Mapping Glacier Basal Sliding Applying Machine Learning

Josefine Umlauft, Christopher W. Johnson, Philippe Roux, Daniel Taylor Trugman, Albanne Lecointre, Andrea Walpersdorf, Ugo Nanni, Florent Gimbert, Bertrand Rouet-Leduc, Claudia Hulbert, et al.

JGR Earth Surface

RESEARCH ARTICLE
10.1029/2023JF007280

Mapping Glacier Basal Sliding Applying Machine Learning

Josefine Umlauft[1], Christopher W. Johnson[2], Philippe Roux[3], Daniel Taylor Trugman[4], Albanne Lecointre[3], Andrea Walpersdorf[3], Ugo Nanni[5], Florent Gimbert[6], Bertrand Rouet-Leduc[7], Claudia Hulbert[8], Stefan Lüdtke[9], Sascha Marton[10], and Paul A. Johnson[2]

[1]ScaDS.AI – Center for Scalable Data Analytics and Artificial Intelligence, Leipzig University, Leipzig, Germany, [2]Los Alamos National Laboratory, Los Alamos, NM, USA, [3]ISTerre – Institut des Sciences de la Terre, Maison des Geosciences, Grenoble, France, [4]Nevada Seismological Laboratory, University of Nevada, Reno, NV, USA, [5]Department of Geosciences, University of Oslo, Oslo, Norway, [6]IGE - Institut de Geophysique de l'Environnement, Grenoble, France, [7]Disaster Prevention Research Center, Kyoto University, Kyoto, Japan, [8]Ecole Normale Superieure, Paris, France, [9]Institute for Visual & Analytic Computing, University of Rostock, Rostock, Germany, [10]Institute for Enterprise Systems, University of Mannheim, Mannheim, Germany

Abstract During the RESOLVE project ("High-resolution imaging in subsurface geophysics: development of a multi-instrument platform for interdisciplinary research"), continuous surface displacement and seismic array observations were obtained on Glacier d'Argentière in the French Alps for 35 days in May 2018. The data set is used to perform a detailed study of targeted processes within the highly dynamic cryospheric environment. In particular, the physical processes controlling glacial basal motion are poorly understood and remain challenging to observe directly. Especially in the Alpine region for temperate based glaciers where the ice rapidly responds to changing climatic conditions and thus, processes are strongly intermittent in time and heterogeneous in space. Spatially dense seismic and Global Positioning System (GPS) measurements are analyzed applying machine learning to gain insight into the processes controlling glacial motions of Glacier d'Argentière. Using multiple bandpass-filtered copies of the continuous seismic waveforms, we compute energy-based features, develop a matched field beamforming catalog and include meteorological observations. Features describing the data are analyzed with a gradient boosting decision tree model to directly estimate the GPS displacements from the seismic noise. We posit that features of the seismic noise provide direct access to the dominant parameters that drive displacement on the highly variable and unsteady surface of the glacier. The machine learning model infers daily fluctuations and longer term trends. The results show on-ice displacement rates are strongly modulated by activity at the base of the glacier. The techniques presented provide a new approach to study glacial basal sliding and discover its full complexity.

Plain Language Summary Alpine glaciers are a major component in the dynamic cryospheric environment. They are characterized by a multitude of processes occurring side by side, including but not limited to melt water flow, crevasse formation, and frictional basal sliding of the ice mass over the rigid and obstructive bedrock. Each of these processes generates distinctive acoustic signals that can be recorded by seismic instruments and the changing on-ice motions are resolvable with Global Positioning System. Considering the rapidly changing glacial environment, there is an increasing need for reliable models to predict glacial dynamics to properly assess any associated hazard. Understanding basal sliding is of particular interest to this problem. Investigated here is how to overcome the challenge of describing glacier sliding using seismic signals since the records often contains multiple "loud" signals originating from associated surface processes within the glacier. To uncover specific processes occurring at the ice-bedrock interface, we design a machine learning model to incorporate signals recorded on the glacier to predict the on-ice surface motions. The results provide valuable insights into the spatiotemporal dynamics of an active Alpine glacier with the potential to contribute to a better understanding of the driving mechanisms of glacier sliding.

Key Points:

- Seismic and Global Positioning System (GPS) data are examined to study physical processes controlling glacial basal motion
- Decision tree model uses beamforming catalog and statistical features of time series to constrain correlations with GPS recorded motions
- Model features indicate glacial on-ice velocity is modulated by basal motions

Josefine Umlauft and Christopher W. Johnson contributed equally to this work.

Correspondence to:

J. Umlauft,
josefine.umlauft@uni-leipzig.de

Citation:

Umlauft, J., Johnson, C. W., Roux, P., Trugman, D. T., Lecointre, A., Walpersdorf, A., et al. (2023). Mapping glacier basal sliding applying machine learning. *Journal of Geophysical Research: Earth Surface*, *128*, e2023JF007280. https://doi.org/10.1029/2023JF007280

Received 5 JUN 2023
Accepted 23 OCT 2023

Author Contributions:

Conceptualization: Josefine Umlauft, Christopher W. Johnson, Philippe Roux, Daniel Taylor Trugman, Bertrand Rouet-Leduc, Paul A. Johnson
Data curation: Josefine Umlauft, Christopher W. Johnson, Philippe Roux, Albanne Lecointre, Andrea Walpersdorf, Ugo Nanni
Formal analysis: Josefine Umlauft, Christopher W. Johnson, Philippe Roux, Albanne Lecointre, Andrea Walpersdorf, Ugo Nanni, Stefan Lüdtke, Sascha Marton
Funding acquisition: Florent Gimbert, Paul A. Johnson

1. Introduction

The cryosphere is one of the most rapidly changing environments on Earth and transformations are accentuated by the ongoing evolution of climatic conditions. In mountainous regions, glacier dynamics can be used as a local marker of climate change, and can cause major damage to human infrastructure, so it is of common social interest to study spatiotemporal processes within the ice with high resolution (Faillettaz et al., 2015). The rapidly

Investigation: Josefine Umlauft, Christopher W. Johnson
Methodology: Josefine Umlauft, Christopher W. Johnson, Philippe Roux, Bertrand Rouet-Leduc, Claudia Hulbert, Stefan Lüdtke, Sascha Marton, Paul A. Johnson
Project Administration: Philippe Roux, Florent Gimbert, Paul A. Johnson
Visualization: Josefine Umlauft, Christopher W. Johnson
Writing – original draft: Josefine Umlauft, Christopher W. Johnson
Writing – review & editing: Philippe Roux, Daniel Taylor Trugman, Andrea Walpersdorf, Ugo Nanni, Florent Gimbert, Bertrand Rouet-Leduc, Claudia Hulbert, Paul A. Johnson

emerging field of "cryoseismology" addresses processes within the glacial environment, such as crevassing, hydrofracturing, failure and calving of ice fragments or supraglacial, englacial and subglacial water discharge via the analysis of continuous seismic records (Podolskiy & Walter, 2016). Special emphasis has been put on the investigation of glacier sliding, which is still not completely understood, but affects large-scale ice flow, ice sheet stability, and thus ultimately sea level rise (Ritz et al., 2015).

Glaciers flow via two processes, internal deformation (or "creep") and basal sliding (Cuffey & Paterson, 2010). The stress-strain relationship for internal deformation of the glacier itself describes viscous deformation associated with ice creep and can be approximated by "Glen's flow law" (Glen, 1955). Basal sliding is responsible for fast flow of ice-streams; "sliding" is used as an umbrella term here for actual sliding of the ice sole and deformation of soft subglacial till beds (e.g., Helanow et al., 2021). In view of steep, unstable ice tongues, it is of great interest to scientists and stakeholders to understand the physical basis of glacier sliding given that catastrophic break-off events threaten mountain communities world-wide (Faillettaz et al., 2015; Shugar et al., 2021).

The first theoretical concept of glacier sliding was introduced by postulating that normal forces on undeformable bed undulations produce local shear resistance (Weertman, 1957). Here, a frictionless glacier bed was considered with sliding driven by enhanced deformation and regelation around stiff bed obstacles. Weertman's theory of "hard" bed sliding (Weertman, 1957) was modified to account for subglacial water cavity formation (Gagliardini et al., 2007; Iken, 1981; Schoof, 2005) and deformable subglacial till layers (Murray, 1997). Both mechanisms can explain observations of melt-water enhanced ice flow and basal sliding (Cuffey & Paterson, 2010). Modern sliding theories (e.g., Schoof, 2005; Zoet & Iverson, 2020) are still influenced by these concepts. However, recent cryoseismological studies show that glacier sliding is not always smooth, but interrupted by distinct slip events (Aster & Winberry, 2017). This points to frictional processes, where sudden shear failure at the glacier bed emits seismic waves, analogous to the behavior of tectonic faults. Such stick-slip motion cannot be explained by traditional Weertman-type or soft-bed theories, which describe sliding as a continuous, slow, and smooth process. Instead, frictional processes add to the complexity of basal sliding and thus ice flow. A pivotal challenge in glaciological research is to formulate new or extend existing sliding laws, including conventional concepts but also considering glacier frictional sliding as an additional flow mechanism (e.g., Lipovsky & Dunham, 2017; Lipovsky et al., 2019; Sergienko et al., 2009; Winberry et al., 2011; Zoet & Iverson, 2020).

Evidence from polar and non-polar ice masses suggests that microseismic stick-slip motion is a widespread (see Podolskiy & Walter, 2016, and references therein) and potentially pervasive form of basal sliding (Barcheck et al., 2019; Gräff et al., 2021; Hudson et al., 2020; Kufner et al., 2021; McBrearty et al., 2020; Walter et al., 2020). Individual microseismic stick-slip events are very small with negative magnitudes and shear displacements on millimeter scales or less (Helmstetter et al., 2020). Successive events may coalesce into sustained ice-tremor resulting in ice-stream wide sliding episodes with surface displacements of tens of centimeters per day. The spectral signature of the sliding tremor is characterized by spectral peaks at frequencies corresponding to the inverse of inter-event times between individual stick-slip events (Lipovsky & Dunham, 2016). First detected at rapid Antarctic ice streams, sliding tremor may be a widespread phenomenon with observational evidence for these sliding tremors beneath Greenlandic (McBrearty et al., 2020) and Alpine glacier ice (Umlauft et al., 2021), and the slip displacement may be measurable at the ice surface. Detection of these tremors with conventional on-ice seismometers is challenging because the signals can be masked by the extensive glacial noise from other cryoseismic sources, especially englacial and subglacial water flow (Eibl et al., 2020; McBrearty et al., 2020; Röösli et al., 2014; Umlauft et al., 2021). Thus, in Alpine regions, with temperate glacier ice and high meltwater production, frictional sliding in the form of microseismic stick-slip tremors may be completely overlooked and far more predominant than presently understood.

Analogous to tectonic faults, stick slip motion across glacial faults emits seismic energy and is commonly measured by seismometers (Podolskiy & Walter, 2016). The frictional state of a tectonic fault and information about the current position within its seismic cycle are still challenging to access. As the fault's rupture, nucleation and magnitude, and future earthquake occurrence are directly controlled by the fault frictional state, its quantification is of interest for understanding the underlying physics (Marone, 1998). Numerous theoretical simulations and laboratory experiments contributed to the determination of frictional characteristics (e.g., Dorostkar et al., 2017; Kaproth & Marone, 2013; Madariaga & Ruiz, 2016; Rabinowicz, 1956; Rubinstein et al., 2004; Scholz, 1968). Recently, analyses of seismic signals from laboratory faults (Rouet-LeDuc, Hulbert, Bolton, et al., 2018) and faults in earth (Johnson & Johnson, 2021) applying machine learning have yielded remarkable results indicating

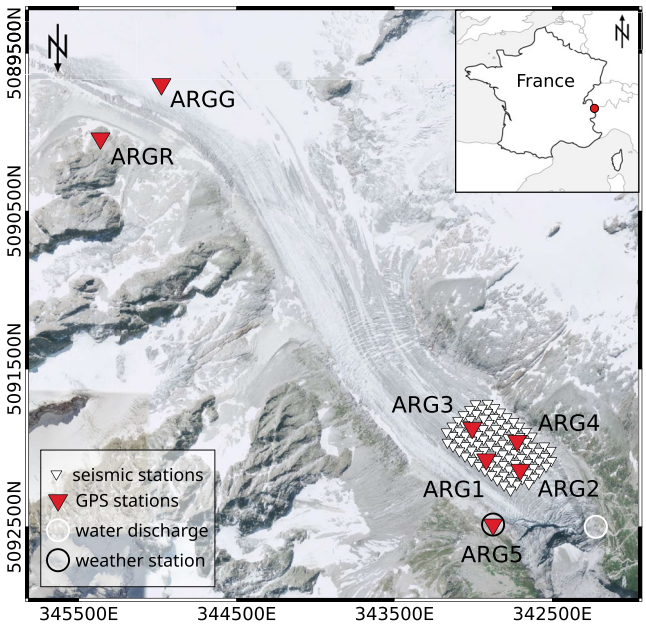

Figure 1. Overview map of Glacier d'Argentière together with the RESOLVE sensor infrastructure (Gimbert et al., 2021) including the locations of the seismic nodes (white triangles), the Global Positioning System (GPS) stations (red triangles, ARGx), the weather station (black circle around ARG5) and the location of the borehole for measurements of water discharge (white circle). The GPS stations ARG1, ARG2, ARG3, ARG4, and ARGG were installed on the surface of the glacier (on-ice stations), GPS stations AGR5 and ARGR were installed on solid ground/bedrock (off-ice stations).

that the seismic waves contain information about the fault characteristics at all times.

We use this analogy to guide the choice of research methodology to monitor the physical state of the glacier. So far, direct and continuous quantification of fault friction cannot be achieved using conventional geophysical approaches, whereas supervised machine learning models are suitable to directly quantify instantaneous fault friction in laboratory experiments and fault properties in tectonic environments (Hulbert et al., 2019; Johnson & Johnson, 2021; Ren et al., 2020; Rouet-LeDuc, Hulbert, Bolton, et al., 2018; Rouet-LeDuc, Hulbert, & Johnson, 2018; Wang et al., 2021, 2022).

In laboratory experiments it was demonstrated that frictional properties can be accessed through the statistical characteristics of continuous seismic records (range of the data, root mean square, variance, skewness, kurtosis, quantile ranges) and that even different modes of slip along these laboratory faults were captured, which demonstrates that seismic data are a rich archive that allows one to directly observe the physical state of a fault (Hulbert et al., 2019; Rouet-LeDuc, Hulbert, Bolton, et al., 2018). These processes are similar to basal motion in the glacial environment where the displacement takes place at the ice-bed-interface.

With the aim to uncover the signals related to sliding that are not directly observable, we applied a decision tree model to a new data set from a dense on-ice network on Glacier d'Argentière (French Alps) comprising continuous measurements of local seismicity, surface velocities, and meteorological observations. Due to the highly variable and noisy glacial environment, extensive preprocessing of the seismic and geodetic measurements is essential for a robust feature space with the goal of directly estimating glacier sliding behavior from the surface of the ice and hence, to monitor its dynamics.

2. Methods

2.1. Resolve Data Collection

As part of the RESOLVE project "High-resolution imaging in subsurface geophysics: development of a multi-instrument platform for interdisciplinary research"), researchers from ISTerre and IGE Grenoble (France) and ETH Zürich (Switzerland) installed a unique sensor infrastructure at the surface of Glacier d'Argentière (Figure 1) (Gimbert et al., 2021). A dense seismic monitoring array with 98 geophones, 7 Global Positioning System (GPS) stations, a meteorological station, and a water discharge station were operational during approximately 1 month in May 2018 (24 April 2018 to 27 May 2018). Five of the GPS stations were installed directly on the surface of the ice (ARG1–ARG4, ARGG) with four of them integrated with the seismic array (ARG1–ARG4). The remaining two stations (ARG5, ARGR) were placed on solid bedrock next to the glacier near the seismic array.

The GPS derived displacement rate (velocity) was computed using a centered moving time window of size ±3 hr with a 1 hr time step for east, north, and vertical components, and the combined horizontal components (east + north). This sampling was found to provide the best agreement between errors and signal-to-noise ratio.

Seismic observations were continuously recorded at a sample rate of 500 Hz in a grid-like dense array (Ø 700 m). The stations were deployed into snow about 30 cm below the surface within the ablation zone of Glacier d'Argentière (see Gimbert et al., 2021, for specific details). Signal preprocessing includes removing the instrument response, detrending and demeaning the continuous waveforms.

Temperature and precipitation were monitored at a 10 min sampling rate using one station situated on solid bedrock about a kilometer to the north of the array. Water discharge was measured every 15 min by the Emosson power supply company in excavated tunnels below the glacier tongue (Gimbert et al., 2021; Vincent & Moreau, 2016).

2.2. Matched Field Processing

Matched field processing (MFP) is the natural extension of plane wave beamforming and yields for the location of seismic noise sources in range, depth and azimuth by analyzing spherical waves in the close environment of the underlying seismic array (Bucker, 1976). The approach was originally developed in ocean acoustics (Baggeroer et al., 1993; Kuperman & Turek, 1997), but a broad spectrum of applications can be found in environmental seismology to study near-surface processes on the exploration scale (Corciulo et al., 2012; Cros et al., 2011; Umlauft & Korn, 2019; Vandemeulebrouck et al., 2010) and the rapidly emerging special research field of cryoseismology to better understand dynamics within for example, Alpine glacial ice (Nanni et al., 2021, 2022; Umlauft et al., 2021; Walter et al., 2015, 2020).

Assuming the spatial coherence of the wave field across the array, a systematic correlation of portions of continuous seismic field records and the model-based Green's function (replica) is performed at various candidate source positions. The approach is performed in the frequency domain and can be considered as an equivalent of shift-and-stack techniques in the time domain. For a certain frequency, replica parameterization allows improved data fitting by velocity inversion (Gradon et al., 2019) or polarity optimization for the location of double-couple sources (Umlauft et al., 2021). The procedure is aimed to estimate phase matches between the data wave field and the replica field with the beam power maximum representing the most probable source location.

2.3. Data Features

Data features are statistics of the continuous seismic records that are input to model (e.g., Johnson & Johnson, 2021; Johnson et al., 2020; Rouet-LeDuc, Hulbert, & Johnson, 2018) from a five-node subarray with high signal-to-noise-ratio, meteorological and water discharge measurements, and events spatially binned from a beamforming catalog (see Figure 2 for station locations and a snapshot of the beamforming catalog). An overview of the features is provided in Table A1. Statistical features were computed for the continuous seismic record of five selected stations shown as inverted white triangles in Figure 2. We made four copies of the records using a bandpass filter between 10 and 50 Hz: 10–20, 20–30, 30–40, and 40–50 Hz to cover the frequency bands related to the most dominant processes in glacial ice, such as water flow, crevassing, icequakes or stick-slip tremors (Podolskiy & Walter, 2016). A moving time window of 1 hr is applied to compute the variance, kurtosis, mean, root mean square, skewness, range and interquantile ranges (0.025, 0.25, 0.5, 0.75, and 0.95) using ± 3 hr before and after the respective time stamp. This sampling matches the GPS data sampling resolution. Hence, statistical features at every hour reflect the distribution of the seismic data within the same 6-hr-windows as the averaged GPS data. The meteorological data (temperature and precipitation) and water discharge measurements are applied by computing the average of 30 data points (meteorological data)/24 data points (water discharge measurements) corresponding to 6 hr of seismic data (1 data point is the average of the data during the previous 10 min/15 min) to obtain consistent feature time windows.

We extracted information from an extensive beamforming catalog which was developed using an advanced MFP localization scheme based on a gradient-decent optimization that meets the challenging, seismically "loud" environment. A complete detailed description on the methodology and the MFP implementation can be found in (Nanni et al., 2022). We used four sub-catalogs with center frequencies of 5, 10, 15, and 20 Hz. Each catalog was limited in x, y, z with respect to the dimension of the array and the depth of the glacier. The seismic velocities were limited to 1, 300–3, 800 $\frac{m}{s}$ and we expect that range to cover Rayleigh wave, P- and S-wave velocities within glacial ice (Podolskiy & Walter, 2016). We additionally reduced each catalog to normalized beam power values between 0.2 and 1.0. Figure 2 shows a 1 hr snapshot of a 10 Hz-catalog together with the ice surface and the bedrock topography. To use the high-resolution catalog results as features in the gradient tree boosting model for predicting displacement rates on the surface of Glacier d'Argentière, we spatially binned the MFP derived sources within 8 predefined source regions of the same ice volume (voxels V1–V8). Voxels 1–4 capture the deeper part of the glacier, close to its base, and voxels 5–8 capture the surface equivalent. For each voxel we sum the number of sources and sum their beam power respectively. For consistency with the other data, we apply a moving time window of 1 hr using ± 3 hr before and after the respective time stamp to match previous feature and label sampling. Virtual cut surfaces and voxel notations are indicated in Figure 2.

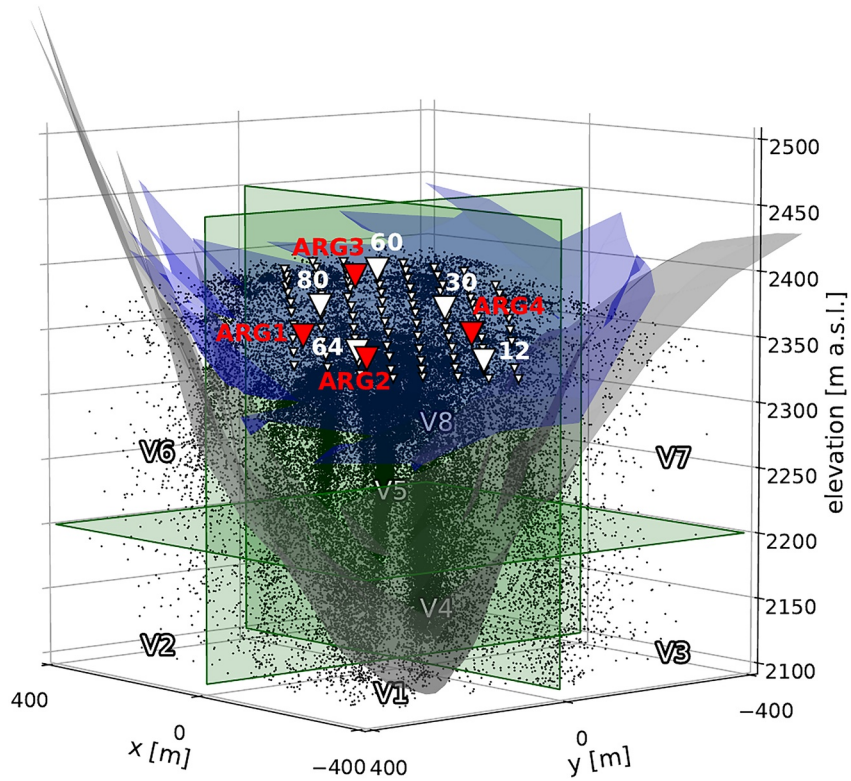

Figure 2. Snapshot of the thresholded beamforming catalog together with the drone derived ice surface (shades of blue) and the bedrock topography measured by radio-echo sounding (shades of gray). Black dots represent seismic source locations during 1 hr (temporal resolution of 1 s), for a center frequency of 10 Hz and beam power values between 0.2 and 1.0. The white triangles indicate the seismic array with the five heightened ones being the selected stations for the computation of the statistical features (12, 30, 60, 64, and 80). The red triangles display Global Positioning System (GPS) stations (ARG1–ARG4) situated within the seismic array. The green planes indicate the cut surfaces that divide the glacier into eight voxels (V1–V8) with V1–V4 capturing the lower part close to the glacier bed and V5–V8 encompassing portions of the ice surface.

2.4. eXtreme Gradient Boosting Model for Glacier d'Argentière

Gradient tree boosting (Friedman et al., 2000) is a widely used and scalable supervised machine learning approach. It is a very powerful tool that is based on, but usually outperforms, decision tree ensembles (Breiman, 2001; Chen & Guestrin, 2016). Decision tree ensembles use multiple shallow trees that can be built in a serial manner, in parallel or even independently from each other and combined in a next step in order to enhance model performance. The ensemble learner can be used for classification or regression problems. In order to predict a target variable (label), a model is trained based on simple decision rules learned from the data (feature). Depending on the purity of the individual leaves of the tree, the prediction is weighted through a comparison with the respective label. The deviation is represented by an arbitrary loss function. The model is trained sequentially by adding a gradient term to the current decision tree model iteration, with the aim to minimize the loss function for the weighted ensemble of all previous decision trees. Usually, trees that are added in each iteration are shallow (weak learners), but the full ensemble contains a large number of them in total quantity. Once the model is trained, the feature importance and SHapley Additive exPlanations (SHAP) values (Lundberg & Lee, 2017) can be evaluated to provide insight into the model predictions, thus allowing one to learn which input observations yield the best estimates on the output label. The SHAP values provide an independent assessment of the contribution for each feature as they sum to total the prediction value.

To estimate the GPS velocity (on-ice displacement rate) on the surface of Glacier d'Argentière, we develop a gradient boosted tree regression model using the features extracted from the data. Specifically, we use the XGBoost package and routines form scikit-learn (Chen & Guestrin, 2016; Pedregosa et al., 2011). When optimizing model hyperparameters the choice of data split and feature preprocessing is performed iteratively by solving the model using fivefold cross validation on the training data set to minimize the average

mean-squared-error for each validation fold. A Bayesian optimizer is implemented for a search space to select the best hyperparameters (Head et al., 2018). The procedure randomly selects hyperparameters for 100 iterations, then gradient descent is applied to converge on the best selection for an additional 100 iterations. Initially the search space is large, then expanded or narrowed for specific parameters to avoid final values converging at the upper and lower limits. For each optimization run the evolution of parameters is viewed to update the search space, then the procedure is repeated. The workflow is distributed on a GPU server to train multiple models with different hyperparameters simultaneously to select the final model based on convergence.

2.5. Model Development and Optimization

To analyze the ability of the model to perform predictions for data with a temporal limit to the training data, we experimented with different train/test splits. During the model development we experimented with a random train/test split, such that for each test sample, training samples in close temporal proximity are available. We subsequently increased the length of consecutive train/test intervals from 24 to 96 hr. As the time intervals increase the model needs to learn time-invariant, globally valid features to achieve good performance. For the more time dependent cases, we experimented with 50%/50% splits using the first half of the time series is used for training and the second half as test data, and 80%/20% splits where the first 80% of the time series is used for training and the last 20% for testing. Splitting the data to best represent the data proved critical to assess the model improvements due to non-stationary signals in the time series and avoid a testing data set that is out-of-distribution.

Different techniques applied to the feature preprocessing included standard scaling (S), quantile transformation (Q; $n\_quantiles = 50$), principal component analysis (P; $n\_components = 50$, $whiten = True$), and a random forest regressor (R; $n\_estimators = 200$, max $\_depth = 3$, $n\_features\_to\_select = 20$, $step = 1$). All routines are available in the scikit-learn package (Buitinck et al., 2013; Pedregosa et al., 2011). We optimized the hyperparameters for each type of split on the training fraction using the original data and each possible combination of S, Q, P, and R. The results show that the best-fit model hyperparameters with the lowest loss, hence, the best model, strongly depends on the choice and combination of data split and feature preprocessing. For each GPS velocity time series we evaluate the type of split with the choice of preprocessing and accordingly apply the respective model hyperparameters which yield the highest possible prediction score. Comparison of data and best-fit model are expressed through the coefficient of determination (R^2), the root-mean-squared-error (RMSE), and the correlation coefficient (CC). These metrics are applied to allow direct comparison between models and do not reflect the absolute quality of the results. To further improve predictions, we tested different applications of a low-pass filter to the GPS velocity time series to reduce the high-frequency "spiky" fluctuations inherent to the time series. The cutoff frequency was optimized to maximize the evaluation score.

3. Results

Results are presented to show the capability of the model to predict the velocity time series for all available RESOLVE GPS stations and specifically highlight the model predictions for three GPS stations that yield the highest scores (ARG2, ARG3 and ARGG in Figure 1). Additionally, we provide details on the best-fit model hyperparameters for station ARG3 considering the implementation of different data splits and feature preprocessing.

We assess the different types of data splitting for model evaluation and provide results for a direct comparison of the performance and robustness of each technique. In Figure 3 we show the prediction scores for CC and RMSE using the testing fraction of the GPS velocity time series of ARG3 for different short-term and long-term splits, and the dependence on feature smoothing window length between 1 and 24 hr to reduce bias from temporary signals. The results indicate the model performance is strongly dependent on the type of data split and that it generally improves when larger smoothing windows are applied. The best results are observed when using random splits, where the entire time series is shuffled before selecting the training and test data, with a CC > 0.8 for a smoothing window >5 hr. Increasing the smoothing window length further improves the predictions towards CC $= 1$ (Figure 3a). Similar metrics are observed with the RMSE < 1 using a window length > 7 hr and further decreasing to RMSE $= 0.25$ with longer windows (Figure 3b). These results show the best model fit but this split does not encourage the model to learn time-invariant features, as no predictions for data with large temporal distance to the training data have to be made.

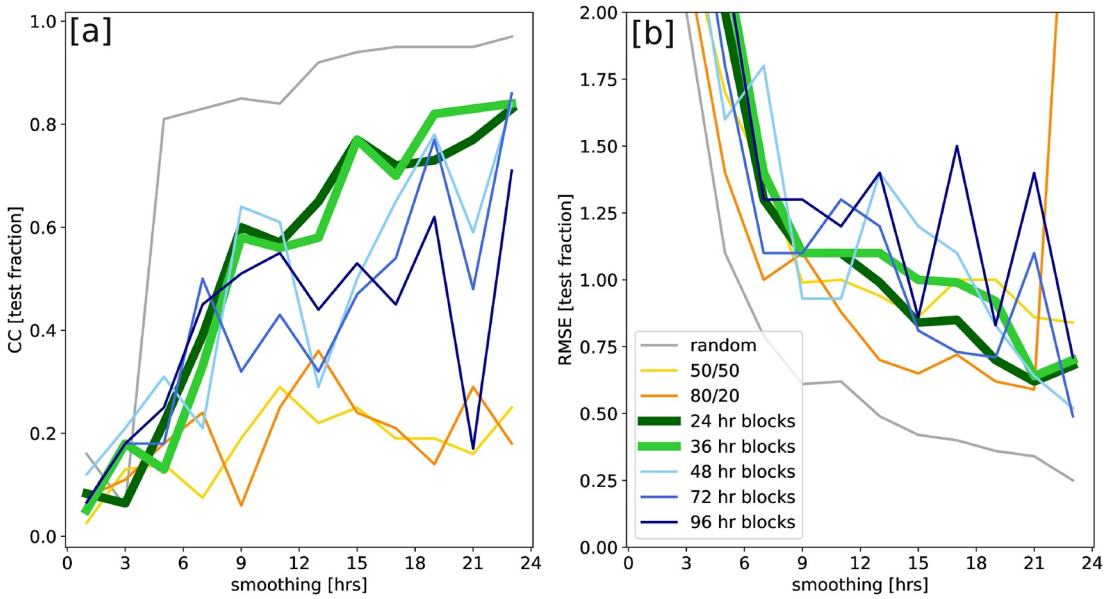

Figure 3. Model prediction scores ((a) CC and (b) RMSE) of the testing velocity time series of Global Positioning System (GPS) station ARG3 in dependence on the degree of smoothing window duration applied to the features. Blocks refer to the lengths of the train/test intervals. No additional feature preprocessing was applied.

To maintain temporal sequencing, we split the data into uniform temporal blocks with sizes between 24 and 96 hr for the entire duration of the data. The best results are found using temporal block sizes of 24 and 36 hr with CC > 0.6 (RMSE < 1.2) for a window length >9 hr. The results improve to CC = 0.8 (RMSE = 0.75) for the largest smoothing windows tested and increase with an approximate linear trend (Figure 3). Applying block sizes >36 hr show inconsistent, alternating behavior with little improvement above CC = 0.6 (RMSE < 0.6). This is consistent when using even larger fractions of the data for the long-term splits (80%/20% split, 50%/50% split). In general, these models show a CC < 0.2 (RMSE > 0.65) with the maxima derived using a smoothing window between 9 and 13 hr.

The sensitivity of the best-fit model hyperparameters when applying different data splits and feature preprocessing is illustrated in Figure 4 and show the variations in the hyperparameters are strongly dependent on and significantly differ for the type of split and the choice of feature preprocessing. Except for *n_estimators* ranging between about 1,000 and 1,200 for all types of split, no trend can be observed for other hyperparameters and types of split (Figure 4a). Just as the data split alters the model, different choices and combinations of feature preprocessing produce inconsistent model hyperparameters. Values of min_*child_weight* seem to be lower when less preprocessing is applied, but overall the response of the model hyperparameters shows no clear pattern for different choices of preprocessing. Comparing the models of three equally 50%/50% split GPS stations (ARG2, ARG3, ARGG) in terms of hyperparameters, preprocessing, and low-pass filtering, the results indicate that the requirements for the best prediction score (CC = 0.25 − 0.46) are fundamentally different and significantly influence the model performance (Figure 1).

3.1. Short-Term Sliding Predictions

The model predictions for the velocity time series of GPS station ARG3 using three different types of train/test split are shown in Figure 5. For the testing set, data versus model correlations are shown in the inset. Without any additional preprocessing applied to the features except smoothing to suppress noise (smoothing window of 15 hr), the random split yields outstanding performance (Figure 5a) with RMSE = 0.42, CC = 0.94, and $R^2 = 0.88$. The model is able to capture hourly fluctuations by randomly training and testing on the time sampling domain of the data (1 hr).

Next, we increase the length of train/test intervals in the range of 24 and 96 hr. The most robust and performant model with a reasonable agreement between smoothing and prediction score was achieved using blocks of 36 hr and smoothed features with a smoothing window of 15 hr (Figure 5b). Compared to using blocks of 24 hr, the

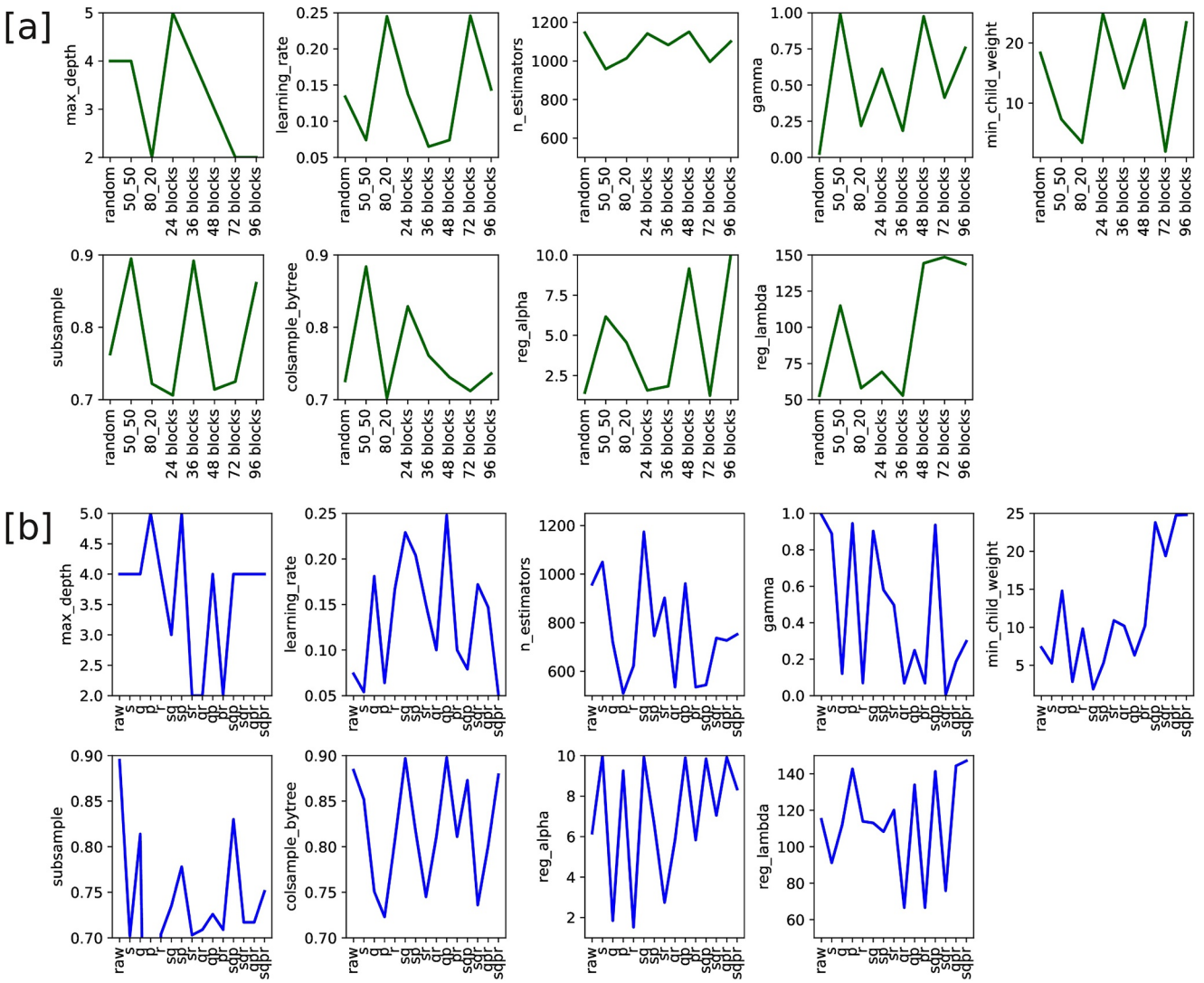

Figure 4. (a) Best-fit model hyperparameters optimized for different types of data split applied to the model of Global Positioning System (GPS) station ARG3. The data was used in the original format with no additional preprocessing applied. (b) Best-fit model hyperparameters optimized for the raw data and all available combinations of preprocessing S (standard scaling), Q (quantile transformation), P (principal component analysis), and R (random forest regressor). Data were split 50%/50%.

36 hr block split model shows slight deficiencies expressed by a lower RMSE in the range of about 0.2, which is not reflected by the CC. This marginal shortcoming is counterbalanced by the gain in block size from 24 to 36 hr, leading to a gain in prediction horizon of 12 hr which serves the scientific motivation of this study. Without any additional feature preprocessing applied, the model scores with RMSE = 0.84, CC = 0.75, and $R^2 = 0.5$ (Figure 5b). Apart from some infrequent failures and not fully capturing the amplitude at all times, the model is able to predict fluctuations with daily resolution.

3.2. Long-Term Sliding Predictions

With the aim to stretch the prediction horizon, we apply a 50%/50% split. Since the model seems to be less sensitive towards smoothing than the one using the 80%/20% split we consider it more robust (Figure 3). We train the model on the first half of the data and test it on the remaining half. Analogous to the short-term splits, we use the raw features and only apply a smoothing window length of 15 hr in a first iteration, which results in a significantly lower prediction score (CC < 0.4, RMSE > 0.9, see also Figure 3). Extensive feature preprocessing involving S, Q, P, and R, and the additional application of a low-pass filter with a cutoff frequency of 16.5 hr improves

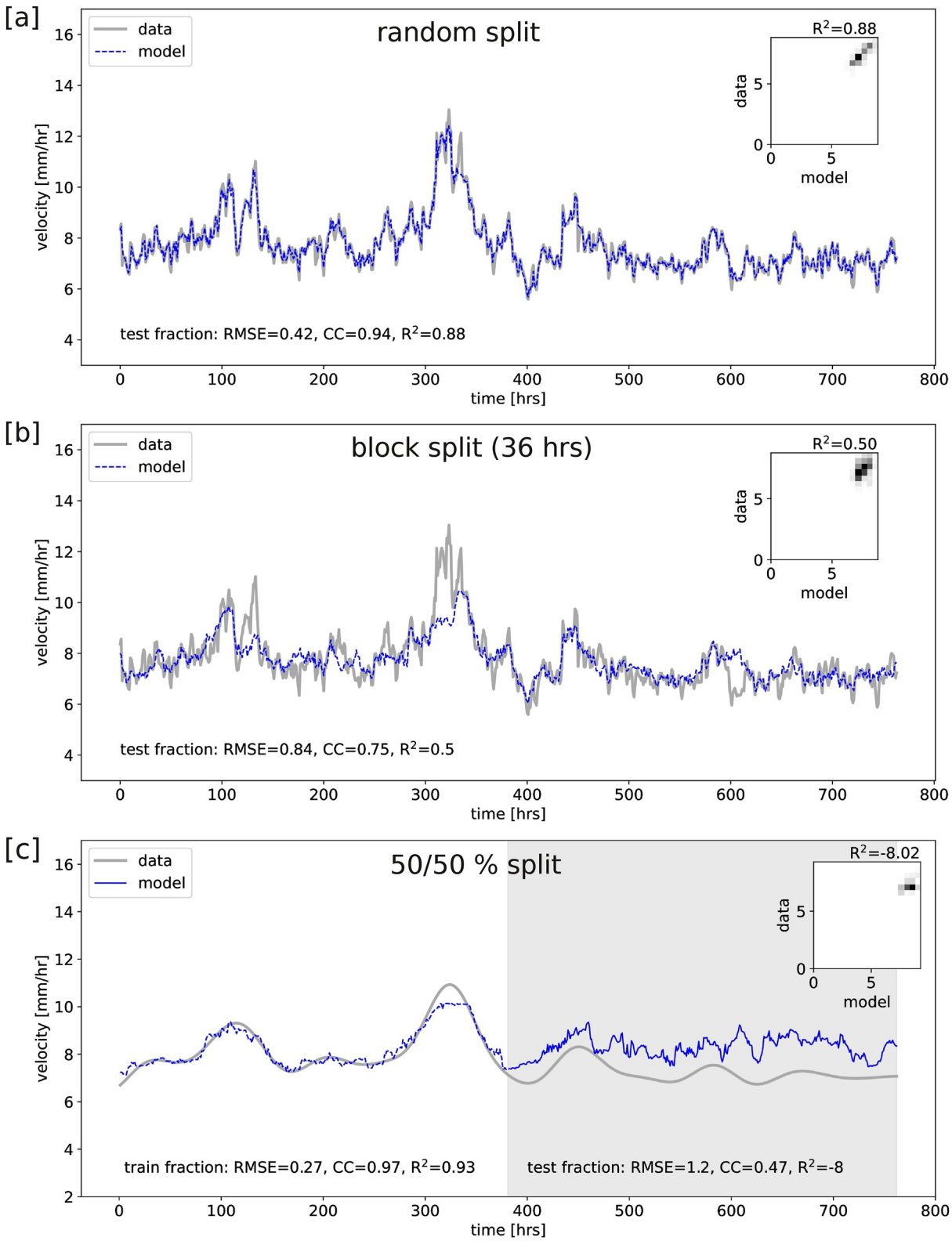

Figure 5.

Table 1
Overview of Best-Fit Model Hyperparameters, Choices of Preprocessing and Low-Pass Filters Applied to Global Positioning System (GPS) Stations ARG2, ARG3, and ARGG Using a 50%/50% Split.

		GPS stations		
		ARG2	ARG3	ARGG
Hyperparameters	max_depth	5	4	3
	learning_rate	0.052	0.051	0.052
	n_estimators	514	752	527
	gamma	0.816	0.298	0.696
	min_child_weight	1.28	24.803	23.946
	subsample	0.708	0.751	0.738
	colsample	0.771	0.879	0.732
	reg_alpha	9.849	8.349	1.929
	reg_lambda	100.271	147.171	134.623
Preprocessing low-pass filter (hr)		P	S, Q, P, R	–
		2.5	16.5	1.39
Correlation coefficient		0.25	0.46	0.37

the correlation coefficient up to CC = 0.47 (Figure 5c). While short-term dynamics cannot be captured by the model, it is able to predict the long-term behavior of the GPS velocity, notable the varying trend but with a static offset.

3.3. Sliding Predictions Across Glacier d'Argentière

Next we evaluate short-term model predictions (block split, 36 hr) and long-term model predictions (50%/50% split) for three GPS stations (ARG2, ARG3, and ARGG in Figure 1), which yield the highest prediction scores within the network. Unlike the feature preprocessing (S, Q, P, and R) and an additional lowpass filter for long-term predictions to enhance the model performance, here we only apply a smoothing window of 15 hr with no additional preprocessing.

For GPS station ARG2, which was located within the seismic array and situated close to seismic node 64 (Figure 2), we derive a model score of CC = 0.25 using a 50%/50% split with preprocessing P and a low-pass filter with a cutoff frequency of 2.5 hr applied (see Table 1 for best-fit model hyperparameters). The SHAP values show that statistical features of node 12 contributed most to the model (Figure A2a). It is important to note that node 12 was situated on the north-western flank of the glacier while ARG2 was located in the central-north close to the glacier tongue. The interstation distance and the model's decision "against" favoring features from the closest node 64 posit that the long-term behavior of the surface velocity of the ice is likely not locally driven by for example, an opening crevasse, but rather controlled by some seismic activity along the north-western flank. As displayed in Figure 6a short-term predictions (36 hr blocks, smoothing window of 15 hr) for ARG2 yield an increase in CC by a factor of 2.64. The SHAP values (Figure 7a) show that beamforming features replace statistical features when analyzing shorter time windows. Explicitly, the low-frequency source locations (5 Hz) within lower voxel V1 (Figure 2) contributed most to the model predictions followed by the skewness of seismic node 12 in the 30–40 Hz filter band.

GPS station ARG3 was situated in line with ARG2 and integrated with the seismic array as well. More precisely it is located next to seismic node 60. For long-term predictions (50%/50% split), we derive a model score of CC = 0.46 with preprocessing S, Q, P, and R and a low-pass filter with a cutoff frequency of 16.5 hr applied (see Table 1 for best-fit model hyperparameters). When comparing to ARG2, more preprocessing and a stronger filter are applied. As a result of the smoother GPS data from the low-pass filter data, the model shows a CC value with an increase by a factor of two. As revealed by the SHAP values and as for ARG2, statistics from node 12, situated at the north-western flank of the glacier, are of upmost importance (Figure A2b). Figure 6b shows the equivalent short-term predictions (36 hr blocks, smoothing window of 15 hr) for ARG3 which result in CC = 0.75. Again, statistical features important for long-term predictions are here replaced by low-frequency beamforming features (5 Hz) from the lower voxels V1 and V2 (Figures 2 and 7b). Additionally, the 0.5 interquartile range of the 30–40 Hz filtered record of seismic node 80 strongly contributes to the model predictions.

For GPS station ARGG situated within the accumulation zone of the glacier <3 km north-west from the seismic array, the best long-term model score based on a 50%/50% split is CC = 0.37 with a low-pass filter with a cutoff frequency of 1.39 hr applied (see Table 1 for best-fit model hyperparameters). The data features were best suited in the original format (no preprocessing) using the filter to suppress short-term dynamics. As for ARG2 and ARG3, the SHAP values again indicate the importance of statistical features from seismic node 12 (Figure A2c). For short-term predictions (36 hr block split) we derive a CC = 0.6, which is mostly dependent

Figure 5. Performance of the eXtreme Gradient Boosting (XGB) model to predict surface velocity (mm/hr) trained on Global Positioning System (GPS) station ARG3 (see Figures 1 and 2), which was situated within the seismic array in the ablation zone of Glacier d'Argentière. The model performance is compared for different types of split applying the respective best-fit model hyperparameters. The data curve is displayed in gray and the model predictions in blue. The inset plots in each panel show the bi-dimensional histogram plots comparing data and model. (a) The model was trained and tested on random samples. No additional preprocessing was applied. (b) The model was trained and tested using blocks of 36 hr. No additional preprocessing was applied. (c) The model was trained on 50% of the velocity time series (white facecolor) and tested on the remaining 50% (gray facecolor). Preprocessing involved S, Q, P, R, and a low-pass filter with a cutoff frequency = 16.5 hr was applied.

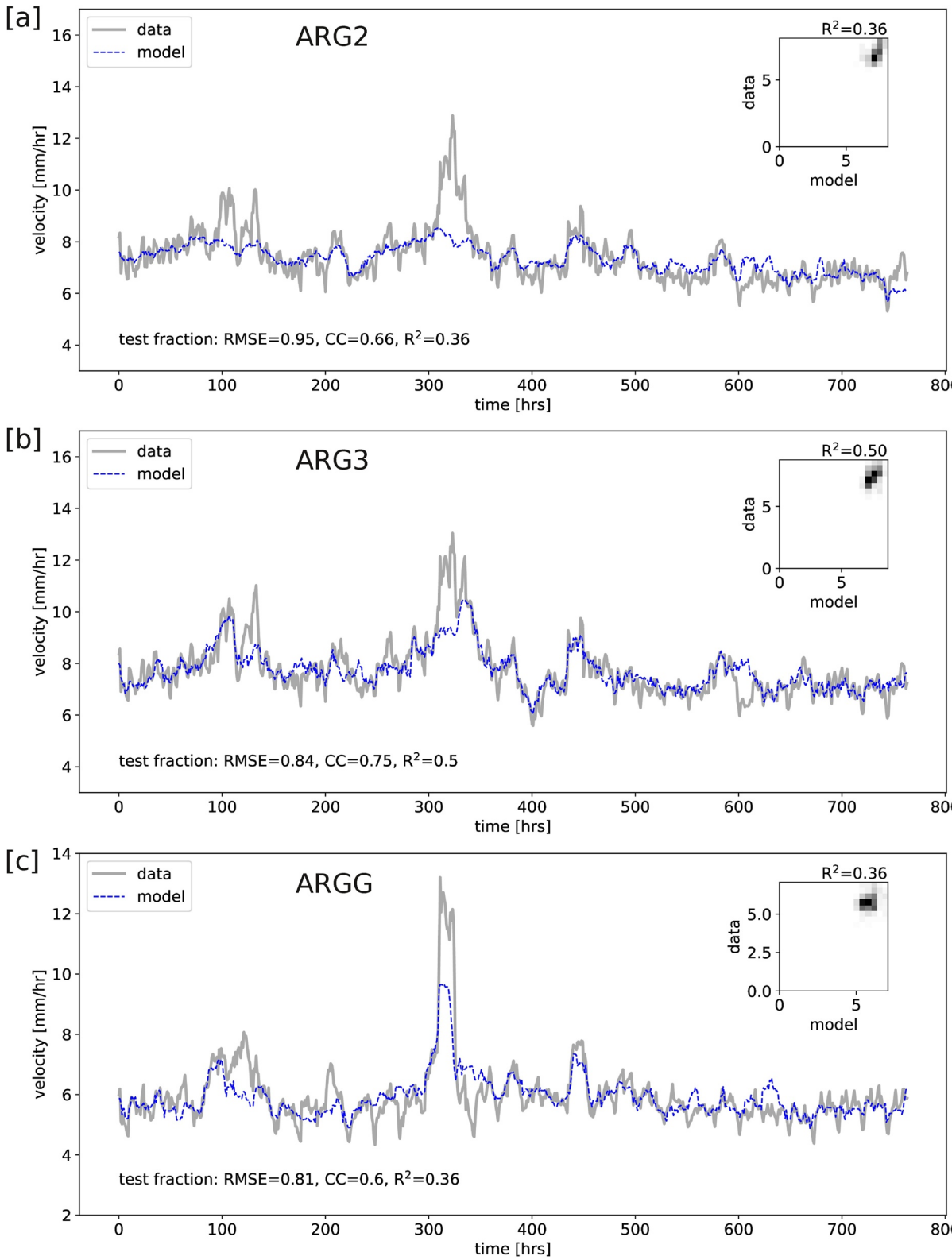

Figure 6. Performance of the eXtreme Gradient Boosting (XGB) model to predict surface velocity (mm/hr) using blocks of 36 hr for training and subsequent testing. The model was trained and tested on Global Positioning System (GPS) station ARG2 (a), ARG3 (b), and ARGG (c) (see Figures 1 and 2). The best-fit model hyperparameters were optimized for all three models and data were smoothed over 15 hr. No additional preprocessing was applied.

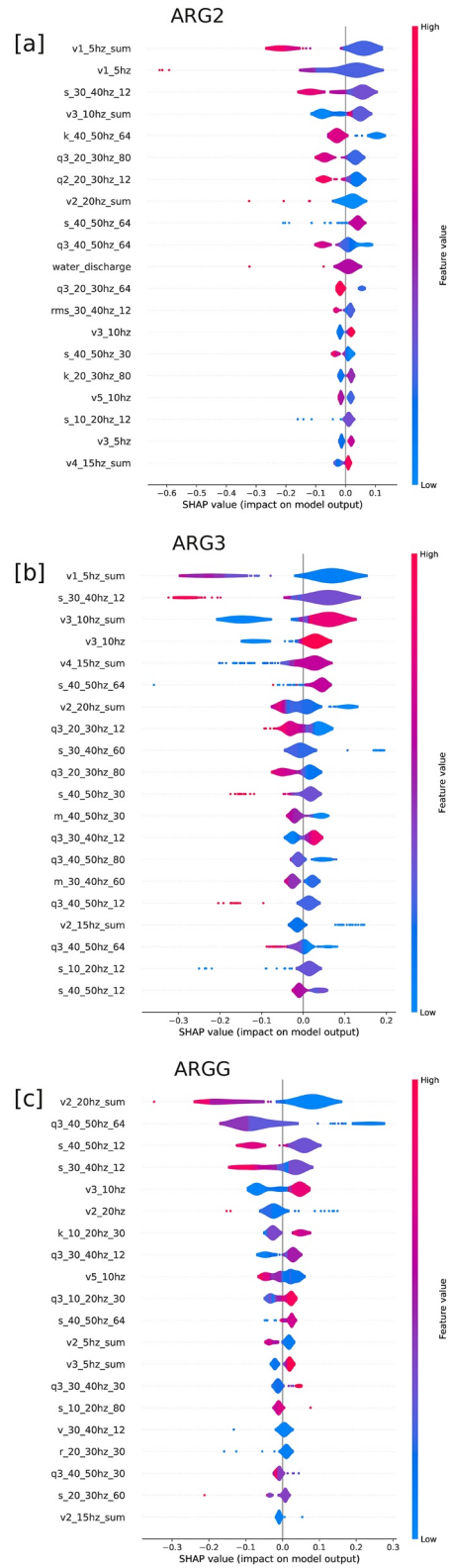

Figure 7. SHapley Additive exPlanations (SHAP) values in descending order by importance for model predictions using a block sampling size of 36 hr as shown in Figure 6: (a) ARG2, (b) ARG3, and (c) ARGG. The model prediction for each input is a summation of the SHAP values. The color indicates the high or low feature value that contributes to the model prediction. The feature abbreviations are explained in Table A1.

on beamforming features of the lower voxel V2 (Figure 2) in the 20–30 Hz filter band followed by the 0.5 interquartile range of the 40–50 Hz filtered record of seismic node 64 and the skewness of the 40–50 Hz filtered record of seismic node 12.

The long-term model results and the related feature ranks for the three GPS stations analyzed show consistent results that suggest glacial surface velocity is being controlled by activity at the north-western flank of the glacier. Interestingly, the meteorological features and surface beamforming voxels generally play a subordinate role for the model estimates. For short-term model predictions we observe that beamforming features of the lower voxels close to the glacier bed are most important followed by high-frequency statistical features (30–50 Hz), such as the 0.5 interquantile range and the skewness.

4. Discussion

The application of machine learning using continuous seismic records continues to show success in describing physical processes of complex natural systems. While the glacier motion model predictions are not as robust as those for laboratory stick-slip studies (Corbi et al., 2019; Jasperson et al., 2021; Rouet-LeDuc, Hulbert, Bolton, et al., 2018; Shokouhi et al., 2021; Wang et al., 2021), slow slip in Earth (Hulbert et al., 2020), future prediction (Laurenti et al., 2022; Wang et al., 2022), or stick-slip processes in Earth (Johnson & Johnson, 2021), they are nonetheless predictive for the long-term sliding behavior and especially performant for short-term variations. Ice deformation is considered mostly aseismic through viscous creep (Gimbert et al., 2021), which is inherent to the material properties. The data features are designed to capture such deformation using information in the continuous signal emitted from internally deforming slip boundaries during viscous flow, which occurs at a range of pressures and temperatures. The glacial system dynamics are highly complex and variations in signals produced by the sources of noise appear to be more heterogeneous than in a laboratory system or an earthquake fault.

This study shows that on-ice glacial surface displacement rates can be linked to distinct areas, and even in-depth activity, of a temperate Alpine glacier based on the seismic beamforming features. The addition of seismic beamforming as a data feature provides additional information to the model space and enables the estimate of surface displacement rates on Alpine glacial ice in an highly dynamic and noise-prone environment, and the ability to locate its driving process. To our current state of knowledge, basal motion is most likely the driver for deep cryoseismogenic processes which drive the displacement rates at the surface of Glacier d'Argentière and outrivals internal deformation through viscous creep due to its strong seismic fingerprint (Podolskiy & Walter, 2016).

Data splits strongly influence the decision tree models outcome with sample-wise or short-term train/test fractions leading to the highest prediction scores and longer train/test fractions to a subsequent decrease in performance together with a loss in robustness. Even though the short-term models outperform the long-term models in terms of evaluation metrics, they provide less insights into the physics and dynamics of glacial sliding. Hence, there is a tradeoff between model performance and long-term predictions. We found that the best agreement between prediction horizon and model performance is given by using block splits with block sizes of 24–36 hr (CC = 0.75). Those models are robust towards feature smoothing, meaning that within each block the dynamics and processes are similar and hence "understandable" for the model.

The best long-term model captures the long wavelength characteristics, suggesting that the highly variable temporal fluctuations are generated by a number of incoherent processes and the model cannot isolate into these unique characteristics in the feature space. A possible cause is the seismic features contain a combination of information from multiple weak processes and expanding the feature space might improve the high frequency estimates. With the current best model and features, the surface ice velocity can be predicted with an accuracy of up to 46% for the longer term behavior in the range of 16.5 hr.

Intensively studying the hyperparameter space and the dependence on data split, different choices of preprocessing and low-pass filters shows that each station-related model has to be tuned independently and model settings may not be generalized in the Alpine cryospheric environment. We found that individual station estimates generally score better than averages of multiple on-ice velocity time series and that bedrock stations were less suited for the analysis.

Overall, we observe that the relevant features for model predictions differ for GPS stations that were situated in the noise-prone ablation zone (ARG2, ARG3) compared to ARGG, which was situated in the accumulation zone. For ARGG, less influenced by cryoseismic sources as for example, crevassing or water flow, which can potentially mask in-depth activity of the glacier, the long-term model and short-term model can both pick up processes at the glacier's base relevant for sliding (lower beamforming voxels V2 and V4, 20–30 Hz). For long-term model predictions of ARG2 and ARG3 those features are revised by statistical features, as they potentially reflect the dominant local sources such as crevassing or water flow. The short-term models of ARG2 and ARG3 however capture in-depth activity. We observe consistent results for both stations in favoring low-frequency beamforming features from the bottom voxels V1 and V2 (5 Hz).

The RESOLVE experiment design was most advantageous for capturing the spatiotemporal seismic and geodetic behavior driven by glacial processes in the 1 month of data collection. Limitations to the seismic and geodetic measurements as applied to this analysis include the discrepancy in sampling rate (500 Hz for seismic observations vs. 1 hr for geodetic observations). This mismatch requires several steps of preprocessing to properly align the data features and labels, specifically the moving time window analysis and smoothing of the time series data or the compilation of the highly resolved beamforming catalog. Those procedures come with a potential loss of information regarding short-term variations of the glacier's activity. Furthermore, seismic observations were solely collected in the ablation zone of the glacier, while GPS station coverage spanned over the entire length of Glacier d'Argentière (<3 km). The accumulation zone of temperate based glacial ice is typically less active than the ablation zone. The ablation zone, however, is characterized by a multitude of physical processes such as crevasse formation, meltwater flow or avalanches and rockfalls provoked by increasing temperatures in lower altitudes (Nanni et al., 2022). Even though the geodetic observations show coherent behavior across the network and the glacier's extent (Figure A1), model predictions of distant stations which were situated in the accumulation zone may be challenged due to regime differences. Compared to predictions made on GPS stations which were integrated with the seismic array, model estimates of high-altitude geodetic observations show reasonable performance, but might have benefited from nearby seismic observations.

The mild power threshold of the beamforming catalog (0.2–1.0) subsequently leads to the integration of poorly resolved seismic sources in our analysis which poses the risk to decrease the model performance due to random, physically unconstrained locations. However, in view of the high noise level in Alpine glacial environments, locations with a lower resolution likely carry relevant information from deep processes at the glacier bed, as for example, basal stick-slip (Umlauft et al., 2021) or subglacial water flow (Nanni et al., 2020). As revealed by the feature importance for model estimates of GPS ARGG (Figure 7c) the 20 Hz beamforming catalog as applied to this analysis carries information enabling the best model prediction. The surface displacement itself but also the center frequency of the catalog reasons that glacier basal motion, potentially coupled with subglacial water flow, is most likely the driving mechanism for the displacement of ARGG, as pure subglacial water flow is characterized by lower frequencies (3–7 Hz) (Nanni et al., 2020) and does not ultimately lead to surface displacement.

We have learned that this line of analysis could potentially contribute to an improvement of glacial sliding laws by considering relevant drivers for model parameterizations that are revealed by the feature importance.

5. Conclusions

A profound understanding and the formulation of sliding laws for glacier basal motion are still a major challenge for the scientific community and needed for hazard assessment and the generation of new prediction models. Especially for temperate glaciers in Alpine regions, sliding is difficult to monitor with conventional geophysical approaches. On-ice seismological records prove to be a very rich archive of glacial activity, but due to glacial noise from other cryoseismic sources, stick-slip events and tremors are often masked and remain unnoticed. New approaches are needed which involve on-ice seismological measurements densely sampled in space and time, as well as modern tools that efficiently analyze such large data sets and reveal previously hidden signals.

We applied a supervised ML approach gradient tree boosting to a seismic array data set acquired in course of the RESOLVE project on Glacier d'Argentière and showed its general suitability for the identification of seismic

signatures of ice beds in the presence of melt-induced microseismic noise. The analysis is designed to verify if model estimates are driven by basal motion. Our results demonstrate that gradient tree boosting is a suitable tool to estimate ice surface displacement rates from seismic data collected at glaciers and that information about basal processes can be accessed from on-ice seismometers, analogous to frictional behavior of tectonic fault zones, at least at long period. We have learned that other than for quiet laboratory faults (Rouet-LeDuc, Hulbert, Bolton, et al., 2018) or reasonably long monitoring time series along tectonic faults (Johnson & Johnson, 2021; Rouet-LeDuc, Hulbert, & Johnson, 2018), using only statistical properties of continuous seismic records are not sufficient to describe glacial environments. We adapted the ML model by creating expressive beamforming features using array processing that meet the challenging, seismically "loud" environment. As revealed by the feature importance, the spatiotemporal compilation of seismic source locations provides the essential information for the model to relate estimates of surface velocities to in-depth activity.

Appendix A: RESOLVE GPS Analysis

The RESOLVE GPS (Global Navigation Satellite System (GNSS)) analysis has been performed by a static, differential positioning using the GAMIT software (Herring et al., 2018) in a network combining the five RESOLVE GNSS stations (ARG1–4 on the glacier and ARG5 beside the glacier on the bedrock), plus the ISTerre long-term station ARGG on Glacier d'Argentière outside the RESOLVE network, with 14 permanent and stable RENAG (http://renag.resif.fr) stations in less than 180 km distance (including ARGR on bedrock close to Glacier d'Argentière at 3 km distance from the RESOLVE network). This set of stations has been analyzed in 6-hr-sessions (corresponding to 30–40 mm of displacement of stations on Glacier d'Argentière) shifted by 1 hr to obtain hourly positions for each of the stations. The formal uncertainties of each of the position estimates are 2–3 mm on the horizontal components. The positioning of the bedrock site ARG5, close to the glacier stations and therefore in a comparable environment, indicates a dispersion of 4–6 mm. This value is probably a realistic estimate of the hourly positioning precision of the glacier stations.

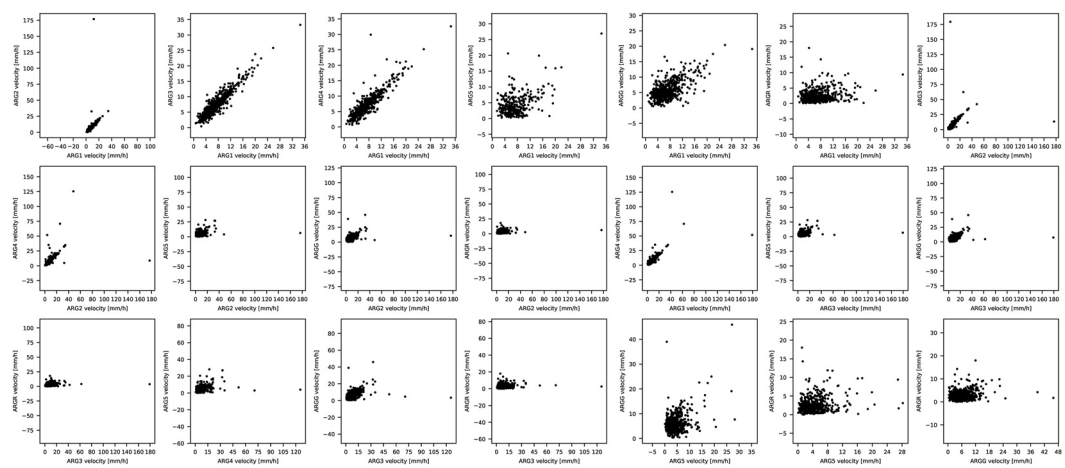

Figure A1. Station-wise correlation of geodetic observations.

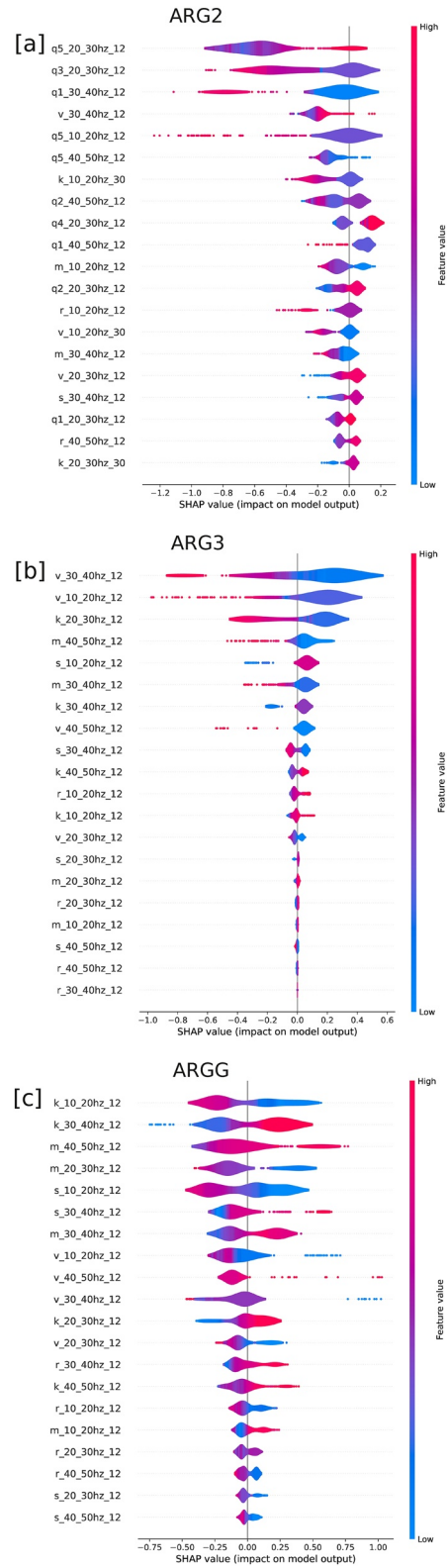

Figure A2. Feature importance (SHapley Additive exPlanations (SHAP) values) for model predictions using a 50%/50% split on the velocity time series of (a) ARG2, (b) ARG3, and (c) ARGG. The feature abbreviations are explained in Table A1.

Journal of Geophysical Research: Earth Surface

10.1029/2023JF007280

Table A1
Overview of All Features Used as Input for the eXtreme Gradient Boosting (XGB) Model

	Abbreviation	Parameter	Frequency bands (F1–F2) in Hz	Locations (L)
Statistical features	v_*F1*_*F2*hz_*L*	Variance	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
	k_*F1*_*F2*hz_*L*	Kurtosis	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
	m_*F1*_*F2*hz_*L*	Mean	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
	s_*F1*_*F2*hz_*L*	Skewness	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
	rms_*F1*_*F2*hz_*L*	Root mean square	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
	r_*F1*_*F2*hz_*L*	Range	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
	q1_*F1*_*F2*hz_*L*	Interquantile range [0.05]	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
	q2_*F1*_*F2*hz_*L*	Interquantile range [0.25]	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
	q3_*F1*_*F2*hz_*L*	Interquantile range [0.5]	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
	q4_*F1*_*F2*hz_*L*	Interquantile range [0.75]	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
	q5_*F1*_*F2*hz_*L*	Interquantile range [0.95]	10–20, 20–30, 30–40, 40–50	Seismic stations 12, 30, 60, 64, 80
Beamforming features	v1_*F1*_*F2*hz	MFP output	10–20, 20–30, 30–40, 40–50	Beamforming voxel 1
	v2_*F1*_*F2*hz	MFP output	10–20, 20–30, 30–40, 40–50	Beamforming voxel 2
	v3_*F1*_*F2*hz	MFP output	10–20, 20–30, 30–40, 40–50	Beamforming voxel 3
	v4_*F1*_*F2*hz	MFP output	10–20, 20–30, 30–40, 40–50	Beamforming voxel 4
	v5_*F1*_*F2*hz	MFP output	10–20, 20–30, 30–40, 40–50	Beamforming voxel 5
	v6_*F1*_*F2*hz	MFP output	10–20, 20–30, 30–40, 40–50	Beamforming voxel 6
	v7_*F1*_*F2*hz	MFP output	10–20, 20–30, 30–40, 40–50	Beamforming voxel 7
	v8_*F1*_*F2*hz	MFP output	10–20, 20–30, 30–40, 40–50	Beamforming voxel 8
Meteorological features	Temperature	Air temperature		Weather station
	Precipitation	Precipitation		Weather station
	Water discharge	Water discharge		Borehole gauging station

Data Availability Statement

The MFP source codes are described and available via https://lecoinal.gricad-pages.univ-grenoble-alpes.fr/resolve/ (last access: 11/11/2021) under a creative commons attribution 4.0 international license. The data derived from the MFP analysis (i.e., 29 sources localizations per second over 34 days and for 20 frequency bands) together with 1 day of raw seismic signal recorded over the 98 seismic stations are available via https://doi.org/10.5281/zenodo.5645545 under a creative commons attribution 4.0 international license (Nanni et al., 2021). The complete set of raw seismic data can be found at https://doi.org/10.15778/resif.zo2018 under a creative commons attribution 4.0 international license. The processed GPS (GNSS) velocity time series data is stored at https://zenodo.org/records/10013212 under a creative commons attribution 4.0 international license.

Acknowledgments

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany and by the Sächsische Staatsministerium für Wissenschaft Kultur und Tourismus in the program Center of Excellence for AI-research "Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig", project identification number: ScaDS.AI. CWJ and PAJ acknowledge support by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under Grant 89233218CNA000001. CWJ also acknowledges Institutional Support (Laboratory Directed Research and Development) at the Los Alamos National Laboratory. This work has been conducted in the framework of the RESOLVE Project (https://resolve.osug.fr/; LabEx OSUG@2020, Investissement d'avenir—ANR10LABX56 and IDEX Université Grenoble Alpes). The seismic nodes and GPS stations were provided and operated by the French seismological and geodetic network Rèsif-Epos. The computations of the beamforming catalog were performed using the GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr), which is supported by the Grenoble research communities, and with the CiGri tool (https://github.com/oar-team/cigri) that was developed by GRICAD, Grid5000 (https://www.grid5000.fr) and LIG (https://www.liglab.fr/). The machine learning model was developed using the resources of the Leipzig University Computing Centre. Open Access funding enabled and organized by Projekt DEAL.

References

Aster, R., & Winberry, R. (2017). Glacial seismology. *Reports on Progress in Physics*, *80*(12), 126801. https://doi.org/10.1088/1361-6633/aa8473

Baggeroer, A., Kuperman, W., & Mikhalevsky, P. (1993). An overview of matched field methods in ocean acoustics. *IEEE Journal of Oceanic Engineering*, *18*(4), 401–424. https://doi.org/10.1109/48.262292

Barcheck, C., Tulaczyk, S., Schwartz, S., Walter, J., & Winberry, J. (2019). Implications of basal micro-earthquakes and tremor for ice stream mechanics: Stick-slip basal sliding and till erosion. *Earth and Planetary Science Letters*, *486*, 54–60. https://doi.org/10.1016/j.epsl.2017.12.046

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Bucker, H. (1976). Use of calculated sound fields and matched-field detection to locate sound sources in shallow water. *Journal of the Acoustical Society of America*, *59*(2), 368–373. https://doi.org/10.1121/1.380872

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., et al. (2013). API design for machine learning software: Experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for data mining and machine learning*, Prague, Czech Republic, September 2013 (pp. 108–122). ffhal-00856511f

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). ACM.

Corbi, F., Sandri, L., Bedford, J., Funiciello, F., Brizzi, S., Rosenau, M., & Lallemand, S. (2019). Machine learning can predict the timing and size of analog earthquakes. *Geophysical Research Letters*, *46*(3), 1303–1311. https://doi.org/10.1029/2018gl081251

Corciulo, M., Roux, P., Campillo, M., Dubucq, D., & Kuperman, W. (2012). Multiscale matched-field processing for noise-source localization in exploration geophysics. *Geophysics*, *77*(5), KS33–KS41. https://doi.org/10.1190/geo2011-0438.1

Cros, E., Roux, P., Vandemeulebrouck, J., & Kedar, S. (2011). Locating hydrothermal acoustic sources at Old Faithful Geyser using matched field processing. *Geophysical Journal International*, *187*(1), 385–393. https://doi.org/10.1111/j.1365-246x.2011.05147.x

Cuffey, K., & Paterson, W. (2010). *The physics of glaciers*. Academic Press.

Dorostkar, O., Guyer, R., Johnson, P., Marone, C., & Carmeliet, J. (2017). On the micromechanics of slip events in sheared, fluid-saturated fault gouge. *Geophysical Research Letters*, *44*(12), 6101–6108. https://doi.org/10.1002/2017gl073768

Eibl, E., Bean, C., Einarsson, B., Pàlsson, F., & Vogfjörd, K. (2020). Seismic ground vibrations give advanced early-warning of subglacial floods. *Nature Communications*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-15744-5

Faillettaz, J., Funk, M., & Vincent, C. (2015). Avalanching glacier instabilities: Review on processes and early warning perspectives. *Reviews of Geophysics*, *53*(2), 203–224. https://doi.org/10.1002/2014rg000466

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *Annals of Statistics*, *28*(2), 337–407. https://doi.org/10.1214/aos/1016218223

Gagliardini, O., Cohen, D., Råbback, P., & Zwinger, T. (2007). Finite-element modeling of subglacial cavities and related friction law. *Journal of Geophysical Research*, *112*, 1–11. https://doi.org/10.1029/2006JF000576

Gimbert, F., Nanni, U., Roux, P., Helmstetter, A., Garambois, S., Lecointre, A., et al. (2021). A multi-physics experiment with a temporary dense seismic array on the Argentière glacier, French Alps: The RESOLVE project. *Seismological Society of America*, *92*(2A), 1185–1201. https://doi.org/10.1785/0220200280

Glen, J. (1955). The creep of polycrystalline ice. *Proceeding of the Royal Society of London, Series A*, *288*, 519–538. https://doi.org/10.1098/rspa.1955.0066

Gradon, C., Moreau, L., Roux, P., & Ben-Zion, J. (2019). Analysis of surface and seismic sources in dense array data with match field processing and Markov chain Monte Carlo sampling. *Geophysical Journal International*, *218*(2), 1044–1056. https://doi.org/10.1093/gji/ggz224

Gräff, D., Knöpfli, M., Lipovsky, B., Selvadurai, P., Farinotti, D., & Walter, F. (2021). Fine structure of microseismic glacial stick-slip. *Geophysical Research Letters*, *48*(22), e2021GL096043. https://doi.org/10.1029/2021gl096043

Head, T., MeachCoder, Louppe, G., Shcherbatyi, I., fcharras, Vinícius, Z., et al. (2018). *scikit-optimize/scikit-optimize: v0.5.2*. Zenodo.

Helanow, C., Iverson, N., Woodard, J., & Zoet, L. (2021). A slip law for hard-bedded glaciers derived from observed bed topography. *Science Advances*, *7*(20), 1–8. https://doi.org/10.1126/sciadv.abe7798

Helmstetter, A., Nicolas, B., Comon, P., & Gay, M. (2020). Basal icequakes recorded beneath an Alpine glacier (Glacier d'Argentière, Mont Blanc, France): Evidence for stick-slip motion? *Science*, *368*(6486), 76–78.

Herring, T., King, R., Floyd, M. A., & McClusky, S. (2018). *GAMIT reference manual, GPS analysis MIT, release 10.7*. Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology.

Hudson, T., Brisbourne, A., Walter, F., Gräff, D., White, R., & Smith, A. (2020). Icequake source mechanisms for studying glacial sliding. *Journal of Geophysical Research: Solid Earth*, *125*(11), e2020JF005627. https://doi.org/10.1029/2020jf005627

Hulbert, C., Rouet-LeDuc, B., Johnson, P., Ren, C., Riviere, J., Bolton, D., & Marone, C. (2019). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, *12*(1), 69–74. https://doi.org/10.1038/s41561-018-0272-8

Hulbert, C., Rouet-Leduc, B., Jolivet, R., & Johnson, P. (2020). An exponential build-up in seismic energy suggests a months-long nucleation of slow slip in Cascadia. *Nature Communications*, *11*(1), 4139. https://doi.org/10.1038/s41467-020-17754-9

Iken, A. (1981). The effect of the subglacial water pressure on the sliding velocity of a glacier in an idealized numerical model. *Journal of Glaciology*, *27*(97), 407–421. https://doi.org/10.3189/s0022143000011448

Jasperson, H., Bolton, D., Johnson, P., Guyer, R., Marone, C., & de Hoop, M. (2021). Attention network forecasts time-to-failure in laboratory shear experiments. *Journal of Geophysical Research: Solid Earth*, *126*(11), e2021JB022195. https://doi.org/10.1029/2021JB022195

Johnson, C. W., Ben-Zion, Y., Meng, H., & Vernon, F. (2020). Identifying different classes of seismic noise signals using unsupervised learning. *Geophysical Research Letters*, *47*(15), e2020GL088353. https://doi.org/10.1029/2020GL088353

Johnson, C. W., & Johnson, P. A. (2021). Learning the low frequency earthquake daily intensity on the central San Andreas Fault. *Geophysical Research Letters*, *48*(15), e2021GL092951. https://doi.org/10.1029/2021GL092951

Kaproth, B., & Marone, C. (2013). Slow earthquakes, preseismic velocity changes, and the origin of slow frictional stick-slip. *Science*, *341*(6151), 1229–1232. https://doi.org/10.1126/science.1239577

Kufner, S., Brisbourne, A., Smith, A., Hudson, T., Murray, T., Schlegel, R., et al. (2021). Not all icequakes are created equal: Basal icequakes suggest diverse bed deformation mechanisms at Rutford Ice Stream, West Antarctica. *Journal of Geophysical Research: Earth Surface*, *126*(3), 1–17. https://doi.org/10.1029/2020jf006001

Kuperman, W., & Turek, G. (1997). Matched field acoustics. *Mechanical Systems and Signal Processing*, *11*(1), 141–148. https://doi.org/10.1006/mssp.1996.0066

Laurenti, L., Tinti, E., Galasso, F., Franco, L., & Marone, C. (2022). Deep learning for laboratory earthquake prediction and autoregressive forecasting of fault zone stress. *Earth and Planetary Science Letters*, *598*, 117825. https://doi.org/10.1016/j.epsl.2022.117825

Lipovsky, B., & Dunham, E. (2016). Tremor during ice-stream stick slip. *The Cryosphere*, *10*(1), 385–399. https://doi.org/10.5194/tc-10-385-2016

Lipovsky, B., & Dunham, E. (2017). Slow-slip events on the Whillans Ice Plain, Antarctica, described using rate-and-state friction as an ice stream sliding law. *Journal of Geophysical Research: Earth Surface*, *122*(4), 973–1003. https://doi.org/10.1002/2016jf004183

Lipovsky, B., Meyer, C., Zoet, L., McCarthy, C., Hansen, D., Rempel, A., & Gimbert, F. (2019). Glacier sliding, seismicity and sediment entrainment. *Annals of Glaciology*, *60*(79), 182–192. https://doi.org/10.1017/aog.2019.24

Lundberg, S., & Lee, S. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30* (pp. 1–10).

Madariaga, R., & Ruiz, S. (2016). Earthquake dynamics on circular faults: A review 1970–2015. *Journal of Seismology*, *20*(4), 1235–1252. https://doi.org/10.1007/s10950-016-9590-8

Marone, C. (1998). Laboratory-derived friction laws and their application to seismic faulting. *Annual Review of Earth and Planetary Sciences*, *26*(1), 643–696. https://doi.org/10.1146/annurev.earth.26.1.643

McBrearty, I., Zoet, L., & Anandakrishnan, S. (2020). Basal seismicity of the Northeast Greenland Ice stream. *Journal of Glaciology*, *66*(257), 1–17. https://doi.org/10.1017/jog.2020.17

Murray, T. (1997). Assessing the paradigm shift: Deformable glacier beds. *Quaternary Science Reviews*, *16*(9), 995–1016. https://doi.org/10.1016/s0277-3791(97)00030-9

Nanni, U., Gimbert, F., Roux, P., & Lecointre, A. (2021). Observing the subglacial hydrology network and its dynamics with a dense seismic array. *Proceedings of the National Academy of Sciences*, *118*(28), e2023757118. https://doi.org/10.1073/pnas.2023757118

Nanni, U., Gimbert, F., Roux, P., & Lecointre, A. (2022). Dynamic imaging of glacier structures at high-resolution using source localization with a dense seismic array. *Geophysical Research Letters*, *49*(6), 1–9. https://doi.org/10.1029/2021gl095996

Nanni, U., Gimbert, F., Vincent, C., Gräff, D., Walter, F., Piard, L., & Moreau, L. (2020). Quantification of seasonal and diurnal dynamics of subglacial channels using seismic observations on an Alpine glacier. *The Cryosphere*, *14*(5), 1475–1496. https://doi.org/10.5194/tc-14-1475-2020

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.

Podolskiy, E., & Walter, F. (2016). Cryoseismology. *Reviews of Geophysics*, *54*(4), 708–758. https://doi.org/10.1002/2016rg000526

Rabinowicz, E. (1956). Autocorrelation analysis of the sliding process. *Journal of Applied Physics*, *27*(2), 131–135. https://doi.org/10.1063/1.1722321

Ren, C., Peltier, A., Ferrazzini, V., Rouet-Leduc, B., Johnson, P., & Brenguier, F. (2020). Machine learning reveals the seismic signature of eruptive behavior at Piton de la Fournaise volcano. *Geophysical Research Letters*, *47*(3), e2019GL085523. https://doi.org/10.1029/2019gl085523

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

Röösli, C., Walter, F., Husen, S., Andrews, L., Lüthi, M., Catania, G., & Kissling, E. (2014). Sustained seismic tremors and icequakes detected in the ablation zone of the Greenland Ice Sheet. *Journal of Glaciology*, *60*(221), 563–575. https://doi.org/10.3189/2014jog13j210

Rouet-LeDuc, B., Hulbert, C., Bolton, D., Ren, C., Riviere, J., Marone, C., et al. (2018). Estimating fault friction from seismic signals in the laboratory. *Geophysical Research Letters*, *45*(3), 1321–1329. https://doi.org/10.1002/2017gl076708

Rouet-LeDuc, B., Hulbert, C., & Johnson, P. (2018). Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nature Geoscience*, *12*(1), 75–79. https://doi.org/10.1038/s41561-018-0274-6

Rubinstein, S., Cohen, G., & Fineberg, J. (2004). Detachment fronts and the onset of dynamic friction. *Nature*, *430*(2), 1005–1009. https://doi.org/10.1038/nature02830

Scholz, C. (1968). Microfracturing and the inelastic deformation of rock in compression. *Journal of Geophysical Research*, *73*(4), 1417–1432. https://doi.org/10.1029/jb073i004p01417

Schoof, C. (2005). The effect of cavitation on glacier sliding. *Proceedings of the Royal Society*, *461*(2055), 609–627. https://doi.org/10.1098/rspa.2004.1350

Sergienko, O., MacAyeal, D., & Bindschadler, R. (2009). Stick–slip behavior of ice streams: Modeling investigations. *Annals of Glaciology*, *50*(52), 87–94. https://doi.org/10.3189/172756409789624274

Shokouhi, P., Girkar, V., Rivière, J., Shreedharan, S., Marone, C., Giles, C., & Kifer, D. (2021). Deep learning can predict laboratory quakes from active source seismic data. *Geophysical Research Letters*, *48*(12), e2021GL093187. https://doi.org/10.1029/2021GL093187

Shugar, D., Jaquemart, M., Shean, D., Bhushan, S., Upadhyay, K., Sattar, A., et al. (2021). A massive rock and ice avalanche caused the 2021 disaster at Chamoli, Indian Himalaya. *Sciences*, *373*(6552), 300–306. https://doi.org/10.1126/science.abh4455

Umlauft, J., & Korn, M. (2019). 3D fluid channel location from noise tremors using matched field processing. *Geophysical Journal International*, *219*(3), 1550–1561. https://doi.org/10.1093/gji/ggz385

Umlauft, J., Lindner, F., Roux, P., Mikesell, D., Haney, M., Korn, M., & Walter, F. (2021). Stick-slip tremor beneath an Alpine glacier. *Geophysical Research Letters*, *48*(2), 1–10. https://doi.org/10.1029/2020gl090528

Vandemeulebrouck, J., Roux, P., Gouédard, P., Legaz, A., Revil, A., Hurst, A., et al. (2010). Application of acoustic noise and self-potential localization techniques to a buried hydrothermal vent (Waimangu Old Geyser site, New Zealand). *Geophysical Journal International*, *180*(2), 883–890. https://doi.org/10.1111/j.1365-246x.2009.04454.x

Vincent, C., & Moreau, L. (2016). Sliding velocity fluctuations and subglacial hydrology over the last two decades on Argentière glacier, Mont Blanc area. *Journal of Glaciology*, *62*(235), 805–815. https://doi.org/10.1017/jog.2016.35

Walter, F., Gräff, D., Lindner, F., Paitz, P., Knöpfli, M., Chmiel, M., & Fichtner, A. (2020). Distributed acoustic sensing of microseismic sources and wave propagation in glaciated terrain. *Nature Communications*, *11*(2436), 1–10. https://doi.org/10.1038/s41467-020-15824-6

Walter, F., Roux, P., Röösli, C., Lecointre, A., Kilb, D., & Roux, P. (2015). Using glacier seismicity for phase velocity measurements and Green's function retrieval. *Geophysical Journal International*, *201*(3), 1722–1737. https://doi.org/10.1093/gji/ggv069

Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2021). Predicting fault slip via transfer learning. *Nature Communications*, *12*(1), 7319. https://doi.org/10.1038/s41467-021-27553-5

Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2022). Predicting future laboratory fault friction through deep learning transformer models. *Geophysical Research Letters*, *49*(19), e2022GL098233. https://doi.org/10.1029/2022GL098233

Weertman, J. (1957). On the sliding of glaciers. *Journal of Glaciology*, *21*(3), 33–38. https://doi.org/10.3189/s0022143000024709

Winberry, J., Anandakrishnan, S., Wiens, D., Alley, R. B., & Christianson, K. (2011). Dynamics of stick-slip motion, Whillans Ice stream, Antarctica. *Earth and Planetary Science Letters*, *305*(3–4), 283–289. https://doi.org/10.1016/j.epsl.2011.02.052

Zoet, L., & Iverson, N. (2020). A slip law for glaciers on deformable beds. *Science*, *368*(6486), 76–78. https://doi.org/10.1126/science.aaz1183