Geophysical Journal International

Event recognition in marine seismological data using Random Forest machine learning classifier

Przemyslaw Domel [ORCID],[1] Clément Hibert [ORCID],[2] Vera Schlindwein[3,4] and Andreia Plaza-Faverola[1]

[1]*Department of Geosciences, UiT The Arctic University of Norway, Dramsvegen 201, 9010 Tromsø, Norway. E-mail: przemyslaw.domel@uit.no*
[2]*ITES/Institut Terre et Environnement de Strasbourg, CNRS UMR7063 CNRS – Université de Strasbourg, 5 rue Descartes, F-67084 Strasbourg, France*
[3]*Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Alten Hafen 26, 27568 Bremerhaven, Germany*
[4]*Faculty of Geosciences, University of Bremen, Klagenfurter Straße 2-4, 28359 Bremen, Germany*

SUMMARY

Automatic detection of seismic events in ocean bottom seismometer (OBS) data is difficult due to elevated levels of noise compared to the recordings from land. Popular deep-learning approaches that work well with earthquakes recorded on land perform poorly in a marine setting. Their adaptation to OBS data requires catalogues containing hundreds of thousands of labelled event examples that currently do not exist, especially for signals different than earthquakes. Therefore, the usual routine involves standard amplitude-based detection methods and manual processing to obtain events of interest. We present here the first attempt to utilize a Random Forest supervised machine learning classifier on marine seismological data to automate catalogue screening and event recognition among different signals [i.e. earthquakes, short duration events (SDE) and marine noise sources]. The detection approach uses the short-term average/long-term average method, enhanced by a kurtosis-based picker for a more precise recognition of the onset of events. The subsequent machine learning method uses a previously published set of signal features (waveform-, frequency- and spectrum-based), applied successfully in recognition of different classes of events in land seismological data. Our workflow uses a small subset of manually selected signals for the initial training procedure and we then iteratively evaluate and refine the model using subsequent OBS stations within one single deployment in the eastern Fram Strait, between Greenland and Svalbard. We find that the used set of features is well suited for the discrimination of different classes of events during the training step. During the manual verification of the automatic detection results, we find that the produced catalogue of earthquakes contains a large number of noise examples, but almost all events of interest are properly captured. By providing increasingly larger sets of noise examples we see an improvement in the quality of the obtained catalogues. Our final model reaches an average accuracy of 87 per cent in recognition between the classes, comparable to classification results for data from land. We find that, from the used set of features, the most important in separating the different classes of events are related to the kurtosis of the envelope of the signal in different frequencies, the frequency with the highest energy and overall signal duration. We illustrate the implementation of the approach by using the temporal and spatial distribution of SDEs as a case study. We used recordings from six OBSs deployed between 2019 and 2020 off the west-Svalbard coast to investigate the potential link of SDEs to fluid dynamics and discuss the robustness of the approach by analysing SDE intensity, periodicity and distance to seepage sites in relation to other published studies on SDEs.

Key words: Machine learning; Arctic region; Computational seismology; Seismic noise; Wave propagation; Time-series analysis.

1 INTRODUCTION

Detecting seismicity in ocean bottom seismometer (OBS) data has additional challenges when compared to the seismicity recorded on land. In a marine setting a wide range of signals not encountered on land is present (e.g. ship noise, mammal calls and ocean current tremor). In addition, the general level of ambient noise is much higher. In the low frequency band (0.05–0.1 Hz), persistent signals are related to Rayleigh waves created by ocean surface waves interacting with both shallow and deep ocean floor (primary and secondary micro seismic peak, e.g. Sutton *et al.*1965; Barstow *et al.*1989; Hilmo & Wilcock 2020). In the higher frequency band (above 1 Hz), noise can be generated by underwater currents (e.g. Stähler *et al.*2018; Ramakrushana Reddy *et al.*2020; Essing *et al.*2021), marine mammals vocalizations (e.g. McDonald *et al.*1995; Soule & Wilcock 2013; Løviknes *et al.*2021), active seismic exploration and marine traffic, especially on shipping routes (Hildebrand 2009). Spurious signals caused by electronic malfunction, mass centring and tilt correction for broadband sensors also generate unwanted signals on the seismogram (Sutton & Latham 1964).

Additionally, there is a type of signal reported from OBS data that often outnumber the recorded earthquakes by a large margin. These signals consist of a short duration, high amplitude pulses with no discernible seismic phases and have been referred in literature as short duration events (SDE). Some studies associate SDEs with fluid migration and sedimentary fracturing processes in shallow sediments (e.g. Tary *et al.*2012; Hsu *et al.*2013; Batsi *et al.*2019). However, what exactly is at the origin of SDEs remains a matter of debate (Tary *et al.*2012). An interesting observation comes from the review of the currently available SDE studies, namely the ubiquitous use of either manual picking or classical short-term/long-term average (STA/LTA) ratio described by Allen in 1982 (e.g. Buskirk *et al.* 1981; Sohn *et al.* 1995; Díaz *et al.* 2007; Embriaco *et al.*2014; Franek *et al.* 2014; Ugalde *et al.*2019; Sgroi *et al.* 2021).

The problem of microseismicity detection has been studied in great detail over several decades and the non-exhaustive list includes methods based on: energy-ratios (e.g. Allen 1982; Baer & Kradolfer 1987), autoregression modelling (e.g. Sleeman & van Eck 1999), statistical parameters of the signal (e.g. Saragiotis *et al.*2004; Baillard *et al.* 2013), fuzzy logic theory (e.g. Chu & Mendel 1994), shallow neural networks (e.g. McCormack *et al.*1993; Gentili, & Michelini 2006), cross correlation (e.g. Molyneux & Schmitt 1999; De Meersman *et al.*2009) wavelet transforms (e.g. Anant & Dowla 1997; Bogiatzis, & Ishii 2015; Mousavi *et al.*2016) or combinations of methods (e.g. Gelchinsky & Shtivelman 1983; Diehl *et al.*2009; Nippress *et al.*2010). Many of these approaches are sensitive to the high noise level which is a case for a marine setting (Withers *et al.*1998; Mousavi *et al.*2016; Guan & Niu 2017). Other, such as polarity analysis (e.g. Vidale 1986; Jurkevics 1988), rely on the knowledge of the horizontal component geographical orientation which not always can be established for OBS data. Network based methods cannot be deployed to systematically detect SDEs since their signal is typically seen only locally on individual stations and not across a network. Impulsive nature of the SDEs makes the detection using relatively simple STA/LTA possible and justified in the past studies, however the detector itself does not allow to separate SDEs from other signal sources. Constructing reliable event catalogues that discriminate between earthquakes and SDEs from STA/LTA alone is challenging and require manual verification of the entire data set or specific time intervals of interest (e.g. Meier

*et al.*2021; Jeddi *et al.* 2021; Domel *et al.*2022). The need of reducing the amount of manual processing makes SDEs an interesting case study for the classification problem now commonly addressed by the use of machine learning.

Recent years led to a rapid development of the earthquake detection and phase picking methods based on deep learning (e.g. Mousavi *et al.*2020; Ross *et al.* 2018a, b; Zhu & Beroza 2018). Many of these methods rely on the training of the corresponding models using very large databases (thousands of examples and more) of manually curated earthquakes. Several such models exist for land recordings (e.g. Münchmeyer *et al.*2022a), and in some scenarios they can be applied to OBS data without retraining (Chen *et al.*2022). However, differences in the noise sources and levels can lead to a poor performance of the land-based earthquake detectors in marine records (we demonstrate the performance of popular machine learning models on our data in the text S1 in the electronic supplement).

When it comes to the recognition of other signals, such as SDEs, the detection and classification problem remains underdeveloped. STA/LTA method can be fine-tuned to detect SDEs but still either misses signals at high detection thresholds or contains false detections upon sensitive tuning due to the abundance of stochastically varying over time noise in marine settings (e.g. Tary *et al.*2012; Batsi *et al.*2019; Ugalde *et al.* 2019; Domel *et al.*2022). The number of false detections in the catalogues created with this approach can outnumber the signal of interest by orders of magnitude and require time-consuming quality control, which is simply not feasible for large data sets. Previously described machine learning based approaches could not be used yet for SDE detection, as no large enough, curated SDE catalogues exist for model training. Models specifically trained for earthquake detection consider SDEs as a noise and do not trigger on them (see electronic supplement). Moreover, a careful comparison between the signals from different data sets would need to be made to properly define the common characteristics of SDEs that still can vary in between the publications.

Supervised machine learning-driven approaches based on the Random Forest classifier (Breiman 2001) have shown over the years to be highly efficient in the recognition of different types of signals, ranging from earthquakes to volcanic tremors, avalanches, and landslides (e.g. Hibert *et al.*2017, 2019; Provost *et al.*2017; Chmiel *et al.*2021; Wenner *et al.*2021). These methods usually rely on curated databases of events to extract the necessary amount of information that would allow the separation between distinct categories of signals. Compared to the mentioned deep learning approaches, Random Forest can provide similar or even better results with dozens to hundreds of examples to train from per signal type (Hibert *et al.*2017, 2019; Provost *et al.*2017). This makes it well suited to microseismic studies, where there is a need of recognizing new signals with a limited set of examples to train from. It is also shown to work well in finding rare events of interests in very large data sets dominated by noise or superfluous signals (Wenner *et al.*2021). To our knowledge, Random Forest has not yet been tested in marine seismological records.

Here, we propose a new approach based on the Random Forest supervised machine learning algorithm that reduces the amount of data processing and manual verification needed compared to manual screening of STA/LTA detections. The aim of the study is to test whether the method introduced in Provost *et al.* (2017) and Hibert *et al.* (2017), for rockfalls, landslides and volcanic earthquakes on land is suitable for classifying earthquakes, SDEs and noise present in marine seismograms. We use data from six OBSs deployed in 2019 to study microseismicity and seafloor seepage off the west

coast of Svalbard (Bünz 2023a). We show the initial training process for a new data set that relies on preparation of a small number of manual examples. We study the trained model performance on one station and subsequently progress with the data processing on other OBSs, extending the training data set to properly accommodate the variation of observed noisy signals. We discuss the observed variation in the method performance and provide the metrics of the classification accuracy where feasible. Finally, we demonstrate the method validity by using the trained model to obtain a catalogue of SDEs on all OBS stations and briefly discuss their periodicity, intensity and potential link to the gas seepage sites in the area.

2 DATA

2.1 Deployment

As part of a study on seafloor seepage dynamics off the west coast of Svalbard (Bünz 2023a), we deployed seven long-term OBSs along the length of Vestnesa Ridge (Fig. 1) to monitor potential micro seismicity. We positioned most devices within an area of a known, active methane seepage on the ridge crest (OBS25-29), with two seismometers located further away: one of them at the ridge bend separating western and eastern part of Vestnesa ridge (OBS-24) and another at the south-eastern termination of the structure (OBS-30). Three of them (OBS25, 28 and 29) surrounded a seabed pockmark broadly investigated due to repeated methane release observations in the water column (e.g. Panieri *et al.*2017; Himmler *et al*. 2019). Stations in the pockmark area were separated by 800–900 m (around the pockmark), up to 2–2.5 km (OBS-26 and OBS-27). The distance between OBS-24 and the closest station (OBS-26) was 14.8 km, and for station OBS-30, the distance from OBS-29 was 18.3 km (Fig. 1). The experiment began in early July 2019 and ended with the successful recovery of all instruments in August 2020 (Bünz 2023b).

Deployment by free fall took place from H/V Helmer Hanssen at water depths of roughly 1200 m (Table 1). To establish their true location on the seafloor, we conducted an active source seismic experiment, acquiring several seismic lines within the area. Most of the OBSs drifted in either NW or NE direction with an average horizontal drift of 223 m.

Each OBS recorded seismicity using a short-period three-channel geophone with a corner frequency of 4.5 Hz. Two types of recorders were used: KUM '6d6' for OBS-24 and OBS-25, and Geomar 'Ge-olog' for the remaining instruments. We set the sampling frequency to 250 Hz for OBS-24 and OBS-25.

OBS-24 and OBS-25 recorded a full year of data until July 2020 (Table 1). However, we encountered several problems with the data on other stations. OBS-29 stopped recording after one day of deployment. The data on OBS-26, OBS-27, OBS-28 and OBS-30 was sampled erroneously at 500 Hz due to an internal error, leading to a much quicker storage use and shorter duration of the recording. For OBS-26, OBS-27 and OBS-30 this led to a stop of the record approximately 3 months after the deployment (4–7.10.2019). We had equipped OBS-28 with a twice as big memory card (128 GB), therefore its recording lasted roughly twice as long, until middle of January 2020.

2.2 Most frequently observed event types

Earthquakes

Most of the seismicity in the area comes from the nearby oceanic ridges (Fig. 1, inset). The earthquakes originating from there have most of their energy in the frequency range below 10–15 Hz and have a P–S phase separation of >8 s for our station positions (Fig. 2a). Their duration is relatively short (30–60 s on average) and there is often a *T*-wave train visible in the data.

SDEs

Short duration events in this data set are characterized by a single, strong impulse with energy from 5 Hz up to 25 Hz and more (Fig. 2b). P and S phases cannot be recognized. Most of the SDE events last about 1–2 s. They occur in groups of several SDEs or as single events (Fig. 2b). Their relative amplitude can be stronger than even the largest of local earthquakes. Unlike earthquakes, they are visible mostly only on the vertical channel of the geophone. Only for the stronger ones, we see energy on horizontal components and even hydrophone records in some cases.

Undesired signals

In an underwater setting, it is difficult to list all the possible types of noise present. The most prevalent one in our case, is most likely related to current-induced instrument shaking and is referred to as harmonic tremor (Fig 2c; Stähler *et al.*2018; Ramakrushana Reddy *et al.*2020; Essing *et al.*2021). We consider this type of signal as a noise occurring in repeating patterns lasting up to several hours. Even though it is usually less prominent from the background noise than other signals, individual wave 'packets' can easily exceed the amplitude threshold in the STA/LTA detector. Unlike earthquakes and SDEs, tremors are not observed in hydrophone data. This type of noise is likely the largest factor in lowering the number of earthquake records (both visible and detected) on different OBSs in this study.

In addition to this signal, different high-amplitude events can be included in the noise category. Some of them are related to instrumentation problems, others to marine mammals, seismic acquisition or marine traffic and some do not have an explanation and may simply be stochastic in nature. For the purposes of the study, we do not make a specific differentiation between different origins of the noise and do not try to make different subcategories of it.

3 METHODS

3.1 Detection and extraction of signals for classification

We are interested in separating local earthquakes from SDEs and different types of noise present in the data set. We aim to improve and build upon the classically used search routine based on the STA/LTA detector (Allen 1982). We do it by using STA/LTA as a tool to extract time slices from continuous data and then assign them to different classes of events based on their signal characteristics. To account for differences between SDEs and earthquakes, we used STA/LTA detector twice and adjusted the parameters to improve its sensitivity to SDEs (short, abrupt signals) and earthquakes (long, more emergent signals). Even with finely tuned settings, the STA/LTA detector will trigger on noise and other sources of signal, and in a typical processing workflow, events of interest have to be manually extracted. We down sampled all data sets to 50 Hz to equalize the data from different stations. We used an STA window
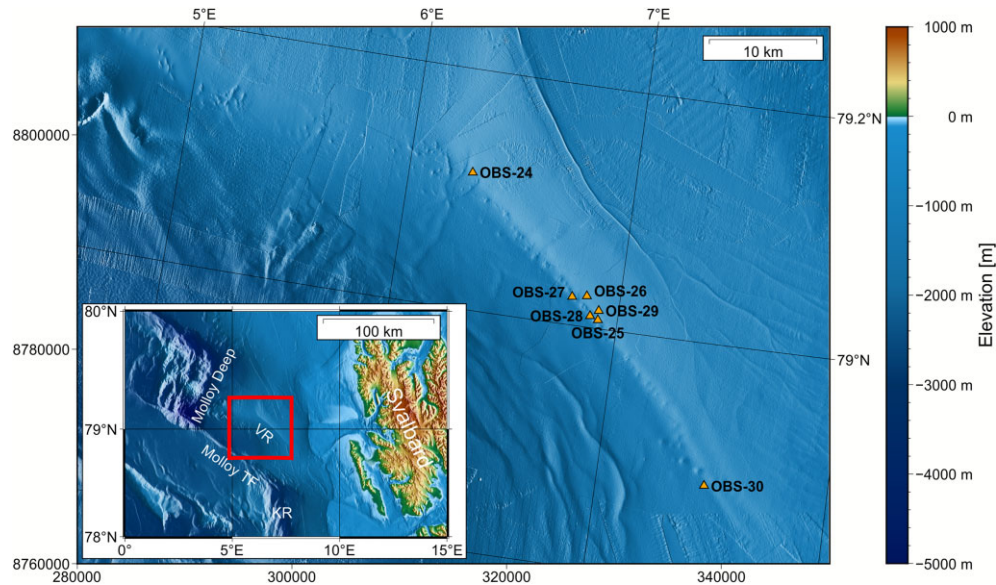
Figure 1. Map of the study area and location of the instrumentation. The triangles represent true ocean bottom seismometer (OBS) positions (after relocation) deployed at Vestnesa Ridge. Inset: Overview map of the Fram Strait with location of Vestnesa Ridge, off the west Svalbard margin and major tectonic features labelled. Regional bathymetry from the IBCAO (International Bathymetric Chart of the Arctic Ocean) (Jakobsson *et al.*2020), see Data Availability. VR-Vestnesa Ridge; Molloy TF-Molloy Transform Fault; KR- Knipovich Ridge.

Table 1. Overview of all the stations in the experiment, with their calculated position at the sea bottom, respective water depth, duration of the recording, sampling frequency and drift (difference between the calculated and deployed location) with its direction.

Station	Longitude [°E]	Latitude [°N]	Water depth [m]	Data record start [dd/mm/yyyy]	Data record end [yyyy/mm/dd]	Sampling frequency [Hz]	Drift [m]/direction
OBS-24	6.2790	79.1141	1243	07/07/2019	09/07/2020	250	202/NE
OBS-25	6.9153	79.0073	1207	07/07/2019	09/07/2020	250	271/NW
OBS-26	6.8532	79.0258	1230	07/07/2019	07/10/2019	500	286/NW
OBS-27	6.7899	79.0236	1210	07/07/2019	07/10/2019	500	238/NW
OBS-28	6.8790	79.0097	1203	07/07/2019	19/01/2020	500	226/NW
OBS-29	6.9134	79.0150	1223	07/07/2019	08/07/2019	500	238/NW
OBS-30	7.4681	78.8821	1135	07/07/2019	04/10/2019	500	101/SW

length of 0.8 s with an LTA window length of 45 s for earthquakes, and an STA window length of 0.35 s with an LTA window length of 8 s for SDEs. With the STA/LTA detection ratio set at 7, we ran the detector on the 1 Hz high pass filtered vertical channels of the geophones. Each detection was set to end at the ratio of 1.5 (Fig. 3b). We selected the lowest possible value of the trigger threshold that would not lead to continuous detections of the noise of the data set. In the earthquake-adjusted search, we removed all the detections with a duration shorter than 4 s (value determined empirically, Fig 3a). Conversely, for the SDE-adjusted search, we only kept the detections shorter than or equal to 4 s. Additionally in earthquake-adjusted search, to reduce the number of separate detections of P and S phases from the same earthquake we merged the detections that were less than 10 s from each other. We also used a kurtosis-based picker (Baillard *et al.*2014; Hibert *et al.*2014) to automatically improve the initial onset of all earthquakes found in this run. This method relies on the fact that while random noise and seismic signal have, statistically speaking, a normal amplitude distribution, the change between them does not. By computing a characteristic function based on the kurtosis (4th moment) of the signal provided by Baillard *et al.* (2014), the onset of the event can be found. We computed the function in several sliding windows for

different frequency ranges and then determined the onset from the sum of all functions. We used the sliding windows of 1, 2, 3 and 5 s, and computed characteristic functions in the frequency bands of: 1–5 Hz, 5–10 Hz, 10–20 Hz and 20–25 Hz. In our case, the picker was used on a window from −10 to +1 s relative to original pick (window range determined by trial and error; Fig. 3c). In the test we noted that the detection of many earthquakes ended too abruptly, hence for each detection, we also calculated a new end time based on a mean signal amplitude in the window from the start of STA/LTA detection to 60 s after the pick. When the amplitude of the current sample dropped below 1.5 of this mean amplitude (for samples after initial detection end), we shifted the end time to this position (Fig. 3d). As the final step, we merged the detections from earthquake-adjusted and SDE-adjusted runs and kept only one copy of the detection in the case of overlapping. We subsequently used the created catalogue to cut the windows around the signals in question and use them as an input to the subsequent classification.
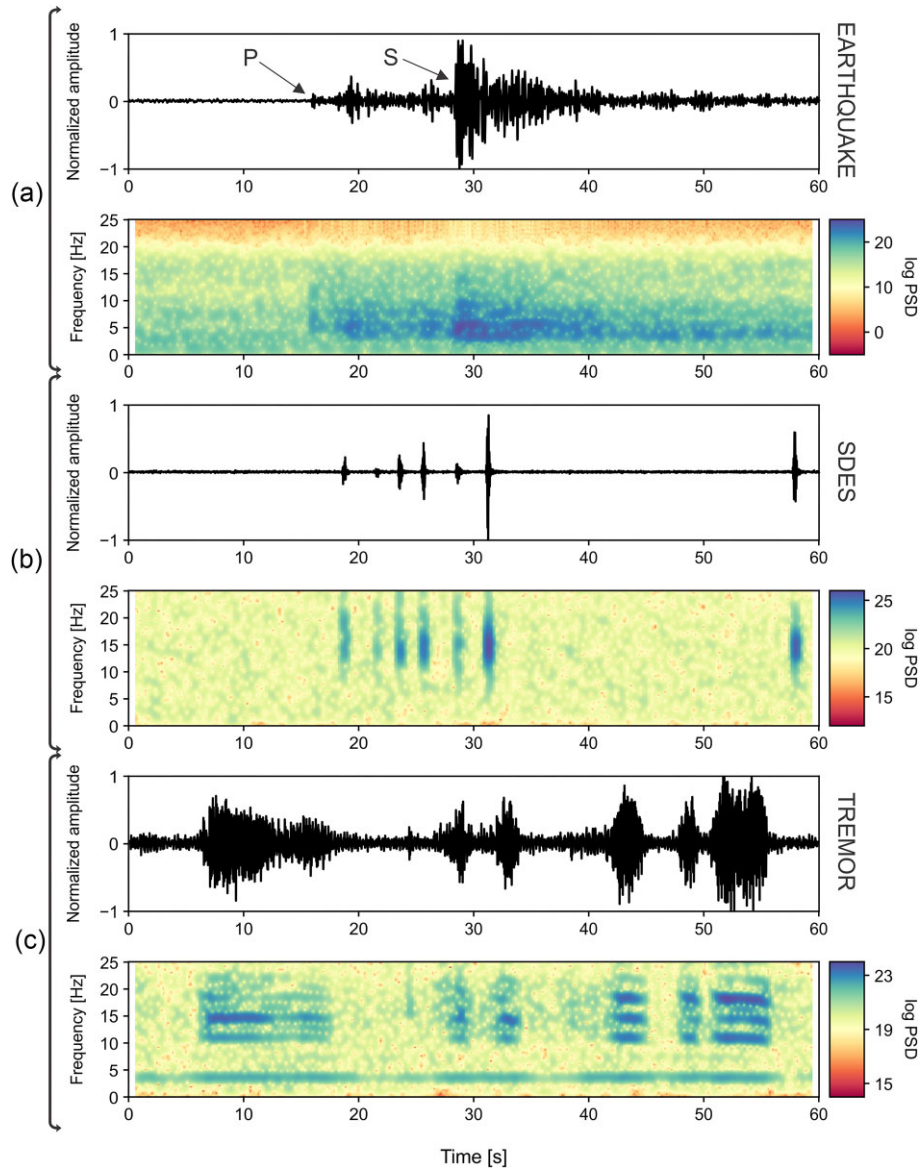
Figure 2. Examples of distinct types of events present in data belonging to different classes used in the machine learning classification approach with their corresponding spectrograms (window length: 1 s, overlap 90 per cent, Hanning taper). All examples from the vertical channel of a geophone. (a) Local ridge earthquake. (b) SDEs. (c) Harmonic tremor.

3.2 Random forest and seismic signal features

After extraction of each example, we use the Random Forest algorithm to differentiate between events of interest and noises. The implementation is provided by the scikit-learn python library (Pedregosa *et al.* 2011). This approach is based on computing a large number (1000 in our case) of decision trees, where each decision tree takes a random subset of parameters used in training to conduct a vote assigning an event to a class (Breiman 2001). Based on the majority of votes from all trees, the final decision of the outcome class of event is made. We specifically use the approach previously applied successfully to classify environmental sources (Hibert *et al.* 2017; Provost *et al.* 2017; Chmiel *et al.* 2021; Wenner *et al.* 2021) and volcanic seismicity (Hibert *et al.* 2014; Maggi *et al.* 2017; Malfante *et al.* 2018; Falcin *et al.* 2021). With the exception of the number of trees used, we retain the default parameters of the method provided by scikit-learn (see electronic supplement).

We trained a model with a purpose of recognizing three classes: EQ (earthquakes), SDE and NOISE.

Supervised machine learning algorithms such as the Random Forest require that each event is described by the same features. To do so, we transform each seismic signal in the data set into an array of curated features that are designed to extract information that is similar to what human operators use in distinguishing between different classes of seismic sources. For the features used as an input for the training, we provided 24 waveform-based features, 17 related to the frequency content and 17 pseudo-spectrogram features calculated for single channel data, plus four parameters related to the polarity computed from all three channels. We followed the approach and features proposed by Provost *et al.* (2017) (excluding the network-based features included there). We worked on the signals filtered down to 50 Hz sampling rate (Nyquist frequency of 25 Hz) and we modified the frequency windows used to calculate the parameters accordingly. Instead of calculating the features only on the
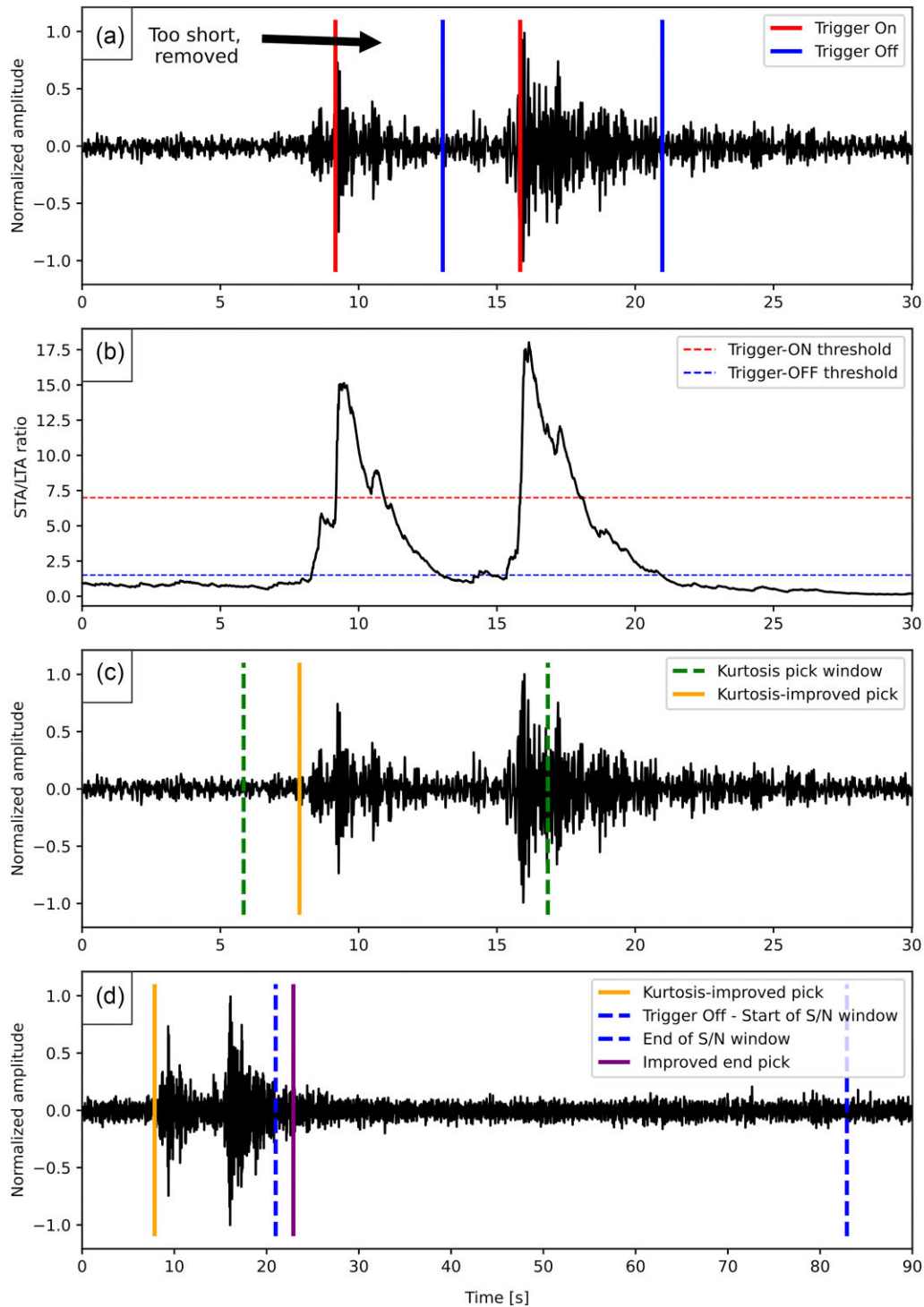
Figure 3. Graphical explanation of the steps taken to obtain detection time windows used later in the classification. (a) An example earthquake on which STA/LTA detector triggered twice, with the first detection removed due to the length s) (b) STA/LTA ratio for the event in (a) with the thresholds used for onsets and terminations of detections used in this study. (c) Time window for the kurtosis-based pick with the final improved onset pick for the remaining detection from (a). (d) Time window used to estimate the background amplitude of the signal and adjust the end pick when the ratio drops below certain threshold (1.5 in this case), with the improved onset pick from step (c) and a new position of the detection end.

channel with highest signal-to-noise ratio (SNR), we computed the features on three components of each OBS. This gives in total 174 features computed on single channel data, plus four features using three channels simultaneously. We provide the list of features with corresponding formulae in Table A1.

3.3 Iterative model building

We manually screened the catalogue to prepare the initial set of examples used in training the machine learning classification. From all stations, we selected 100 earthquakes representing a class EQ, 100 SDEs representing a class SDE and 100 examples of different

types of noise that triggered the STA/LTA detector that were together grouped into one class labelled NOISE. We used all the samples in the training step and assessed the performance of the model on the continuous data detections. We present the results of the training phase in the form of a confusion matrix (Fig. 4). We define an accuracy of model as a total number of correct predictions in a class divided by all predictions in said class. Throughout the text we also use precision as a measure of the amount of false positive detections in each corresponding class. We compute it as a ratio between correct detections in the class versus the sum of correct (true positive) and incorrect (false positive) detections. In an ideal model scenario, all training samples should end up on the diagonal of the matrix.

Initial training set—model A

The initial classification test performed well with only one example in the entire test data set being mislabelled. This shows that the differences between types of signals are potentially easily recognizable with this approach (Fig. 4, Model A). Encouraged by the results, we decided to check the model performance on a full data set from one of the stations. Based on the results, we intended to either directly transfer the model to analyse the subsequent stations or iteratively increase the training data set with incorrectly labelled examples, retrain the model and then verify how well the model transfers between different OBSs. We followed the numbering order of the stations, starting with OBS-24 and concluding with OBS-30.

OBS-24—model B

We used the Reviewed Bulletin of the International Seismological Centre (ISC; www.isc.ac.uk) to list all earthquakes recorded at the closest land stations on Svalbard and Greenland. This catalogue subsequently served as a reference to which we compared the results obtained from our detector. We downloaded all records of earthquakes located between 70°N and 90°N in latitude and 25°W to 40°E in longitude that occurred during the duration of the survey. Not all events seen on land were recorded by the OBS and vice versa. We manually verified 206 earthquakes on OBS-24 from the 393 reported in the reviewed version of the ISC catalogue. Out of these 206 earthquakes, our method detected 191 and labelled 181 as an earthquake. All the remaining visible events had a low SNR, and it would be impossible to correctly pick P and S phases for location purposes. We manually screened 5 months of data to see how many additional detections were properly labelled as earthquakes. We found that 117 new earthquakes have been detected (bringing the total amount of earthquakes to 298), but also 406 detections labelled as an earthquake turned out to be incorrect (Table 2). Due to the very large number of detections for SDE (16 442) and NOISE (2677) classes, we were not able to check all of them individually, but we observed that while there are some events classified as SDE that we consider noise and vice versa, there were no immediately visible earthquakes that were attributed to these classes. To improve the model performance, we added the cleanest examples of the different classes of events into the training set (137 EQ, 41 SDE and 109 NOISE), and obtained an updated model, which had a slightly lower accuracy during training (99 per cent for EQ class, 99 per cent for SDE class, 98 per cent for NOISE class; Fig. 4, Model B). After this procedure, we ran the detector again on OBS-24 and ended up with 324 correctly identified earthquakes and only 18 false positives (caused by noise) during the entire duration of the experiment (precision of 94.7 per cent; Table 2).

OBS-25—model C

After using this newly trained model (Model B) on a next station, OBS-25, we observed that new distinct types of noisy signal led to much a lower accuracy of the approach. While in the subset of earthquakes seen simultaneously at OBS-24, we have found no incorrectly labelled earthquakes (102 earthquakes seen in total for OBS-25), a large number of false picks of noise (220) led to an overall precision of 31.7 per cent for this data set (Table 2). In an effort to train the model to correctly identify wrongly classified examples, we added all correct and wrong examples from the EQ and NOISE classes to the appropriate groups in the training data set, with additional examples of SDEs included as well (Fig. 4, Model C). We found remarkably similar, high performance in recognition of all training events compared to previous models, with all SDEs from this data set recognized properly. After testing the new model on continuous data from OBS-25, we again found that providing additional events, specifically new examples of noise from OBS-25, leads to drastically improved results compared to the first attempt on this station. This iteration resulted in 98 correctly labelled earthquakes (4 seen previously have been wrongly classified) and 14 incorrect detections labelled as earthquakes (EQ class precision of 87.5 per cent; Table 2). We retained almost all correctly labelled earthquakes while removing 206 false positive earthquake detections.

OBS-26—model D

Due to the overall shorter duration of the OBS-26 data set (roughly 3 months in total), we decided to manually verify all detections and labels after the classification with the current version of the model (Fig. 4, Model C). We found that 452 earthquakes had a correct label, but 2253 noise-type signals were also present in the EQ category (precision equal to 16.7 per cent; Table 2). On the other hand, we confirmed our observations of correct recognition between SDEs and noise. Out of 97 SDE detections, 7 of them were incorrectly labelled earthquakes, with 4 additional events difficult to assign manually (precision equal to 89.8 per cent). We found 452 correctly marked noise events, with 9 examples of SDEs and only two earthquakes in the NOISE class (precision of 97.6 per cent). Since, we checked all detections on this data set manually, we were also able to compute how many true detections for the class we actually missed (recall). We use the definition of recall that it is a ratio between the true positive detections divided by the sum of true positive and false negative detections. The recall value for each class on OBS-26 is as following: EQ—98.0 per cent, SDE—90.9 per cent and NOISE—16.7 per cent). This confirmed our observations from previous inspections that almost all earthquakes and SDEs are being labelled correctly and the only outstanding problem is a large number of noisy events that are labelled as earthquakes or SDEs. As with previous stations, we took the verified examples and added them to the training data set, providing another iteration of the model (Fig. 4, Model D).
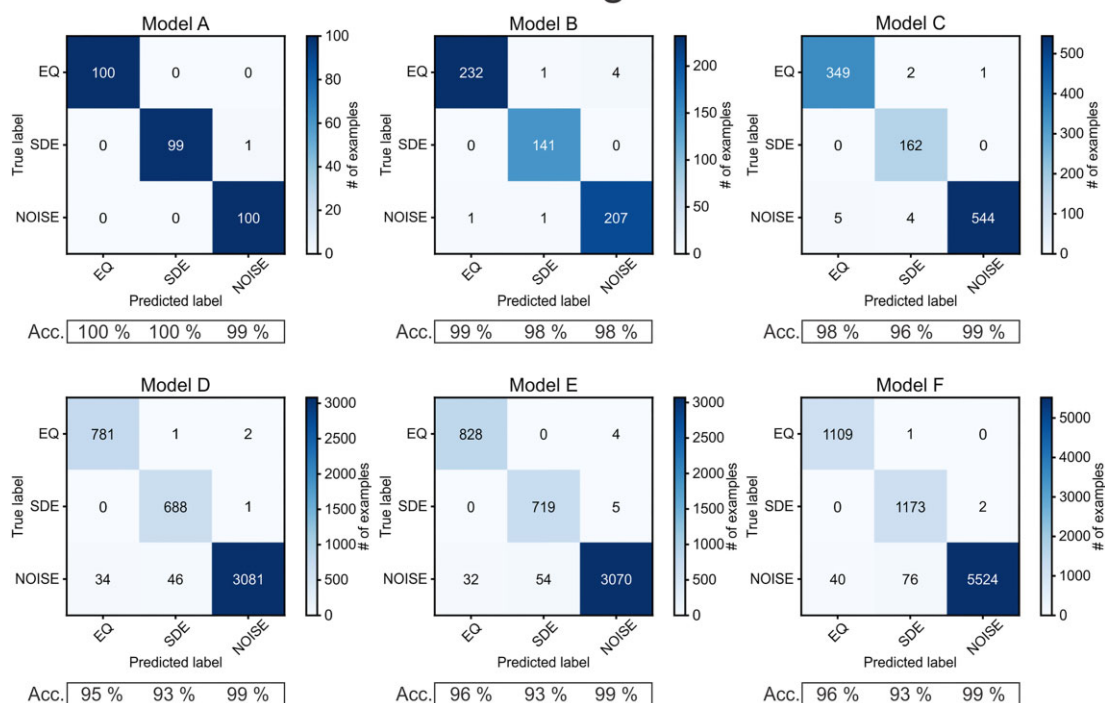
Model training results

Figure 4. Confusion matrices showing how well the classification algorithm recognized events used in the training after subsequent additions of new waveforms to the training data set. Accuracy given below is the number of correct predictions in each class divided by the total number of predictions assigned to this class.

Table 2. Performance of the trained models on the continuous data—classifier precision for all events assigned to EQ class.

Model/station	Correct detections—EQ class	Incorrect detections—EQ class	Precision [true positives/ (true positives + false positives)]
Model A/OBS-24	298	406	42.3 per cent
Model B/OBS-24	324	18	94.7 per cent
Model B/OBS-25	102	220	31.7 per cent
Model C/OBS-25	98	14	87.5 per cent
Model C/OBS-26	452	2253	16.7 per cent
Model D/OBS-27	264	562	31.9 per cent
Model D/OBS-28	346	601	36.5 per cent
Model F/OBS-30	236	94	71.5 per cent

OBS-27 and OBS-28—model E

Observations up until this point indicated that a relatively small number of earthquakes and SDEs is improperly classified, even if there is a large number of false events in EQ class. Due to the large amount of data and detections remaining, we only checked events in the EQ class from this point onward. Results for the next two stations are as follows: 346 correctly recognized earthquakes and 601 incorrectly labelled signals (precision of 36.5 per cent) for OBS-28, and 264 earthquakes and 562 cases of noise (precision of 31.9 per cent) for OBS-27 (Table 2). After this step, we again provided the highest quality examples as an additional input to recompute the machine learning model (Fig. 4, Model E).

OBS-30–model F

For the last station (OBS-30), we wanted to test whether the training data set is robust enough to be applied on a new station without having any prior knowledge about the types of signals present. For this

purpose, we removed all examples of signals from OBS-30 present in the initial training data set and all subsequent iterations. This meant the removal of 19 earthquakes from Model E. There were no examples of NOISE and SDE class from OBS-30 used in any iteration so far. This final model (Fig. 4, Model F) correctly recognized 236 earthquakes on OBS-30 while also outputting 94 false detections (precision of 71.5 per cent; Table 2). Additionally, for 32 events we had a difficulty in manually assigning them to any of the classes. In the final comparison with the ISC land-based detections, out of 132 earthquakes listed for the OBS-30 survey duration, the STA/LTA detector triggered on 62 of them and the classifier subsequently correctly labelled 57. We checked the remaining ISC earthquakes listed, and they were either not present in the data at all or their quality was too low for further investigation. This concluded the training process of the classifier.
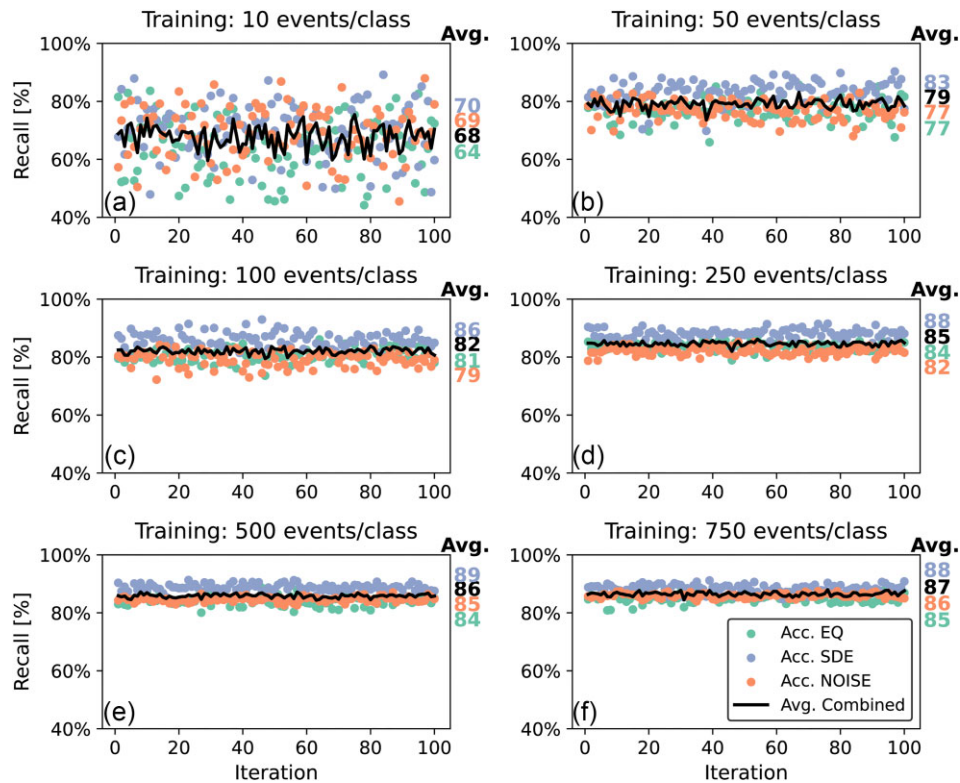
Figure 5. Recall value of the Random Forest classifier for recognizing events belonging to EQ, SDE and NOISE class for 100 iterations of training and validating of the model, when training data set equals: (a) 10 events/class, (b) 50 events/class, (c) 100 events/class, (d) 250 events/class, (e) 500 events/class and (f) 750 events/class. The validation of the model is conducted on the remainder of the events from final training data set. Average recall from all iterations per class presented in corresponding colour on the right side. The black line shows average values for all classes per iteration and the black number on the right total average for all iterations.

3.4 Performance of the model

To evaluate how well the Random Forest algorithm is suited to recognize different classes of the events, we repeated the training using 10, 50, 100, 250, 500 and 750 events per each class from all event examples used to train final model (Model F). We established the performance of each attempt using each corresponding model on the remaining samples from the data set. For each size of training data set, we repeated the procedure 100 times and calculated the overall average recall value and the average recall value per class. As previously mentioned, recall informs about the percentage of the samples from each class labelled correctly by the model (true positive rate, sensitivity).

We also investigated how well the model performs when we have a training data set with a large difference of event examples per class (i.e. as it is in our case, with strongly overfitted NOISE class). To do so, we performed two rounds of testing, one in which all classes are represented by the same number of examples and one in which the ratio between the samples in each class reflects the ratio of examples in our final model. We took randomly between 1–30 per cent of all samples from the final model for each class (which means between 79 and 2337 events for each category) and trained the model while verifying its recall value on all remaining examples. For classes EQ and SDE this leads to using the same examples multiple times, since the final model contains 1110 and 1175 examples for these classes, respectively. The data set for the final model contains 5640 examples for NOISE class. For each percentage, we repeated the training 100 times and averaged the results. In the second scenario, we repeated the same training process, but we used the percentage

of all samples belonging to their respective class. This means taking between 11 and 333 examples for the EQ class, between 12 and 352 examples for the SDE class and between 56 and 1692 examples for the NOISE class. This way, the ratio between the samples in each class remained constant and equal to the ratio in the final model.

3.5 Importance of features

With all the signal features provided, it is important to determine which of the selected features were most important during the classification of events. The Random Forest approach allows to calculate the importance of each feature, obtained by comparing the performance and the error of the model with given features randomly permuted for all the examples versus the classification result without the permutation (Breiman 2001). Arbitrary swapping of values of the selected feature allows to estimate how critical this parameter was in achieving the final result. We used the built-in feature_importances_ function within the scikit-learn library for this purpose (Pedregosa *et al.* 2011).

4 RESULTS/BUILDING A RELIABLE CLASSIFICATION MODEL

4.1 Testing of the final model (model F)

The results of repeated training using increasingly larger number of events show that even with only 10 events per class the average recall reaches 68 per cent (Fig. 5a). Gradually increasing a number of examples leads to less varying results per iteration and overall
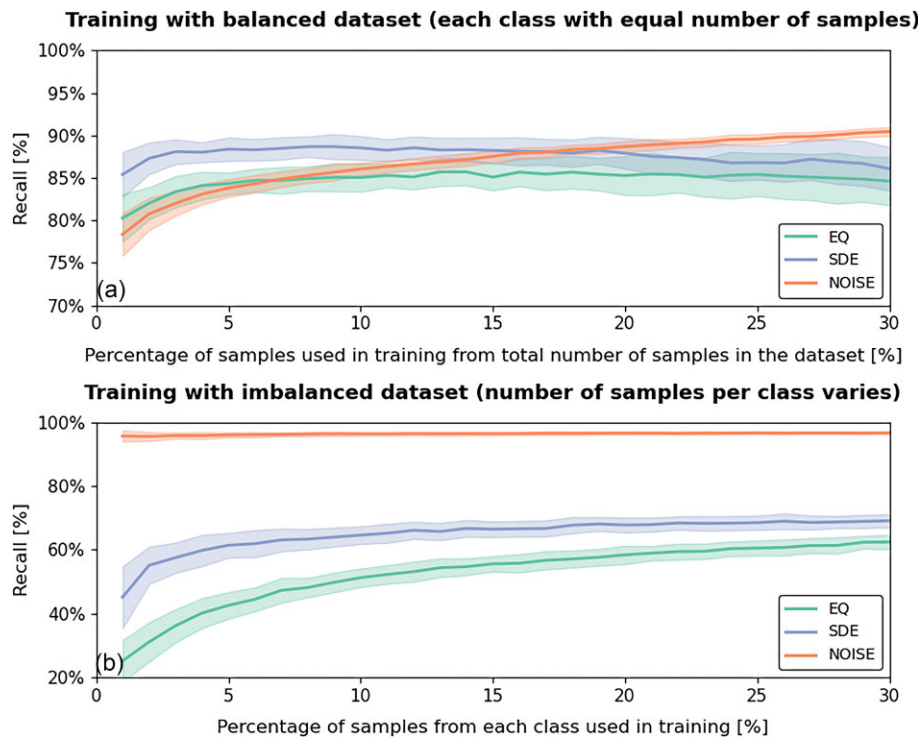
Figure 6. Recall value of the Random Forest classifier in recognizing different classes with the increase in the number of samples included in the training data set depending on whether the training data sets for each class are equal (i.e. balanced model, a) or some classes are overfitted (as NOISE in the final model, i.e. unbalanced model, b). Training was done on a given percentage of balanced or unbalanced data set and verified on the remainder of all samples. For each percentage, the model was trained 100 times and recall values were averaged. Shaded areas denote standard deviation obtained from 100 iterations of training and validation.

higher recall value in each of the classes (Figs 5b–f). The SDE class seems to be the least dependent on the number of samples needed for training, reaching a recall value of 86 per cent already with 100 examples (Fig. 5c), and increasing only by 2–88 per cent recall value when 750 examples were used (Fig. 5f). On the other hand, the NOISE class requires more examples overall in training to reach a similar performance (86 per cent recall value at 750 events used for training) (Fig. 5f). The results for the EQ class indicate that with 100 events the recall value is already at 81 per cent (Fig. 5c), there is no larger increase in the performance of the model and the overall recall value at 750 examples is high (85 per cent), but the lowest of all the classes (Fig. 5f). The final average recall value for all classes is equal to 87 per cent at 750 examples per class used in training, but it is already at 85 per cent with much smaller subset of 250 events per class (Figs 5d and f).

When using the same number of events per class for the training process, the recall value in detecting earthquakes and SDEs is already high with 1 per cent of total examples used for each class (80 and 85 per cent, respectively, Fig. 6a), whereas the NOISE class benefits the most on the increase in the number of samples used for training (accuracy of 78 per cent at 1 per cent of total data used, 90 per cent when using 30 per cent of total training samples available, Fig. 6a). The recall values of classes EQ and SDE began to decrease, when the model was trained with the same examples used multiple times. The entire training data set consists of 7925 examples (1149 EQ, 1250 SDE and 5526 for NOISE; Model F in Fig. 4). When we use the percentage of all 7925 examples to train the model and compute the recall, the number of examples for EQ and SDE class becomes greater than the actual number of examples available for these classes at the value of around 15 per cent. Above that number,

an increasing number of examples from the EQ and SDE classes is used more than once for the training.

Testing the model with the ratio between classes similar to the one for the entire training data set we have leads to a very high recall value for NOISE class that remains constant regardless of the number of samples used (*ca*. 95 per cent from 1 to 30 per cent of NOISE class examples used, Fig. 6b). Due to the much lower numbers of examples used initially for classes EQ and SDE (only 11 and 12 examples at 1 per cent of each class, respectively), the initial recall value is low (25 per cent for earthquakes, 45 per cent for SDEs, Fig. 6b). It increases steadily, however, reaching 62 per cent for EQ class and 69 per cent for SDE class when 30 per cent of examples from each class are used (Fig. 6b).

4.2 Importance of features for the final model

The ten most important features during the training of the final model are presented in Fig. 7 (all the features with their corresponding importance can be found in a text file provided with the electronic supplement to this article). Seven out of ten features are from the vertical channel of the geophone. Out of them, four of the five most important are related to the kurtosis of the signal computed for different frequency ranges. The remaining one from the top five, on the third place, is the frequency with the maximum energy in the frequency spectrum. Three of the top ten features are related to the duration of the signal (on the seventh, eighth and tenth place), and all of them have equal values, since the signal in each sample is cut into even-length windows on all channels. The remaining features on the list are the root-mean-square (RMS) difference between the decreasing part of the signal and the straight-line approximation (on the sixth place) and the energy in the first quarter of frequency band
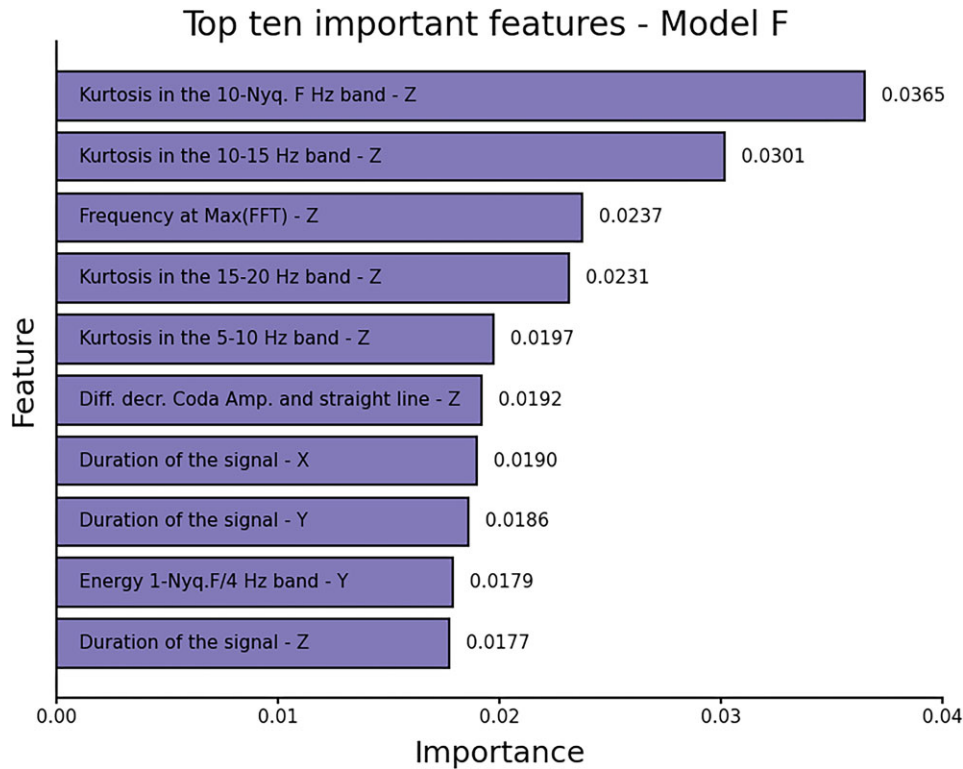
Top ten important features - Model F

Figure 7. Top 10 most important features from the final trained model with their corresponding importance, based on the mean and the standard deviation of accumulation of the impurity decrease within each tree in the trained model (1000 trees used), obtained by randomly permuting values of all features and comparing results with the results without permutation (Pedregosa *et al.*2011).

from 1 Hz to the Nyquist Frequency (25 Hz in this case) on one of the horizontal channels (on the ninth place).

5 DISCUSSION

5.1 Training process for the new model and its transferability between data sets

The detection and classification approach presented in this study aimed to improve the processing of OBS data sets which, due to the much higher noise levels than in data recorded on land, poses a more difficult challenge. To properly classify different types of signals, we needed to ensure that the time windows extracted using an STA/LTA detector correctly encapsulated and separated different types of signals. By testing and adjusting the parameters of the STA/LTA detector separately for SDEs and earthquakes, we were able to capture two types of events with the same method, but at the expense of the overall detection number. We found that SDEs, due to their high SNR, are usually correctly detected and extracted from the continuous data, but earthquake extraction proved more problematic.

Adjusting STA/LTA parameters alone does not ensure a proper detection of full earthquake signals. We often observed a presence of two separate detections for P and S phases and in the case of weaker signals, the STA/LTA detector responded only to the S phase, usually the strongest part of a waveform. It was therefore important to adjust the onset of each detection to properly capture the full signal if the detection contained an earthquake. Using the kurtosis-based picker designed to recognize slowly emergent signal in a noisy data set (Baillard *et al.*2014; Hibert *et al.*2014) proved satisfactory for this purpose. While not always the exact onset of the *P*-wave arrival was

selected, the Random Forest classifier very rarely had a difficulty with correctly labelling an event as an earthquake even with an only partially present *P*-wave phase (Fig. 8).

The initial verification of our approach using a curated catalogue of events (Model A, Fig. 4) showed that the method is very well suited to the problem and features computed from the extracted windows provide enough information to discriminate between earthquakes, SDEs and noisy signals. The test on the continuous data set from OBS-24 using publicly available ISC catalogue showed that 95 per cent of the detections made by the STA/LTA detector were assigned correctly to the EQ category and the remaining earthquakes listed in the catalogue were either not present or too weak to be recognized by automatic detection. We also found that SDE classification is working correctly, but we were not able to manually verify all of the detections due to the large size of the resulting catalogue.

This step also made us realize that the data set containing examples of noise is likely lacking examples of many different spurious signals that we can encounter in the OBS data. These signals can be sometimes traced to the issues with the instrumentation itself, but can be related, for example to the ocean currents inducing shaking on the frame (e.g. Stähler *et al.*2018; Essing *et al.*2021), marine mammals (e.g. McDonald *et al.*1995; Soule & Wilcock 2013; Løviknes *et al.*2021) or can include seismic operations at sea, which we also conducted during the first days of the survey. Providing the additional examples to the model allowed us to not only find 200 more local earthquakes, but also regain high level of correct detections in EQ class (94.7 per cent, Table 2; Model B, Fig. 4).

We repeatedly observed that applying the Random Forest model trained on one set of data to a new OBS requires retraining to
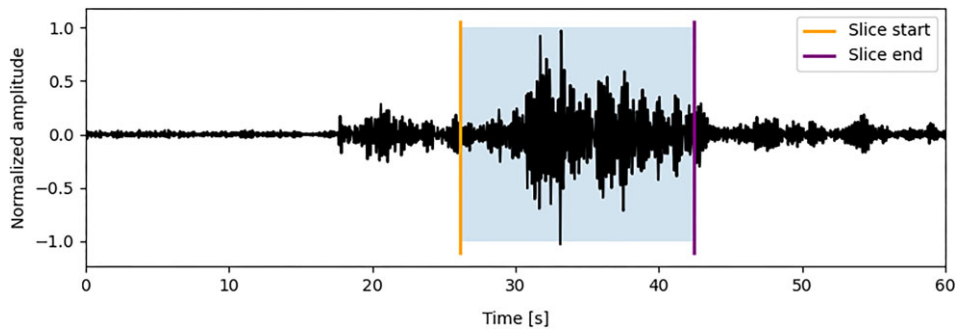
Figure 8. An example of the correct classification of an earthquake recorded on OBS-30 by the final model (Model F) for a slice that do not encompasses the earthquake waveform fully. Blue outline shows the part of the signal used in the classification, which lacks a majority of the P-phase within.

achieve high accuracy in the earthquake detections, but our classifier correctly recognizes between earthquakes and SDEs. On all subsequently trained and verified iterations of the model (Model C-F, Fig. 4), new types of noise examples were present and labelled as earthquake signals. We attribute these issues to the random nature and/or very localized origin of the noise and the differences in the signal characteristics between the different stations (separated by up to 30 km between the farthest OBS positions). Different recorders used in experiment also could have influenced the results. After making sure that the amount of improperly labelled earthquakes remains small, we found that even with the manual processing involved, checking additional few hundreds of examples leads to a significant speed-up of the data processing compared to the laborious scanning through the tens of thousands of the STA/LTA detections in the database. As for resulting SDE detections, even if the manual verification is still not a feasible task, our observations indicate that the resulting catalogue already is a notable improvement over the STA/LTA detection studies without any subsequent processing steps, which was the routine approach so far (e.g. Franek *et al.*2017; Batsi *et al.*2019; Domel *et al.*2022).

By gradually increasing the number of examples in the training set for the NOISE class, we hoped to achieve a high enough level of completeness that would lead to a small number of incorrect detections for the subsequent OBS stations we checked. While starting from station OBS-26, we found a larger number (452, Table 2) of correctly labelled earthquakes present (OBS-25 had small number of earthquakes compared to other data sets; 102, Table 2), the overall accuracy of detections when reapplying a model to a new data set was decreasing. It is important to mention that each station can pose unique challenges. In the example of OBS-26 we found an order of magnitude higher number of incorrect detections, which is probably explained by the disproportionately large amount of noise on this station. We suspect the strength of underwater currents for this OBS was much larger than in any other station possibly due to its location with respect to the seafloor morphology. For stations OBS-27 and OBS-28, a third of detections in the EQ category were retained as correct, same as for OBS-25 when using the model without including additional examples for a given station after manual checking (Table 2). For the last data set (OBS-30), we found that the final model (Fig. 4, Model F), with the five times more examples of noise than samples in the other two categories, shows a significant improvement in the reduction of incorrect detections in EQ category. With the precision of 71.5 per cent for EQ class (Table 2), the majority of distinct types of noise were properly recognized by this approach. We intend to use the final model in future processing of different OBS data sets to reduce the amount of manual labour needed.

5.2 Final model performance and its training data set

The results shown in confusion matrices inform us whether the model is well suited to recognize events using the features provided (Fig. 4), however by training the model on all examples available, these plots do not tell us about the robustness of the model in recognizing new, unknown samples (generalization). Therefore, it was important to conduct rigorous testing using even, increasingly large training sets and verify the model performance on samples of different types of events that were not included in the training phase (Fig. 5). We see that it takes a small number of 50 examples per class to reach the average recall value of 79 per cent and that a gradual increase of training sets leads to a smaller performance improvement. The recall value of detecting SDEs remains highest among all the categories throughout the testing process, which is confirmed by the verification of the SDE class content for OBS-26. With 750 events per class used for training, the recall value of independent recognition of both earthquakes and noise is high (85 and 86 per cent, respectively). In the manual verification we observed a higher recall in the recognition of earthquakes (with only two earthquakes labelled as noise and 452 correctly labelled examples of noise in OBS-26), but a lower recall value for the recognition of noise (2253 noise examples in EQ category for the same OBS). We attribute this discrepancy to a significant portion of noise being mislabelled due to the data set not fully capturing the broad range of types and features noise that a marine setting can exhibit. At this point in the model development, we opted against the creation of separate categories for different noise sources.

By providing more examples for the NOISE class (that has the largest expected variability of sources and signal shapes), we attempted to counteract the generalization of event types in this class and enforce correct recognition of most of the noise sources. Training with the increased percentage of the overall data set highlights that the NOISE class benefits the most with the provision of additional samples (Fig. 6a). The perceived drop of the recall for the remaining classes by 15 per cent mark is caused by reusing examples in these classes during the testing phase. When we train the model keeping the relative ratio between the number of examples in each category the same as for the entire, final training data set, we see that the recall value in recognizing noise is excellent regardless of the percentage of samples used for training (Fig. 6b). Therefore, we believe that when one category consists of a variety of signals with different origins, unbalancing the model to compensate for this diversity is a valid, but not necessarily the best approach. It would be beneficial to spend more time in differentiating between signals that obscure the events of interest in OBS data sets and potentially train the model using separate categories of different noise types.
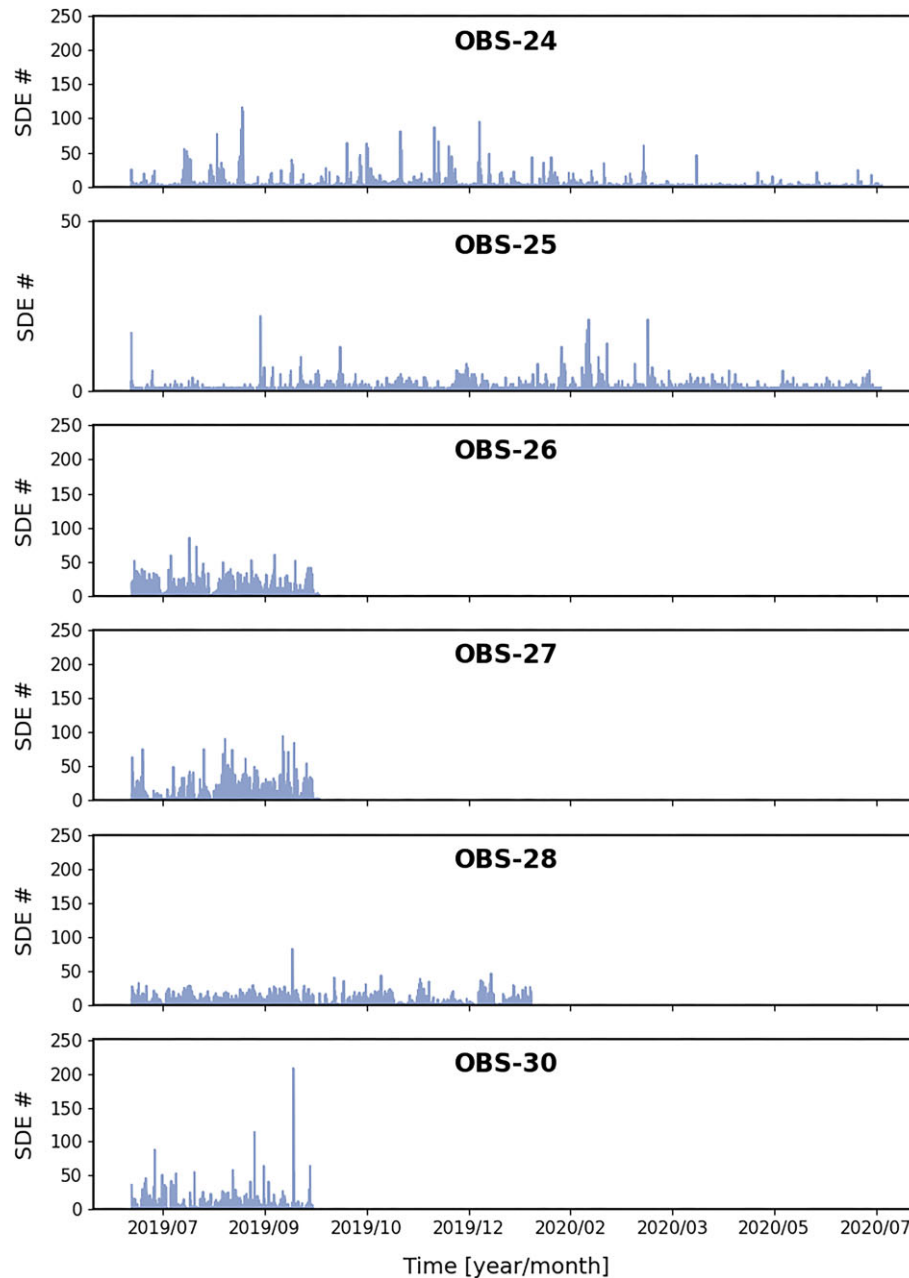
Figure 9. Time-series of hourly SDE occurrence on OBS stations obtained using random forest classification with Model F. Note the difference in vertical axis for OBS-25. The difference in the duration of data sets is explained in more detail in Sections 2 and 3.

5.3 Most important signal characteristics

In the recognition of earthquakes, SDEs and noise, the kurtosis of the signal and signal duration played a key role, judging by their prevalence in the top ten most important features (Fig. 7). This is expected, based on observed differences between the different signal waveforms generated by each source, as explained below. Kurtosis informs about the deviation of the random variable distribution from normal distribution (Hibert *et al.*2014) and the curve for different frequency ranges will have different shapes depending on the event type. We have initially a low amplitude signal (P-phase), followed by an increase (S-phase) and slow decay for earthquakes that in our data set usually not exceeds one minute of duration. SDEs on the other hand, generally show sharp spikes with an immediate termination of the signal and the usual individual duration under 5 s. In the NOISE

class, we contained an assortment of both longer and shorter events, some continuous, emergent signal as in ocean current tremor, but generally more varied than two other classes. The differences in their length and shape of the kurtosis proved to be significant enough for a high level of recognition from the remaining two classes. The importance of the way in which amplitude decreases towards the end of the signal is also reflected by the presence of feature related to the coda of the signal deviating from the straight line (feature 24 in Table 2; on sixth place in Fig. 7).

SDEs are high amplitude events, with most of their energy in the upper part of the studied spectrum (10–25 Hz, Fig. 2), similar to tremor noise present in the NOISE class. Therefore, the frequency at which maximum energy was found (third most important feature, Fig. 7), also proved useful in discerning these classes from the earthquakes, for which we expect highest amounts of energy below
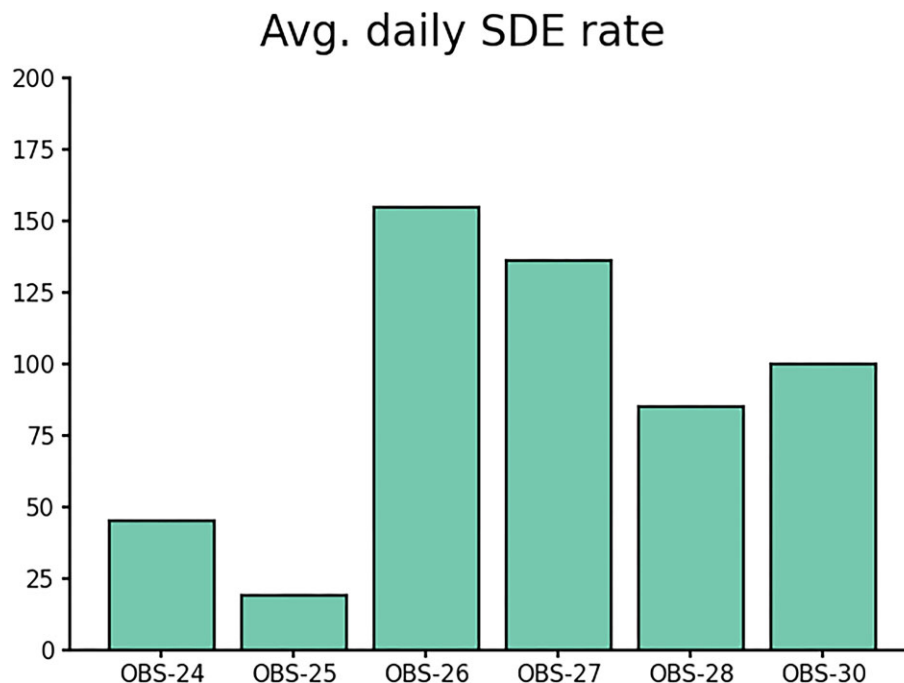
Figure 10. Average daily rate of SDE occurrence obtained from the Random Forest classification method using the final trained model (model F) for all OBS stations used in this study.
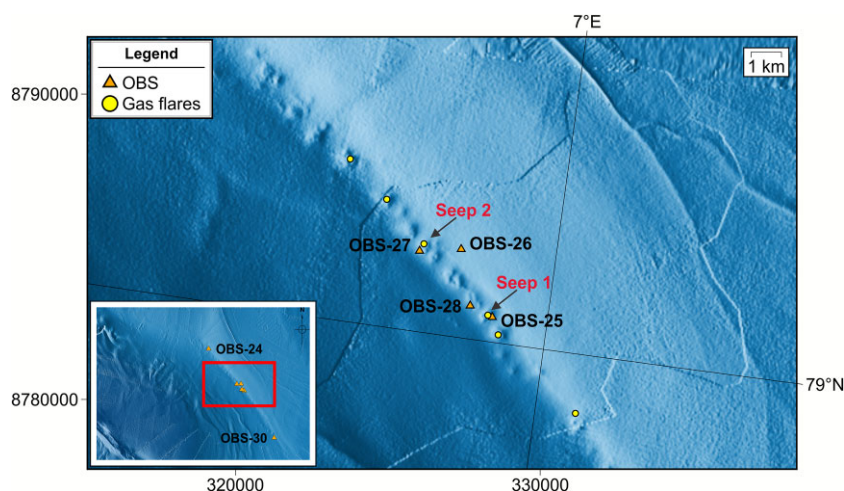

Figure 11. Positions of OBS-25, OBS-26, OBS-27 and OBS-28 at the crest of Vestnesa Ridge with relation to proven, continuous gas seepage sites. Seep sites discussed in the text labelled in red with the arrows. Inset shows which part of the map from Fig. 1 is presented here in a greater detail, with positions of OBS-24 and OBS-30 further away from seep sites.

10 Hz. This differentiation is linked with the energy in the lower quarter of spectrum used (1–6.25 Hz in this case), which is also present on the list of the most important features (on the ninth position, Fig. 7).

6 APPLICATION OF THE TRAINED MODEL ON SDE TIME SERIES ANALYSIS

With the Model F established to classify between event types with high accuracy on all stations, we have tested the capability of the method to build a SDE catalogue useful for subsequent analysis, alongside its earthquake detecting capability. Out of a total number of 180 000 STA/LTA triggers on all seven stations of the network,

model F classified 2374 as earthquakes, almost 73 000 as SDEs and the remainder as noise.

These numbers clearly underline the importance of SDEs in marine environments where they can outnumber earthquake detections, in our deployment setting by a factor of 30. We present their hourly occurrence throughout the duration of the experiment (Fig. 9). The overall intensity varies greatly between the stations, with OBS-25 having the least amount of the detections (3580—4.9 per cent of total SDE count), despite one of the longest duration of recordings. OBS-28 recorded the greatest amount of SDEs total over its roughly five-month-long record—16 822 (23 per cent of total SDE count). The total detection numbers for remaining stations are: OBS-24–16 425 (22.5 per cent of total SDE count), OBS-26–14 431 (19.8 per cent of total SDE count), OBS-27–12 674 (17.4
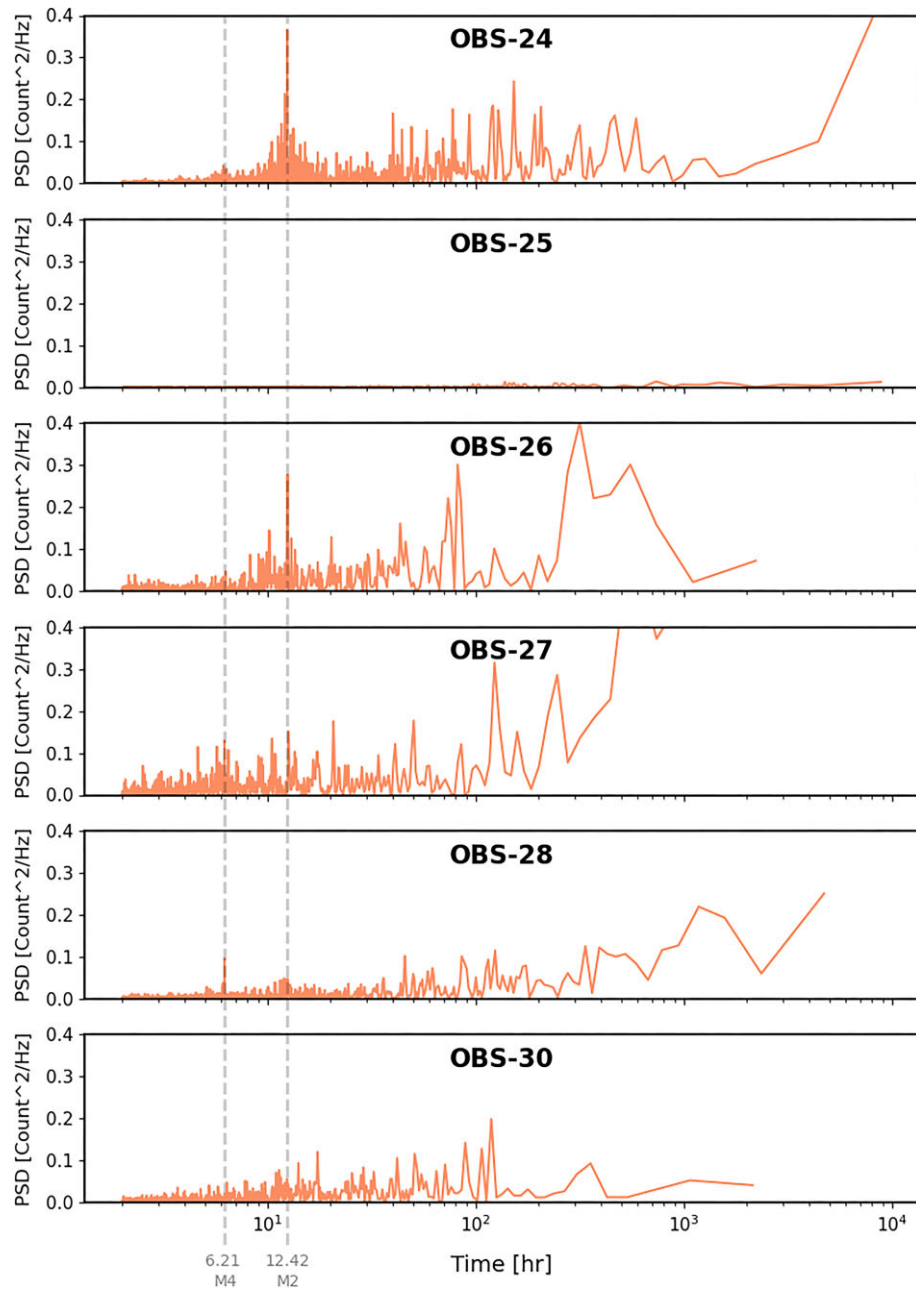
Figure 12. Periodograms (power spectrum density) plots of hourly SDE occurrences at investigated OBS stations. Two dashed lines correspond to principal lunar semidiurnal tidal component (M2–12.42 hr) and first overtide of M2 component (M4–6.31 hr).

per cent of total SDE count), and OBS-30–9032 (12.4 per cent of total SDE count). To better reflect the SDE occurrence intensity across the stations with different data set lengths, we computed their average daily rate of SDE occurrence, which is as follows: 45 for OBS-24, 19 for OBS-25, 155 for OBS-26, 136 for OBS-27, 85 for OBS-28 and 100 for OBS-30 (Fig. 10). SDEs typically occur in a form of seemingly irregular bursts, separated by periods of relative quiescence, with a varying frequency between stations.

The OBS data we use for the testing of Random Forest machine learning approach in a marine environment was collected as part of a study on fracture-controlled seepage dynamics along Vestnesa Ridge (e.g. Singhroha *et al.*2020; Plaza-Faverola *et al.* 2015). SDEs documented along continental margins globally have been linked to fluid migration and gas release from the seafloor via cracking of subseabed sediment in response to external factors (e.g. tides, e.g. Hsu *et al.*2013; Bayrakci *et al.*2014; Embriaco *et al.*2014; Hilmo & Wilcock 2020; Domel *et al.*2022).

The origin of SDEs is a matter of ongoing debate and goes beyond the scope of the present study. However, with the classification approach presented in this study we were able to quickly produce a data set of SDE detections that should be a more accurate depiction of SDE intensity than unfiltered STA/LTA picker results utilized so far (e.g. Tary *et al.*2012; Batsi *et al.*2019; Domel *et al.*2022; Ugalde *et al.* 2019). The resulting catalogue of SDEs includes six OBS stations distributed at varying distances from gas seepage pits and underlying structures of gas migration (Bunz *et al.*2012; Panieri *et al.*2017; Fig 11).

Preliminary analyses of SDE counts with respect to their distribution along Vestnesa Ridge fluid migration system reveal interesting observations with implications for advancing knowledge on SDEs:

(i) We observe the absolute lowest daily rate of SDEs at OBS-25, which is the station closest to one of the gas pits in the area (*ca.* 150 m from Seep 1 in Fig. 11). OBS-28 located 650 m away from the same seep has higher, but lower than other nearby stations, daily rate of SDE occurrence (85 per day; Fig. 10).

(ii) The two stations with the highest average daily count of SDEs (OBS-26 and OBS-27; Fig. 10) are located *ca.* 1200 and 650 m, respectively, from another seep site (Seep 2 in Fig. 11).

(iii) Two OBSs located *ca.* 11 and 14 km west and east from the cluster of seeps (OBS-24 and OBS-30, respectively; Fig. 11), recorded a sustained number of SDEs throughout the experiment duration.

Overall, this data shows that SDE intensity does not necessarily increase with the decreasing distance to the active seepage pits. On one hand, these observations challenge the perception that SDEs are associated with gas bubble rise at the seafloor (e.g. Tary *et al.* 2012; Hsu *et al.* 2013; Batsi *et al.* 2019). On the other hand, an increase of SDEs in a closed fluid flow system (shallow gas accumulations without an active gas release at the seafloor) supports observations from piezometer data where sub-seabed excess pore pressures fluctuate in response to the sea level changes in the region (Sultan *et al.* 2020).

A correlation between SDEs and tidal cycles is partially visible in periodograms of hourly counts of SDEs (Fig. 12). There are visible peaks of the SDE periodicity corresponding to semidiurnal tide component M2 (12.42 hr period) in the case of stations OBS-24 and OBS-26, and potentially a small peak of the first overtide of this component (M4–6.21 h period) in the case of OBS-28 (Fig. 12). Since this group represents the stations both within the area with an active seepage (OBS-26 and OBS-28) and away from it (OBS-24), no conclusive links can be established between seepage, tides and intensity of SDEs for the results of this experiment.

7 CONCLUSION

We applied a Random Forest classifier used previously in seismology on land to recognize signals in marine data from OBSs. We focused specifically on recognition of earthquakes and signals referred to as short duration events from the noise. These events are an interesting case study due to their high incidence rate, relatively similar characteristic between events, high/signal to noise ratio and a prevalence of STA/LTA based research with only some degree of manual verification in literature. to recognize between earthquakes, short duration events and noise present in the data from OBSs. We used the set of features previously tested on seismological data from land and found it robust enough to differentiate between the distinct classes of events with high accuracy during the training phase. During the verification on the continuous data sets, we encountered two major challenges. First, the resulting earthquake catalogues contained a large number of non-earthquake events. This is due to the high percentage of noise present. We attempted to mitigate this by including a progressively larger number of noise examples during the training phase for the noise class. This led to an eventual improvement with the percentage of correct earthquake detections rising, but not reaching the high recall values suggested by the initial testing of the recognition between the classes. The second challenge

lies in the categorization of noise in a marine realm. Combining different types of signals into a single NOISE category hampers the possibility of applying the same training model to data sets from different instruments. This can be potentially solved by creating separate categories for noisy signals of different origin (e.g. bottom currents, ship noise, instrument vibrations and mammal signals).

Overall, we achieved a mean accuracy of 87 per cent for event classification (final model; after training using examples from six OBS stations). The final results show that the signal characteristics that play the most important role in training the algorithm match the features we use in visual differentiation of events. Using the final iteration of the model we were able to compile a catalogue of 73 000 SDEs, close to two orders of magnitude larger than the amount of SDE examples used for training. With this new catalogue, we studied the variability of SDE between the stations during the duration of experiment. The data set we obtained shows an apparent relation between the daily SDE rate and the distance of active seep pits. We also observe some of the same periodic patterns reported from other studies about SDEs. This reinforces the approach presented as an efficient and potentially more precise method of SDE event detection and recognition in the absence of large enough SDE databases to train deep-learning models.

Additionally, this supervised machine learning method can potentially be used to discern other sources of signal in the marine seismological data, such as mammal vocalizations, volcanic tremors or ship noises. This requires further investigation and compilation of a larger database of different types of events to achieve a better generalization of the approach.

DATA AVAILABILITY

Raw OBS data will be uploaded to University of Tromsø open research repository (DataverseNO; https://dataverse.no/) in 2023. The land data earthquake catalogue was obtained from the International Seismological Centre (ISC) Seismological Dataset Repository (https://doi.org/10.31905/6TJZECEY). The ObsPy package was used for processing of the seismological data (Krischer *et al.* 2015). Plotting waveforms and spectrograms was done using Pyrocko (Heimann *et al.* 2017). Figs 1 and 11 were created using Generic Mapping Tools (GMT; Wessel *et al.* 2019). Bathymetry of the region provided by the International Bathymetric Chart of the Arctic Ocean (IBCAO) (Jakobsson *et al.* 2020). Supplemental material for this article includes a text describing the performance of popular machine learning approaches in discussed data set with figures showing examples of their performance. Additional text file contains information about the final model, with the order of importance of the signal characteristics after training.

REFERENCES

Allen, R., 1982. Automatic phase pickers: their present use and future prospects, *Bull. seism. Soc. Am.,* **72**(6B), S225–S242.

Anant, K.S. & Dowla, F.U., 1997. Wavelet transform methods for phase identification in three-component seismograms, *Bull. seism. Soc. Am.,* **87**(6), 1598–1612.

Baer, M. & Kradolfer, U., 1987. An automatic phase picker for local and teleseismic events, *Bull. seism. Soc. Am.,* **77**(4), 1437–1445.

Baillard, C., Crawford, W. C., Ballu, V., Hibert, C. & Mangeney, A., 2013. An Automatic Kurtosis-Based P- and S-Phase Picker Designed for Local Seismic Networks, *Bull. seism. Soc. Am.,* **104**(1), 394–409.

Baillard, C., Crawford, W.C., Ballu, V., Hibert, C. & Mangeney, A., 2014. An automatic kurtosis-based P- and S-phase picker designed for local seismic networks, *Bull. seism. Soc. Am.,* **104**(1), 394–409.

Barstow, N., Sutton, G.H. & Carter, J.A., 1989. Particle motion and pressure relationships of ocean bottom noise: 3900 M depth; 0.003 to 5 Hz. *Geophys. Res. Lett.,* **16**(10), 1185–1188.

Batsi, E. *et al.*, 2019. Nonseismic signals in the ocean: indicators of deep sea and seafloor processes on ocean-bottom seismometer data. *Geochem. Geophys. Geosyst.,* **20**(8), 3882–3900.

Bayrakci, G. *et al.*, 2014. Acoustic monitoring of gas emissions from the seafloor. Part II: a case study from the Sea of Marmara, *Mar. Geophys. Res.,* **35**(3), 211–229.

Bogiatzis, P. & Ishii, M., 2015. Continuous wavelet decomposition algorithms for automatic detection of compressional- and shear-wave arrival times, *Bull. seism. Soc. Am.,* **105**(3), 1628–1641.

Breiman, L., 2001. Random forests, *Mach. Learn.,* **45**(1), 5–32.

Bünz, S., 2023a. CAGE19-1 cruise report: passive and active ocean-bottom seismic surveys at Vestnesa Ridge, west-Svalbard margin within the framework of the SEAMSTRESS project, CAGE – Centre for Arctic Gas Hydrate, Environment and Climate Report Series 7.

Bünz, S., 2023b. CAGE20-5 cruise report: tectonic stress studies and seismic surveys on the West-Svalbard margin, CAGE – Centre for Arctic Gas Hydrate. Environment and Climate Report Series 8.

Bünz, S., Polyanov, S., Vadakkepuliyambatta, S., Consolaro, C. & Mienert, J., 2012. Active gas venting through hydrate-bearing sediments on the Vestnesa Ridge, offshore W-Svalbard, *Mar. Geol.,* **332-334**, 189–197.

Buskirk, R.E., Frohlich C. Latham G.V. Chen, A.T. & Lawton J. 1981. Evidence that biological activity affects ocean bottom seismograph recordings, *Marine Geophysical Researches,* **5**(), 189–205.

Chen, H., Yang, H., Zhu, G., Xu, M., Lin, J. & You, Q., 2022. Deep outer-rise faults in the Southern Mariana Subduction Zone indicated by a machine-learning-based high-resolution earthquake catalog, *Geophys. Res. Lett.,* **49**, e2022GL097779.

Chmiel, M., Walter, F., Wenner, M., Zhang, Z., McArdell, B.W. & Hibert, C., 2021. Machine learning improves debris flow warning, *Geophys. Res. Lett.,* **48**(3), e2020GL090874.

Chu, C. K. & Mendel, J., 1994. First break refraction event picking using fuzzy logic systems: IEEE Transaction on Fuzzy Systems, **2**, 255–266.

De Meersman, K., Kendall, J.M. & van der Baan, M., 2009. The 1998 Valhall microseismic data set: an integrated study of relocated sources, seismic multiplets, and S-wave splitting, *Geophysics,* **74**(5), B183–B195.

Díaz, J., Gallart, J. & Gaspà, O., 2007. Atypical seismic signals at the Galicia Margin, North Atlantic Ocean, related to the resonance of subsurface fluid-filled cracks, *Tectonophysics,* **433**(1–4), 1–13.

Diehl, T., Deichmann, N., Kissling, E. & Husen, S., 2009. Automatic S-wave picker for local earthquake tomography, *Bull. seism. Soc. Am.,* **99**(3), 1906–1920.

Domel, P., Singhroha, S., Plaza-Faverola, A., Schlindwein, V., Ramachandran, H. & Bünz, S., 2022. Origin and periodic behavior of short duration signals recorded by seismometers at Vestnesa Ridge, an active seepage site on the West-Svalbard continental margin, *Front. Earth Sci.,* **10**, doi:10.3389/feart.2022.831526.

Embriaco, D. *et al.*, 2014. Monitoring of gas and seismic energy release by multiparametric benthic observatory along the North Anatolian Fault in the Sea of Marmara (NW Turkey), *Geophys. J. Int.,* **196**(2), 850–866.

Essing, D., Schlindwein, V., Schmidt-Aursch, M.C., Hadziioannou, C. & Stähler, S.C., 2021. Characteristics of current-induced harmonic tremor signals in ocean-bottom seismometer records, *Seismol. Res. Lett.,* **92**, 3100–3112.

Falcin, A. *et al.*, 2021. A machine-learning approach for automatic classification of volcanic seismicity at La Soufrière Volcano, Guadeloupe, *J. Volc. Geotherm. Res.,* **411**, doi:10.1016/j.jvolgeores.2020.107151.

Franek, P., Mienert, J., Buenz, S. & Géli, L., 2014. Character of seismic motion at a location of a gas hydrate-bearing mud volcano on the SW Barents Sea margin, *J. geophys. Res.: Solid Earth,* **119**(8), 6159–6177.

Franek, P., Plaza-Faverola, A., Mienert, J., Buenz, S., Ferré, B. & Hubbard, A., 2017. Microseismicity linked to gas migration and leakage on the Western Svalbard Shelf, *Geochem. Geophys. Geosyst.,* **18**(12), 4623–4645.

Gelchinsky, B. & Shtivelman, V., 1983. Automatic picking of first arrivals and parametrization of traveltime curves, *Geophys. Prospect.,* **31**(6), 915–928.

Gentili, S. & Michelini, A., 2006. Automatic picking of P and S phases using a neural tree, *J. Seismol.,* **10**(1), 39–63.

Guan, Z. & Niu, F., 2017. An investigation on slowness-weighted CCP stacking and its application to receiver function imaging, *Geophys. Res. Lett.,* **44**(12), 6030–6038.

Heimann, S. *et al.*, 2017. Pyrocko - an Open-Source Seismology Toolbox and Library. V. 0.3. GFZ Data Serv.

Hibert, C. *et al.*, 2014. Automated identification, location, and volume estimation of rockfalls at Piton de la Fournaise volcano, *J. geophys. Res.,* **119**(5), 1082–1105.

Hibert, C., Michéa, D., Michéa, D., Provost, F., Malet, J. P. & Geertsema, M., 2019. Exploration of continuous seismic recordings with a machine learning approach to document 20 yr of landslide activity in Alaska, *Geophys. J. Int.,* **219**(2), 1138–1147.

Hibert, C., Provost, F., Malet, J.-P., Maggi, A., Stumpf, A. & Ferrazzini, V., 2017. Automatic identification of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise volcano using a Random Forest algorithm, *J. Volc. Geotherm. Res.,* **340**, 130–142.

Hildebrand, J.A., 2009. Anthropogenic and natural sources of ambient noise in the ocean, *Mar. Ecol. Prog. Ser.,* **395**, 5–20.

Hilmo, R. & Wilcock, W.S.D., 2020. Physical sources of high-frequency seismic noise on cascadia initiative ocean bottom seismometers, *Geochem. Geophys. Geosyst.,* **21**(10), doi:10.1029/2020GC009085.

Himmler, T., Sahy, D., Martma, T., Bohrmann, G., Plaza-Faverola, A., Bünz, S., *et al.*, 2019. A 160,000-year-old history of tectonically controlled methane seepage in the Arctic, *Science Advances,* **5**(8), eaaw1450.

Hsu, S.-K., Wang, S.-Y., Liao, Y.-C., Yang, T.F., Jan, S., Lin, J.-Y. & Chen, S.-C., 2013. Tide-modulated gas emissions and tremors off SW Taiwan, *Earth planet. Sci. Lett.,* **369-370**, 98–107.

Jakobsson, M. *et al.*, 2020. The International Bathymetric Chart of the Arctic Ocean Version 4.0, *Sci Data,* **7**(1), 176.

Jeddi, Z., Ottemöller, L., Sørensen, M. B., Rezaei, S., Gibbons, S. J. & Strømme, M. L., 2021. Improved Seismic Monitoring with OBS Deployment in the Arctic: A Pilot Study from Offshore Western Svalbard, *Seismol. Res. Lett.,* **92**(5), 2705–2717.

Jurkevics, A., 1988. Polarization analysis of three-component array data, *Bull. seism. Soc. Am.,* **78**(5), 1725–1743.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C. & Wassermann, J., 2015. ObsPy: a bridge for seismology into the scientific Python ecosystem, *Comput. Sci. Discov.,* **8**(1), doi:10.1088/1749-4699/8/1/014003.

Løviknes, S., Jensen, K.H., Krafft, B.A., Anthonypillai, V. & Nøttestad, L., 2021. Feeding hotspots and distribution of fin and humpback whales in the Norwegian Sea from 2013 to 2018, *Front. Mar. Sci.,* **8**, doi:10.3389/fmars.2021.632720.

Maggi, A., Ferrazzini, V., Hibert, C., Beauducel, F., Boissier, P. & Amemoutou, A., 2017. Implementation of a multistation approach for automated event classification at Piton de la Fournaise Volcano, *Seismol. Res. Lett.,* **88**(3), 878–891.

Malfante, M., Dalla Mura, M., Mars, J.I., Métaxian, J.-P., Macedo, O. & Inza, A., 2018. Automatic classification of volcano seismic signatures, *J. geophys. Res.,* **123**(12), 10 645–10 658.

McCormack, M.D., Zaucha, D.E. & Dushek, D.W., 1993. First-break refraction event picking and seismic data trace editing using neural networks, *Geophysics,* **58**(1), 67–78.

McDonald, M.A., Hildebrand, J.A. & Webb, S.C., 1995. Blue and fin whales observed on a seafloor array in the northeast pacific, *J. acoust. Soc. Am.,* **98**(2 Pt 1), 712–721.

Meier, M., Schlindwein, V., Scholz, J.R., Geils, J., Schmidt-Aursch, M.C., Kruger, F., Czuba, W. & Janik, T., 2021. Segment-scale seismicity of the ultraslow spreading Knipovich Ridge, *Geochem. Geophys. Geosyst.,* **22**(2), e2020GC009375, doi:10.1029/2020GC009375.

Molyneux, J.B. & Schmitt, D.R., 1999. First-break timing; arrival onset times by direct correlation, *Geophysics,* **64**(5), 1492–1501.

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, *Nat. Commun.,* **11**(1), 3952. https://www.ncbi.nlm.nih.gov/pubmed/32770023.

Mousavi, S.M., Langston, C.A. & Horton, S.P., 2016. Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform, **81**(4), V341–V355.

Münchmeyer, J. *et al.*, 2022a. Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers, *J. geophys. Res.,* **127**(1), doi:10.1029/2021JB023499.

Nippress, S.E.J., Rietbrock, A. & Heath, A.E., 2010. Optimized automatic pickers: application to the ANCORP data set, *Geophys. J. Int.,* **181**(2), 911–925.

Panieri, G. *et al.*, 2017. An integrated view of the methane system in the pockmarks at Vestnesa Ridge, 79 degrees N, *Mar. Geol.,* **390**, 282–300.

Pedregosa, F. *et al.*, 2011. Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.,* **12**, 2825–2830.

Plaza-Faverola, A., Bünz, S., Johnson, J. E., Chand, S., Knies, J., Mienert, J. & Franek, P., 2015. Role of tectonic stress in seepage evolution along the gas hydrate-charged Vestnesa Ridge, Fram Strait, *Geophys. Res. Lett.,* **42**(3), 733–742.

Provost, F., Hibert, C. & Malet, J.P., 2017. Automatic classification of endogenous landslide seismicity using the Random Forest supervised classifier, *Geophys. Res. Lett.,* **44**(1), 113–120.

Ramakrushana Reddy, T., Dewangan, P., Arya, L., Singha, P. & Kamesh Raju, K.A., 2020. Tidal triggering of the harmonic noise in ocean-bottom seismometers, *Seismol. Res. Lett.,* **91**(2A), 803–813.

Ross, Z.E., Meier, M.-A. & Hauksson, E., 2018b. P-wave arrival picking and first-motion polarity determination with deep learning. *J. geophys. Res.,* **123**(6), 5120–5129.

Ross, Z.E., Meier, M.A., Hauksson, E. & Heaton, T.H., 2018a. Generalized seismic phase detection with deep learning, *Bull. seism. Soc. Am.,* **108**(5A), 2894–2901.

Saragiotis, C.D., Hadjileontiadis, L.J., Rekanos, I.T. & Panas, S.M., 2004. Automatic P phase picking using maximum kurtosis and /spl kappa/-statistics criteria, *IEEE Geosci. Remote Sens. Lett.,* **1**(3), 147–151.

Sgroi, T., Polonia, A., Beranzoli, L., Billi, A., Bosman, A., Costanza, A. *et al.*, 2021. One Year of Seismicity Recorded Through Ocean Bottom Seismometers Illuminates Active Tectonic Structures in the Ionian Sea (Central Mediterranean), *Front. Earth Sci.,* 9.

Singhroha, S., Bünz, S., Plaza-Faverola, A. & Chand, S., 2020. Detection of gas hydrates in faults using azimuthal seismic velocity analysis, Vestnesa Ridge, W-Svalbard Margin, *J. gophys. Res.,* **125**(2), doi:10.1029/2019JB017949.

Sleeman, R. & van Eck, T., 1999. Robust automatic P-phase picking: an online implementation in the analysis of broadband seismogram recordings, *Phys. Earth planet. Inter.,* **113**(1), 265–275.

Sohn, R.A., Hildebrand, J.A., Webb, S.C. & Fox, C.G., 1995. Hydrothermal microseismicity at the megaplume site on the southern Juan de Fuca Ridge, *Bull. seism. Soc. Am.,* **85**(3), 775–786.

Soule, D.C. & Wilcock, W.S., 2013. Fin whale tracks recorded by a seismic network on the Juan de Fuca Ridge, Northeast Pacific Ocean, *J. acoust. Soc. Am.,* **133**(3), 1751–1761.

Stähler, S.C., Schmidt-Aursch, M.C., Hein, G. & Mars, R., 2018. A self-noise model for the German DEPAS OBS pool, *Seismol. Res. Lett.,* **89**(5), 1838–1845.

Sultan, N., Plaza-Faverola, A., Vadakkepuliyambatta, S., Buenz, S. & Knies, J., 2020. Impact of tides and sea-level on deep-sea Arctic methane emissions, *Nat. Commun.,* **11**(1), doi:10.1038/s41467-020-18899-3.

Sutton, G.H. & Latham, G.V., 1964. Analysis of a feedback-controlled seismometer, *J. geophys. Res.,* **69**(18), 3865–3882.

Sutton, G.H., McDonald, W.G., Prentiss, D.D. & Thanos, S.N., 1965. Ocean-bottom seismic observatories, *Proc. IEEE,* **53**(12), 1909–1921.

Tary, J.B., Géli, L., Guennou, C., Henry, P., Sultan, N., Çağatay, N. & Vidal, V., 2012. Microevents produced by gas migration and expulsion at the seabed: a study based on sea bottom recordings from the Sea of Marmara, *Geophys. J. Int.,* **190**(2), 993–1007.

Ugalde, A., Gaite, B., Ruiz, M., Villaseñor, A. & Ranero, C.R., 2019. Seismicity and noise recorded by passive seismic monitoring of drilling operations offshore the eastern Canary Islands, *Seismol. Res. Lett.,* **90**(4), 1565–1576.

Vidale, J.E., 1986. Complex polarization analysis of particle motion, *Bull. seism. Soc. Am.,* **76**(5), 1393–1405.

Wenner, M., Hibert, C., van Herwijnen, A., Meier, L. & Walter, F., 2021. Near-real-time automated classification of seismic signals of slope failures with continuous random forests, *Nat. Haz. Earth Syst. Sci.,* **21**(1), 339–361.

Wessel, P., Luis, J.F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W.H.F. & Tian, D., 2019. The Generic Mapping Tools version 6, *Geochem. Geophys. Geosyst.,* **20**, 5556–5564.

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.

Zhu, W. & Beroza, G.C., 2018. PhaseNet: a deep-neural-network-based seismic arrival time picking method, *Geophys. J. Int.,* **216**(1), doi:10.1093/gji/ggy423.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

FigureS1. Example 1 of the earthquake detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

FigureS2. Example 2 of the earthquake detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S3. Example 3 of the earthquake detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S4. Example 4 of the earthquake detection on data recorded in a marine setting. The upper plot shows a waveform from the

vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S5. Example 5 of the earthquake detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S6. Example 5 of the earthquake detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S7. Example 1 of the SDE detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S8. Example 2 of the SDE detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S9. Example 3 of the SDE detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S10. Example 4 of the SDE detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S11. Example 5 of the SDE detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability

for EQTransformer; P and S phase probability for the remaining two).

Figure S12. Example 6 of the SDE detection on data recorded in a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S13. Example 1 of the detections recorded in noise examples for data from a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S14. Example 2 of the detections recorded in noise examples for data from a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S15. Example 3 of the detections recorded in noise examples for data from a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S16. Example 4 of the detections recorded in noise examples for data from a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

Figure S17. Example 5 of the detections recorded in noise examples for data from a marine setting. The upper plot shows a waveform from the vertical channel of a seismometer used for the detection (all methods use three-component data). Below presented are characteristic functions computed by each approach (Detection, P and S phase probability for EQTransformer; P and S phase probability for the remaining two).

supplement_attribute_importances.txt

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

APPENDIX

Table A1. Signal features calculated from the examples used to train machine learning classifier [features calculation following Provost *et al.* (2017) and Hibert *et al.* (2017)]. Numbers-channel pairs represent order of the features in the array used for classification. DFT stands for discrete fourier transform.

Number-seismometer channel	Description	Formula
Waveform-based features		
1-Z, 63-Y, 121-X	Duration of the signal	$t_j - t_i$, where t_i and t_j: start and end of the signal
2-Z, 64-Y, 122-X	Ratio of the max and the mean of the normalized envelope	$\max[e(t)]/\mathrm{mean}[e(t)]$
3-Z, 65-Y, 123-X	Ratio of the max and the mean of the normalized envelope	$\max[e(t)]/\mathrm{median}[e(t)]$
4-Z, 66-Y, 124-X	Ascending time/decreasing time of the envelope	$\frac{t_{max}-t_i}{t_j-t_{max}}$, t_{max}: time of the largest amplitude
5-Z, 67-Y, 125-X	Kurtosis (peakness) of the raw signal	$\frac{m_4}{\sigma^4}$, m_4: fourth moment, σ: standard deviation
6-Z, 68-Y, 126-X	Kurtosis of the signal envelope	see 5
7-Z, 69-Y, 127-X	Skewness of the raw signal	$\frac{m_3}{\sigma^3}$, m_3: third moment
8-Z, 70-Y, 128-X	Skewness of the signal envelope	see 7
9-Z, 71-Y, 129-X	Number of peaks in the autocorrelation function of the raw signal	-
10-Z, 72-Y, 130-X	Energy in the first 1/3 of the autocorrelation function	$\int_0^{\frac{T}{3}} C(\tau)\mathrm{d}\tau$, T: signal duration, C: autocorrelation function
11-Z, 73-Y, 131-X	Energy in the remaining part of the autocorrelation function	See 10
12-Z, 74-Y, 132-X	Ratio of the 10 and 11	-
13–17-Z, 75–79-Y, 133–137-X	Energy of the signal in: 1–5 Hz, 5–10 Hz, 10–15 Hz, 15–20 Hz, 10-Nyquist frequency	$ES_i = \log_{10} \int_0^T y_f(t)\mathrm{d}t$, y_f: filtered signal in the frequency range
18–22-Z, 80–84-Y, 138–142-X	Kurtosis of the signal in: 1–5 Hz, 5–10 Hz, 10–15 Hz, 15–20 Hz, 10-Nyquist frequency	see 5
23-Z, 85-Y, 143-X	RMS between the coda of the signal and the straight line $l(t) = Y_{max-}\frac{Y_{max}}{t_f-t_{max}}t$	$\sqrt{\overline{Y(t)-l(t)}^2}$, Y: envelope of the signal
24-Z, 86-Y, 144-X	Ratio between maximum of the envelope of the signal and signal duration	-
Frequency-related features		
25-Z, 87-Y, 145-X	Mean of the DFT	DFT: discrete Fourier transform
26-Z, 88-Y, 146-X	Maximum of the DFT	-
27-Z, 89-Y, 147-X	Frequency at the maximum of DFT	-
28-Z, 90-Y, 148-X	Frequency of the spectral centroid	$f(\gamma_1)$, see 39
29-Z, 91-Y, 149-X	Central frequency of the 1st quartile	-
30-Z, 92-Y, 150-X	Central frequency of the 3rd quartile	-
31-Z, 93-Y, 151-X	Median of the normalized DFT	-
32-Z, 94-Y, 152-X	Variance of the normalized DFT	-
33-Z, 95-Y, 153-X	Number of peaks (>0.75 DFT_{MAX})	DFT_{MAX}: maximum of the DFT
34-Z, 96-Y, 154-X	Mean peaks value from 33	-
35–38-Z, 97–100-Y, 155–158-X	Energy in 0–0.25*Nyq., 0.25–0.5*Nyq., 0.5–0.75*Nyq., 0.75–1*Nyq., where Nyq. -Nyquist frequency of the signal	$\int_{f_1}^{f_2} DFT(f)\mathrm{d}f$, f_1,f_2: corresponding frequency range
39-Z, 101-Y, 159-X	Spectral centroid	$\gamma_1 = \frac{m_2}{m_1}$, m_1, m_2: first and second moment
40-Z, 102-Y, 160-X	Spectral gyration radius	$\gamma_2 = \sqrt{\frac{m_3}{m_2}}$, m_3: third moment
41-Z, 103-Y, 161-X	Spectral centroid width	$\sqrt{\gamma_1^2-\gamma_2^2}$
Pseudo spectrogram features (calculated with DFTs of 10-s length and an overlap of 90 per cent)		
42-Z, 104-Y, 162-X	Kurtosis of the envelope of the maximum energy on the spectrograms	Kurtosis $[\overset{max}{\underset{t\,=\,0,\ldots,T}{}}(SPEC(t,\ f)]$, SPEC (t,f): spectrogram
43-Z, 105-Y, 163-X	Kurtosis of the envelope of the median energy on all spectrograms	see 42
44-Z, 106-Y, 164-X	Mean ratio between the maximum and the mean of all DFTs	mean $\left[\frac{\max(SPEC)}{\mathrm{mean}(SPEC)}\right]$
45-Z, 107-Y, 165-X	Mean ratio between the maximum and the median of all DFTs	see 44
46-Z, 108-Y, 166-X	Number of the peaks in the curve showing the temporal evolution of the DFTs maximum	-
47-Z, 109-Y, 167-X	Number of the peaks in the curve showing the temporal evolution of the DFTs mean	-
48-Z, 110-Y, 168-X	Number of the peaks in the curve showing the temporal evolution of the DFTs median	-
49-Z, 111-Y, 169-X	Ratio between 46 and 47	-
50-Z, 112-Y, 170-X	Ratio between 46 and 48	-
51-Z, 113-Y, 171-X	Number of peaks in the curve of the temporal evolution of the DFTs central frequency	-

Table A1. Continued

Number-seismometer channel	Description	Formula
52-Z, 114-Y, 172-X	Number of peaks in the curve of the temporal evolution of the DFTs maximum frequency	-
53-Z, 115-Y, 173-X	Ratio between 51 and 52	-
54-Z, 116-Y, 174-X	Mean distance between the curves of the temporal evolution of the DFTs maximum frequency and mean frequency	-
55-Z, 117-Y, 175-X	Mean distance between the curves of the temporal evolution of the DFTs maximum frequency and median frequency	-
56-Z, 118-Y,176-X	Mean distance between the 1st quartile and the median of all DFTs as a function of time	-
57-Z, 119-Y, 177-X	Mean distance between the 3rd quartile and the median of all DFTs as a function of time	-
58-Z, 120-Y, 178-X	Mean distance between the 3rd quartile and the 1st quartile of all DFTs as a function of time	-
Polarity features (all channels used)		
59	Rectilinearity	$1 - \frac{\lambda_{11} + \lambda_{22}}{2\lambda_{33}}$, $\lambda_{33} \gg \lambda_{22} \gg \lambda_{11}$
60	Azimuth	$\arctan(\lambda_{23}/\lambda_{13}) \times 180/\pi$
61	Dip	$\arctan(\lambda_{33}/\sqrt{\lambda_{23}^2 + \lambda_{13}^2}) \times 180/\pi$
62	Planarity	$1 - \frac{2\lambda_{11}}{\lambda_{33} + \lambda_{22}}$