Automating the classification of surging 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, 0316 Oslo, Norway
[2]ISTerre, Univ. Grenoble Alpes, Grenoble INP, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, 38000 Grenoble, France

*Corresponding address: colili@uio.no

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. The detailed processes that control this intermittent behaviour remain elusive. We develop a machine learning framework to classify surge-type glaciers, based on their location, exposure, geometry, surface 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, surface 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. Our methodology could be extended to classify surge-type glaciers in other areas of the world.

Plain Language Summary

Around 1% of the glaciers in the world exhibit intermittent phases of accelerated motion, called surge. These accelerations are not fully understood. They may lead to dramatic glacier advances over rivers and damming up lakes that are then prone to a sudden and possibly catastrophic drainage. Glacier surge dynamics also contributes to uncertainties concerning sea-level rise projections. The Svalbard archipelago, located in the high Arctic, hosts more than one hundred surging glaciers. By combining several datasets and analysing them statistically using several machine learning techniques, we calculate the probability for every glaciers to experience surge events. Our results show that some specific combinations of the surface and bed slopes values, 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 of surge dynamics. We finally produce the first probabilistic map of surging for all the glaciers in Svalbard and suggest that our method is applicable to other areas in the world.

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 quantify the relative importance of relevant features on the surge probability.
- We compute the first map of glacier surge probability in Svalbard.

1 Introduction

Glacier instabilities, such as surges, are primary contributors to uncertainties of future sea-level rise projections (Pörtner et al., 2019). Surge-type glaciers exhibit long periods of quiescence and short periods of accelerated motion, often leading to rapid ice loss (Meier & Post, 1969; Cuffey & Paterson, 2010). They represent approximately 1%

of the glaciers in the world (Sevestre & Benn, 2015) and a considerable hazard potential (Kääb et al., 2018, 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 (Cuffey & Paterson, 2010; Meier & Post, 1969; Bhambri et al., 2017). 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 the hydro-mechanical interactions between a glacier and its bed (Cuffey & Paterson, 2010; Thøgersen et al., 2019; Benn et al., 2019). Due to the limited accessibility of subglacial environments, the physical processes at the ice-bed interface are difficult to measure. Recently, two approaches have been proposed to unify the theories of glacier instabilities (Fig. 11). 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, 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 (L. Zoet et al., 2020; L. K. Zoet & Iverson, 2020). Based on these two approaches, enthalpy budget and rate-and-state friction, 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: 1) they are more likely to be longer and/or wider (Clarke et al., 1986; Clarke, 1991; Jiskoot et al., 1998; Barrand & Murray, 2006) than non-surging glaciers; 2) their bed and surface slopes are likely to appear as important features but are often highly correlated with other features (Clarke et al., 1986; Clarke, 1991); 3) their bed contains more likely younger and mechanically weaker lithologies than hard beds (Jiskoot et al., 1998, 2000); 4) they are clustered in climatic envelopes between cold-dry and warm-humid environments (Sevestre & Benn, 2015); and 5) they are more likely polythermal (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 Consortium (2017) into five surging categories, from no surge to surge-type. However, statistical studies of glacier surges have two limitations: 1) they use integrated features for entire glaciers, and 2) except for Barrand and Murray (2006), none of them exposes the used methodology to competing methods. 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 investigate the surging probability of glaciers in Svalbard and identify the controlling features. 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 (Consortium, 2017). We propose a framework to regularize the evaluation of several machine learning models for determining glacier surge probability. Selecting the best performing model, the Extreme Gradient Boosting (XGBoost) (T. 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 easily applied for assessing the surge probability of glaciers in other regions of the world, when new data are available and/or can be adapted to other fields (e.g. landslides, earthquakes dynamics).

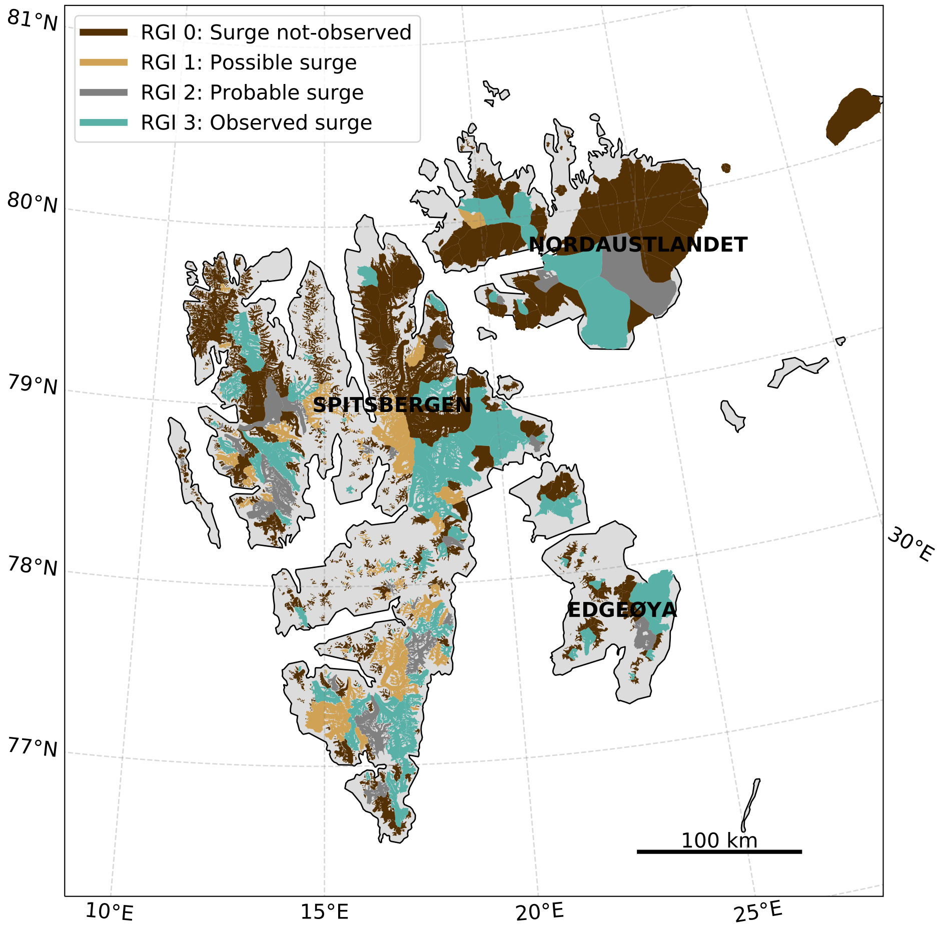

Figure 1: Classification of glaciers in Svalbard in the Randolph Glaciological Inventory database (Consortium, 2017). This database contains five classes that characterize the surge potential of glaciers (Sevestre & Benn, 2015) : Not observed (0), Possible (1), Probable (2), Observed (3), Not assigned (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 (Consortium, 2017), geometrical features (Maussion et al., 2019; Fürst et al., 2018), 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.

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 (X.-w. Chen & Jeong, 2007), and finally we use the Shapley Additive values (Lundberg & Lee, 2017). The sketch in Figure 2 illustrates our framework.

2.1 Data

2.1.1 Randolph Glacier Inventory features

The Randolph Glacier Inventory (Consortium, 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 (Fig. 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 and 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 (Consortium, 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 (Fig. 2). Other features present in the Randolph Glacier Inventory, such as the surface area and the length of glaciers, are not used because these are integrated features across each glaciers while we focus in this study on discretized variables along glacier centerlines.

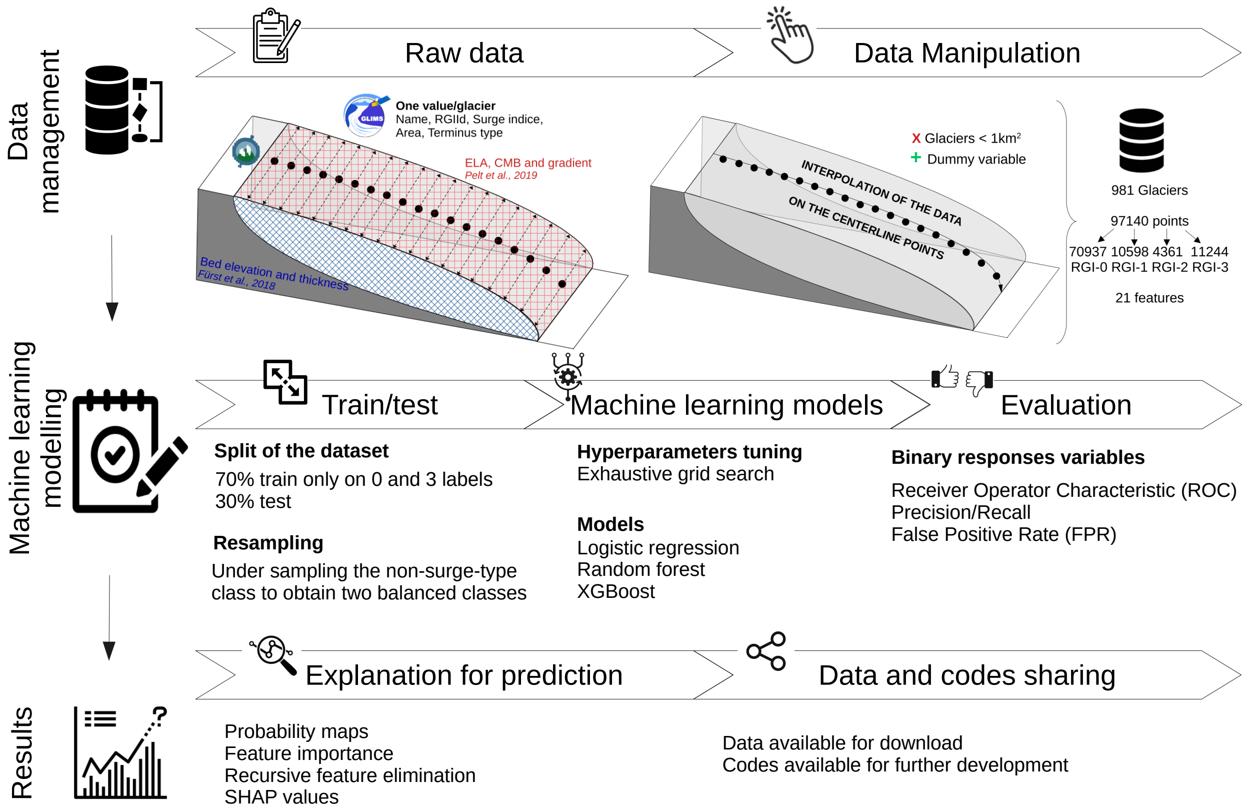

Figure 2: Workflow of the machine learning methods used to classify surge-type glaciers. Once the raw data are collected, the features are interpolated along the centerlines points. The database is then filtered and separated into a train data-set and a test data-set. Data are re-sampled to obtain balanced classes between surge-type and non-surge-type glaciers. The machine learning model 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.

2.1.2 Geometric features

Many studies investigating glacier surges have highlighted the importance of geometrical features (Sevestre & Benn, 2015; Hamilton & Dowdeswell, 1996; Jiskoot et al., 1998, 2000; Clarke et al., 1986; Clarke, 1991; Björnsson et al., 2003; Barrand & Murray, 2006). 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 (Maussion et al., 2019). 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. In total, it corresponds to an accumulated length of 700 km of measured thickness profiles. 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, i.e. runoff and surface mass balance. Pelt et al. (2019) created a long-term (1957-2018) dataset of surface mass balance for the glaciers, snow conditions, and runoff with a 1 km $\times$ 1 km spatial resolution and 3-hours 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 surface mass balance 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 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 km x 1km, we exclude all the glaciers with a surface area less than 1 km^2 and a length less than one kilometer.

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 database contains thirteen features: the Randolph Glacier Inventory identifier (1), the corresponding glacier name (2), the surging class (3), the bed elevation and slope (4, 5), the surface elevation and slope (6, 7), the thickness (8), the surface mass balance (9), the glacier width (10), the width divided by the thickness (11), and the driving stress (12). A random number is also added as a dummy feature (13) that does not have a physical interpretation and is used here to verify that the model is not taking it into consideration during the classification.

We exclude the glacier identifier and the glacier name features from the analysis resulting in eleven features for training the machine learning models. Figure 3 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 towards 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.

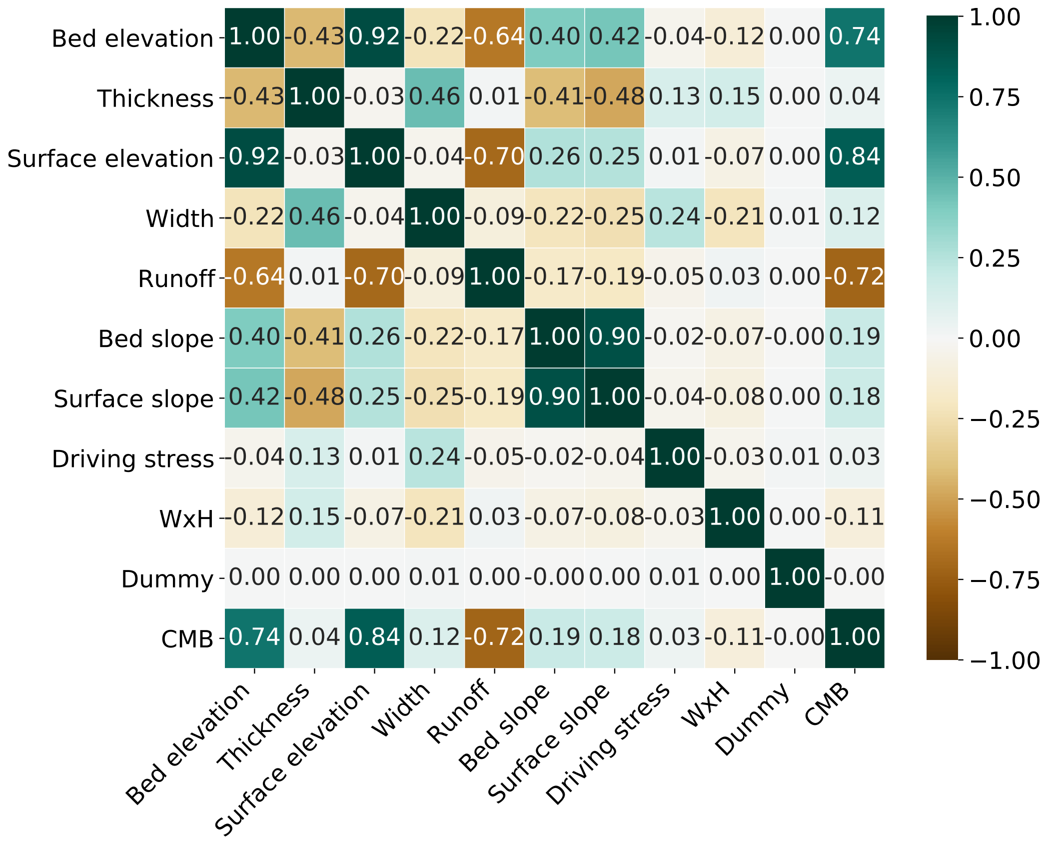

Figure 3: Correlation matrix between the most important features using only the training data. The colors shows the value of the coefficient of correlation.

2.3 Machine learning modelling

2.3.1 Training and testing data-sets

The training data-set is organized in the following ways: 1) only glaciers classified as Not-Observed surge (class 0) or Observed surge (class 3) in the Randolph Glacier Inventory are used; 2) the training data-set is resampled such that it contains an equal number of surge-type and non-surge-type glaciers; and 3) the training and testing 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).

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 using 50 bootstrap simulations using 50 different training and testing sets, the split in the data-set using this proportion is considered acceptable (Fig. 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). 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 (Supplementary Material, Section Appendix B).

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):

$$\ell = \log \frac{p}{1-p} = \beta_0 + \boldsymbol{\beta}\mathbf{X} \tag{1}$$

where β and β_0 are parameters that weight the impact of the input features $\mathbf{X}$. $p = P(Y = 1)$ 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 1x10^{-5} and the penalty to L2 (Table B1, Section Appendix B).

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 (Table B1, Section Appendix B).

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) (T. Chen & Guestrin, 2016). XGBoost is an implementation of a stochastic gradient boosting machine (Friedman, 2001, 2002; T. Chen et al., 2015; T. Chen & Guestrin, 2016). XGBoost can use a variety of learners as its base learners such as linear models or decision trees (T. 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. $F_m(X)$ is the current iteration fitted to the previous iterations residuals r_m. F_0 is the base estimate.

We implemented XGBoost using the xgboost library in Python (T. Chen & Guestrin, 2016). The objective is the logistic regression, we define 20000 boosting learners, trees have a maximum depth of 2, and the minimum child weight is 1 (Tab. B1, Sect. Appendix B).

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 ROC curve is the true positive rate $\mathrm{TP}_{\mathrm{rate}}$ against the false positive rate $\mathrm{FP}_{\mathrm{rate}}$:

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

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

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 labelled as surge-type while the true label should have been non-surge-type. The true positives correspond to surge-type glaciers that have been labelled 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:

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

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

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

2.4 Explanation for prediction

Over the thirteen features present in our custom-built database, we do not use the glacier name and the glacier number in the Randolph Glaciological Inventory. We examine how the eleven remaining features impact model decision in several ways: 1) we compute the relative feature importances across all models and compare them, 2) we examine two feature importances relevant to XGBoost (gain and weight), the model that has the highest performance, 3) we calculate the Shapely Additive values for the XGBoost model.

We compare the feature importances of three models in a stacked diagram (Fig. 7). 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 the Supplementary Material, Section Appendix C). For comparison purpose, we normalised all the scores using a min-max normalisation. 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 1) a global interpretation of the predictions by analyzing how much each predictor contributes positively or negatively to the target feature; 2) 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, e.g. a high value of a certain feature has a positive impact on the model output meaning that a high value will influence the model towards 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 Fig. 1. The average probability per glacier calculated in the present study is shown in the map of Figure 5. We also examine a subset of discretized glacier centerlines in Nordaustlandet Island (Fig. 5, 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 (Fig. 4 a.) 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, Fig. 4 b.) and lowest False Positive Rate (0.23, Fig. 4 d.) of all the models. This model demonstrates a lower recall (0.63) than logistic regression (0.68) and random forest (0.69) (Fig. 4 c.). 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.

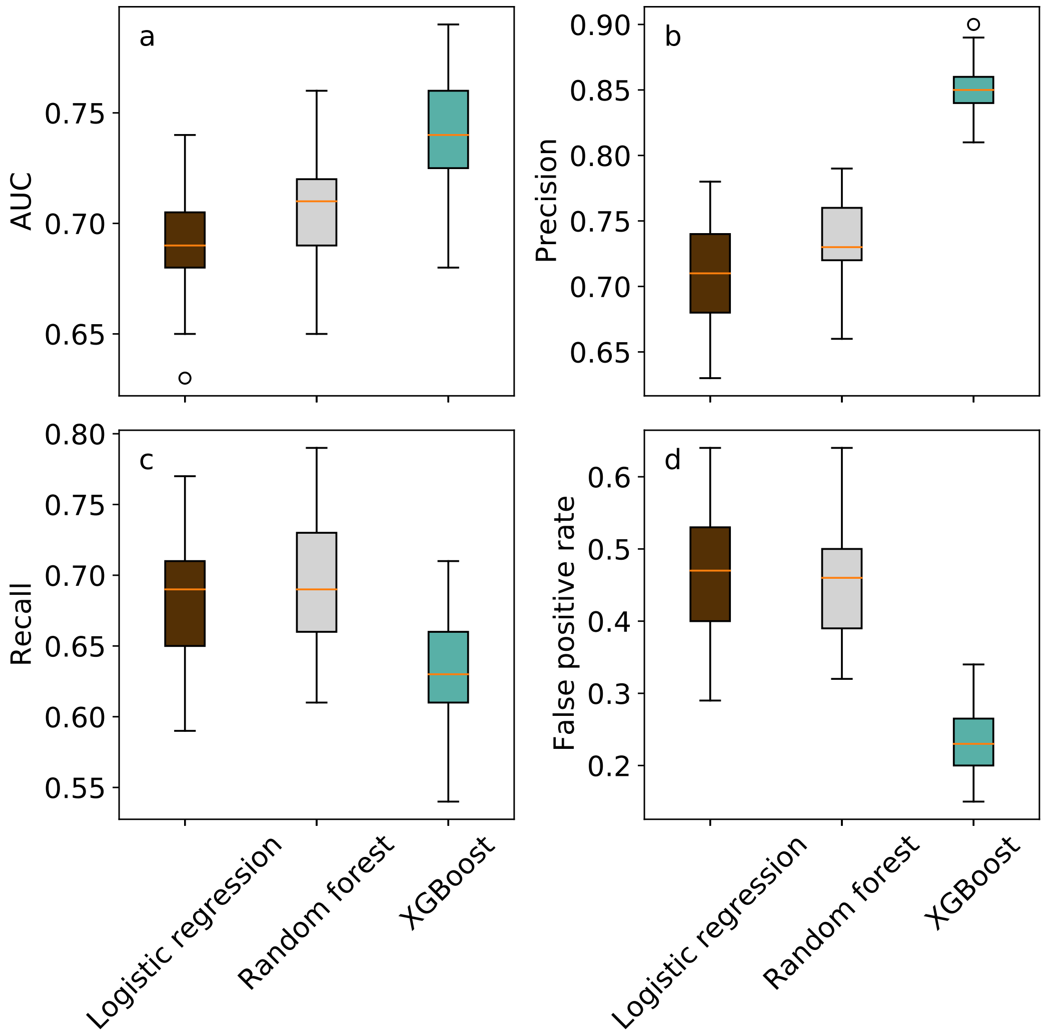

Figure 4: Boxplot representing the a) Area Under the Curves (AUC), b) precision score, c) recall score, and d) false positive rate for each of the three machine learning models. The scores have been calculated for 50 different training and testing data-sets. The orange line corresponds to the median 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.

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 (Fig. 5) 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 the supplementary material, Section Appendix D, Fig. D1. However, preferential zones of surge can be identified in e.g. Nordaustlandet island, Torell Land. Other areas, e.g. Nordensköld Land, Andree Land, do not gather a significant number of surge-

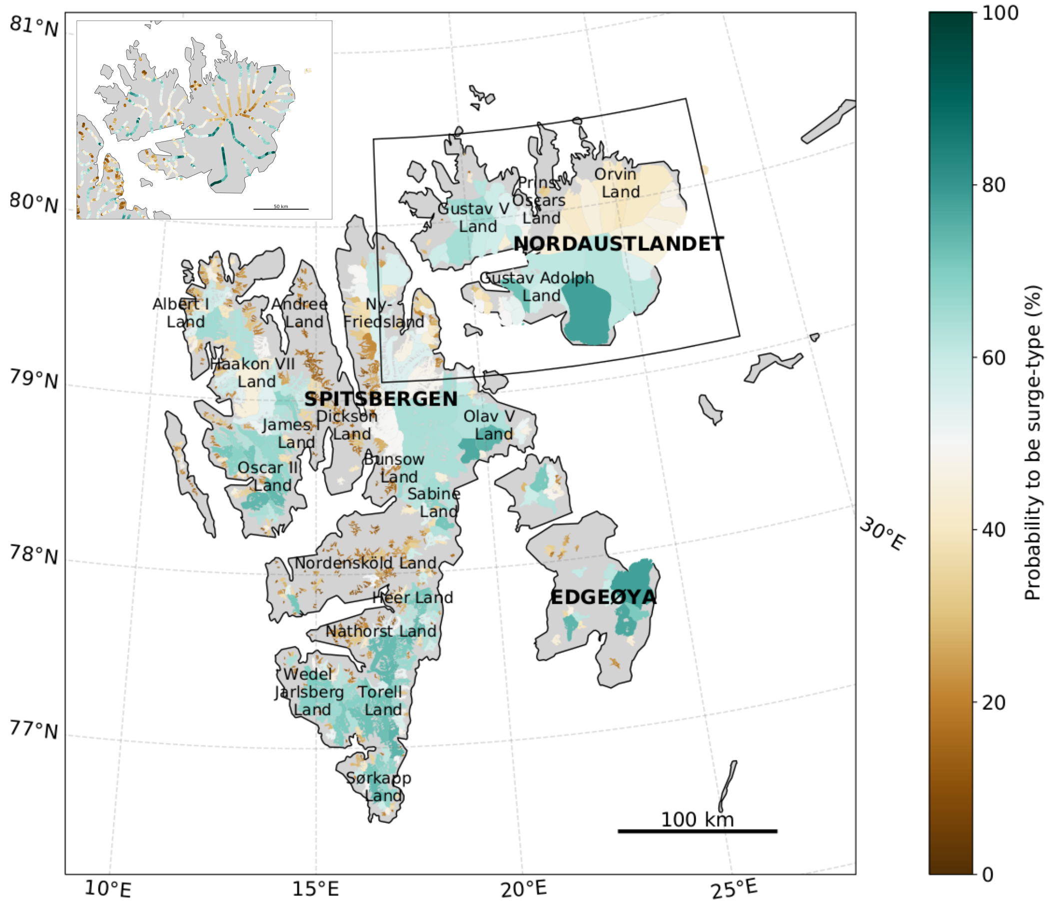

Figure 5: Averaged probability map for each glacier to be classified as surge-type in the XGBoost model. The zoom in the Austfonna ice cap shows how the average probability has been computed. First, a probability is calculated at every point along the centerline of every glacier. Then, we average the probabilities to surge of every point along the centerline to obtain an average surging probability for a given glacier.

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 (Fig. 5, inset).

In addition to the probability map, we compare our results to the existing Randolph Glacier Inventory classifications for surge-type glaciers. Figure 6a) 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). These results are supported by the histogram in the inset of Fig. (6b)) 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.

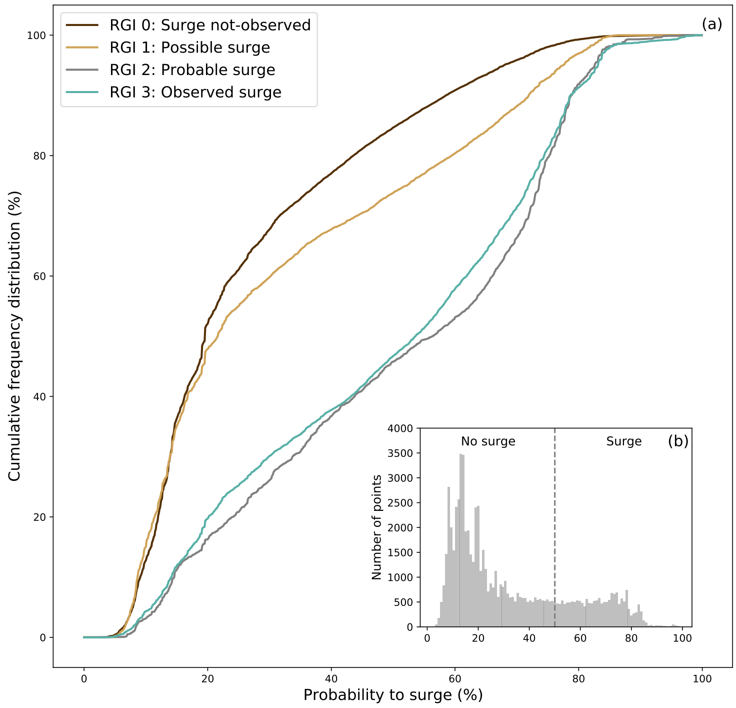

Figure 6: Cumulative frequency distribution for a glacier to be surge-type labeled by the classes defined in the Randolph Glacier Inventory. The inset shows the distribution of the probabilities. The vertical line indicates a 50% probability from which we separate surge-type from non surge-type glaciers.

3.3 Importance of geometrical and climatic features

Figure 7 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 surface mass balance, 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 8 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 surface mass balance, 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 (Fig. 9). To predict the surging potential of a glacier in Svalbard, the surface and bed slope, thickness, surface mass balance, 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.

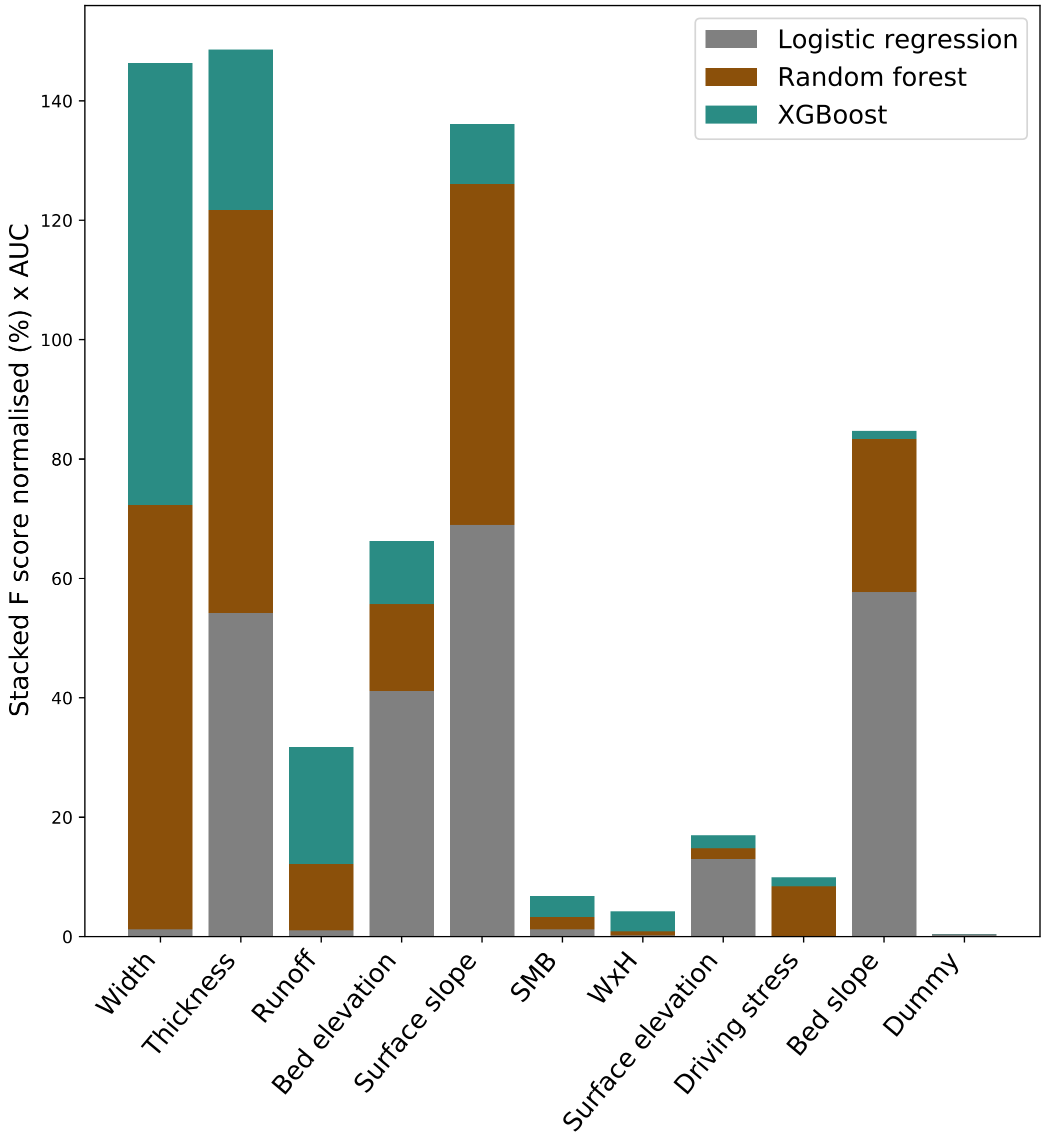

Figure 7: Feature importances for the logistic regression, random forest and XGBoost models stacked together. The F-score of each features has been normalised and multiplied by the AUC value. The features are organized from left to right from the most to the least important, according to XGBoost score.

3.4 Feature values and local impact on prediction

Using SHAP value analysis (Fig. 10), we find that some features have clear patterns. Higher values of glacier surface slopes, surface mass balance, 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 surface mass balance are likely to increase the probability of a glacier to be classified as surge-type. Some features do not show clear separation between the values and the correspond-

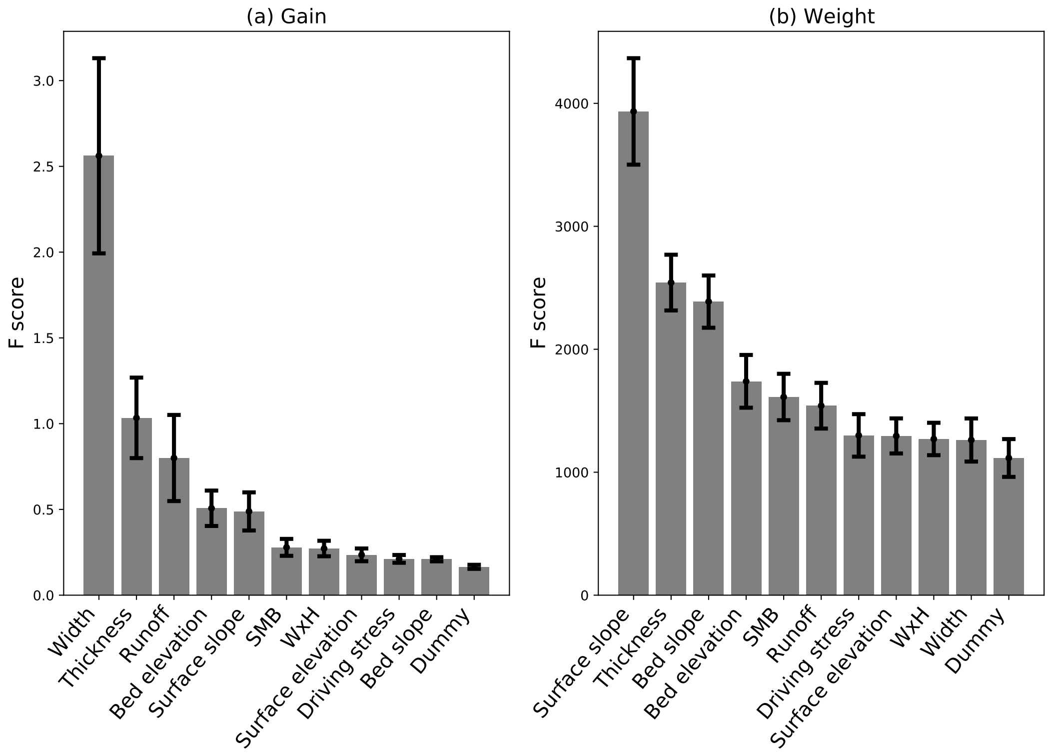

Figure 8: Feature importance of XGBoost model: (a) gain, (b) weight.

ing 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, surface mass balance, 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 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 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, i.e. univariate or multivariate regression (Clarke et al., 1986; Clarke, 1991; Hamilton & Dowdeswell,

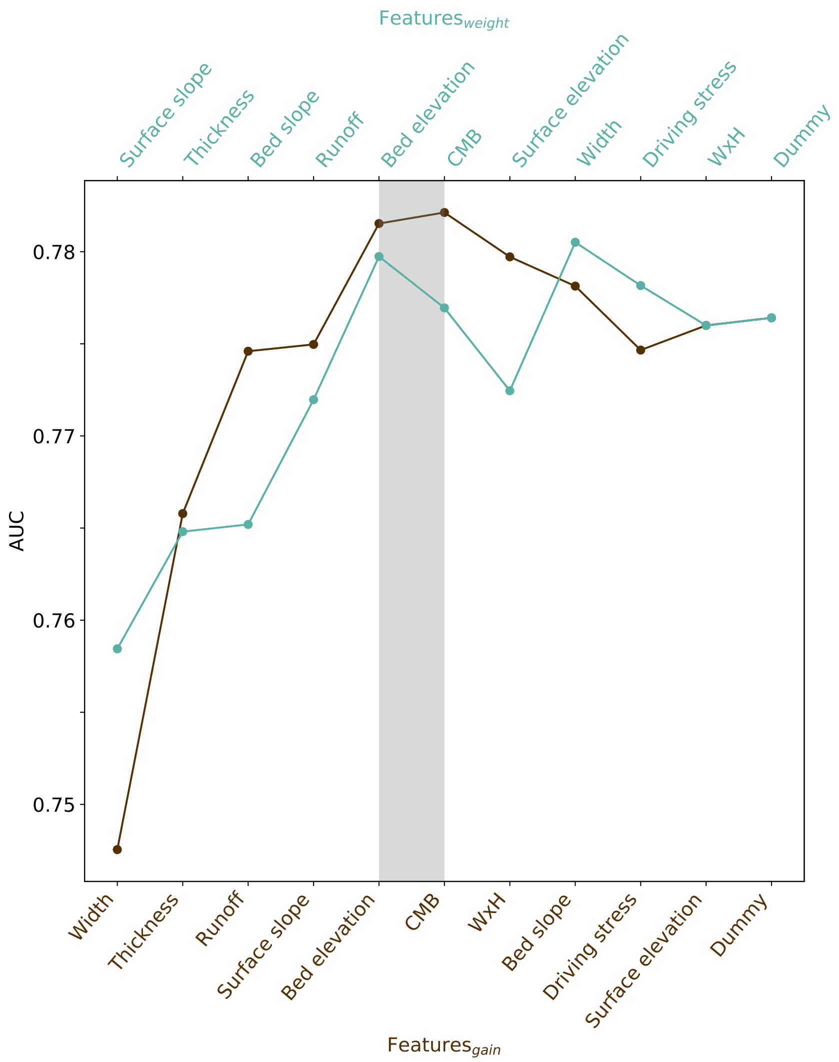

Figure 9: Results of recursive feature elimination show that four to five features explain most of the gain of information in the classification of surge-type glaciers. The number of features is added according to their order in the feature importance.

1996; Jiskoot et al., 2000; Barrand & Murray, 2006) 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 model 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., Fig 7).

As 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 bias and errors. 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. The resulting AUC of XGBoost probably cannot be higher because the training is biased by the input features.

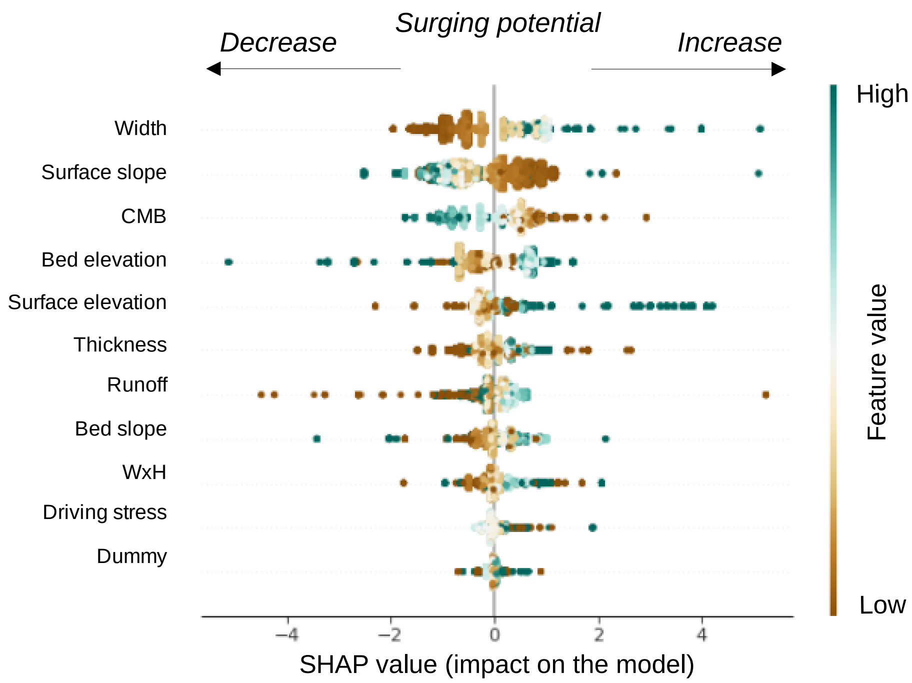

Figure 10: Summary of the SHAP values for every features at each glacier centerline point. The x position of each dot is the impact of the feature on the prediction of the model and the color of the dot represents the contribution of that value at each point.

In addition, the spatial resolution of individual features differs, e.g., the bed elevation has been computed on a 100m resolution whereas the runoff and surface mass balance have been calculated on a 1km 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.

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 surface mass balance to a smaller extent (Fig. 11). The width and the thickness have been shown to be important in previous statistical studies (Clarke, 1991; Barrand & Murray, 2006; Jiskoot et al., 2003) together with the surface slope (Sevestre & Benn, 2015; Jiskoot et al., 1998, 2000, 2003). Although XGBoost models predict that lower slope will drive the prediction towards 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 (Kamb, 1987; Fowler, 1989; Benn et al., 2019; 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 modelling 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 10).

To a smaller extent, the surface mass balance 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 (Fig. 9), as found in Jiskoot et al. (2000). The surface mass balance is highly correlated to other features in the model, e.g. the surface elevation and the runoff, meaning that the effect of the surface mass balance 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 have 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. Therefore, the geometrical features capture already a major part of the climatic influence. The climatic features do not appear primarily important because their effect is already captured into geometrical features.

In Svalbard, we expect that climatic features are not playing a central role on the prediction since, 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 on assessing the underlying physical processes leading to glacier surge.

Our framework can be used in other regions of the world. Our model has been trained on Svalbard and shows good performance scores. One could apply the trained model to a testing data-set in another region of world. If the performance score decreases drastically, one could infer that the predictions should be controlled by different features than in Svalbard, highlighting another surging mechanism or a different relation between geometric features and climatic conditions.

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 the Table 1 have a probability higher than 50% to be surge-type in our model, while in the Randolph Glacier Inventory they are currently labelled 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.

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)

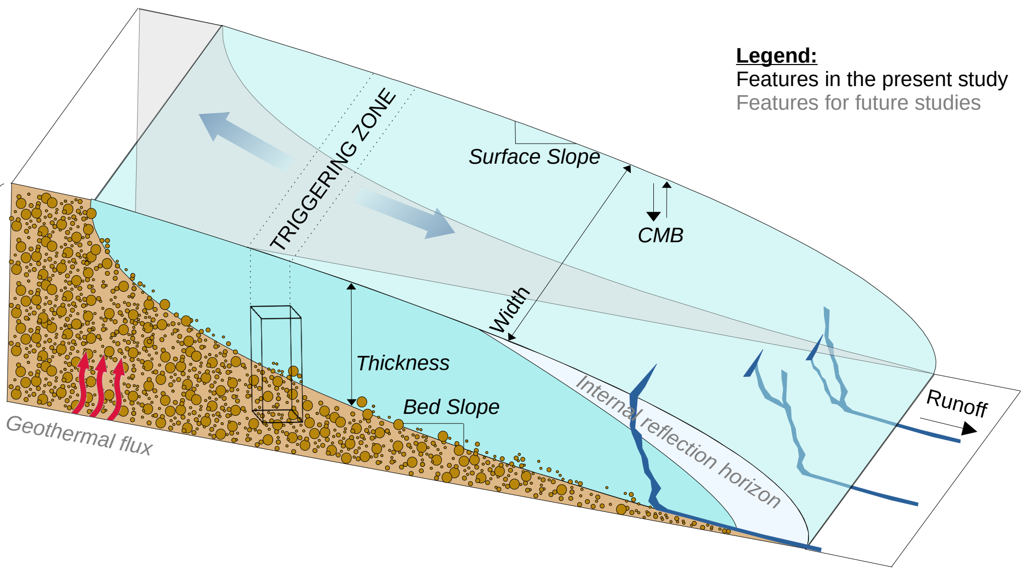

Figure 11: Sketch of the features that have been implemented in our model and new features that could be implemented for evaluating the surging potential of glaciers.

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, i.e. the width, the thickness, the runoff, the surface and bed slopes and the surface mass balance 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.

To complement theories of glacier surge, new features might be added to our framework, i.e. 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 towards this goal (Fig. 11). 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).

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 data and code are available in the repository `https://doi.org/10.5281/zenodo.5657088` (Bouchayer, 2021). The authors declare no conflict of interest.

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. Retrieved from `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., . . . Jin, W. (2021). Increasing glacial lake outburst flood hazard in response to surge glaciers in the karakoram. *Earth-Science Reviews*, *212*, 103432. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0012825220304785` doi: 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.

Bhambri, R., Hewitt, K., Kawishwar, P., & Pratap, B. (2017). Surge-type and surge-modified glaciers in the karakoram. *Scientific reports*, *7*(1), 1–14. Retrieved from `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. Retrieved from `https://doi.org/10.3189/172756403781816365`

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

Breiman, L. (1999). Prediction games and arcing algorithms. *Neural computation*, *11*(7), 1493–1517.

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

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

Clarke, G. K. (1991). Length, width and slope influences on glacier surging. *Journal of Glaciology*, *37*(126), 236–246. Retrieved from `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: Solid*

Earth, *91*(B7), 7165–7180. Retrieved from https://doi.org/10.1029/JB091iB07p07165

Consortium, R. (2017). *Randolph glacier inventory (rgi) - a dataset of global glacier outlines: Version 6.0.* Retrieved 2021 - 02 - 01, from http://www.glims.org/RGI/randolph60.html

Cuffey, K. M., & Paterson, W. S. B. (2010). *The physics of glaciers.* Academic Press. Retrieved from 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. Retrieved from https://www.sciencedirect.com/science/article/pii/004019519290055B doi: 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., ... others (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, *537*, 45–60.

Fowler, A. (1989). A mathematical analysis of glacier surges. *SIAM Journal on Applied Mathematics*, *49*(1), 246–263. Retrieved from 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.

Fürst, J. J., Navarro, F., Gillet-Chaulet, F., Huss, M., Moholdt, G., Fettweis, X., ... others (2018). The ice-free topography of svalbard. *Geophysical Research Letters*, *45*(21), 11–760.

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.

Hamilton, G. S., & Dowdeswell, J. A. (1996). Controls on glacier surging in svalbard. *Journal of Glaciology*, *42*(140), 157–168. Retrieved from 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. Retrieved from 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., ... Slangen, A. B. (2019). Glaciermip–a model intercomparison of global-scale glacier mass-balance models and projections. *Journal of Glaciology*, *65*(251), 453–467.

Hosmer Jr, D. W., 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. Retrieved from 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. Retrieved from 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. Retrieved from https://doi.org/10.3189/172756403781816220

Kääb, A., Jacquemart, M., Gilbert, A., Leinss, S., Girod, L., Huggel, C., ... others (2021). Sudden large-volume detachments of low-angle mountain glaciers–more frequent than thought? *The Cryosphere*, *15*(4), 1751–1785. Retrieved from https://doi.org/10.5194/tc-15-1751-2021

Kääb, A., Leinss, S., Gilbert, A., Bühler, Y., Gascoin, S., Evans, S. G., ... others (2018). Massive collapse of two glaciers in western tibet in 2016 after surge-like instability. *Nature Geoscience*, *11*(2), 114–120. Retrieved from https://doi.org/10.1038/s41561-017-0039-7

Kamb, B. (1987). Glacier surge mechanism based on linked cavity configuration of the basal water conduit system. *Journal of Geophysical Research: Solid Earth*, *92*(B9), 9083–9100.

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. Retrieved from 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. Retrieved from https://tc.copernicus.org/articles/15/4901/2021/ doi: 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.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon et al. (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

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

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

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (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., ... others (2019). A long-term dataset of climatic mass balance, snow conditions, and runoff in svalbard (1957–2018). *The Cryosphere*, *13*(9), 2259–2280.

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

Pörtner, H.-O., Roberts, D. C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., & Weyer, N. (2019). *The ocean and cryosphere in a changing climate.* Geneva: Intergovernmental Panel on Climate Change.

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. Retrieved from 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*(in review). Retrieved from 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. Retrieved from https://doi.org/10.1038/s41467-019-10506-4

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.

Zoet, L. K., & Iverson, N. R. (2020). A slip law for glaciers on deformable beds. *Science*, *368*(6486), 76–78.

Appendix A Boostrapping results

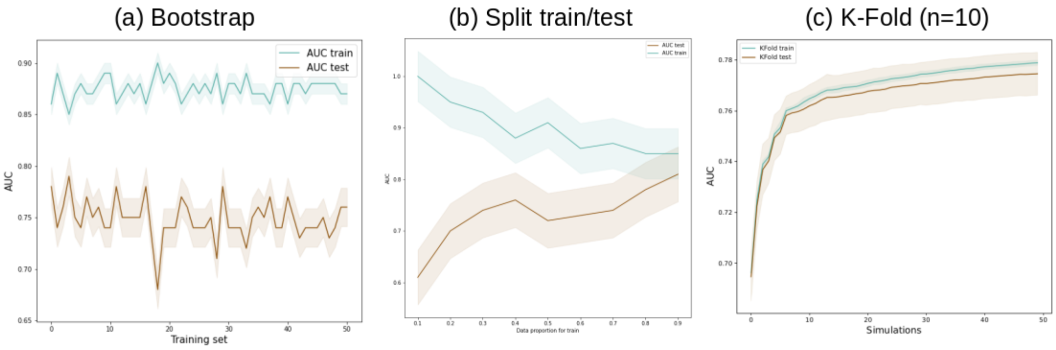

Figure A1: Evaluation of model consistency. (a) We perform the model calculation on fifty different training sets selected randomly and calculate the Area Under the Curve (AUC) for the training and the testing models. (b) The training and testing sets can be separated following different proportions. We performed different splits from 10% of the data-set belonging to the training set to 90% of the data-set belonging to the training set. We calculated the AUC. (c) K-Fold cross validation performed on 10 folds and the associated AUC for the train and the test set.

Appendix B 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 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 B1 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.

Table B1: 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	$1\text{x}10^{-5}$	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	

Appendix C 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 analysed 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 D Centerline probability

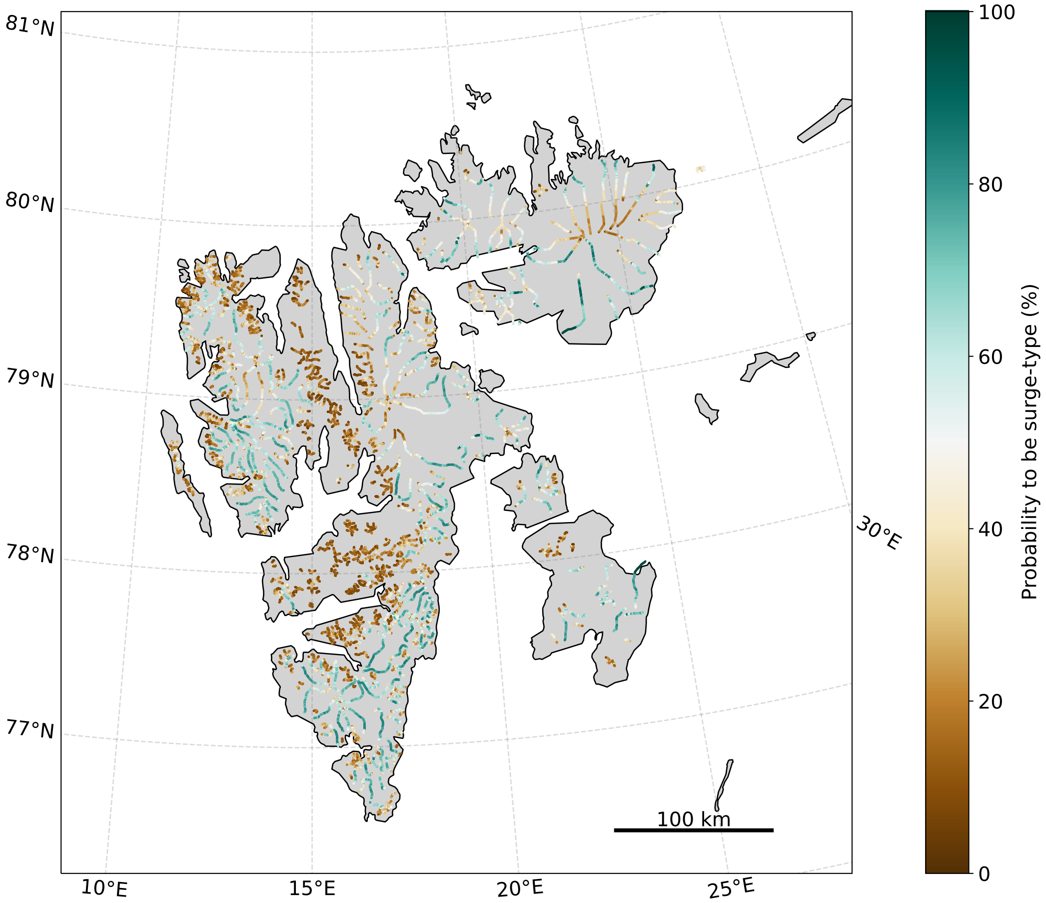

Figure D1: Probability map for each point of glacier centerlines to be classified as surge-type in the XGBoost model.