolph Glacier Inventory surge-type classes are shown in the map of Figure 1. The average probability per glacier calculated in the present study is shown in the map of Figure 6. We also examine a subset of discretized glacier centerlines in Nordaustlandet Island (Figure 6, inset). This step is useful to show that surge probabilities are varying along the centerline of a glacier, highlighting the potential triggering zone where a surge may develop.

3. Results

3.1. Machine Learning Models Evaluation

All three models (logistic regression, random forest, XGBoost) perform better than random chance (Figure 5a) with testing AUC values ranging from 0.69 to 0.74 (mean AUCs calculated for 50 different training and testing data-sets). XGBoost shows the highest precision (0.85, Figure 5b) and lowest False Positive Rate (0.23, Figure 5d) of all the models. This model demonstrates a lower recall (0.63) than logistic regression (0.68) and random forest (0.69) (Figure 5c). Given these superior fit statistics for XGBoost, we choose this model to calculate the probabilities along glacier centerlines and to produce the surge-type glacier classification map. We perform complementary evaluations of the performance stability of XGBoost (Appendix B, Figures B1 and B2). The AUC is consistent using different training data sets picked randomly and stays stable when we perform the K-Fold cross validation. XGBoost is the best performing model in this study and its performance is stable regardless of the points used to train the models. Because of the stability of the performance, the final results are not computed using the K-Fold cross validation since it will not improve tremendously the performance of the model and we choose to establish the model stability as presented in Appendix B.

3.2. Surge Probability Map of Svalbard Glaciers

Using the XGBoost model, we compute the surge probability of glaciers in Svalbard. The predicted probability map (Figure 6) indicates the presence of surge-type glaciers in the entire archipelago. This map has been computed from averaging the probability of every point along the centerline for each glacier. The map with centerline points can be found in Appendix E (Figure E1). However, preferential zones of surge can be identified in for example, Nordaustlandet island, Torell Land. Other areas, for example, Nordenskköld Land, Andree Land, do not gather a significant number of surge-type glaciers. The XGBoost model classifies 162 glaciers as surge-type out of 981 (see Section 2.1 for more details on the data set). While some glacier centerlines present a uniform probability distribution, some others see their probabilities for surging evolve along the centerline (Figure 6, inset).

In addition to the probability map, we compare our results to the existing Randolph Glacier Inventory classifications for surge-type glaciers. Figure 7a shows the cumulative frequency distribution of probabilities to surge calculated by the XGBoost model. The cumulative frequency distributions of the two classes with low surge potential in the Randolph Glacier Inventory (0: surge not observed, 1: possible surge) appear very similar. The same observation applies for the two classes with high surge potential (2: probable surge, 3: observed surge).

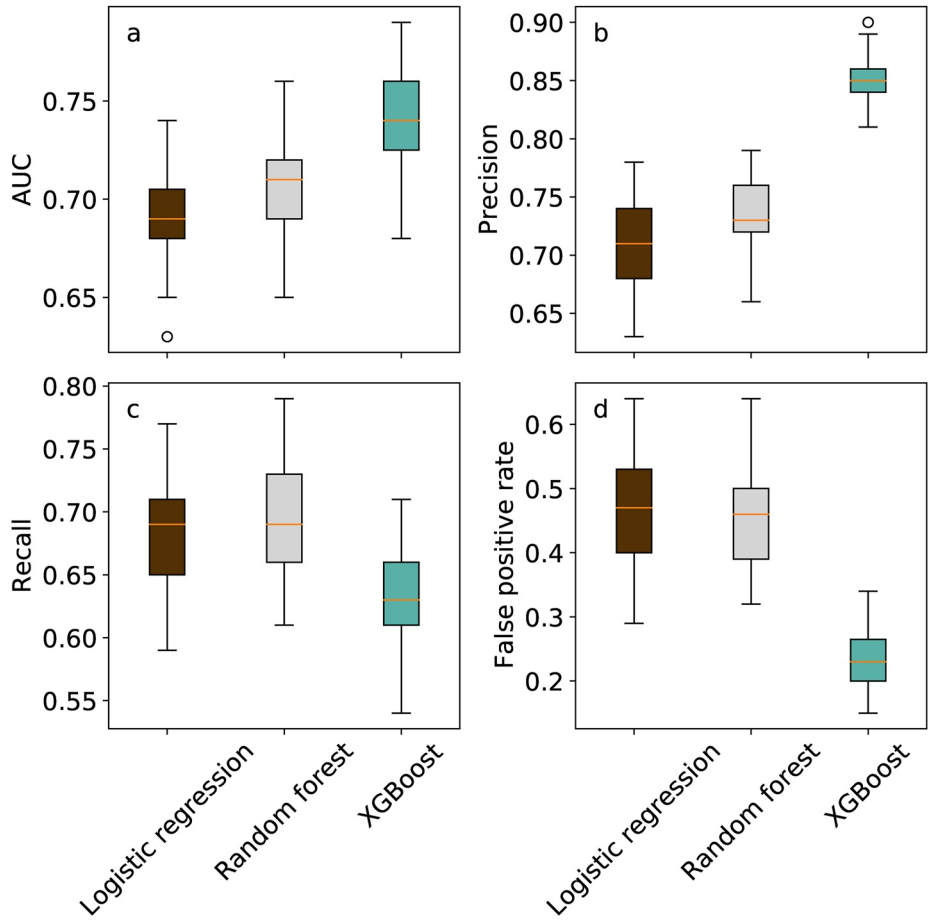

Figure 5. Boxplot representing the (a) Area Under the Curves (AUC), (b) precision score, (c) recall score, and (d) false positive rate for each of the three machine learning models. The scores have been calculated for 50 different training and testing data-sets. The orange line corresponds to the median while the box corresponds to the interquartile range. Both extremes indicates the minimum and the maximum value and the dots indicate the presence of outliers.

These results are supported by the histogram in the inset of Figure 7b which shows two distinct classes, non-surge type and surge-type glaciers. The non surge-type class is however better defined than the surge-type class.

3.3. Importance of Geometrical and Climatic Features

Figure 8 shows the combined importance for each feature used in each model (logistic regression, random forest, XGBoost). For all three models the width, thickness, and surface slope are the most important features explaining most of the models' predictions. The CMB, the width × height (W × H), the surface elevation, the driving stress, and the dummy features do not have a high impact on the model prediction. The runoff, bed elevation, and bed slope explain partially the predictions.

Beyond the comparison of features between each model, we also examine the feature importances for the best-fit model, XGBoost. Figure 9 shows the feature importance scores computed with the gain and weight implementation for the XGBoost model. The width of the glacier adds a considerable amount of information when it is selected on the trees, while the surface slope and the thickness are the features that are selected the most. The thickness, runoff, and the bed elevation add more information than the CMB, W × H, surface elevation, driving stress that are equally not significantly important to assess the surging potential of glaciers in Svalbard. The dummy feature appears in all cases to be the least important feature, as expected.

Using recursive feature elimination, we find that five to six features are needed for the model to reach the highest AUC values (Figure 10). To predict the surging potential of a glacier in Svalbard, the surface and bed slope,

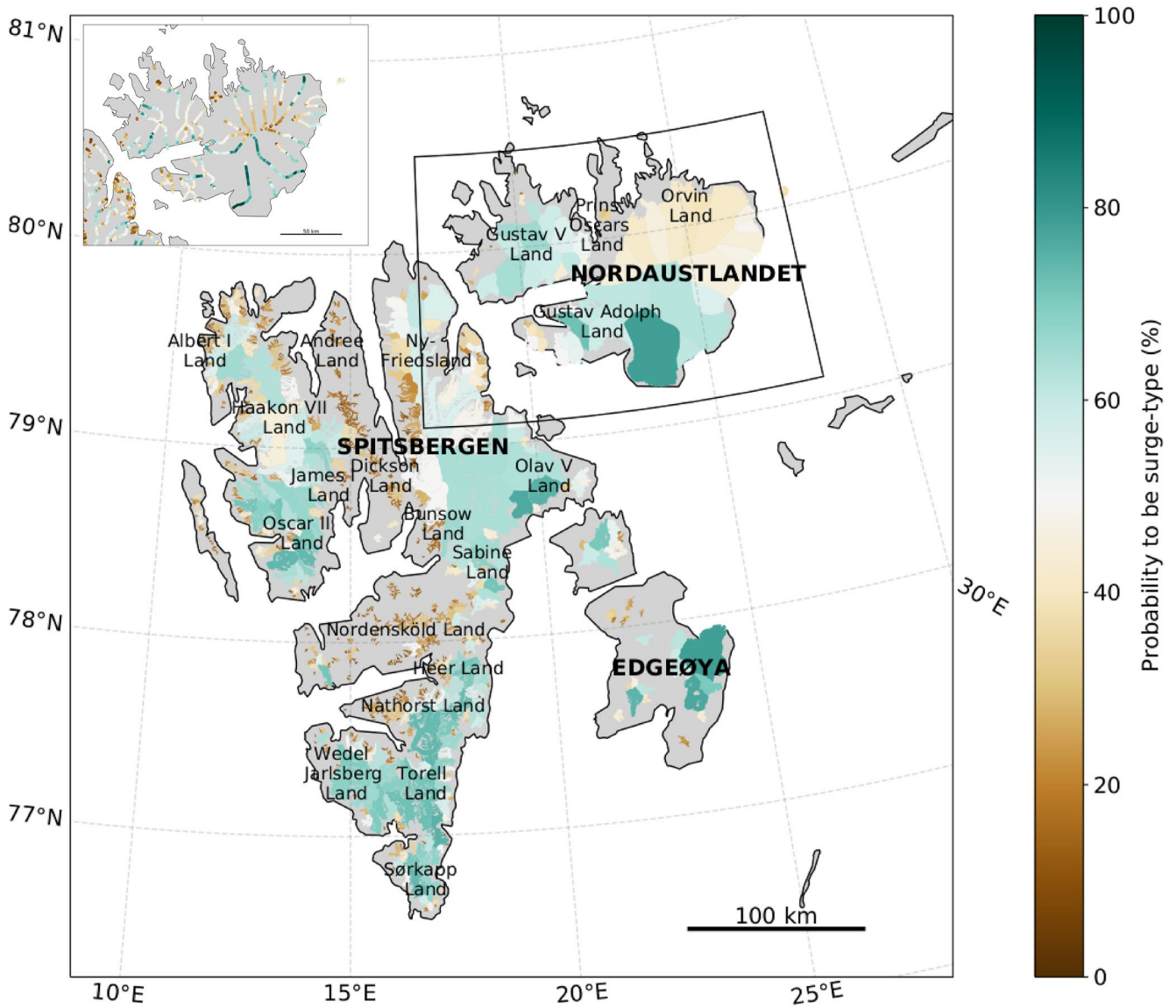

Figure 6. Averaged probability map for each glacier to be classified as surge-type in the XGBoost model. The zoom in the Austfonna ice cap shows how the average probability has been computed. First, a probability is calculated at every point along the centerline of every glacier. Then, we average the probabilities to surge of every point along the centerline to obtain an average surging probability for a given glacier.

thickness, CMB, runoff and width need to be considered. The driving stress, surface elevation and the dummy features do not have a significant impact on the model performance. All the eleven features are however used for the model final decision.

3.4. Feature Values and Local Impact on Prediction

Figure 11 shows the summary of the SHAP value analysis. While some feature values have a clear impact on the model decision, that is, the surface slope, width, others, that is, the bed and surface elevation, do not show a clear relationship between the feature values and the impact on the classification. Higher values of glacier surface slopes, CMB, and in some cases bed elevations all decrease the probability to be classified as a surge-type. Lower values of width, bed elevation, surface elevation, thickness, run off, bed slope, and W × H also decrease the probability for a glacier to be classified as surge-type. In contrast, high values of width, in some cases bed elevation, surface elevation, thickness, and W × H increase the probability for a glacier to be surge-type. Low values of surface slope and CMB are likely to increase the probability of a glacier to be classified as surge-type. Some

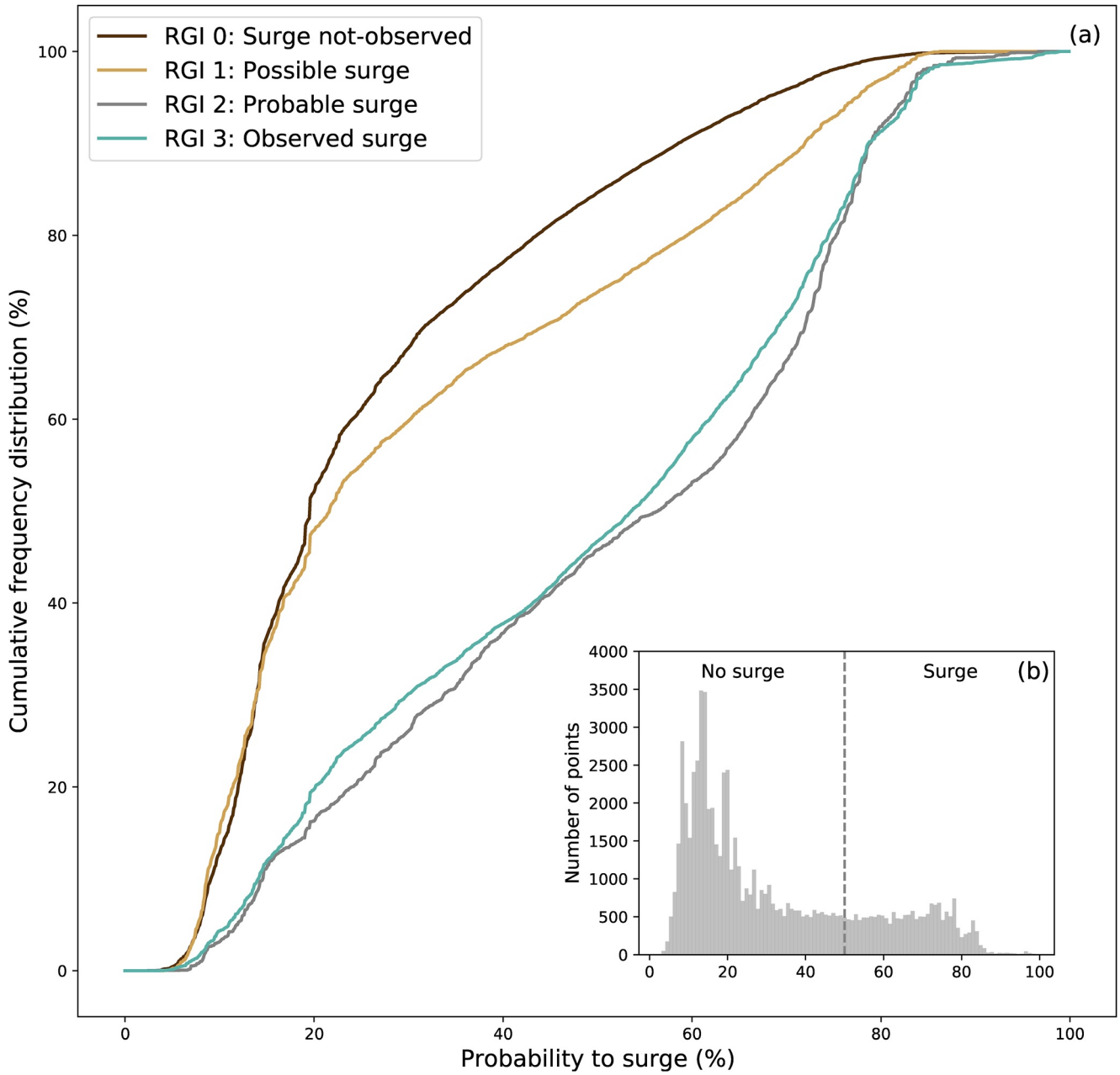

Figure 7. Cumulative frequency distribution for a glacier to be surge-type labeled by the classes defined in the Randolph Glacier Inventory. The inset shows the distribution of the probabilities. The vertical line indicates a 50% probability from which we separate surge-type from non surge-type glaciers.

features do not show clear separation between the values and the corresponding impact on the model: the bed and surface elevation, the driving stress and the dummy feature (this should be expected for the dummy feature). To summarize, a thicker and wider glacier with a low surface slope, CMB, and high runoff has more potential to be classified as surge-type.

4. Discussion

4.1. Evaluation of the Surge-Type Classification Framework

We calculate surging probabilities for glaciers in Svalbard after an evaluation of the best performing model, XGBoost. The discretization of the glaciers along centerlines increases the number of data points used to train

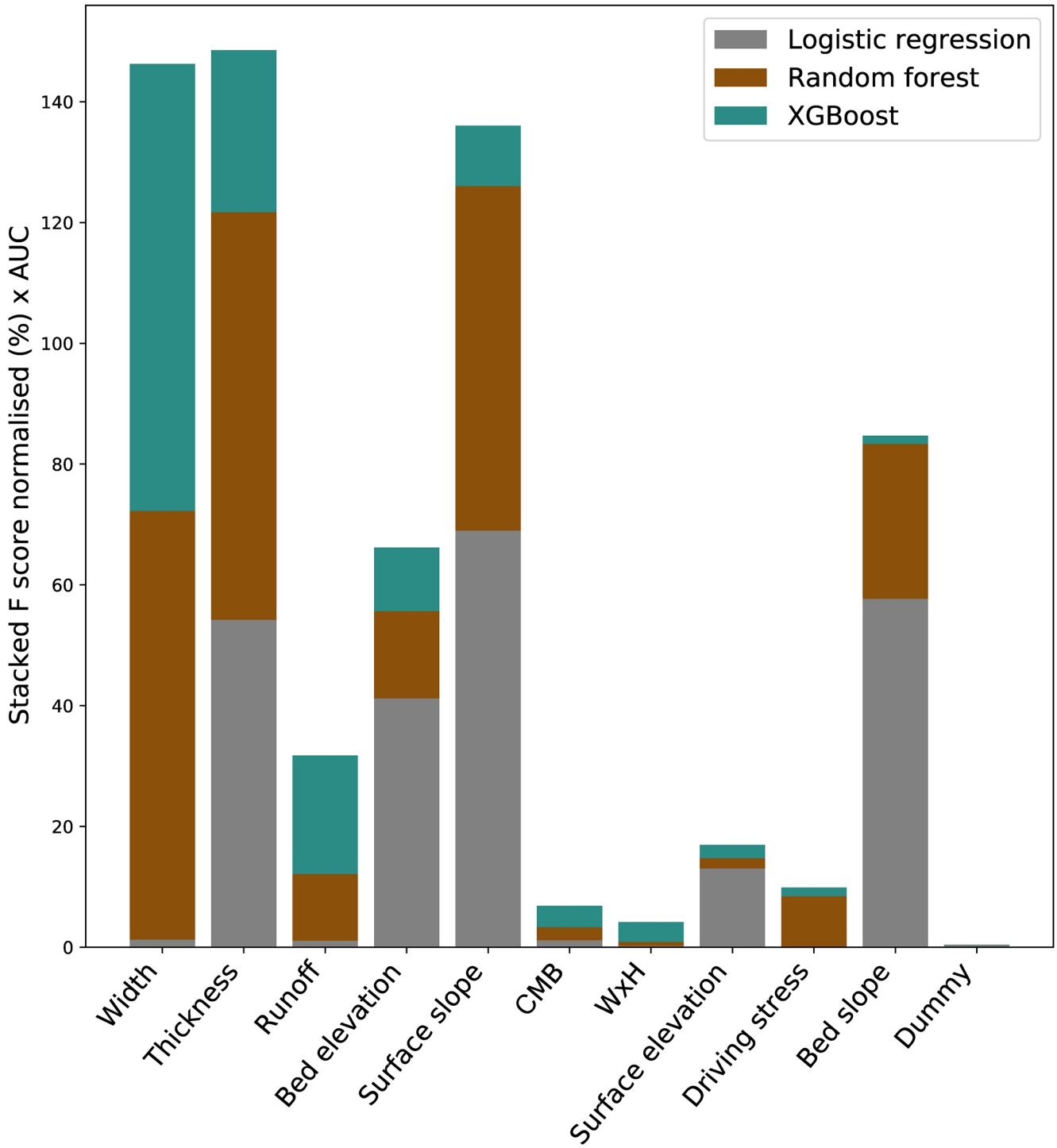

Figure 8. Feature importances for the logistic regression, random forest and XGBoost models stacked together. The F-score of each features has been normalized and multiplied by the AUC value. The features are organized from left to right from the most to the least important, according to XGBoost score.

and test models (981 glaciers corresponding to 97,140 points). The availability of data provides better insights into the relationships between features used as input for machine learning algorithms (Halevy et al., 2009). As explored in other fields (e.g., Fatichi et al. (2016)), complex natural systems cannot always be simplified using integrated features. For example, the glacier surface slope can vary along the centerline, which will change the driving stress. Averaging the slope in this case would misinform the model on changes that could impact model

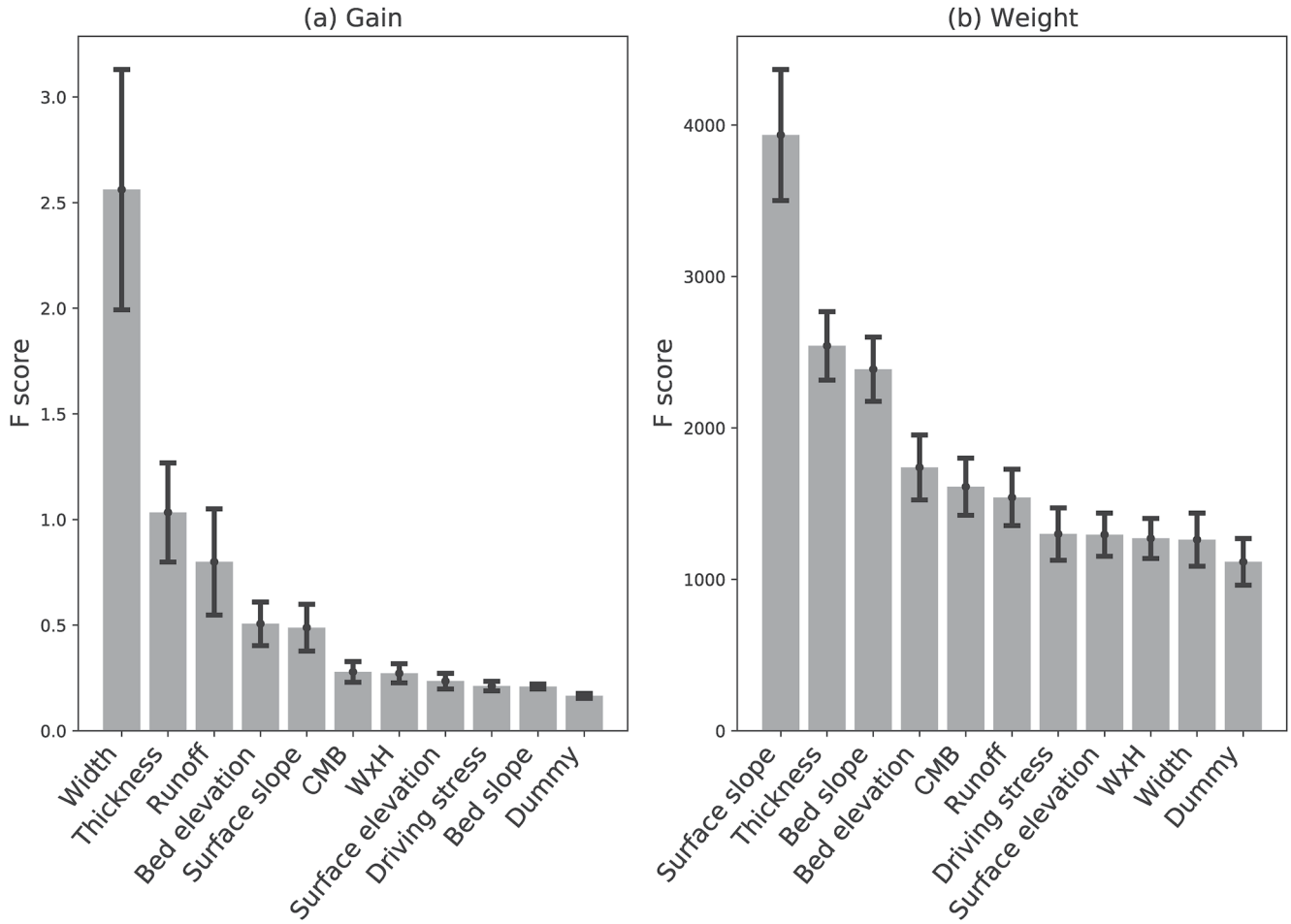

Figure 9. Feature importance of XGBoost model: (a) gain, (b) weight.

classification. In addition, discretized features enhance the spatial variability of the glaciers. A longer glacier will be constituted by a higher number of points along its centerline than a smaller glacier.

The framework presented here uses a model comparison and an evaluation method grounded in the best machine learning practices (Hastie et al., 2009). Previous statistical studies aiming at understanding surging glaciers used only one model, that is, univariate or multivariate regression (Barrand & Murray, 2006; Clarke, 1991; Clarke et al., 1986; Hamilton & Dowdeswell, 1996; Jiskoot et al., 2000) or maximum entropy (Sevestre & Benn, 2015). To our knowledge, this is the first study that compares the performances of several machine learning models to classify surge-type glaciers. Comparing models provides more confidence on the results of the best one. Numerical modeling studies have compared several models to determine the most accurate one for a defined task (e.g. (Hock et al., 2019),). The approach we take is similar. While a single model, XGBoost, is used in the final production of the classification map, we rely on the plurality of model results to support our understanding of what the models learned (e.g., Figure 8).

As with every machine learning model, the performance of XGBoost is tied to the quality of the input features. We use mostly features resulting from numerical simulations, and therefore, by nature, containing uncertainties. Pelt et al. (2019) highlights that the CMB and runoff values contain uncertainties due to (a) elevation offsets, (b) the input climatic parameters used for the simulations and (c) the modeling itself. Fürst et al. (2018) produced an error estimate map for the thickness computation that ranges from 1 to 500m. XGBoost is then trained with features that are not considered ground-truth (as opposed to field measurements). In addition, the error associated with the modeling data is unknown. Therefore, the range of resulting AUC values of XGBoost is controlled by possible bias in the input features used for training.

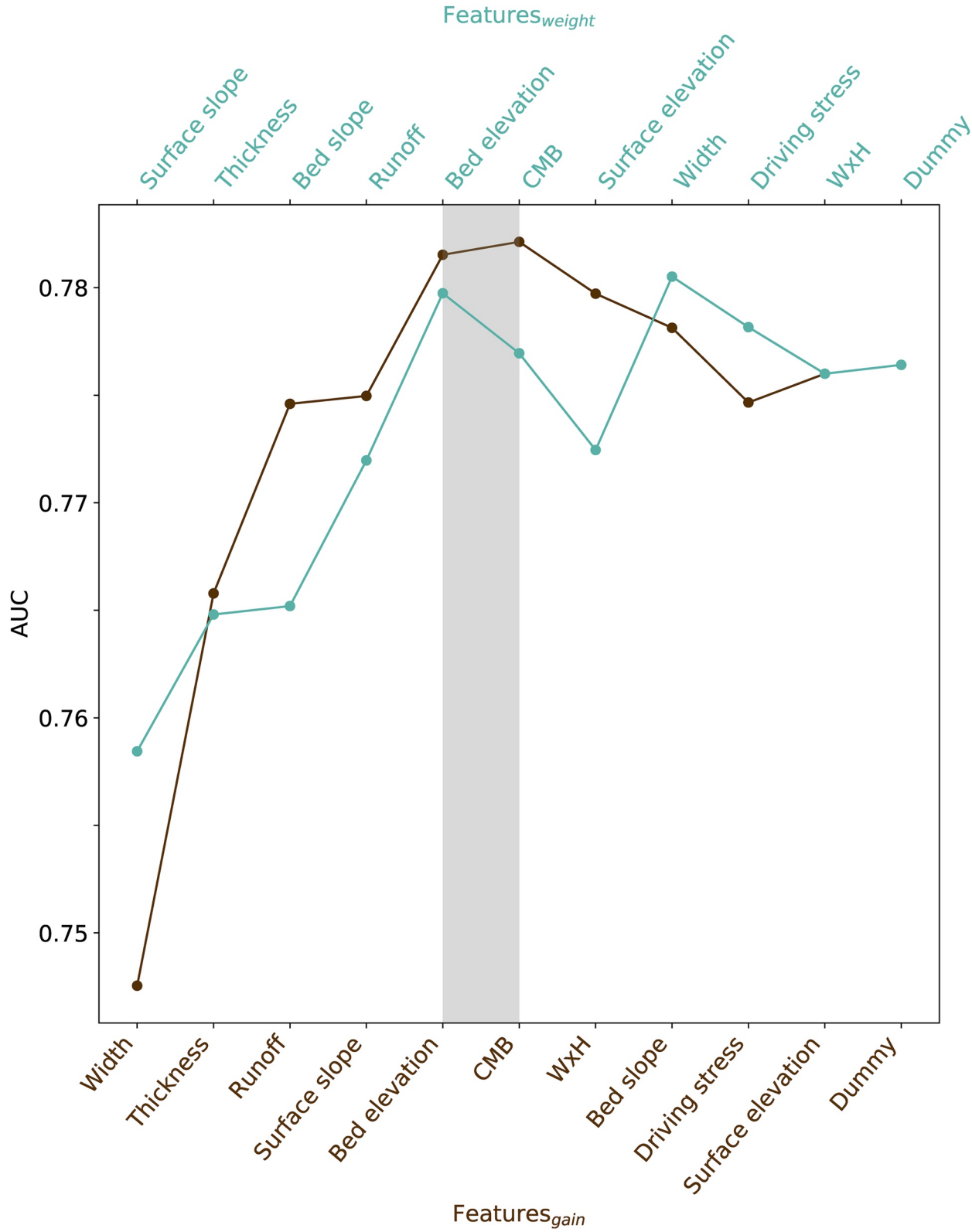

Figure 10. Results of recursive feature elimination show that four to five features explain most of the gain of information in the classification of surge-type glaciers. The number of features is added according to their order in the feature importance.

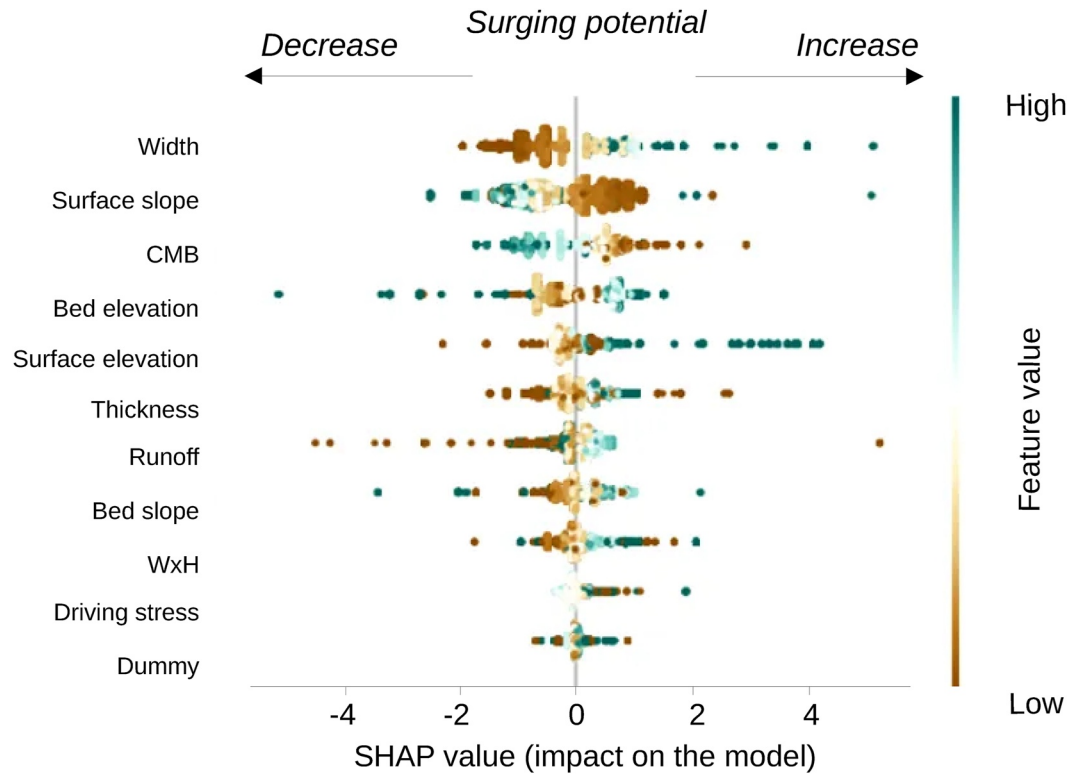

Figure 11. Summary of the SHAP values for every features at each glacier centerline point. The *x* position of each dot is the impact of the feature on the prediction of the model and the color of the dot represents the contribution of that value at each point.

In addition, the spatial resolution of individual features differs, for example, the bed elevation has been computed on a 100m resolution whereas the runoff and CMB have been calculated on a 1 km resolution grid. The features used in the models represent as well a snapshot for a particular point in time and may therefore represent different stages in the surge-cycle. While the topographic data (Fürst et al., 2018) represent the state of Svalbard in 2010, the climatic data (Pelt et al., 2019) represents the year 2018. Although we want to capture which features could cause a transient behaviour while using a snapshot in time, we consider that there are no better data available since simulating surges in real glacier geometry represents a real challenge. The current framework does not inform on the timing of the surging glacier, for example, pre-surge, post-surge. To do so, time series should be included in the framework by adding complementary types of data for example, remote sensing products. Using time series could potentially highlights an increase in probability when a glacier becomes closer to the development of a surge in the triggering zone where the surge will start to develop. However, the probability for triggering zone should always be higher than 50% since the model bases its classification on the potential of the glacier to be surge-type due to its geometrical and climatic configuration.

4.2. Feature Importance Informs on Theories of Glacier Surging

The important features in our models are the glacier width, the ice thickness, the surface and bed slopes, the runoff, and the CMB to a smaller extent (Figure 2). The width and the thickness have been shown to be important in previous statistical studies (Barrand & Murray, 2006; Clarke, 1991; Jiskoot et al., 2003) together with the surface slope (Jiskoot et al., 1998, 2000, 2003; Sevestre & Benn, 2015). Although XGBoost models predict that lower slope will drive the prediction toward increasing the probability for the point to be surge-type as in Sevestre and Benn (2015), Jiskoot et al. (1998) found the opposite. The surface and bed slopes, and the ice thickness are features controlling the dynamic of a glacier through the hydraulic gradient and the driving stress. Both are known to play a crucial role in surging theories (Benn et al., 2019; Fowler, 1989; Kamb, 1987; Thøgersen et al., 2019). Although the features controlling the surge classification are in good agreement with previous statistical studies,

the features we use in our model rely on more recent observations or modeling studies. In addition, by discretizing the features along the centerlines of the glaciers, we significantly increase the number of points, permitting a more robust statistical analysis. Thøgersen et al. (2021) highlight that in the context of a velocity weakening regime, the friction along the glacier margins is less important with an increasing glacier width. Therefore, wider glaciers should be more likely to be of surge-type which is in good agreement with the SHAP summary result (Figure 11).

To a smaller extent, the CMB is also influencing the classification. However, we are not considering that this feature is important on assessing the surge probability for glaciers due to the negligible increase of the AUC during the recursive feature elimination (Figure 10), as found in Jiskoot et al. (2000). The CMB is highly correlated to other features in the model, for example, the surface elevation and the runoff, meaning that the effect of the CMB is likely captured already into other features. However, in the interior parts of Svalbard, glaciers in drier areas show lower probabilities to be surge-type as opposed to the higher probabilities observed on the coast, in areas with more precipitation. A large glacier has a large accumulation area, receiving more precipitation than a small glacier with a small accumulation basin. Thus, the size of glaciers depends as well on the amount of precipitation they receive. Therefore, the geometrical features already account for part of the climatic influence. The relationship between the climatic and topographic variables is learned by the model and is specific to Svalbard for the trained model used in this study.

In Svalbard, we expect that climatic features are not playing a central role on the prediction compared to other regions in the world, because the climate is relatively homogeneous within the archipelago.

If more observations at the interface between the ice and the bed would become available, they could be incorporated directly into our framework, helping to assess the underlying physical processes leading to glacier surges. The framework allows to incorporate till properties if available. The change in rheological parameters, that is, the permeability and porosity of the sediment, can be responsible for a surge enhancement (Minchew & Meyer, 2020).

The complete framework develops in the present study can be directly used in other regions of the world. However, the models have been trained on Svalbard and show good performance scores. Applying the model to different regions may require a re-training. Indeed, the model in Svalbard has learned specific relation between the features that might be valid in Svalbard but different in other regions of the world. For the trained model to be directly tested in other regions of the world, every climatic-related features should be removed from the training data set in Svalbard, for example, the surface elevation which is a proxy for the mass balance and so is directly linked to the runoff as indicated previously. Due to the nature of these models, we cannot directly infer the predictive power of the topographic against the climatic features that implies a limitation for the trained-model to be directly applicable in another region.

4.3. Quantification of Surging Probabilities

To our knowledge, we produce the first map aiming at quantifying the surge probability of glaciers. The map together with the associated probabilities add new information to the Randolph Glacier Inventory surging classes. Beyond the previous binary distinction between surge-type or non surge-type glaciers, our approach quantifies these classes along a continuous scale with robust statistical methods. We propose that the four qualitative classes in Svalbard can be combined into two statistically informed classes: glaciers that have a probability to surge equal or larger than 50% can be classified as surge-type, and glaciers that present a probability lower than 50% can be classified as non-surge-type.

Our results suggest that some glaciers are misclassified in the Randolph Glacier Inventory. The glaciers listed in Table 1 have a probability higher than 50% to be surge-type in our model, while in the Randolph Glacier Inventory they are currently labeled as Not observed, Possible or Probable surge. Recent field observations have shown that all these glaciers have been seen surging, confirming the high probabilities computed in our model.

In the present study, we are averaging the probability of every points along the flowline to calculate a glacier-wide probability to be surge-type. As seen in Figure 6, some glaciers have a high probability to be surge-type on their low altitude range while other parts of the glacier have a very low probability to surge, for example, in Austfonna ice cap. In that case, to average the probability along the centerline might lead to a mis-classification of the glacier due to the averaging procedure. Further studies should be conducted on these glaciers to understand how the local

Table 1
Comparison of Several Glaciers Where Surge has Been Observed, Their Corresponding Label in the Randolph Glacier Inventory (RGI) Classification, and the Probability Estimates of Our XGBoost Model

RGIId	Name	RGI 6.0 (classes)	Probability XGBoost	Reference
RGI60–07.00276	Arnesenbreen	Possible	67%	Leclercq et al. (2021)
RGI60–07.00296	Strongbreen	Probable	72%	Leclercq et al. (2021)
RGI60–07.00440	Svalisbreen	Not Observed	64%	Leclercq et al. (2021)
RGI60–07.00241	Penckbreen	Possible	65%	Leclercq et al. (2021)
RGI60–07.00501	Aavatsmarkbreen	Possible	70%	Luckman et al. (2015)
RGI60–07.00296	Morsnevbreen	Probable	72%	Benn et al. (2019)
RGI60–07.00027	Austfonna Basin 3	Probable	71%	Schellenberger et al. (2017)

predictions are made for each point and then unravel if the evolving probability along the centerline corresponds to for example, partial surge-type glaciers and triggering zones where the surge may develop. Here, we consider that the points that have a high probability for surging correspond to zone where an instability can be triggered due to the geometrical and climatic settings. Going beyond this conclusion would require more investigations on the model prediction.

Furthermore, to average the probability along the centerline is a very basic way on representing the overall glacier potential for surging. More sophisticated methods should be investigated to represent in the best way the potential for a glacier to be surge-type for example, meta-learning techniques could be considered but would require a prior extensive data analysis.

5. Conclusions and Perspectives

We present a framework based on machine learning models as well as a newly combined database to perform probabilistic glacier hazard mapping. The framework involves discretizing features along glacier centerlines. The most important features that explain glacier surge, that is, the width, the thickness, the runoff, the surface and bed slopes and the CMB are aligned with theories of glacier surge. Our framework allows a quantitative assessment of the surge potential of glaciers in Svalbard, that complements the previously established classification in the Randolph Glacier Inventory. Several new glaciers have been identified as surging glaciers with our model and confirmed by independent observations, which strengthens the robustness of our approach. The framework includes the comparison between different machine learning models and presents an extensive evaluation of the best model stability.

To complement theories of glacier surge, new features might be added to our framework, that is, the thickness of the underlying till, the internal reflection horizons imaging the transition between cold and temperate ice, the basal temperature and geothermal gradient, and the lithology of the underlying bed. Monitoring efforts are encouraged to be pursued toward this goal (Figure 2). The present study represents a snapshot in time but future studies may include multi-temporal data of certain features to enable inclusion of a time component. Finally, future work should focus on the development of an accurate method to compute a glacier-wide probability from the centerline probabilities.

Our method to compute probabilistic glacier hazard mapping based on machine learning methods and a discretized database could also be applied to other regions of the world and/or adapted to other field (e.g., landslides and earthquakes dynamics).

Appendix A: Proportion Between the Training and Testing Data Sets

The more data is used to train the model, the better the model will become. We performed an analysis to choose the best split in the database between the points that will be used as training data and testing data. We are looking for the split that will provide the best performance of the model without overfitting. As seen in Figure A1, a split at 40% for the training data set gives a relatively high AUC value (0.77). However, a considerable proportion of ground-truth data are then dropped to train the model. A split at 70% for the training data set gives a relatively

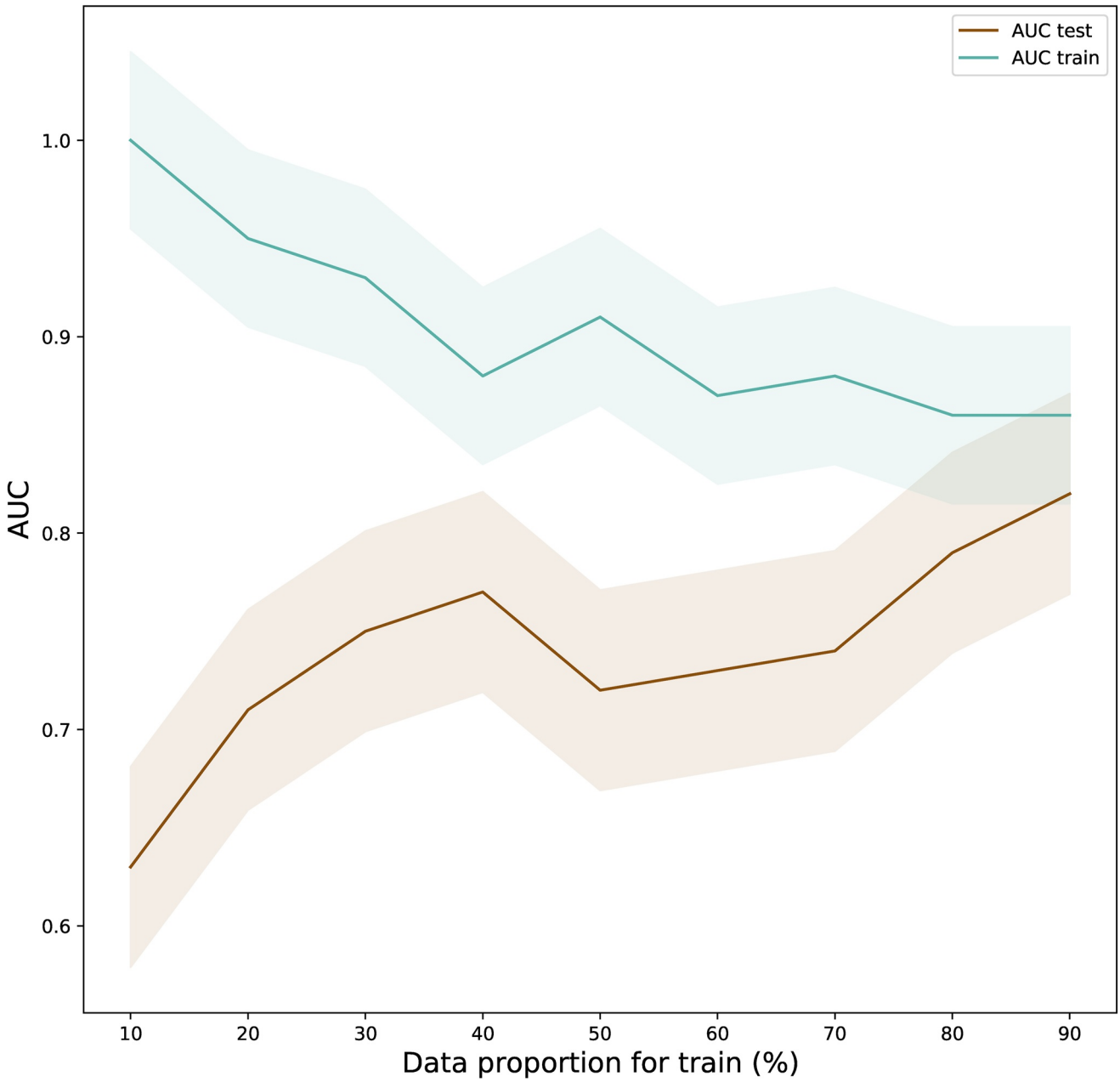

Figure A1. Evaluation of the proportion of the database to use to train the model. We trained ten models using different proportion between the training and the testing set ranging from 10% to 90%. For each models, we calculate the AUC for the training data set (blue curve) and the testing data set (brown curve).

high AUC (0.74) and use more data to train the model. If a bigger proportion of the database is used for training, the model is more sensitive to overfitting. We observe in that case that the testing AUC values might be higher that the training AUC values. For this study, we are then using a 70%–30% split for the training/testing data sets which represents a good compromise to reach good model performance and avoid overfitting.

Appendix B: Evaluation of the Model Stability

B1. Bootstrap Simulations

We performed fifty bootstrap simulations to study the consistency of the model performance among the fifty simulations. The model is trained by picking randomly the points that will be used for training. We set the

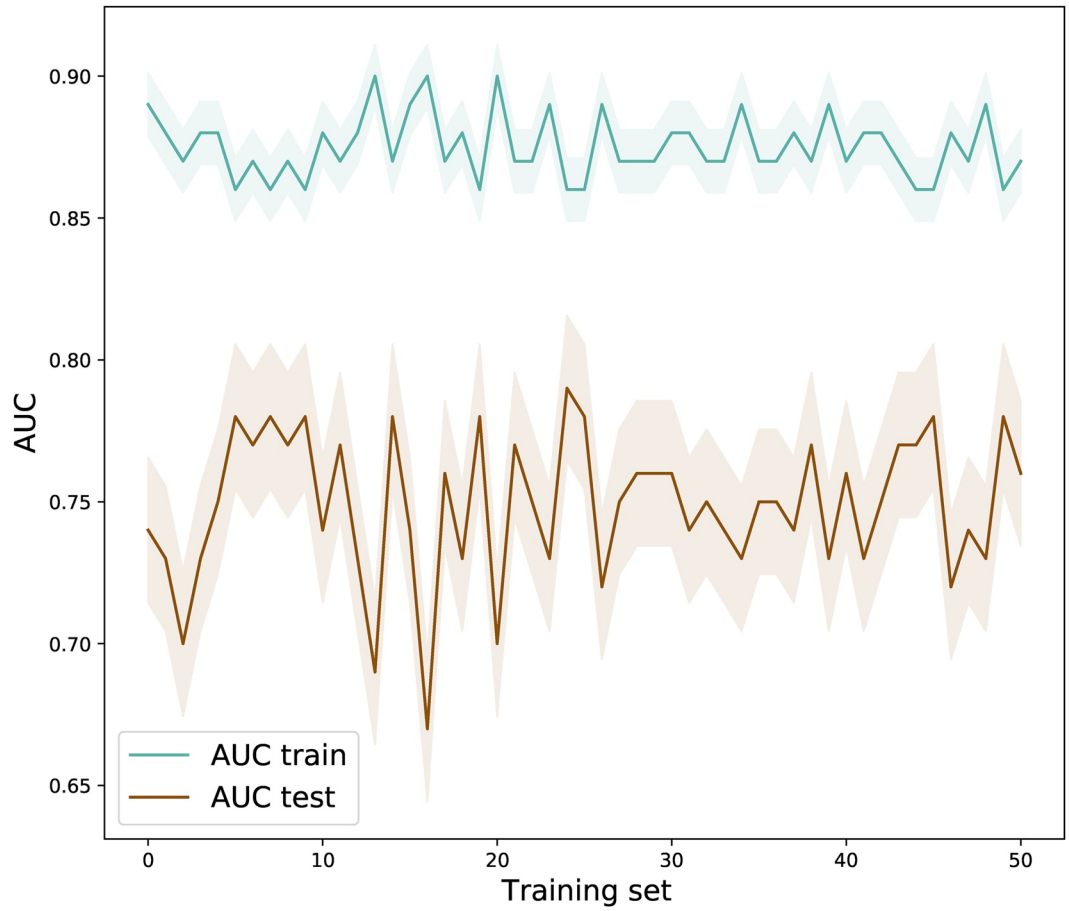

Figure B1. Evaluation of the impact of selecting fifty different training data sets on the model performance. We run the model fifty times with fifty randomly picked training data sets and computed the trained (blue) and tested (brown curve) AUCs for each of these models.

proportion for the training set at 70% of the database. To ensure that the model performance is not dependent of the training data sets that have been picked, we are training fifty different models on fifty different, randomly picked training data set and calculate the AUC for both the training data set and the testing data set. As display on Figure B1, the model remains stable with a very consistent AUC in between all the training sets.

B2. K-Fold Cross Validation

Cross-validation is a common practice in machine learning to evaluate the model performance and its performance stability. We performed a ten K-Fold cross validation for XGBoost model and calculated the corresponding AUC. In Figure B2, we observe that the model performance is relatively stable. XGBoost is robust and the performances are not affected by the quality of the training data set.

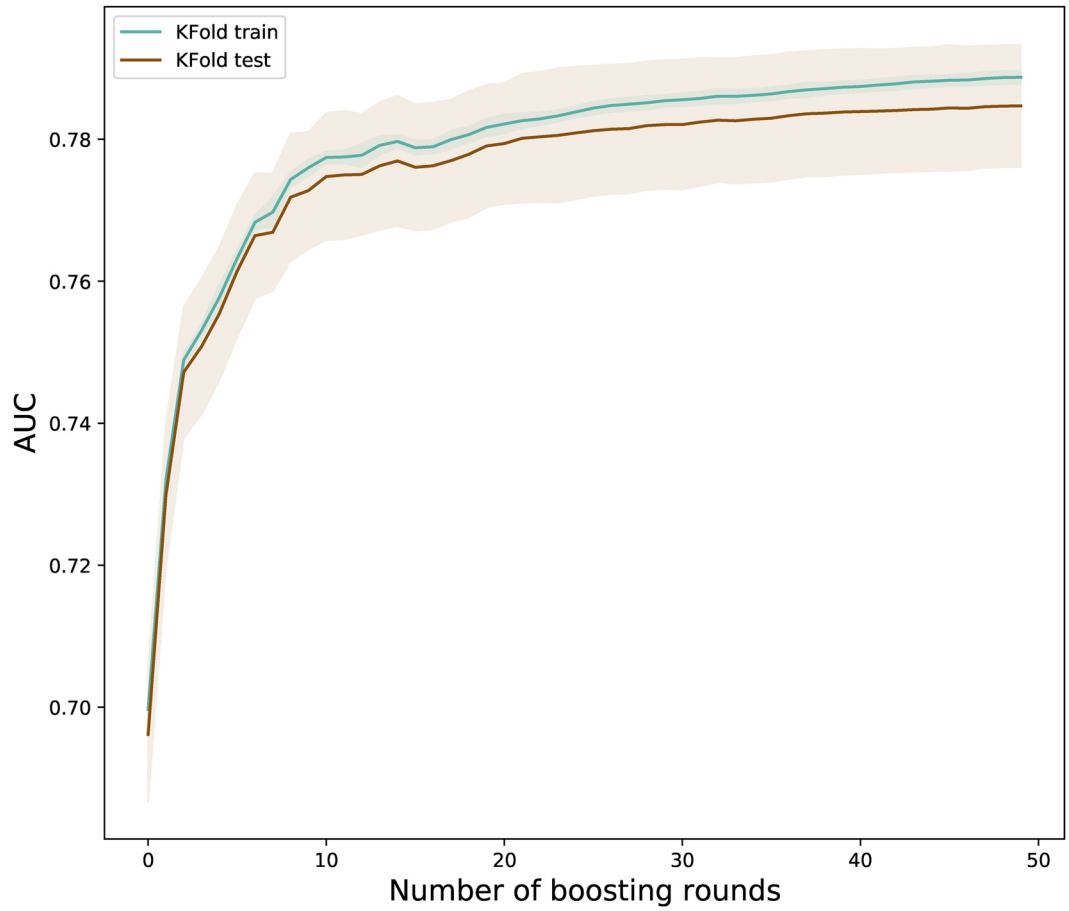

Figure B2. Evaluation of the model performances using a K-Fold cross-validation. We train the model using different 10 folds and calculated the corresponding AUC. The model uses 50 boosting rounds.

Appendix C: Exhaustive Grid search for Hyper Parameters Tuning

Hyperparameters define the structure of a model. For example, the hyperparameters of a random forest model would describe how many trees to grow, the depth of those trees, and the algorithm to use to grow the trees. Hyperparameters are separate from the data used to train the model and their values cannot be estimated from the data while they need to be set before the learning process begins. To optimize the hyperparameters we used the exhaustive grid search method. It considers several possibilities for each hyperparameters and try every combination possible before choosing the combination that returns a lower error score. This method should be guided

Table C1
Results of the Exhaustive Grid Search for Hyperparameter Tuning and the Corresponding Area Under the Curve (AUC)

Method	Hyperparameter	Best value	AUC
Logistic regression	C	1×10^{-5}	0.70
	Penalty	L2	
Random forest	Number of trees	1000	0.71
	Maximum depth	2	
XGBoost	Maximum depth	2	0.75
	Minimum child weight	1	

by cross-validation on the training set. The exhaustive grid search is run using the scikit-learn library of Python (Pedregosa et al., 2011). Table C1 displays the three different models evaluated in our study and the hyperparameters that have been selected by the exhaustive grid search. The best values for these hyperparameter are shown in the last column.

Appendix D: Detailed Description of Feature Importances

To better understand how the surge probabilities are calculated and can be correlated to surging mechanisms, the relative contribution of each feature can be analyzed by calculating the feature importance. Each one of the three machine learning models we used calculate differently the feature importances. We detail mostly how feature importance is implemented in XGBoost since this model performs the best for the assessment of surging potential for Svalbard glaciers. In XGBoost, after the trees are built, the model reports directly the feature importance instead of the coefficient values commonly reported in logistic regression. Each time a feature is used in a tree, the tree will split optimally to a certain location to increase the accuracy, so-called the gain. For each specific feature, the feature importance corresponds to the average gain across all decision making. Different implementation are proposed to estimate the contribution of each feature in the model decision. We focused on the gain and weight. The gain is the improvement in accuracy brought by a feature to the branches. A higher value implies that the feature is more important for generating a prediction. The weight corresponds to the number of times a feature is used to split the data across the tree. To assess how many features are needed to maximize the AUC, we performed a recursive feature elimination. Initially, the model is trained and tested with the features that had the highest feature importance score. Then, at every iteration, the model is trained and tested adding one more feature and this process is repeated until the maximum number of features is reached. The AUC is saved at every iteration and the recursive feature elimination is performed using the scores computed with the weight and the gain method.

Appendix E: Centerline Probability

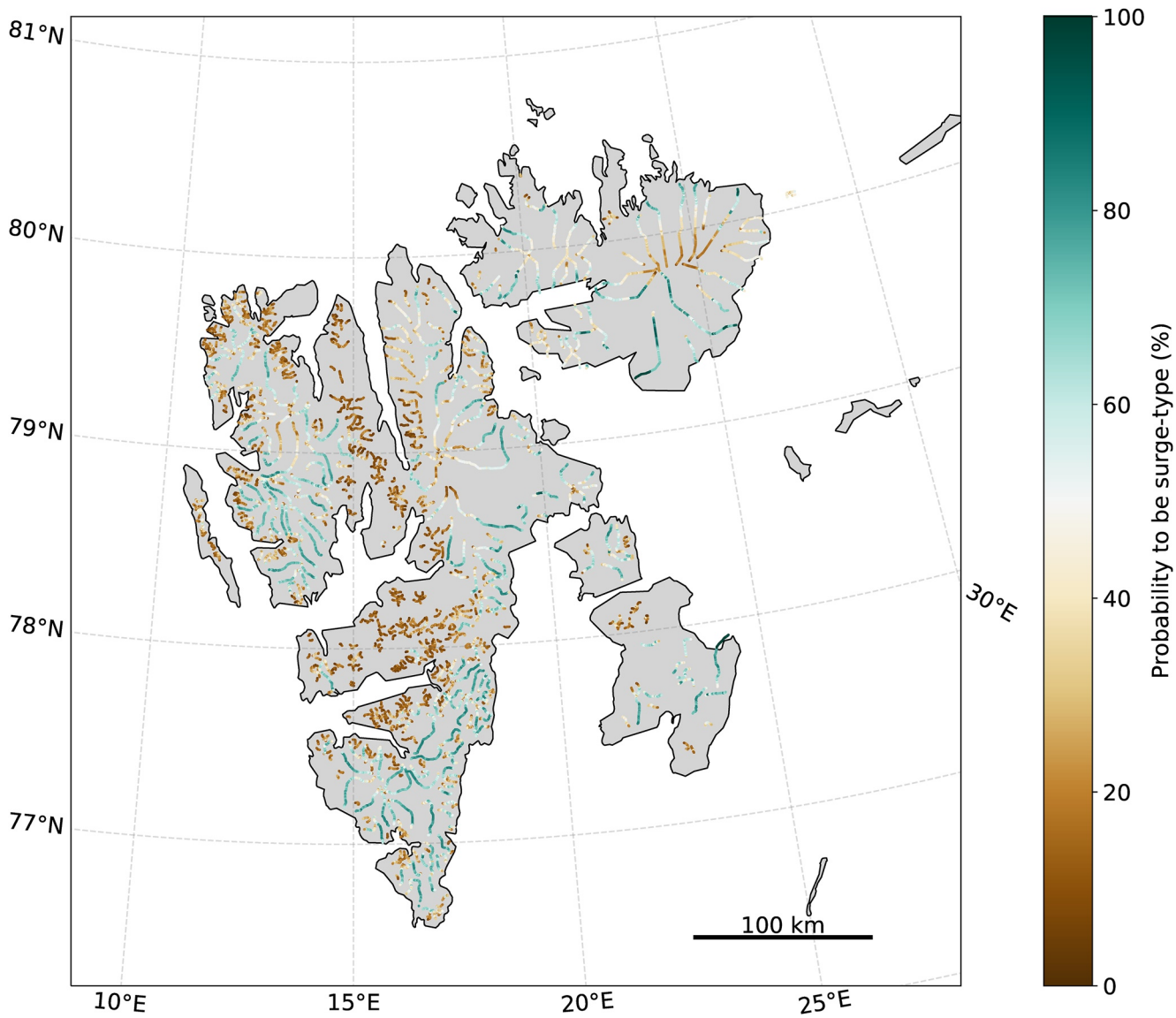

Figure E1. Probability map for each point of glacier centerlines to be classified as surge-type in the XGBoost model.

Data Availability Statement

The data and code are available in the repository https://doi.org/10.5281/zenodo.5657088 (Bouchayer, 2021).

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Acknowledgments
This project has received support from the Norwegian Agency for International Cooperation and Quality Enhancement in Higher Education (DIKU), which supports the Center for Computing in Science Education, from the Research Council of Norway through the project MAMMAMIA (Grant No. 301837) and from the Faculty of Mathematics and Natural Sciences at the University of Oslo through the strategic research initiative EarthFlows. We thank Dr. PiM Lefeuvre and Dr. Simon Filhol for constructive insights. The authors declare no conflict of interest. The authors thank reviewer Martin Truffer, one anonymous reviewer, the Associate Editor and the Editor, Olga Sergienko, for thorough reviews that improved the manuscript.

References

Barrand, N. E., & Murray, T. (2006). Multivariate controls on the incidence of glacier surging in the karakoram himalaya. *Arctic Antarctic and Alpine Research*, *38*(4), 489–498. https://doi.org/10.1657/1523-0430(2006)38[489:mcotio]2.0.co;2

Bazai, N. A., Cui, P., Carling, P. A., Wang, H., Hassan, J., Liu, D., et al. (2021). Increasing glacial lake outburst flood hazard in response to surge glaciers in the karakoram. *Earth-Science Reviews*, *212*, 103432. https://doi.org/10.1016/j.earscirev.2020.103432

Benn, D., & Evans, D. J. (2014). *Glaciers and glaciation*. Routledge.

Benn, D., Fowler, A. C., Hewitt, I., & Sevestre, H. (2019). A general theory of glacier surges. *Journal of Glaciology*, *65*(253), 701–716. https://doi.org/10.1017/jog.2019.62

Bhambri, R., Hewitt, K., Kawishwar, P., & Pratap, B. (2017). Surge-type and surge-modified glaciers in the karakoram. *Scientific Reports*, *7*(1), 1–14. https://doi.org/10.1038/s41598-017-15473-8

Björnsson, H., Pálsson, F., Sigurđsson, O., & Flowers, G. E. (2003). Surges of glaciers in Iceland. *Annals of Glaciology*, *36*, 82–90. https://doi.org/10.3189/172756403781816365

Bouchayer, C. (2021). Colinebouch/glacier_ML: Glacier surging potential repo. *Zenodo*. https://doi.org/10.5281/zenodo.5657088

Breiman, L. (1999). Prediction games and arcing algorithms. *Neural Computation*, *11*(7), 1493–1517. https://doi.org/10.1162/089976699300016106

Chen, & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Chen, & Jeong, J. C. (2007). Enhanced recursive feature elimination. In *Sixth international conference on machine learning and applications* (pp. 429–435).

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

Clarke, G. K. (1991). Length, width and slope influences on glacier surging. *Journal of Glaciology*, *37*(126), 236–246. https://doi.org/10.3189/s0022143000007255

Clarke, G. K., Schmok, J. P., Ommanney, C. S. L., & Collins, S. G. (1986). Characteristics of surge-type glaciers. *Journal of Geophysical Research*, *91*(B7), 7165–7180. https://doi.org/10.1029/JB091iB07p07165

Cuffey, K. M., & Paterson, W. S. B. (2010). *The physics of glaciers*. Academic Press. https://doi.org/10.1016/c2009-0-14802-x

Dieterich, J. H. (1992). Earthquake nucleation on faults with rate-and state-dependent strength. *Tectonophysics*, *211*(1), 115–134. https://doi.org/10.1016/0040-1951(92)90055-B

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., et al. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, *537*, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Fowler, A. (1989). A mathematical analysis of glacier surges. *SIAM Journal on Applied Mathematics*, *49*(1), 246–263. https://doi.org/10.1137/0149015

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/s0167-9473(01)00065-2

Fürst, J. J., Navarro, F., Gillet-Chaulet, F., Huss, M., Moholdt, G., Fettweis, X., et al. (2018). The ice-free topography of svalbard. *Geophysical Research Letters*, *45*(21), 11–760. https://doi.org/10.1029/2018gl079734

Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering*, *2*(4), 42–47.

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, *24*(2), 8–12. https://doi.org/10.1109/mis.2009.36

Hamilton, G. S., & Dowdeswell, J. A. (1996). Controls on glacier surging in svalbard. *Journal of Glaciology*, *42*(140), 157–168. https://doi.org/10.3189/s0022143000030616

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.

Hock, R., Bliss, A., Marzeion, B., Giesen, R. H., Hirabayashi, Y., Huss, M., et al. (2019). Glaciermip–a model intercomparison of global-scale glacier mass-balance models and projections. *Journal of Glaciology*, *65*(251), 453–467. https://doi.org/10.1017/jog.2019.22

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (Vol. 398). John Wiley & Sons.

Jiskoot, H., Boyle, P., & Murray, T. (1998). The incidence of glacier surging in svalbard: Evidence from multivariate statistics. *Computers & Geosciences*, *24*(4), 387–399. https://doi.org/10.1016/s0098-3004(98)00033-8

Jiskoot, H., Murray, T., & Boyle, P. (2000). Controls on the distribution of surge-type glaciers in svalbard. *Journal of Glaciology*, *46*(154), 412–422. https://doi.org/10.3189/172756500781833115

Jiskoot, H., Murray, T., & Luckman, A. (2003). Surge potential and drainage-basin characteristics in east Greenland. *Annals of Glaciology*, *36*, 142–148. https://doi.org/10.3189/172756403781816220

Kamb, B. (1987). Glacier surge mechanism based on linked cavity configuration of the basal water conduit system. *Journal of Geophysical Research*, *92*(B9), 9083–9100. https://doi.org/10.1029/jb092ib09p09083

Kienholz, C., Rich, J., Arendt, A., & Hock, R. (2014). A new method for deriving glacier centerlines applied to glaciers in Alaska and northwest Canada. *The Cryosphere*, *8*(2), 503–519. https://doi.org/10.5194/tc-8-503-2014

Leclercq, P. W., Kääb, A., & Altena, B. (2021). Brief communication: Detection of glacier surge activity using cloud computing of sentinel-1 radar data. *The Cryosphere*, *15*(10), 4901–4907. https://doi.org/10.5194/tc-15-4901-2021

Luckman, A., Benn, D. I., Cottier, F., Bevan, S., Nilsen, F., & Inall, M. (2015). Calving rates at tidewater glaciers vary strongly with ocean temperature. *Nature Communications*, *6*(1), 1–7. https://doi.org/10.1038/ncomms9566

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett. (Eds.). *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. Retrieved from https://urldefense.com/v3/__https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf__;!!N11eV2iwtfs!oRJsIOuE_6L1YRA_NoSYpMObcUTMcdfBQje1BwWFkyejc0nSpZ3R-Tk3DuUIt7GYsZCdVEBHYXin13YEq-saTh8V6zaF1$

Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., et al. (2019). The open global glacier model (oggm) v1. 1. *Geoscientific Model Development*, *12*(3), 909–931. https://doi.org/10.5194/gmd-12-909-2019

McBeck, J. A., Aiken, J. M., Mathiesen, J., Ben-Zion, Y., & Renard, F. (2020). Deformation precursors to catastrophic failure in rocks. *Geophysical Research Letters*, *47*(24), e2020GL090255. https://doi.org/10.1029/2020gl090255

Meier, M. F., & Post, A. (1969). What are glacier surges? *Canadian Journal of Earth Sciences*, *6*(4), 807–817. https://doi.org/10.1139/e69-081

Minchew, B. M., & Meyer, C. R. (2020). Dilation of subglacial sediment governs incipient surge motion in glaciers with deformable beds. *Proceedings of the Royal Society A*, *476*(2238), 20200033. https://doi.org/10.1098/rspa.2020.0033

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A*, *135*(3), 370–384. https://doi.org/10.1007/978-1-4612-4380-9_39

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pelt, W. V., Pohjola, V., Pettersson, R., Marchenko, S., Kohler, J., Luks, B., et al. (2019). A long-term dataset of climatic mass balance, snow conditions, and runoff in svalbard (1957–2018). *The Cryosphere*, *13*(9), 2259–2280. https://doi.org/10.5194/tc-13-2259-2019

Pfeffer, W. T., Arendt, A. A., Bliss, A., Bolch, T., Cogley, J. G., Gardner, A. S., et al. (2014). The randolph glacier inventory: A globally complete inventory of glaciers. *Journal of Glaciology*, *60*(221), 537–552. https://doi.org/10.3189/2014jog13j176

RGI, C. (2017). Randolph glacier inventory (rgi) - a dataset of global glacier outlines. Version 6.0. Retrieved from http://www.glims.org/RGI/randolph60.html

Ritz, C., Edwards, T. L., Durand, G., Payne, A. J., Peyaud, V., & Hindmarsh, R. C. (2015). Potential sea-level rise from Antarctic ice-sheet instability constrained by observations. *Nature*, *528*(7580), 115–118. https://doi.org/10.1038/nature16147

Schellenberger, T., Dunse, T., Kääb, A., Schuler, T. V., Hagen, J. O., & Reijmer, C. H. (2017). Multi-year surface velocities and sea-level rise contribution of the basin-3 and basin-2 surges, austfonna, svalbard. *The Cryosphere Discussions*, 1–27.

Sevestre, H., & Benn, D. I. (2015). Climatic and geometric controls on the global distribution of surge-type glaciers: Implications for a unifying model of surging. *Journal of Glaciology*, *61*(228), 646–662. https://doi.org/10.3189/2015jog14j136

Thøgersen, K., Gilbert, A., Bouchayer, C., & Schuler, T. V. (2021). Glacier surges controlled by the close interplay between subglacial friction and drainage. *Journal of Geophysical Research* (in review). https://doi.org/10.31223/x5jg87

Thøgersen, K., Gilbert, A., Schuler, T. V., & Malthe-Sørenssen, A. (2019). Rate-and-state friction explains glacier surge propagation. *Nature Communications*, *10*(1), 1–8. https://doi.org/10.1038/s41467-019-10506-4

Truffer, M., Kääb, A., Harrison, W. D., Osipova, G. B., Nosenko, G. A., Espizua, L., et al. (2021). Glacier surges. In *Snow and ice-related hazards, risks, and disasters* (pp. 417–466). Elsevier.

Zoet, L., Ikari, M., Alley, R. B., Marone, C., Anandakrishnan, S., Carpenter, B., & Scuderi, M. M. (2020). Application of constitutive friction laws to glacier seismicity. *Geophysical Research Letters*, *47*(21), e2020GL088964. https://doi.org/10.1029/2020gl088964