

Enhancing seismic calving event identification in Svalbard through empirical matched field processing and machine learning

A. Köhler¹, E.B. Myklebust¹ and S. Mæland²

¹NORSAR, PO Box 53, 2027 Kjeller, Norway. E-mail: andreas.kohler@norsar.no

²SimulaMet, Pilestredet 52, 0167 Oslo, Norway

Accepted 2022 March 18. Received 2022 March 15; in original form 2021 November 12

SUMMARY

Seismic signals generated by iceberg calving can be used to monitor ice loss at tidewater glaciers with high temporal resolution and independent of visibility. We combine the empirical matched field (EMF) method and machine learning using convolutional neural networks (CNNs) for calving event detection at the Spitsbergen (SPITS) seismic array and the single broad-band station KBS on the Arctic Archipelago of Svalbard. EMF detection with seismic arrays seeks to identify all signals generated by events in a confined target region similar to single *P* and/or *S* phase templates by assessing the beam power obtained using empirical phase delays between the array stations. The false detection rate depends on threshold settings and therefore needs appropriate tuning or, alternatively, post-processing. We combine the EMF detector at the SPITS array, as well as an STA/LTA (short term average/long term average) detector at the KBS station, with a post-detection classification step using CNNs. The CNN classifier uses waveforms of the three-component record at KBS as input. We apply the methodology to detect and classify calving events at tidewater glaciers close to the KBS station in the Kongsfjord region in Northwestern Svalbard. In a previous study, a simpler method was implemented to find these calving events in KBS data, and we use it as the baseline in our attempt to improve the detection and classification performance. The CNN classifier is trained using classes of confirmed calving signals from four different glaciers in the Kongsfjord region, seismic noise examples and regional tectonic seismic events. Subsequently, we process continuous data of six months in 2016. We test different CNN architectures and data augmentations to deal with the limited training data set available. Targeting Kronebreen, one of the most active glaciers in the Kongsfjord region, we show that the best performing models significantly improve the baseline classifier. This result is achieved for both the STA/LTA detection at KBS followed by CNN classification, as well as EMF detection at SPITS combined with a CNN classifier at KBS, despite of SPITS being located at 100 km distance from the target glacier in contrast to KBS at 15 km distance. Our results will further increase confidence in estimates of ice loss at Kronebreen derived from seismic observations which in turn can help to better understand the impact of climate change in Svalbard.

Key words: Glaciology; Arctic region; Neural networks, fuzzy logic; Earthquake monitoring and test-ban treaty verification.

1 INTRODUCTION

Routine seismic monitoring increasingly relies on automatic detection and discrimination of events or single seismic phase arrivals. Detection is traditionally performed using characteristic functions of the waveform data (e.g. short term average over long term average, STA/LTA, Allen 1982; Withers *et al.* 1998), either presented as single channels or, in case of seismic arrays, as beams of all channels steered towards a defined direction (backazimuth) with

a fixed apparent propagation velocity (Schweitzer *et al.* 2012). A well-established recipe for building complete seismic event bulletins is to split this task into seismic phase detection and classification followed by event association using multiple seismic stations (Le Bras *et al.* 1994). However, there are a lot of application cases where one is only interested in detecting specific repeating events, either to sort them out to reduce the analyst's workload, for example, aftershocks and mine blasts (Gibbons *et al.* 2016), or to use those events for dedicated studies or monitoring tasks. Seismic

events in the cryosphere can fall into the latter category since they allow us to better understand glacier dynamics (Podolskiy & Walter 2016; Aster & Winberry 2017), for example by monitoring iceberg calving at tidewater glaciers (O'Neel *et al.* 2010; Walter *et al.* 2012; Köhler *et al.* 2015). Such records can not only be used to monitor relative changes in calving activity, but also to quantify the ice loss (Bartholomäus *et al.* 2015; Köhler *et al.* 2016, 2019a; Minowa *et al.* 2019; Sergeant *et al.* 2019), providing essential observations on how glaciers react in a warming climate and contribute to sea level rise (Gardner *et al.* 2013; Vaughan *et al.* 2013; Huss & Hock 2015).

Master event cross-correlation detection using single or multiple channels is a well-established method for processing repeating events (e.g. Gibbons & Ringdal 2006). A challenge arises if signals originating from the confined source area to be monitored show considerable waveform variability. This issue can be addressed by using multiple master event templates and more advanced methods such as subspace (Harris 2006) or cone detectors (Carmichael 2016). Cross-correlation detectors can also be applied to seismic arrays which allows for additional evaluation of the reliability of the detection by array processing methods (Gibbons & Ringdal 2011). Empirical matched field processing (EMFP, Harris & Kvaerna 2010; Gibbons *et al.* 2017a) is another array detection method suitable for repeating events. EMFP does not rely on waveform similarity but uses characteristic phase delays between the array sensors and waveform coherency across the array to generate a detection statistic. A single event template is therefore often sufficient even in case of varying source mechanisms. Separate templates are however required for different seismic arrivals (e.g. *P* and *S* waves) since phase delays depend on the propagation velocity.

The common issue for all the methods above relying on characteristic functions, beam power or similarity statistics is to set an appropriate detection threshold, either fixed or noise-adaptive (Carmichael *et al.* 2015). If the number of missed events should be minimized and, hence, a low threshold is being used, false positives may have to be sorted out in a post-processing step, for example, by event association on multiple stations. However, if the seismic event is observed on too few or just a single station, other approaches have to be undertaken.

Machine learning (ML) is now being used for many different tasks in Earth sciences (Sun *et al.* 2022), including seismology (Bergen *et al.* 2019; Kong *et al.* 2019; Yeck *et al.* 2021). Here, we focus only on supervised ML methods which learn to classify unseen data using a set of labelled training data samples. Of particular interest in seismology is the integration of ML methods in the seismic detection and classification processing pipeline. In this context, convolutional neural networks (CNNs) have been shown to be a powerful method for replacing traditional trigger methods by performing detection and classification of signals in continuous seismic data (Ross *et al.* 2018; Zhu & Beroza 2019; Mousavi *et al.* 2020). The advantage of CNNs over other ML methods is that features, being useful for discriminating seismic signals, do not have to be extracted before-hand from the raw (seismic) data stream. The CNN takes the waveforms as input and learns those features automatically using a series of hidden convolutional layers. As an alternative to completely replacing traditional detection schemes such as STA/LTA and array processing, we can also augment those methods with ML in a post-detection step. This approach has for example been shown to be very useful for discrimination of seismic signals originating from glaciers (Gajek *et al.* 2017).

In this study, we use an EMF detector at a single seismic array and an STA/LTA trigger at a single broad-band station to identify repeating calving events at a tidewater glacier on the Arctic archipelago of Svalbard (78.88°N, 12.55°E). Cross-correlation detection is not suitable here because of significant waveform variability due to complex source processes at the ice–ocean interface (Bartholomäus *et al.* 2012; Köhler *et al.* 2015). Furthermore, since different glaciers in the vicinity of the target glacier also generate calving signals, and frequent tectonic earthquakes and noise bursts are observed, a post-detection step is required, for which we test CNN models applied to the three-component waveforms. We will show that this approach performs better than a previously implemented automatic processing pipeline for calving event detection in Svalbard.

2 STUDY SITE AND DATA

Our target in this study is Kronebreen, a grounded, fast-flowing tidewater glacier in the Northwest of Svalbard about 15 km East of Ny-Ålesund (Fig. 1), which is a research station hosting the three-component seismic broad-band station KBS (STS-2 seismometer). Mass loss at Kronebreen is dominated by frontal ablation, that is, dynamic ice loss through calving and frontal melting (Nuth *et al.* 2012; Luckman *et al.* 2015). In recent years, the glacier has experienced a rapid retreat (Schellenberger *et al.* 2015; Vallot *et al.* 2018; Deschamps-Berger *et al.* 2019; Köhler *et al.* 2019a). Calving at Kronebreen is being observed as seismic signals at KBS with up to 100 events per day during the melt season (Köhler *et al.* 2015; Gajek *et al.* 2017). These observations have been successfully used to estimate the ice loss using empirical models calibrated with satellite and terrestrial remote sensing data (Köhler *et al.* 2016, 2019a). The larger events can also be observed at the small-aperture Spitsbergen seismic array (SPITS) at 100 km distance. The SPITS array has an aperture of 1 km and consists of 9 CMG-3T seismometers. While Kronebreen is dominating the calving seismicity at KBS, calving signals are also being recorded from other glaciers in the vicinity of Kronebreen (Köhler *et al.* 2015), that is, Blomstrandbreen, Kongsbreen and Aavatsmarkbreen (Fig. 1).

Using data of temporary seismic networks deployed in the study area over several months in 2013 and 2016 to obtain precise event locations, we had previously compiled a data set of reviewed calving events observed at KBS (Köhler *et al.* 2016, 2019a) which will be the basis for training and testing classifiers in this study. We refer to it as Reference Data Set 1 (RD1, Table 1). RD1 is not complete in the sense that not all calving activity at Kronebreen during the temporary network deployment is captured. Only events clearly observed at KBS (see signal-to-noise ratios, SNR, in Fig. A1) and assignable to one of the four glaciers in the Kongsfjord area are included. Reference Data Set 2 (RD2) is a complete record of all calving events at Kronebreen which generated a seismic signal on the temporary stations close to the terminus during 10 d of observation between 2016 August 24 and September 3. Only about 10 percent of these events are observed at KBS (Köhler *et al.* 2019a) because amplitudes of small events are being attenuated below the noise level with distance. Tectonic earthquakes from the nearby Atlantic mid-ocean ridge and from the Svalbard region are also frequently observed at KBS. Training data for this tectonic event class are obtained from NORSAR's reviewed seismic bulletin (NORSAR 1971a). Seismic background noise conditions at KBS can vary over time depending on local weather conditions and ocean

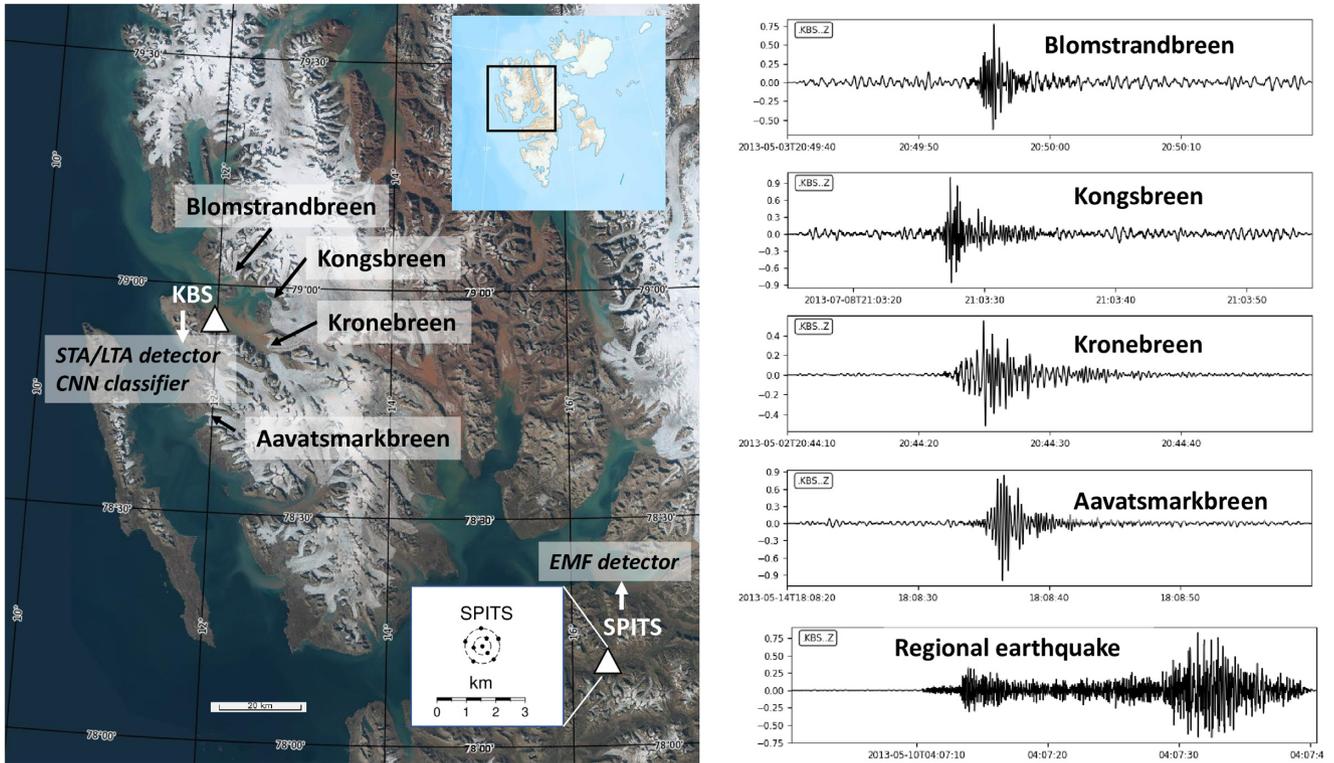


Figure 1. Left-hand panel: study area, location of seismic stations (white triangles) and tidewater glaciers in the vicinity of station KBS. Insets show overview of Svalbard and geometry of SPITS array. Background maps: courtesy of the Norwegian Polar Institute. Right-hand panel: examples of vertical component waveforms recorded at KBS for calving and earthquake signals.

Table 1. Number of calving events at different glaciers in two reference data sets (RD1 and RD2).

Data set	Time period	Number of events			
		Kronebreen	Blomstrandbreen	Kongsbreen	Aavatsmarkbreen
RD1	Apr–Sep (2013 and 2016)	1850	207	150	73
RD2	Aug 25–Sep 3 (2016)	3987	–	–	–

tides (Köhler & Weidle 2019), which is taken into account during training data selection (see below).

3 METHODS

Our objective in this study is to detect calving events at Kronebreen, minimizing the number of missed as well as misclassified seismic signals. We not only want to distinguish between calving events, earthquakes and noise bursts, but also aim to avoid calving at other glaciers to be assigned to Kronebreen. Signal detection in continuous data from KBS was previously performed using an STA/LTA detector in the calving frequency band (see Köhler *et al.* 2016, for details). In that study, we classified events from Kronebreen by first sorting out all known earthquakes included in the NORSAR bulletin and then using a classifier based on a single feature, that is, the average polarization direction computed from the horizontal seismogram components of KBS. If the event was observed at the SPITS array, additional features could be used to select Kronebreen events, that is, the traveltime difference between *P* and *S* waves and the backazimuth. A similar approach has been used by Gajek *et al.* (2017) who also used an STA/LTA detector at KBS, but combined it

with a fuzzy logic algorithm to classify calving based on the signal frequency and energy flow.

We consider the already implemented method of Köhler *et al.* (2016) using only KBS data to be the baseline calving detector which we aim to improve in this study. Our goal thereby is twofold: first, we want to explore if the detector can be run on SPITS using EMF processing and how it performs compared to the STA/LTA trigger at KBS, the station which is located much closer to the target glacier. Secondly, we aim to improve the simple previous classifier by applying CNNs directly on the KBS waveforms for all obtained detections. In the following, we introduce the methods being used in more detail.

3.1 Empirical matched field processing for event detection

The general idea behind beamforming is to align and stack array waveforms so that the SNR of the coherent target signal is increased. Doing so facilitates the detection of the signals using traditional trigger algorithms. Aligning the waveforms is accomplished by applying time-shifts at each array station based on a wave propagation model, an approach known as steering. In the frequency domain,

the beam power $P(\omega)$ can be expressed in vectorised form as:

$$P(\omega) = \mathbf{e}(\omega)^H \mathbf{R}(\omega) \mathbf{e}(\omega), \quad (1)$$

where $\mathbf{e}(\omega)$ is the steering vector, H denotes the Hermitian transpose and $\mathbf{R}(\omega)$ is the spatial co-variance or cross-spectral matrix for frequency ω :

$$\mathbf{R}(\omega) = E\{\mathbf{Y}(\omega)\mathbf{Y}(\omega)^H\}. \quad (2)$$

The vector $\mathbf{Y}(\omega)$ includes the complex spectral values of all array channels. The spectra are estimated for a short time window including the target signal. This is indicated by the expectation operator E . For short time windows and narrow-frequency bands, the fast Fourier transform (FFT) might not provide a stable estimation of the spectrum (Gibbons *et al.* 2017a) and using the multitaper method of Prieto *et al.* (2009) is preferred. The steering vector $\mathbf{e}(\omega)$ is defined as:

$$\mathbf{e}(\omega) = e^{-j\omega\Delta T_i}, \quad (3)$$

where ΔT_i is the time delay for which the beam is supposed to be calculated for array station i . In case of continuous data, the cross-spectral matrix can be computed using sliding time windows for each time step t as $\mathbf{R}(\omega, t)$ resulting in a beam power time-series $P(\omega, t)$

For traditional beamforming including frequency–wavenumber (FK) analysis, the tested steering vectors correspond to plane waves travelling over the array with a given slowness. In case of non-planar waves, the steering vectors can be constructed accordingly from the source location and the velocity model, an approach known as matched field processing. Thus, rather than searching the slowness space to fit an incoming signal as for FK analysis, the geographical location is the target variable.

Furthermore, a steering vector can be computed directly from an observed seismic signal. It can be shown that the principal eigenvector of the cross-spectral matrix can be used as an empirical steering vector maximizing the beam power for the signal under investigation (Harris & Kvaerna 2010; Gibbons *et al.* 2017a). In other words, we can use the steering vector \mathbf{e}_T computed from a template signal to find similar seismic signals by using the beam power as a detection statistic, a method called EMFP:

$$P_{\text{emf}}(\omega, t) = \mathbf{e}_T(\omega)^H \mathbf{R}(\omega, t) \mathbf{e}_T(\omega). \quad (4)$$

Gibbons *et al.* (2017b) suggested to take the time derivative $\dot{P}_{\text{emf}}(\omega, t)$ in order to enhance the signal detectability, a method we follow in this study. The differentiated beam power time-series for all frequencies can be referred to as the EMFP pseudo-spectrogram (Figs 2a–d). Finally, the beam power is normalized by the sum of power spectra, which is the trace of \mathbf{R} , to compute the relative beam power. Furthermore, we can average incoherently over a frequency band from ω_1 to ω_2 to get a scalar detection statistic for each time step (Gibbons *et al.* 2017a):

$$\hat{P}_{\text{emf}}(t) = \sum_{\omega_1 < \omega < \omega_2} d \left[\frac{P_{\text{emf}}(\omega, t)}{\text{trace}\{\mathbf{R}(\omega, t)\}} \right] / dt \quad (5)$$

This EMFP detection statistic is sensitive to signals originating from locations close to the template event. The extent of this geographical footprint depends on the array geometry and aperture (Kvaerna *et al.* 2021). Furthermore, there is a trade-off between reducing the footprint area and not missing weak seismic signals similar to the template when tuning the detection threshold. If we choose a low threshold, post-processing may be required to remove event detections too far away from the target location.

We use the P and S waves of a large calving event at Kronebreen (2016-08-12 05:33:45 UTC at KBS) observed at SPITS as templates for detecting all signals arriving with similar time delays across SPITS. Both templates are 3 s long, which corresponds to the window length for sliding windows processing in continuous data with a step width of 0.5 s. A detection is declared if the relative beam power exceeds fixed thresholds, 0.041 for the P (see Fig. 2) and 0.025 for the S wave, and the time difference between both detections is the same as the for the template event. Note that the threshold values cannot be interpreted as typical coherency or semblance values due to the beam power differentiation. We decided for low thresholds obtained by scaling up the average background noise beam power (0.008 for P -wave statistic and 0.003 for S -wave statistic) by a factor of 5 and 8, and let the post-detection classifier deal with false detections. These factors were found empirically by visually inspecting selected time periods on SPITS and KBS. Fig. 2 demonstrates EMF processing by presenting the detection statistic for the template and a selected detection. In case of the detected event, a second, much smaller calving signal is recorded at KBS one minute after the first one (see Figs 2f and h). Notably, the detection statistics for P and S waves at SPITS exhibit clear peaks with time difference similar to the one of the template, although the second signal is hardly seen in the waveforms. In contrast, a clear P -wave detection before the first event (at about 150 s) is not associated with an S wave at the expected time and is therefore not included as a detection.

3.2 Event classification with convolutional neural networks

CNNs have shown strong performance in earthquake detection and phase picking (Kong *et al.* 2019), but the classification objective can also be source mechanism or location, as in our case. Mathematically speaking, a CNN is a function that maps structured input data, such as images or time-series, onto a set of output values. In the classification setting, output values are bound to the range [0, 1], which indicating the degree of belief that an input data sample belongs to a given class. Because the CNN is also a differentiable function, its free parameters can be optimized from training data through gradient descent.

Classification from a set of data features (such as amplitudes and frequency content) is done by *dense* layers, which multiply and add the learnable parameters to their inputs. The output goes through an activation function, resulting a nonlinear relationship to the input; for instance, a softmax activation function is used in the final layer to ensure the output lies in the [0, 1] interval. The recent, huge progress of neural networks are largely due to the introduction of *convolution* layers, which are essentially performing cross-correlation of the input and short template functions, where the templates are in entirety learned from data. Successive steps of convolution layers eliminate the need for manually computing features, and ensures that the learned features are optimized for the given data set. Activation functions are used also here, and force the output to be non-negative. As for regular cross-correlation, signal detection is independent of the signal's location in time in input data. Being interested only in the presence of signal, and not location in time, the output of each filter in the final convolution layer can be averaged (known as *global average pooling*), before being input to the dense layers that perform classification. The parameters of the convolution and the dense layers are all optimized in the same training procedure. In the following, we take *architecture* to mean

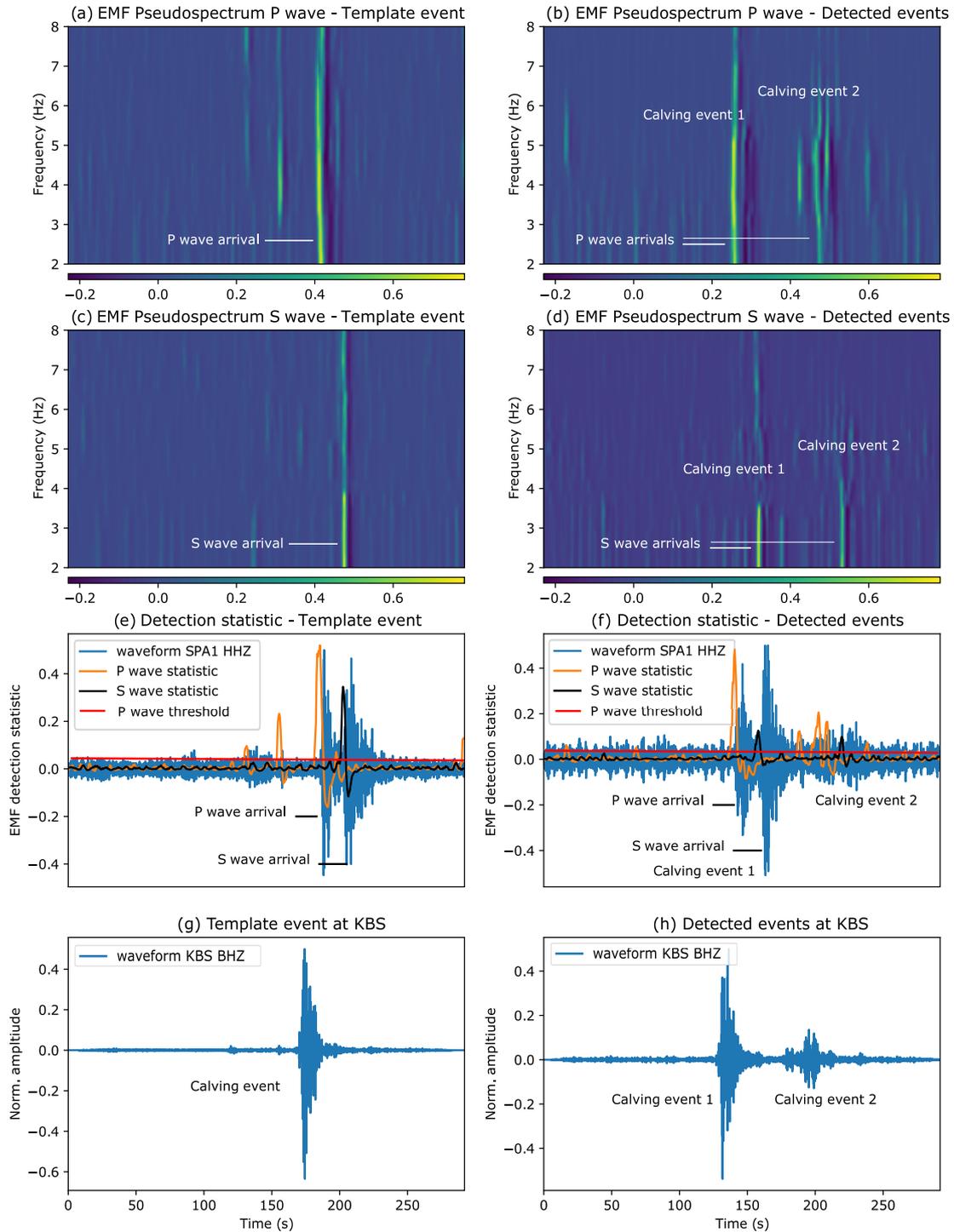


Figure 2. Demonstration of EMFP. (a)–(d) Pseudo-spectrogram for *P* and *S* waves showing the differentiated beam power time-series for different frequencies for calving events at SPITS. Colour scale indicates beam power values. (e) and (f) Scalar detection statistic obtained by summing over frequencies. The event in (e) is used to define *P*- and *S*-wave templates (2016-08-12 05:33:45 UTC at KBS). The events in (f) are examples of detected calving at Kronebreen (2016-08-04 04:37:25 UTC). (g) and (h) Same calving events recorded at KBS. Timescale is valid for all subplots.

the structural composition of the CNN, while we use *model* for a specific CNN instance with optimized parameters.

3.2.1 Training data preparation

The three-component KBS record is cut into 80 s long time windows for all calving events in reference data set RD1 starting 20 s

before the detection time at KBS, and the waveforms are detrended, bandpass filtered between 2 and 15 Hz and normalized with the maximum amplitude. The same is done for 236 regional earthquake records with *P*-wave picks at KBS in the NORSAR bulletin (NORSAR 1971a) between April and September in the years 2013 and 2016. Finally, a noise class is added by randomly selecting 318 noise instances of 80 s length without STA/LTA triggers in the

RD1 time period. In addition, 80 noise samples are added within two time periods of strong ocean wave-generated tremors which are described in Köhler & Weidle (2019): (1) 2016-04-10 13:30:00 until 2016-04-10 16:00:00 UTC and (2) 2016-05-27 02:30:00 until 2016-05-27 06:00:00 UTC. The prepared data are split into training (70 per cent) and test (30 per cent) sets.

3.2.2 Training data augmentation

Training data are augmented due to the low number of ground-true events available and to provide a more generalized representation of events at KBS. First, two duplicates are created of all calving events and regional earthquakes. Real noise recordings at KBS are added to both duplicates to artificially lower the SNRs to 3 and 6, respectively. The noise recordings are drawn from the set of noise class samples described above. The motivation for this augmentation is that low SNR events are underrepresented in RD1 (Fig. A1) because they were less likely to be locatable with the temporary seismic networks used to generate RD1. Since multiple calving events can occur with short inter-event times, we also prepare additional training samples by stacking each calving signal, not including those augmented by noise, and a randomly selected second event at the same glacier, the latter one randomly shifted in the time window.

Random cropping is commonly used in image recognition tasks to limit bias toward irrelevant information, for example, backgrounds (Takahashi *et al.* 2019). In the present work, we use random cropping of the waveforms to decrease the bias toward the small variances in arrival times. We still make sure to retain the majority of the time window including the signal and apply a 10 per cent crop, that is, 10 per cent of the 80 s long signal sampled with 40 Hz is dropped, either at head or tail, or a combination. Hence, the final input layer has 2880 nodes per channel (Fig. 3). No augmented events were used in the final evaluation using the test data set.

3.2.3 CNN architectures

Even after augmentation, the size of our training data limits the CNN model complexity compared to previous studies which used several hundred thousand earthquakes for model training (Ross *et al.* 2018; Zhu & Beroza 2019). Our neural network model consists therefore of only three convolution layers with increasing filter (kernel) sizes, each followed by ReLU activation, followed by a flattening layer, and finally two densely connected layers (Fig. 3). No pooling between convolution layers is included. The network training is regularized by spatial dropout, that is, by randomly dropping a percentage of the convolution filters during training. The models are implemented in the Keras (Chollet *et al.* 2015) framework, using TensorFlow (Abadi *et al.* 2015) as backend.

We start with a one-stage model with six classes, that is, noise, earthquakes and four glaciers (Model 1). Furthermore, two-stage and dual output models are tested. Model 2 includes simply two CNNs as used in Model 1 applied after each other, first to classify the superclasses tectonic earthquakes, calving events and noise, and then to classify subclasses of calving events at the individual glaciers. Inspired by animal image classification (e.g. La Grassa *et al.* 2021), where each taxonomic level can be classified individually, we construct Model 3, a multi-output CNN (fig. 1, La Grassa *et al.* 2021). Here, we have a simpler taxonomy of only two levels, namely, the superclasses as in Model 2 and subclasses for calving. This model uses the same architecture as Model 1 for the convolutional layers to learn latent features. These latent features are then

used in two model branches containing dense layers, as seen in Fig. 3. A small modification needs to be made to the training procedure: when the model predicts the tectonic or noise as superclass, we should not optimize for the subclass output. This is done by zeroing the gradients for these samples (for the subclass branch). In practice, this is done by setting the sample weights to zero.

3.2.4 Class imbalance

When optimizing a model it is beneficial to have equal amount of samples per class. However, this is rare in natural data sets and there is class imbalance in RD1 as seen in Table 1, that is, Kronebreen is dominating the calving classes. For earthquakes and noise we can easily add more samples, which however did not improve results in our case, and we decided therefore for a class size only slightly larger than those for the non-Kronebreen classes. Class imbalance is best solved by over/undersampling techniques, or as in this case the modelling framework accepts class weights as an input, and no re-sampling is required. The weight for an individual class is $N/(C|N_c)$ (King & Zeng 2001), where N_c and N are the number of samples in class c and in the full data set, respectively. C is the set of classes. In addition, we also tested leaving out 50 per cent of the samples of the over-represented Kronebreen class randomly during training, which did not improve results.

3.2.5 Model training

The models are all trained under the same parameters. We use the Adam (Kingma & Ba 2017) optimizer with learning rate 0.002 over 175 epochs with a batch size of 12. As for the free model parameters (i.e. number and dimension of layers), training parameters are specific to the data set and are found by trial and error using a subset of data that is not included during training. We reduce the learning rate by half if there is no improvement to validation accuracy over seven epochs. If the validation accuracy does not improve over 15 epochs we stop training. We validate the models using stratified fivefold cross validations. Stratification ensures identical class distribution among the folds. The final prediction is an average of the five models.

4 RESULTS

4.1 Model evaluation on test data

The trained CNNs are assessed using the test data set which was not used for training. Table 2 presents different common metrics for the classification performance (Berrar 2019) for all trained models, and Table 3 shows the confusion matrices for the best performing models. From Table 2, we can see that class-weighting, augmentation and random cropping improves all models. We see that Model 3 with this setting performs best of the models presented here, however, only marginally. We see only minor reduction in performance when removing half of calving samples from Kronebreen, indicating that these contain some redundant information. Table 3 affirms the same as Table 2; Model 3 performs better than the others but only slightly. Model 2, however, has a bit more problems classifying the events from the Kronebreen class correctly.

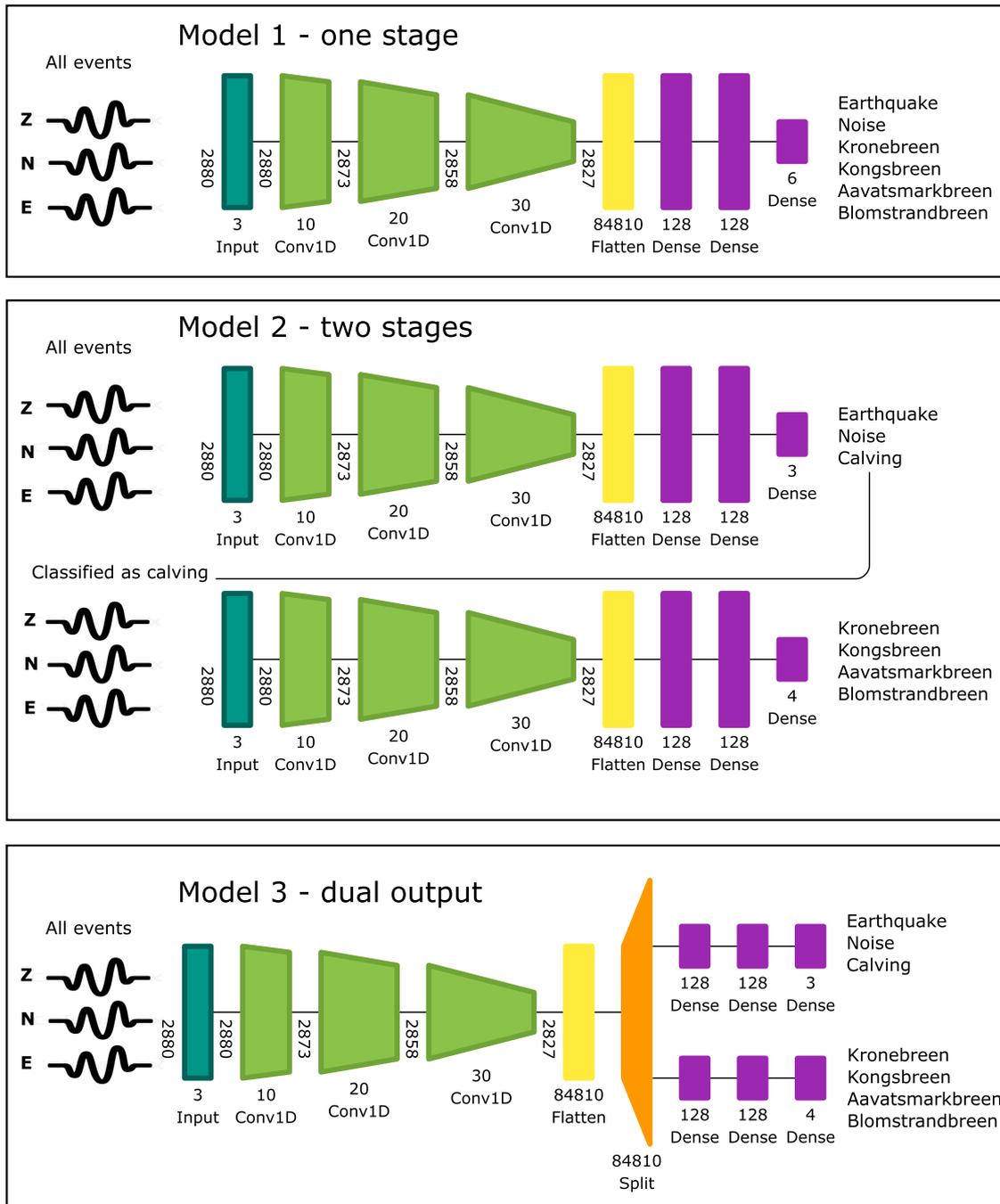


Figure 3. Neural network model architectures used for seismic signal classification. Input nodes are three-component waveforms of station KBS. For the input and Conv 1-D layers, the vertical numbers indicate the length of the data vector, while the horizontal numbers indicate the number of channels. Input has 2880 samples per channel which corresponds to a 80 s time window sampled with 40 Hz and cropped by 10 per cent. After the flattening layer, data are concatenated to 1-D array, with length indicated by the horizontal numbers. Figure created using net2vis (Bäuerle *et al.* 2021).

4.2 Detection and classification of Kronebreen calving in continuous data

The STA/LTA detector at KBS and EMFP at SPITS are applied to 6 months of continuous data from 2016 April until end of September. We use the same STA/LTA processing as described in Köhler *et al.* (2015). First, the KBS waveforms are transformed into a time-series of spectral amplitudes between 2 and 8 Hz computed from 10 s long, sliding time windows. The trigger algorithm is applied to this time-series with STA length of one sample (10 s), LTA length of 25

samples (250 s) and STA/LTA threshold of 1.5. Subsequently, the trained CNNs are applied to KBS waveforms (same preprocessing as training data) using a time window of 80 s length starting 20 s before the trigger time of the STA/LTA detector at KBS or 40 s before the EMFP detection time, taking into account the traveltime from KBS to SPITS for the latter. In addition, we apply the CNNs in a sliding-window processing mode, without the detection step, to continuous KBS data with a step width of 8 s.

Table 4 shows the number of detections and classifications for both detectors for the best performing models (Models 1 and 3).

Table 2. Metrics across the test set (see Berrar 2019, for details). GA is the (unweighted) global accuracy, AUC is the area under the receiver operating characteristic (ROC) curve and F1 is the harmonic mean of precision and recall. AUC and F1 are calculated as the mean of per-class metrics. The three leftmost metrics relate to the event type classification (calving, earthquake and noise), while the rightmost are the metrics for glacier classification (Aavatsmarkbreen, Blomstrandbreen, Kongsbreen and Kronebreen). The training settings are described using cw—class-weighted, a—augmented, cp—crop and kb_x—kronebreen (*x* percent re-sampling).

Training setting	Event type			Glacier		
	GA	AUC	F1	GA	AUC	F1
	Model 1					
	0.98	0.99	0.96	0.91	0.75	0.56
cw	0.98	0.99	0.96	0.94	0.87	0.80
cw, a	0.99	0.99	0.98	0.94	0.88	0.82
cw, a, cp	0.99	1.00	0.99	0.96	0.94	0.88
cw, a, cp, kb ₅₀ percent	1.00	1.00	0.99	0.96	0.94	0.89
	Model 2					
	0.99	0.99	0.98	0.89	0.72	0.48
cw	0.99	0.99	0.97	0.90	0.87	0.73
cw, a	0.99	0.99	0.98	0.93	0.87	0.77
cw, a, cp	1.00	1.00	0.99	0.94	0.95	0.84
cw, a, cp, kb ₅₀ percent	1.00	1.00	0.99	0.92	0.94	0.79
	Model 3					
	0.99	0.98	0.97	0.91	0.75	0.56
cw	0.99	0.99	0.98	0.88	0.85	0.69
cw, a	0.98	0.98	0.95	0.78	0.84	0.63
cw, a, cp	1.00	1.00	1.00	0.96	0.95	0.88
cw, a, cp, kb ₅₀ percent	1.00	1.00	1.00	0.94	0.94	0.83

Table 3. Confusion matrices for Models 1–3 applied to test data set not used for training. All models include class-weighting, data augmentation and random cropping. Columns are predicted and rows are true labels. AB: Aavatsmarkbreen, BB: Blomstrandbreen, KOB: Kongsbreen, KRB: Kronebreen, E: Earthquakes and N: Noise.

	AB	BB	KOB	KRB	E	N	AB	BB	KOB	KRB	E	N
	Test data						Model 1					
AB	19						16	0	0	3	0	0
BB		62					2	59	0	0	0	1
KOB			47				0	7	39	1	0	0
KRB				557			3	9	2	540	2	1
E					71		0	0	0	0	70	1
N						119	0	0	0	0	0	119
	Model 2						Model 3					
AB	17	0	0	2	0	0	17	0	0	2	0	0
BB	2	59	0	0	1	0	2	60	0	0	0	0
KOB	0	7	39	1	0	0	0	7	39	1	0	0
KRB	14	10	2	530	1	0	5	8	4	539	1	0
E	1	0	0	0	70	0	0	0	0	0	71	0
N	0	0	0	0	0	119	0	0	0	0	0	119

Table 4. Total number of detections for STA/LTA and EMFP and number of events assigned to each class for best performing models (Models 1 and 3 with class-weighting, data augmentation and random cropping).

Detector	Detections	Calving				Earthquakes	Noise
		AB	BB	KOB	KRB		
		Model 1					
STA/LTA KBS	31 180	2568	5380	2287	11 171	8530	1244
EMFP SPITS	21 679	667	1557	2203	9371	2806	5177
		Model 3					
STA/LTA KBS	31 180	4235	6011	2906	10 401	6813	814
EMFP SPITS	21 679	971	1760	2656	9581	2863	3850

For the STA/LTA method, detections assigned to the noise class exhibit the lowest number as expected. The detector is unlikely to trigger on pure noise with exception of noise bursts and other local signals not belonging to the calving classes or the earthquakes class.

The EMF detector on the other hand generated a larger number of noise instances since the events might be falsely triggered at SPITS and no signal is observed at KBS at the corresponding time. Kronebreen is the most active glacier in the study region and as

expected exhibits the highest number of classifications. From the STA/LTA detector at KBS we obtain between 10 000 and 11 000 calving events originating at Kronebreen depending on the CNN model, while the number from EMFP at SPITS at a much larger distance is only 7–16 per cent lower.

As for Kronebreen, the STA/LTA and EMFP detectors combined with CNNs produce also similar number of detections for Kongsbreen (Table 4). In contrast, the number of detections for the other two glaciers is clearly lower for EMFP compared to STA/LTA. This result gives an idea about the EMFP footprint. Kronebreen and Kongsbreen are located in very close proximity (see Fig. 1). Hence, the direction and *P*–*S* traveltime difference for calving signals arriving at SPITS is very similar. Consequently, the EMF detection statistic is sensitive to events at both glaciers. Events from Blomstrandbreen and Aavatsmarkbreen are still picked up by EMFP, however, less frequently since backazimuth and distance differ more with respect to the template event.

In order to quantify and evaluate the performance of detecting calving at Kronebreen, we compute different metrics. The first metric is based on matches with the incomplete reference data set RD1. Recall 1 is computed by dividing the number of true positives TP_{RD1} , that is, number of events correctly classified as Kronebreen calving with respect to reference data set RD1, by the total number of Kronebreen events in RD1 N_{RD1} :

$$\text{Recall 1} = \frac{TP_{RD1}}{N_{RD1}}. \tag{6}$$

Maximizing Recall 1 means avoiding false negatives, that is, missed calving events at Kronebreen previously confirmed at station KBS. Recall 2 with respect to RD2 is also computed:

$$\text{Recall 2} = \frac{TP_{RD2}}{N_{RD2}}. \tag{7}$$

Since not all calving events occurring at the glacier (N_{RD2}) are observed at KBS, Recall 2 is expected to be not larger than about 10 per cent as mentioned above. Precision 2 is computed as the ratio of the number of true positives with respect to reference data set RD2 and all events classified as Kronebreen calving by the CNN (true positives TP_{RD2} and false positives FP_{RD2}):

$$\text{Precision 2} = \frac{TP_{RD2}}{(TP_{RD2} + FP_{RD2})}. \tag{8}$$

Since a complete calving event catalog is used as reference (RD2), Precision 2 is an indicator for events falsely classified as Kronebreen calving. It does not make sense to evaluate Precision 1 since RD1 is not complete, that is, there are more calving signals from Kronebreen observed at KBS than included in RD1. All metrics are computed with a tolerance of ± 12 s when matching classified detections and events in the reference data sets.

The results presented in Table 5 show the baseline performance using the method of Köhler *et al.* (2016) with Recall 1 of 81 per cent and a rather low Precision 2 of 58 per cent. Recall 2 with 7.8 per cent reflects the fact that only about 10 per cent of calving at Kronebreen is actually observed at KBS. Without post-detection classification, the STA/LTA at KBS naturally finds all Kronebreen calving events in RD1, however, Precision 2 drops to about 28 per cent since any signal at KBS would then be attributed to this glacier. Essentially, this value tells us that about 28 per cent of all detectable signals at KBS originate at Kronebreen. Similar, Recall 2 around 12 per cent confirms the percentage of observable calving signals from Kronebreen at KBS. Notably, EMFP at SPITS still detects most of Kronebreen calving in RD1 (91 per cent) and Precision 2 without CNN

Table 5. Performance of detectors and classifiers for Kronebreen calving classification. Model modifications: cw—class-weighted, a—augmented, cp—microp. Best preferred model is highlighted (bold).

Classifier	Recall 1 (per cent)	Recall 2 (per cent)	Precision 2 (per cent)
STA/LTA Detection at KBS			
Baseline	80.9	7.8	57.6
Only detection	100.0	11.6	28.1
Model 1	97.1	9.1	74.2
Model 2	96.2	8.9	76.2
Model 3	96.7	8.8	78.5
Model 3 no cp	79.9	7.9	60.3
Model 3 no cp, no a	90.5	9.4	47.5
Model 3 no cp, no a, no cw	99.4	10.0	40.0
EMFP Detection at SPITS			
Only detection	91.0	12.6	36.4
Model 1	90.5	9.1	76.0
Model 2	90.2	9.5	74.5
Model 3	90.5	9.3	76.1
Model 3 no cp	77.1	7.1	73.5
Model 3 no cp, no a	84.3	9.2	66.5
Model 3 no cp, no a, no cw	90.8	10.2	64.2
Continuous time windows at KBS			
Model 3	90.7	27.7	39.7

classification is about 8 per cent higher than with the STA/LTA at KBS. This confirms EMFP to be a powerful detector which is very selective when triggering events at SPITS, that is, has a footprint only around the target area, and is very sensitive to weak seismic arrivals even at 100 km distance.

Classifying the detections decreases Recall 1 only slightly for all models, that is, only a few of the confirmed Kronebreen events end up in a different class. The decrease is a bit more pronounced for Recall 2 which reflects the limited training data available not fully representing the variability of calving signals from Kronebreen. However, a large increase in Precision 2 from 36 per cent up to 75 per cent on average is obtained as expected when classifying detections.

The most important assessment is the comparison with the baseline classifier. Our results in Table 5 show that all CNN models perform significantly better for all three metrics, even EMFP which relies on detections made at 100 km distance. Precision 2 increases from 58 per cent to 75 per cent for both detection methods which ensures a calving record for Kronebreen less contaminated by other glaciers than previously. A further increase towards 100 per cent is of course desirable, but would require extended training data derived from additional local measurements at the terminus. Recall 1 and Recall 2 increase as well compared to the baseline indicating improvement of the completeness of the calving record. Even if Recall 1 is larger for the STA/LTA trigger, EMFP performs slightly better when it comes to Recall 2 and has a similar precision.

While all our CNN models perform similarly better compared to the baseline, the metrics slightly favour Model 1, the one stage classifier, and Model 3, the dual output classifier, over Model 2. If we compare Models 1 and 3 and put more weight on obtaining a higher precision, Model 3 would be the better choice. The results in Tables 3 and 5 also show that training data augmentation, random cropping and class weighting improve performance of the classifiers significantly, especially Precision 2.

5 DISCUSSION

The complexity of seismic calving signals in combination with the short and partly incomplete record of confirmed events is a limiting factor for training our classifiers. Calving at a particular glacier can be considered as a repeating event when it comes to the location on a regional scale (i.e. for EMFP), but not with respect to the source time function dominating the signal at short distance. Learning the signal complexity and variability due to different ice–water interaction mechanisms, styles of calving and location along the terminus would benefit from an extended training data which requires further logistical effort by acquiring more local seismic and direct calving observations. Nevertheless, we find that taking these obstacles into account by designing CNNs with limited size and augmenting the training data, a good classifier performance can be achieved. Using ML clearly improved the simple classifier used previously to detect calving at Kronebreen. This result will therefore increase confidence in the ice loss estimated for Kronebreen from seismic observations at KBS.

Fig. 4 shows the number of Kronebreen calving events per day in the study time period. The temporal variability is very similar for both detectors and the baseline method. Interestingly, EMFP yields lower number of events only in certain time periods, such as in June and July when the number drops to about 50 per cent of the number of STA/LTA detections at KBS. This is most likely due to increased seismic noise levels at SPITS which impair the EMFP detection statistic. During that time period, the snow or ice cover shielding the SPITS borehole sensors from noise generated at the surface has usually disappeared. Furthermore, noise due to fluvial processes and increased icequake activity at close-by mountain glaciers may contribute to decreasing the detection capability.

Overall, the number of Kronebreen events from the baseline method is about 10 per cent larger than from the CNN classification (Fig. 4). This can be explained by the better precision, that is, less events from other glaciers are assigned to Kronebreen. On the other hand, the temporal variability of the calving observations seem to be very similar. However, note that calving at the other glaciers in the study area may follow similar variations due to similar meteorological conditions. As our comparison with the reference data set shows, the baseline method does most likely include more events from other glaciers and misses more events from Kronebreen compared to the CNN classifier. Hence, while using our new method does not affect the conclusions derived from previously published results when it comes to seasonal calving variability and long-term trends (Köhler *et al.* 2016, 2019a), the new classification scheme has the potential to improve the ice loss rates estimated from seismic data in those studies. Since seismic event observations were calibrated with directly measured ice loss empirically in the previous works, the absolute cumulative ice loss will most likely not be affected much. However, we expect that uncertainties of the empirical ice loss estimates can be decreased by providing more precise calving event time-series for Kronebreen. Furthermore, on longer timescales the improved classification may prevent overestimation of the cumulative ice loss at Kronebreen, considering that the new results suggest less calving than the baseline.

We tested different CNN architectures, however, the training data preparation turned out to be much more crucial for achieving good performance than using one stage, two stage or dual output models. One could argue to favour the simplest model in this case (one stage, Model 1). However, the dual output architecture could still be beneficial and may perform better for other applications in seismology where introducing super- and subclasses of seismic events

is appropriate. We therefore consider our study as a successful test of such CNN architectures.

The advantage of CNNs is that features which are best suitable to distinguish the given classes are generated automatically from the raw waveforms. The baseline classifier used the polarization at KBS as the only feature to classify the origin glacier of calving and did not include a separate class for earthquakes and noise. While the learned latent features allowing the CNNs to differentiate between earthquakes, calving and noise probably include information from frequency content, shape of waveforms and polarization, it is reasonable to assume that polarization, that is, the different relative amplitudes on the horizontal components over time, is the main reason why the four different glaciers around KBS are so well separable. Visually, it is hard to tell from the waveforms alone if a calving event recorded at KBS originated from a particular glacier due to the huge in-class variability of waveforms and the overall similar frequency content. This demonstrates how powerful CNNs using waveforms as input are for seismic signal classification. We also tested spectrograms for classification with CNNs instead of using waveforms as input, however, this did not achieve satisfying performance for the glacier classification.

The continuous mode application (see Table 5), where the CNN acts as a detector and classifier, yields a Recall 1 of 91 per cent which is similar as for EMFP and lower than for STA/LTA post-detection classification but still higher than the baseline. However, note that about 46 800 detections are obtained for the Kronebreen class which is 50 per cent higher than all STA/LTA detections at KBS (all classes). In fact, we tested the statistical significance of the occurrence of matches between detected events and ground true data and found that 14 per cent of the matches with RD1 could be explained by coincidence. The same is the case for Recall 2 of 28 per cent where 14 per cent matching could be achieved by randomly distributing the number of events classified as Kronebreen in the RD2 time period. Most importantly, Precision 2 with about 40 per cent is significantly lower than the baseline result. Thus, the continuous classifier seems to classify too many time windows falsely as Kronebreen calving, that is, it struggles to distinguish between noise and calving events, which is a result of the limited training data set. Ideally, one would like to use ML for the full processing pipelines as has been shown in previous studies. However, we find that pre-detection of events indeed helps to overcome the shortage in training examples in our case.

Having demonstrated the usefulness of CNNs for post-detection classification, a question concerning the benefit of using EMFP at SPITS remains. For monitoring calving at Kronebreen, one would prefer the STA/LTA detector at KBS due to its proximity to Kronebreen and therefore slightly better performance when it comes to catalogue completeness. Nevertheless, our results show that the EMF detector at the distant SPITS array is sensitive enough to find a high percentage of the events detected at KBS. This fact is important for the ability to monitor calving at other tidewater glaciers in Svalbard without seismic stations in their proximity. For these glaciers, EMFP at SPITS will be a powerful method once calibration data from local calving observations are available. The detector threshold at SPITS would have to be increased without a seismic station close to the glacier in order to avoid false detections. Furthermore, close glaciers might not always be distinguishable using EMFP at SPITS alone (such as for Kronebreen and Kongsbreen). However, larger events can still be used as a proxy for ice loss as we have demonstrated in Köhler *et al.* (2016) where we also used only calving events observed at SPITS to estimate the ice loss at Kronebreen in addition to the estimation with KBS observations.

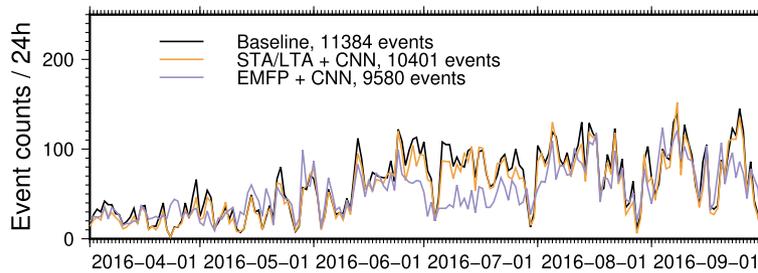


Figure 4. Number of Kronebreen calving events per day for best-performing CNN model (Model 3) applied to continuous detections in SPITS (EMFP) and KBS data (STA/LTA) for six months from 2016 April until September. Results of baseline classifier are shown for comparison. Plot starts at date 2016-04-01. Total number of events are given for each model.

In that study, the events were first detected at KBS. This step can now be replaced by EMFP at SPITS. Furthermore, one could apply post-detection classification with CNNs directly on the SPITS waveforms, that is, using the SPITS beam steered towards the target glacier. Moreover, it could be explored to what extent EMFP can benefit from including the horizontal component record at SPITS (Gibbons *et al.* 2011) and how obtaining the empirical steering vector as an ensemble average over more than one template (Harris & Kvaerna 2010) might further improve the detector. Testing these approaches is beyond the scope of this study and will be subject of future research.

6 CONCLUSIONS

In this study, we combined STA/LTA triggering, the EMF method and CNNs to detect seismic calving events originating at Kronebreen, one of the most active tidewater glaciers on the Arctic Archipelago of Svalbard. We apply the STA/LTA detector on the single three-component station KBS in the vicinity of the glacier and the EMF detector at the SPITS seismic array at 100 km distance. The benefit of EMF is that preferably signals originating in the target region are triggered in contrast to STA/LTA which is not source sensitive. In both cases, we then apply trained CNN classifiers on the three-component KBS waveforms of all detections in a 6 months time period to distinguish between earthquakes, noise and four different active tidewater glaciers close to KBS. We find that this approach works better for both detectors than a previous method implemented for calving monitoring at Kronebreen based on KBS data. The completeness of the calving record as well as false assignment of events originating from other glaciers to Kronebreen could be reduced significantly. Training data augmentation and adjusting for class imbalance was crucial to achieve these results. We introduced a novel CNN architecture for seismic event classification which treats calving at different glaciers as subclasses of a calving class that is distinguished from regional earthquakes and noise records. Our results open new possibilities for calving monitoring all over Svalbard. This will potentially allow us to estimate the ice loss at more tidewater glaciers from seismic observations and, thus, enable us to assess and better understand the impact of a warming climate in Svalbard.

ACKNOWLEDGMENTS

Waveform figures were generated and seismic data processing was done using Obspy (Beyreuther *et al.* 2010). We thank Wojciech Gajek, Tormod Kvaerna and one more anonymous reviewer for their comments and suggestions that helped to improve this manuscript.

This study was partly funded by The Research Council of Norway (project number 311596, GEObyIT). Author contribution: AK did the data processing and wrote the paper. SM implemented and tested CNN models 1 and 2. EM developed CNN model 3 and re-trained and tested all models. SM and EM contributed to the manuscript text.

DATA AVAILABILITY

KBS and SPITS data are available via IRIS (Albuquerque Seismological Laboratory (ASL)/USGS 1988) or the Norwegian EIDA node (Ottemöller *et al.* 2021). All data are stored at NORSAR (1971b). Data from temporary seismic networks to define reference data sets are available through Köhler *et al.* (2019b).

REFERENCES

- Abadi, M. *et al.*, 2015. TensorFlow: large-scale machine learning on heterogeneous systems, <https://www.tensorflow.org>, Software available from tensorflow.org.
- Albuquerque Seismological Laboratory (ASL)/USGS, 1988. Global Seismograph Network - IRIS/USGS [Data set].
- Allen, R., 1982. Automatic phase pickers: their present use and future prospects, *Bull. seism. Soc. Am.*, **72**(6B), S225–S242.
- Aster, R. & Winberry, J., 2017. Glacial seismology, *Rep. Prog. Phys.*, **80**(12), 126801.
- Bartholomäus, T.C., Larsen, C.F., O’Neel, S. & West, M.E., 2012. Calving seismicity from iceberg–sea surface interactions, *J. geophys. Res.*, **117**(F04029).
- Bartholomäus, T.C., Larsen, C.F., West, M.E., O’Neel, S., Pettit, E.C. & Truffer, M., 2015. Tidal and seasonal variations in calving flux observed with passive seismology, *J. geophys. Res.: Earth Surface*, **120**(11), 2318–2337.
- Bäuerle, A., Van Onzenoedt, C. & Ropinski, T., 2021. Net2Vis—a visual grammar for automatically generating publication-tailored CNN architecture visualizations, *IEEE Trans. Visual. Comput. Graph.*, **27**(6), 2980–2991.
- Bergen, K.J., Chen, T. & Li, Z., 2019. Preface to the focus section on machine learning in seismology, *Seismol. Res. Lett.*, **90**(2A), 477–480.
- Berrar, D., 2019. Performance measures for binary classification, *Reference Module in Life Sciences, Encycl. Bioinform. Comput. Biol.*, **1**, 546–560.
- Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. ObsPy: a Python toolbox for seismology, *Seismol. Res. Lett.*, **81**(3), 530–533.
- Carmichael, J.D., 2016. A waveform detector that targets template-decorrelated signals and achieves its predicted performance, Part I: demonstration with IMS data, *Bull. seism. Soc. Am.*, **106**(5), 1998–2012.
- Carmichael, J.D., Joughin, I., Behn, M.D., Das, S., King, M.A., Stevens, L. & Lizarralde, D., 2015. Seismicity on the western Greenland Ice Sheet: surface fracture in the vicinity of active moulins, *J. geophys. Res.: Earth Surface*, **120**(6), 1082–1106.

- Chollet, F. *et al.*, 2015. Keras, <https://keras.io>.
- Deschamps-Berger, C., Nuth, C., Van Pelt, W., Berthier, E., Kohler, J. & Altena, B., 2019. Closing the mass budget of a tidewater glacier: the example of Kronebreen, Svalbard, *J. Glaciol.*, **65**(249), 136–148.
- Gajek, W., Trojanowski, J. & Malinowski, M., 2017. Automating long-term glacier dynamics monitoring using single-station seismological observations and fuzzy logic classification: a case study from Spitsbergen, *J. Glaciol.*, **63**(240), 581–592.
- Gardner, A.S. *et al.*, 2013. A reconciled estimate of glacier contributions to sea level rise: 2003 to 2009, *Science*, **340**(6134), 852–857.
- Gibbons, S.J. & Ringdal, F., 2006. The detection of low magnitude seismic events using array-based waveform correlation, *J. geophys. Int.*, **165**(1), 149–166.
- Gibbons, S.J. & Ringdal, F., 2011. Seismic monitoring of the North Korea nuclear test site using a multichannel correlation detector, *IEEE Trans. Geosci. Remote Sens.*, **50**(5), 1897–1909.
- Gibbons, S.J., Schweitzer, J., Ringdal, F., Kværna, T., Mykkeltveit, S. & Paulsen, B., 2011. Improvements to seismic monitoring of the European arctic using three-component array processing at spits, *Bull. seism. Soc. Am.*, **101**(6), 2737–2754.
- Gibbons, S.J., Kværna, T., Harris, D.B. & Dodge, D.A., 2016. Iterative strategies for aftershock classification in automatic seismic processing pipelines, *Seismol. Res. Lett.*, **87**(4), 919–929.
- Gibbons, S.J., Harris, D.B., Dahl-Jensen, T., Kværna, T., Larsen, T., Paulsen, B. & Voss, P., 2017a. Locating seismicity on the Arctic plate boundary using multiple-event techniques and empirical signal processing, *J. geophys. Int.*, **211**(3), 1613–1627.
- Gibbons, S.J., Harris, D.B., Kværna, T. & Dodge, D.A., 2017b. Comprehensive seismic detection and estimation using matched field processing, Tech. rep., NORSTAR Kjeller Norway.
- Harris, D.B., 2006. Subspace detectors: theory, doi:10.2172/900081.
- Harris, D.B. & Kvaerna, T., 2010. Superresolution with seismic arrays using empirical matched field processing, *J. geophys. Int.*, **182**(3), 1455–1477.
- Huss, M. & Hock, R., 2015. A new model for global glacier change and sea-level rise, *Front. Earth Sci.*, **3**(54), doi:10.3389/feart.2015.00054.
- King, G. & Zeng, L., 2001. Logistic regression in rare events data, *Political Anal.*, **9**, 137–163.
- Kingma, D.P. & Ba, J., 2017. Adam: a method for stochastic optimization, <https://arxiv.org/pdf/1412.6980.pdf>.
- Köhler, A. & Weidle, C., 2019. Potentials and pitfalls of permafrost active layer monitoring using the HVSR method: a case study in Svalbard, *Earth Surf. Dynam.*, **7**(1), 1–16.
- Köhler, A., Nuth, C., Schweitzer, J., Weidle, C. & Gibbons, S.J., 2015. Regional passive seismic monitoring reveals dynamic glacier activity on Spitsbergen, Svalbard, *Polar Res.*, **34**, 1–19.
- Köhler, A., Nuth, C., Kohler, J., Berthier, E., Weidle, C. & Schweitzer, J., 2016. A 15 year record of frontal glacier ablation rates estimated from seismic data, *Geophys. Res. Lett.*, **43**, 12155–12164.
- Köhler, A., Petlicki, M., Lefeuvre, P.-M., Buscaino, G., Nuth, C. & Weidle, C., 2019a. Contribution of calving to frontal ablation quantified from seismic and hydroacoustic observations calibrated with lidar volume measurements, *The Cryosphere*, **13**(11), 3117–3137.
- Köhler, A., Weidle, C. & Nuth, C., 2019b. Glacier dynamic ice loss quantified through seismic eyes (calvingseis) – dataset, *GFZ Data Services*, doi:10.5880/GIPP.201604.1.
- Kong, Q., Trugman, D.T., Ross, Z.E., Bianco, M.J., Meade, B.J. & Gerstoft, P., 2019. Machine learning in seismology: turning data into insights, *Seismol. Res. Lett.*, **90**(1), 3–14.
- Kværna, T., Gibbons, S.J. & Näsholm, S.P., 2021. Ctb seismic monitoring using coherent and incoherent array processing, *J. Seismol.*, **25**(5), 1189–1207.
- La Grassa, R., Gallo, I. & Landro, N., 2021. Learn class hierarchy using convolutional neural networks, *Appl. Intell.*, **51**(10), 6622–6632.
- Le Bras, R., Swanger, H., Sereno, T., Beall, G., Jenkins, R., Nagy, W. & Henson, A., 1994. Global association; final report, *Sci. Appl. Inter. Corp. Tech. Rep.*
- 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 Commun.*, **6**(8566), doi:10.1038/ncomms9566.
- Minowa, M., Podolskiy, E.A., Juvet, G., Weidmann, Y., Sakakibara, D., Tsutaki, S., Genco, R. & Sugiyama, S., 2019. Calving flux estimation from tsunami waves, *Earth planet. Sci. Lett.*, **515**, 283–290.
- Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y. & Beroza, G.C., 2020. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking, *Nature Commun.*, **11**(1), 1–12.
- NORSAR, 1971a. NORSTAR Seismic Bulletins, <https://doi.org/10.21348/b.0001>.
- NORSAR, 1971b. NORSTAR Station Network [Data set], <https://doi.org/10.21348/d.no.0001>.
- Nuth, C., Schuler, T.V., Kohler, J., Altena, B. & Hagen, J.O., 2012. Estimating the long-term calving flux of Kronebreen, Svalbard, from geodetic elevation changes and mass-balance modelling, *J. Glaciol.*, **58**(207), 119–133.
- O’Neil, S., Larsen, C., Rupert, N. & Hansen, R., 2010. Iceberg calving as a primary source of regional-scale glacier-generated seismicity in the St. Elias Mountains, *J. geophys. Res.*, **115** F04034.
- Ottmøller, L., Michalek, J., Christensen, J.-M., Baadshaug, U., Halpaap, F., Natvik, Ø., Kværna, T. & Oye, V., 2021. UiB-NORSAR EIDA node: integration of seismological data in Norway, *Seism. Soc. Am.*, **92**(3), 1491–1500.
- Podolskiy, E.A. & Walter, F., 2016. Cryoseismology, *Rev. Geophys.*, **54**(4), 708–758.
- Prieto, G.A., Parker, R. & Vernon Iii, F., 2009. A Fortran 90 library for multitaper spectrum analysis, *Comput. Geosci.*, **35**(8), 1701–1710.
- Ross, Z.E., Meier, M.-A., Hauksson, E. & Heaton, T.H., 2018. Generalized seismic phase detection with deep learning short note, *Bull. seism. Soc. Am.*, **108**(5A), 2894–2901.
- Schellenberger, T., Dunse, T., Kääh, A., Kohler, J. & Reijmer, C., 2015. Surface speed and frontal ablation of Kronebreen and Kongsbreen, NW Svalbard, from SAR offset tracking, *The Cryosphere*, **9**(6), 2339–2355.
- Schweitzer, J., Fyen, J., Mykkeltveit, S., Gibbons, S., Pirl, M., Kühn, D. & Kværna, T., 2012. Seismic arrays, in Bormann, P.(ed.), *New Manual of Seismological Observatory Practice (NMSOP-2)*. 2nd (revised) edn, <http://ebooks.gfz-potsdam.de/pubman/item/escidoc:43213:7>, pp. 1–80, Potsdam: Deutsches GeoForschungsZentrum GFZ.
- Sergeant, A. *et al.*, 2019. Monitoring Greenland ice sheet buoyancy-driven calving discharge using glacial earthquakes, *Ann. Glaciol.*, **60**(79), 75–95.
- Sun, Z. *et al.*, 2022. A review of earth artificial intelligence, *Comput. Geosci.*, 105034, doi:10.1016/j.cageo.2022.105034.
- Takahashi, R., Matsubara, T. & Uehara, K., 2019. Data augmentation using random image cropping and patching for deep CNNs, *IEEE Trans. Circ. Syst. Video Technol.*, **30**(9), 2917–2931.
- Vallot, D. *et al.*, 2018. Effects of undercutting and sliding on calving: a global approach applied to Kronebreen, Svalbard, *Cryosphere*, **12**(2), 609–625.
- Vaughan, D. *et al.*, 2013. Observations: Cryosphere, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Stocker, T.F., Qin, D., Plattner, G. K. *et al.*, Cambridge University Press Cambridge, United Kingdom and New York, NY, USA.
- Walter, F., Amundson, J.M., O’Neil, S., Truffer, M. & Fahnestock, M. & Fricker, H.A., 2012. Analysis of low-frequency seismic signals generated during a multiple-iceberg calving event at Jakobshavn Isbræ, Greenland, *J. geophys. Res.*, **117**(F01036), doi:10.1029/2011JF002132.
- Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S. & Trujillo, J., 1998. A comparison of select trigger algorithms for automated global seismic phase and event detection, *Bull. seism. Soc. Am.*, **88**(1), 95–106.
- Yeck, W.L. *et al.*, 2021. Leveraging deep learning in global 24/7 real-time earthquake monitoring at the national earthquake information center, *Seism. Soc. Am.*, **92**(1), 469–480.
- Zhu, W. & Beroza, G.C., 2019. PhaseNet: a deep-neural-network-based seismic arrival-time picking method, *J. geophys. Int.*, **216**(1), 261–273.

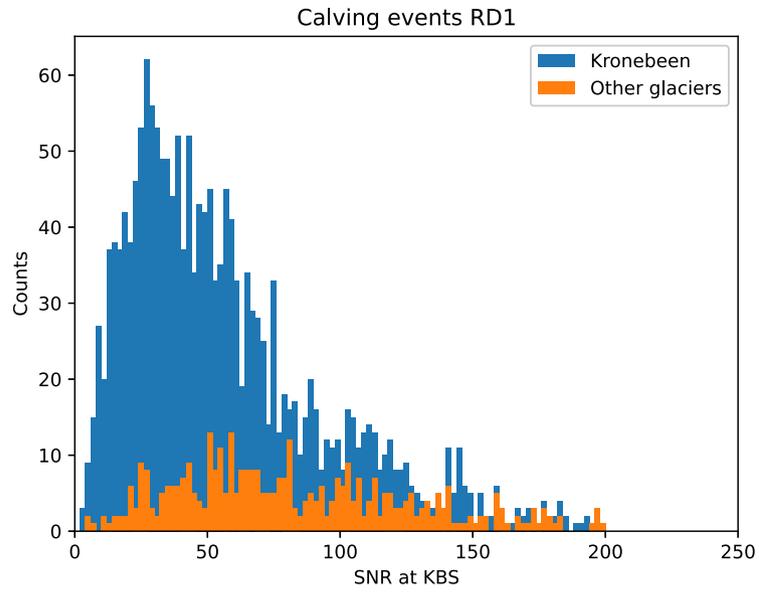


Figure A1. SNRs for calving events in references data set RD1.

Reference data set details