ACCEPTED: Bidirectional recurrent neural networks for seismic event detection [GEO-2020-0806]

Claire Birnie and Fredrik Hansteen, Equinor ASA

ABSTRACT

Real time, accurate passive seismic event detection is a critical safety measure across a range of monitoring applications from reservoir stability to carbon storage to volcanic tremor detection. The most common detection procedure remains the Short-Term-Average to Long-Term-Average (STA/LTA) trigger developed in the 1970's, in part due to its easy implementation and real-time processing capability. However, it has a number of well-documented limitations such as requiring a signal-to-noise ratio greater than one and being highly sensitive to the trigger parameters. Whilst numerous alternatives have been proposed, they often are tailored to a specific monitoring setting and therefore cannot be widely applied, or they are too computationally expensive and therefore cannot be run real time. This work introduces a deep learning approach to event detection that is an alternative to the STA/LTA trigger. A bi-directional, long-short-term memory, neural network is trained solely on synthetic traces. Evaluated on synthetic and field data, the neural network approach significantly outperforms the STA/LTA trigger both on the number of correctly detected arrivals as well as on reducing the number of falsely detected events. Its applicability is proven with 600 traces processed in real time on a single processing unit.

INTRODUCTION

Automated seismic event detection is the vital first step in passive seismic monitoring whether on a global or local scale. In this paper we focus on passive microseismic monitoring - a fundamental element in geothermal, carbon sequestration and hydraulic fracture monitoring. Real time, automated detection procedures play two important roles: primarily for immediate hazard detection (Kanamori, 2005), and secondly to act as a filter prior to performing more computationally expensive procedures such as deriving the source location, magnitude and focal mechanism (Withers et al., 1998). Passive seismic monitoring often focuses on microseismic events, the majority of which are observed below the noise level. Therefore, any detection procedure must be sufficiently able to detect events where the Signal to Noise Ratio (S/N) is below one.

One of the most common techniques for full-field event detection is the Short-Time Average to Long-Time Average (STA/LTA) method (Allen, 1978). However, for the STA/LTA method, and most adaptions of it, the parameters require careful selection and are closely linked to the recording conditions requiring re-calibration should the recording conditions change (Vaezi and Van der Baan, 2015; Kumar et al., 2018; Zheng et al., 2018). Over the years many alternative approaches have been proposed for event detection: waveform template matching (Gibbons and Ringdal, 2006), stacking procedures (Chambers et al., 2010), and machine learning approaches (Yoon et al., 2015). However, these new approaches often proved to be more computationally expensive and time-consuming than traditional detection methods (e.g., Skoumal et al., 2016,and references therein) and therefore are not routinely applied.

Recently, Machine Learning (ML) methodologies, in particular Deep Learning (DL), has seen a resurgence across all fields of seismology, and wider geoscience applications. The majority of applications have focused on the adaptation of computer vision approaches for aiding seismic interpretation. For example, salt body detection (Waldeland et al., 2018), fault detection (Araya-Polo et al., 2017), and horizon detection (Wu and Zhang, 2018). However, other studies have considered how ML methodologies can be utilised for seismic event detection. Chen et al. (2019) proposed the combination of convolutional Neural Networks (NN) with k-means clustering for detection of seismic arrivals, whilst Zhao and Takano (1999) presents one of the earliest studies that considered using NNs for event detection on broadband seismometers. Novelly, they combined three different back-propagation NNs that were trained with short-, mid-, and long-term features, in a manner mimicking the short and long term fundamentals of the STA/LTA detection procedure.

Hochreiter and Schmidhuber (1997) proposed a NN architecture called Long Short Term Memory (LSTM) that allows time series to be recursively passed across cells of the network and therefore has the ability to retain information on temporal trends in the input data. Trained using data derived from rock physics experiments, Zheng et al. (2018) showed that LSTM networks could be used to develop a seismic event detection procedure with a higher detection rate and reduced noise, in comparison to the STA/LTA trigger. Similarly, both Mousavi et al. (2019) and Ross et al. (2019) have investigated the use of networks incorporating LSTM sections for global applications: the former proposing a complex architecture combining both convolutional and recurrent NN elements for seismic arrival detection whilst the latter using an LSTM network to classify pre-detected arrivals on multiple sensors into a single event catalogue. Linville et al. (2019) proposed another recent application of LSTM

networks for the classification of events between earthquakes and quarry blasts. Until recently the majority of NN approaches have focused on time series analysis, however a new school of thought is emerging where methodologies from computer vision are being adapted, i.e., inputs are now being treated as images. For example, Stork et al. (2020) developed an object detection procedure for event detection on dense monitoring arrays. Leveraging the commonly-used UNet architecture of Ronneberger et al. (2015), Birnie et al. (2020b) advanced on the work of Stork et al. (2020) to propose an image segmentation approach that garners additional information on the arrival times at individual stations. Whilst, PhaseNet of Zhu and Beroza (2019) utilises the same UNet architecture to tackle the problem of phase detection at individual stations.

In this paper, focusing on microseismic event detection, we advance on the study of Zheng et al. (2018) by incorporating bidirectionality into the NN and training solely on synthetic data. We show that the incorporation of bidirectionality significantly improves the event detection rate, in particular reducing the number of false positives, and allows for detection of events at lower S/Ns than previously possible. Training on synthetic data allowed the opportunity to create a diverse training set reducing the possibility of overfitting the detection procedure to a specific data set. Our approach is validated by applying the same network to two different field data sets and is shown to accurately detect events with a low false detection rate.

DATA

Supervised machine learning approaches typically require extensive labelled training data sets. These labels become the 'ground truth' for the training of a machine learning model and therefore it is vital that they accurately match their data counterpart. For seismic event detection, labels indicate where a seismic arrival is present within a recording. When using recorded data for training, labels are most commonly generated by manual annotation. However, when the data has been collected as part of a controlled experiment, labels can be automatically generated through knowledge of the experimental conditions, as shown by Zheng et al. (2018).

Two large downsides exist for using recorded data as a training data set for passive seismic event detection. Firstly, it is impossible to determine if all events have been detected within a seismic recording; as such, the labels would only represent events that are currently detectable. Therefore, a model trained on such events would ultimately only be able to detect events at the same detection level. Secondly, the model can only learn from the distribution it has been trained on. For example, if events were previously only detected from events with focal mechanisms resulting in positive-polarity arrivals then it is unlikely a model would detect negative-polarity arrivals as it was not exposed to these during training. To overcome these limitations, in this work we have used synthetic seismic data sets derived from a standard convolutional seismic modelling approach. The use of synthetic seismic data for training gives the opportunity to vary the training data set with a large range of expected event properties, as well as ensuring that no events are mislabelled in the data set.

Synthetic seismic traces are generated following the standard workflow for convolutional seismic traces (Russell, 1988) with the input parameters and noise varying between traces. The central frequency is chosen from a uniform distribution between 20 and 30 Hz whilst

the wavelet polarity is sampled from a Bernoulli distribution with a probability of 0.5. The event arrival time is sampled from a uniform distribution over the trace length. The S/N is sampled between 0.5 and 2.0 from a right-skewed Gaussian distribution to ensure a focus on low S/N events during training. The noise added to the trace is generated from many coloured Gaussian distributions scaled according to previously recorded noise such that the energy within each 2 Hz window is representative of the recorded noise energy distribution, similar to the methodology proposed by Pearce and Barley (1977). Figure 1 illustrates the mean recorded energy across the frequency bands and the similar distribution observed on the generated noise models. After generation, the noise is added to the synthetic trace at the desired S/N, where the S/N is defined as the peak signal amplitude to the root-mean-square of the noise amplitudes.

The labels are generated from the waveform data prior to the noise summation. An event is defined as present within the data where the absolute amplitude of the waveform is greater than 10^{-3} as illustrated in Figure 2. Figure 3 provides examples of traces, and their respective labels, at different S/Ns. The synthetic seismic traces are all bandpassed between 2 and 25 Hz prior to being normalised by the maximum absolute amplitudes.

Ten thousand synthetic traces are generated for the training of the NNs, with a further five thousand synthetic traces generated for initial testing. An important factor to consider is that the network does not learn to always detect an arrival within a trace. As such, some traces are generated that contain only noise. For training, the percentage of traces containing only noise is investigated as described in the methodology section below. In the test data set, i.e., the further 5000 traces, 25% are pure noise traces. A second test data set is created for analysing robustness to noise. This test data set contains 1000 traces where the data has S/N-levels of between 0.2 and 3.0 with a step of 0.1. The wavelet features and event arrival time and polarity follow the same random distributions as previously described.

For a final synthetic analysis, the trained networks are applied to a full waveform synthetic data set generated using a double-couple source where $M_{11} = -1$, $M_{33} = 1$ and the remaining moment tensor components all equal zero. The source is embedded in the Marmousi velocity model recorded by a surface receiver array, as illustrated in Figure 4a. The source emits a Ricker wavelet with a 20 Hz central frequency. The elastic, finite difference modelling is performed using the Madagascar software package (Fomel et al., 2013) and only the vertical component is retained for the arrival detection. The noise-free wavefield data is shown in Figure 4b whilst Figure 4c shows the noisy synthetic data with a moderate S/N. Noise is generated in the same manner as for the previous synthetic data sets. To imitate field recordings that often have increased noise levels around the center of the array due to production equipment (e.g., injection wells), the noise energy is scaled such that it dissipates away from the center of the array by a factor of $1/\sqrt{d}$ where d is the distance from the center of the array.

To conclude, the trained networks are applied to known events from both land and marine data sets to analyse their applicability to field data. The land data set is subset of the hydraulic fracture monitoring undertaken by the Tony Creek Dual Microseismic Experiment (ToC2ME) field program. The ToC2ME array consists of 69 cemented shallow boreholes and in this study we are only considering the vertical component of the 3C geophone, of which there are one per well. The data set has been extensively analysed by Eaton et al. (2018), Poulin et al. (2019) and Rodríguez-Pradilla and Eaton (2020), among

others. We have analysed a ten minute recording beginning at 14:00 on November 1st 2016 and used the event catalogue of Eaton et al. (2018) to guide our analysis. The marine field data set includes an event that was recorded by a Permanent Reservoir Monitoring (PRM) system deployed on the seafloor in the North Sea. The PRM system comprises of 3458 sensors, 3C geophones with a hydrophone, arranged in a gridded-style as shown in Figure 2 of Thompson et al. (2015). Similar to the ToC2ME example, we only utilise the Z-component for event detection. This event is referred to as the "G8 event"and has been previously analysed using a small subset of receivers by Bussat et al. (2018).

METHODOLOGY

Recurrent Neural Networks (RNN) are particularly beneficial for time series processing as they are capable of storing past inputs to produce the currently desired output, i.e., they can use earlier time steps to predict the desired outcome at a later time step. Each training step involves a forward pass of the data through the network to compute the prediction error for a batch of N_{batch} samples,

$$E = \sum_{i=1}^{N_{batch}} \mathcal{L}(y_i, f_\theta(x_i)) \tag{1}$$

where $\mathcal{L}$ denotes the loss function, θ are the networks parameters' (weights and biases), y_i is the true label for the i-th sample, and $f_\theta(x_i)$ is the NNs predicted label from the i-th input x_i. The prediction error is back propagated through the network to compute its gradient with respect to each of the parameters to be updated, $\frac{\partial E}{\partial \theta}$. The update is generally performed by subtracting a scaled version of the gradient from the parameter as follows:

$$\theta \leftarrow \theta - \alpha \frac{\partial E}{\partial \theta} \tag{2}$$

Note that more complicated updates are usually preferred to improve the stability of the optimisation as performed by the likes of the Adam optimizer (Kingma and Ba, 2014) used in this work.

Standard RNNs are vulnerable to the vanishing gradient problem due to the large network sizes (Hochreiter, 1998). This occurs when the gradient computed for determining weight updates becomes smaller and smaller, i.e., $\frac{\partial E}{\partial \theta} \rightarrow 0$, such that the parameters' updates become insignificant and therefore, the network stops learning. LSTM networks are an adaption to standard RNNs that include a forget gate which has the ability to learn which data in a sequence is important to keep or throw away (Hochreiter and Schmidhuber, 1997). These networks consist of cells as illustrated in Figure 5a. Each cell contains three gates: a forget gate, an input gate, and an output gate. Given the input, x_t, and information from the previous units, the forget gate controls what information is forgotten. The following gate, the input gate, controls what of the new information will be encoded into the cell's state. Whilst the final gate, the output gate, controls what of the encoded information is sent to the network as input to the next cell. The presence of the forget gate, along with the additive property of the cells' state gradients, removes the challenges of the vanishing gradient problem.

Initially proposed by Kosko (1988), the inclusion of bidirectionality into RNN architectures allows the network to look both forwards and backwards in time (within a given

window length). This has been shown to greatly improve performance for classification and regression problems (Schuster and Paliwal, 1997). Bi-directional LSTMs (BLSTM) consist of two layers of LSTM cells, one passing information forward whilst the other passing information backwards as illustrated in Figure 5b.

The following section outlines the creation and training of both LSTMs and BLSTMs for the purpose of seismic event detection. As well as, the implementation of an STA/LTA autotrigger for benchmarking the NN approaches against.

Training

Both the LSTM and BLSTM networks are trained with a batch size of 100, validation split of 10%, a maximum number epochs set to 100, and early stopping with a patience of 15. Early stopping is a methodology that helps to avoid overfitting when training NNs by stopping the training when the model performance has stopped improving on the validation data set (Caruana et al., 2001). In this example, if the models performance on the validation data set has not improved within the last 15 epochs, the training is stopped.

The model architecture and its training are dependant on a number of hyperparameters: namely, the length of the time window, the percentage of traces with events, the number of units in the LSTM cell, and the type of loss function. The different options for each hyperparameter is listed in Table 1. The optimum values of the hyperparameters is determined by means of a grid search approach. Note that the parameter *# of units* follows TensorFlow software's terminology and represents the size of the hidden and output states (Abadi et al., 2016). The optimum parameter combination is determined by calculating the F1-score for each parameter combination on a holdout data set, where the F1-score is computed as,

$$F_1 = \frac{TP}{TP + 0.5(FP + FN)} \tag{3}$$

where TP denotes the number of true positives in the results whilst FP and FN denote the number of false positives and false negatives, respectively. The F1-score is particularly useful in problems such as event detection where there is a severe imbalance in the data set classes, i.e., there are significantly fewer timesteps with arrivals than those without.

The results of the gridsearch are illustrated in Figure 6 where the following parameters resulted in the highest F1-score for both the LSTM and BLSTM:

- **percent of traces with signal:** 100,
- **trace window length:** 2s,
- **units:** 100,
- **loss:** binary cross-entropy,

The models trained with these parameters are used in the analysis and benchmarking of the detection procedures detailed in the remainder of the paper. Figure 7 illustrates the accuracy and loss as the training progresses for the models with the optimum training parameters. The close similarity between the training and validation curves indicates the

networks' strong performance on data it wasn't exposed to during training. Although this is likely to happen due to both datasets being generated in the same manner, this result shows that the models have not merely learned the training data but can generalize well on unseen seismic recordings.

Post-processing and evaluation

In the context of seismic event detection, the performance metric of a detection procedure is typically defined as follows: an event is accurately detected within a time window, X. As such, identification of the exact sample indicative of the first-break is often less important than accurate detection of an events arrival. This is slightly contradictory to most out-of-box metrics for evaluating the performance of NNs which compare sample-to-sample between a prediction and label - a resolution that is not needed in this use case. As such, a post-processing workflow is performed that groups time samples into events as outlined in Figure 8.

First, detected arrivals within 50 milliseconds of another arrival are grouped, where an arrival is defined as a number of neighbouring samples that are all identified as containing a seismic arrival. Secondly, any arrivals with a duration of less than 25 milliseconds are removed. The resulting grouped detections are classified as seismic events.

STA/LTA for benchmarking

Throughout this study the STA/LTA algorithm is used for benchmarking purposes. A classic implemenation of the STA/LTA trigger is used via the ObsPy Python package (Beyreuther et al., 2010).

The STA/LTA algorithm is well known to be highly sensitivity to the trigger parameters, see for example, Vaezi and Van der Baan (2015). Zheng et al. (2018) summarised the challenges faced when using the STA/LTA method as: "there have been difficulties encountered in the selections of the parameters, such as window size and threshold, which are the main weaknesses of the STA/LTA method."

To ensure optimum performance and a fair comparison with the DL approaches an exhaustive parameter sweep is performed across the same training data sets as used for the DL model training. Table 2 details the parameter ranges investigated in the parameter sweep. The same evaluation procedure is utilised as in the DL grid search with F1-score being the evaluation metric. Figure 9 details the performance of each parameter combination with the following parameter combination performing the best:

- **short time window:** 0.05s,
- **long time window:** 3.5s,
- **detection threshold:** 5.

This parameter combination is used for the STA/LTA trigger for benchmarking the DL model performances. In addition to the STA/LTA trigger with the above parameter combination, an optimum parameter combination is recomputed for the two different field data

sets. The comparison of the synthetically derived parameters with the optimum parameters serves to illustrate the difference between the optimum STA/LTA performance and that when the parameters are not tuned to the specific recording. For the ToC2ME data set, the optimum parameters are computed using the F1-score for eight events in the ten minutes analysed, guided by the event catalogue of Eaton et al. (2018). For the G8 event, as this is a single event, the optimum parameters are hand-tuned via a trial-and-error approach.

RESULTS

Synthetic results

The STA/LTA, LSTM and BLSTM detection procedures were benchmarked against 5000 synthetic seismic traces with the characteristics as previously described. Four recording scenarios: very low S/N (0.5), low S/N (1.0), moderate S/N (1.5) and no event, along with their detection results are illustrated in Figure 10. The STA/LTA accurately detects both the low and moderate S/N events, however there is also a false detection in three out of the four examples. The LSTM also accurately detects the low and moderate S/N events however the very low S/N event is missed and a false detection is also observed on the low S/N event. The BLSTM creates no false detections and successfully detects the low and moderate S/N events. The red, dashed lines in the LSTM and BLTSM plots indicate the raw output of the NNs. We can observe that the raw outputs for the LSTM show more variability in general indicating that the LSTM network is less certain about the presence of an event. For both NN approaches, in the very low S/N example we observe an increase in the raw output at the events location, however not large enough to cause a trigger.

The results for the 5000 traces as a whole are portrayed in Figure 11. The top plot highlights that overall the BLSTM outperforms both the LSTM and STA/LTA detection procedures. The bottom plots illustrate the high false positive rate arising from the STA/LTA detection procedure, whilst the BLSTM is shown to outperform LSTM in the number of correctly identified events (i.e., the true positive rate).

Figure 12 details how the different detection procedures perform at different S/Ns. For all procedures, the probability of detecting an event below an S/N of one is minimal. For the NN procedures, all events above an S/N of 2.9 are successfully identified. At lower S/Ns (< 0.5) there is little differentiation between the different detection procedures. Above S/N of 0.5, both the LSTM approaches begin to outperform the STA/LTA detection procedure whilst the BLSTM begins to outperform them both around S/N of 0.7. Above S/N of 2.7, the performance difference between the BLSTM and LSTM detection procedures is negligible. Whilst the NN approaches were only trained up to an S/N of three, extending the S/N range up to five showed no change in their performance. In other words, both the triggers experienced the same performance level at higher S/Ns to what they exhibit at an S/N of three.

The Marmousi elastic finite difference results are illustrated in Figure 13 and Figure 14 for the full array and for a number of individual traces , respectively. Considering the full array, all the detection procedures are prone to detect the secondary, stronger arrival and struggle with detections around the center of the array, likely due to the lower amplitude of the arrivals combined with higher noise levels. Arguably, the NN approaches produce more false detections than the STA/LTA in this example however the BLSTM approach

also detects more correct detections than the other two procedures. The individual trace results in Figure 14 provide a close-up illustrating the performance of the different detection procedures, where the NN approaches have accurately detected the first arrival for all three traces despite a reverse in polarity being observed across the array. In general, the raw outputs show higher fluctuations in the LSTM predictions, particularly before the arrival, indicating the LSTM is less certain in the definition between the two classes, i.e., between whether a event is detected or not.

Detection on field data sets

Finally, our study is concluded by testing all three detection procedures against previously detected events from two different field recording environments. The first data set is the ToC2ME land field recording of hydraulic monitoring previously analysed by Eaton et al. (2018), Poulin et al. (2019) and Rodríguez-Pradilla and Eaton (2020). An optimum STA/LTA trigger has been computed using the same parameter sweep approach as before, based on the known arrival times as published by Eaton et al. (2018). The optimum paramaters are a threshold of seven, a short-time window of 150 ms and a long-time window of 2.5 s. Figure 15 shows the difference between the four different triggers across the ToC2ME array for two events with relatively high S/N. The STA/LTA with the original parameters results in a number of false triggers for both events. Considering the clustering of the trigger detections it is likely that the LSTM and BLSTM approaches produce one or two false triggers for the second event. Whilst the optimum STA/LTA does not produce such false triggers, it does likely miss the arrivals on a number of stations. Figure 16 illustrates the different triggers performance on three different traces for the same two events as Figure 15. Again, the original STA/LTA produces what are likely false detections, with the STA/LTA with optimum parameters detecting five out of the possible six arrivals without any additional triggers. Finally, both NN approaches detect only one event per trace, with the timing of the detection agreeing between the two NN procedures. When considering the arrival time across the array, the NN detections of an arrival from Event Two at Station 1 at close to $t = 1s$ is reasonable.

The second field dataset considered is a previously detected event from the Grane PRM system (Bussat et al., 2018). The optimum STA/LTA parameters were determined to be a threshold of three, a short-time window of 15 ms and a long-time window of 4 s. The top row of Figure 17 illustrates the z-component of the seismic data with zoomed-in sections around the noisy center of the array and the quieter edge of the array. The following rows show the event detection results from the three detection procedures: STA/LTA with the original parameters, STA/LTA with the optimum parameters, LSTM, and BLSTM, respectively. Single traces with detections are illustrated in Figure 18. The BLSTM detection outperforms the other detection procedures, with the arrivals clearly detected on the majority of tracesand with the least amount of noise. Considering the raw outputs of the NN approaches (red, dashed lines in Figure 18), the BLSTM detection has less fluctuations than the LSTM detection for the same window indicating the BLSTM model's greater confidence between noise and an event. However, whilst the number of false detections is reduced, they are still present within the BLSTM results. Unfortunately, all detection procedures mislabel the platform noise in the centre of the array as a seismic event, to varying extents.

DISCUSSION

The three main criteria for the development and deployment of seismic event detection procedures are:

- **Automatable**: requiring no parameter adaptation once in place,
- **Reliable over a wide span of S/N, including those less than one**: events must be accurately detected above and below the noise level, and
- **Real time**: computations must be performed faster than data streaming.

Throughout, we have benchmarked the performance of the DL triggers against the STA/LTA method, which is currently one of the most commonly-used detection procedure.

In this study, we have adequately trained the DL approaches such that they do not need parameter tuning (or retraining) when applied to a new data set. The non-trivial nature of STA/LTA parameter setting and the required updating of parameters as the noise environment changes is well documented (Trnkoczy, 2012). This was further highlighted in this study where the optimum STA/LTA parameters for the three different data sets (synthetic, ToC2ME and G8) were shown to drastically differ. In addition to this, the STA/LTA parameters derived from the synthetic data analysis were shown to result in moderately poor performance on both field data sets. On the other hand, the DL approaches which were trained on only the synthetic traces have been successfully applied to two different field data sets - one from a marine environment and one from land.

In this study, we have deemed it unsuitable for detection procedures to require parameter tuning either during the recording span of an array or when applying the detection procedures to a new field site for a similar monitoring application, e.g., microseismic monitoring. Following our approach, in theory a single network could be trained to detect the full range of seismic events ranging from high-amplitude, low-frequency to low-amplitude, high-frequency, provided an adequately varied training data set was provided. This would require a much larger, broader training data set and the resulting model would likely under-perform due to the generic nature of the task at hand. Future work will consider the limitations of a single model trained for such a large range of event properties versus the opportunities provided by transfer learning for reusing elements of a pre-trained network for optimising to a new task (Torrey and Shavlik, 2010).

The STA/LTA trigger is well-known to struggle at detecting events below the noise level (Withers et al., 1998). Both the LSTM and BLSTM approaches showed notable improvements on detection rates across the majority of S/Ns investigated: S/Ns of $0.8 - 3.0$. Again, the use of synthetic data sets provided an advantage here over the use of field or laboratory data, giving us the opportunity to define the S/N distribution of the training set to tailor the models to detect events at low S/Ns. Alternatives to the STA/LTA such as waveform template matching or stacking procedures have much higher resilience to noise and can detect events in low S/N situations. Waveform template matching procedures, for instance, have been shown to detect events at low S/Ns and have been successfully applied for microseismic monitoring of geothermal (Plenkers et al., 2013) and gas (Song et al., 2010) reservoirs. However, they require a-priori information for the waveform template, often extracted from an event catalogue, and therefore are not easily generalisable (Yoon

et al., 2015). Whereas, stacking procedures typically require multiple nearby stations and are therefore not applicable for single station or sparse array applications, ruling out their use for a significant portion of hazard monitoring scenarios.

PhaseNet of Zhu and Beroza (2019) offers a single-station, NN approach that has been shown to perform well at low S/Ns by including all three geophone components and performing the detection in the Fourier domain. Whilst PhaseNet transforms the data domain to provide a better separation between the signal and noise, an alternative approach could be to include a noise suppression procedure prior to detection. However, incorporating a successful automated noise suppression step would be non-trivial due to the continually varying nature of noise requiring continual adaptation of suppression procedures (Birnie et al., 2016, 2017). Moreover, transforming the data into a new domain or performing a noise suppression procedure both introduce additional computational complexity into the detection algorithm. The optimum alternative may be to include realistic noise or prerecorded noise into the training data set, as illustrated by Binder and Chakraborty (2019). The authors hypothesise that the incorporation of platform noise into the training data set may have removed the false detections on the field data set.

Finally, a real concern surrounding the development of any new event detection procedure is that they can be applied real time. Stork et al. (2018) cited this as the reason why new detection approaches are commonly not implemented. In this work, we define real time detection to be when the processing time is less than the duration of the trace window being analysed. Figure 19 illustrates the detection time for a two second window for the three approaches for increasing number of traces with a dashed line at the two second mark indicating when the procedure is no longer real time*. Whilst the BLSTM has a significantly longer compute time, it can perform detection real time on up to 600 traces on the current setup. As the traces are treated independently then the detection could easily be parallelised over a number of machines. The NN approaches require training prior to their implementation, for 10,000 training samples the training time for the LSTM and BLSTM networks was 45 minutes and one hour, respectively. Assuming a training data set representative of the recording setting, this is a one-off computational cost.

Assuming computational resources are not a concern, the BLSTM has been shown to be the more robust of the two DL detection procedures. Considering the raw network predictions, indicated by red, dashed lines in the individual trace results, the BLSTM is shown to have higher degrees of certainty in the arrivals detected. Similarly, the BLSTM detections typically cover more of the wavelets arrival than the LSTM detections. As both networks require an input of two seconds and have been trained to detect arrivals across this full window, there is no additional data requirement for the BLSTM network despite its inclusion of a second, backward-looking layer. However, if a large number of traces require processing at one time, the LSTM approach may be more favourable due to it's significantly reduced computational cost that is accompanied by only a minor decrease in the detection performance.

The use of synthetic data for training of event detection procedures has been employed in training deep learning applications for both global and microseismic monitoring purposes. For example, Ross et al. (2019)'s PhaseLink for earthquake phase association and Binder and Chakraborty (2019)'s NN approach to microseismic detection on DAS. Using

*All compute times are calculated on a Microsoft Azure NC6 virtual machine.

synthetic data sets also offers us the opportunity to have a greater understanding of what the model has been exposed to during training, particularly the S/N and noise characteristics of the training data. Furthermore, evaluating the results on such synthetic data, provides a greater understanding of the networks performance, either by allowing us to conduct a noise sensitivity analysis (i.e., Figure 12) or by aiding our interpretation of the raw NN outputs (the red, dashed lines in Figures 10, 14, 16, and 18). Such knowledge helps to provide confidence limits when the triggers are applied to field data where the event arrival time is unknown. Notably, this relies on the synthetics generated for the analysis being representative of field data, where over-simplistic synthetic data sets may produce misleading results, as illustrated by Birnie et al. (2020a).

This procedure was primarily developed to act as a filtering mechanism for real time selection of data segments to be further analysed by more expensive analysis algorithms, such as those for localisation or focal mechanism determination. Trained using rudimentary, convolutional synthetic data sets containing single P-wave arrivals, the NN approaches were validated using both an elastic finite difference synthetic data set and a field data set. Through the inclusion and separate labelling of S-waves in the training data set, the detection procedure could be extended to a multi-mode classification. Incorporation of coda in the training data set is also likely to further increase the performance of the NN approaches due to the additional signal available within the temporal window of interest.

To conclude, this work has proposed a novel NN-approach to seismic event detection requiring no manual labelling of the data. Whilst it provides a strong competitor to the commonly used STA/LTA event detection procedures, it shares a similar disadvantage in that it is a trace-by-trace approach. Post-processing can be used to attempt to enforce spatial-consistency in detections, for example through the inclusion of a coincidence filter (Joswig, 1993; Trnkoczy, 2012). However, these approaches are used to filter the results as opposed to being an integral part of the detection procedure. Both Binder and Chakraborty (2019) and Stork et al. (2020) illustrated how object detection procedures could be adapted to microseismic event detection to determine a spatial-temporal bounding box of first arrivals on DAS data, incorporating the spatial components into the NN training procedure. Birnie et al. (2020b) extended this approach to an image segmentation procedure therefore identifying the arrival time on each individual sensor. Incorporation of spatial components exponentially increases the number of NN parameters, as does moving from an object detection to segmentation task, and therefore introduces complications around memory requirements and computation time. Birnie and Jarraya (2020) detailed a preliminary study investigating the opportunity of using distributed NN training to allow efficient training of large NNs on seismic data. Future work will continue to investigate the optimum methodologies for incorporating spatial information into the training whilst maintaining a real time detection procedure with reasonable training times.

CONCLUSION

This study highlighted how the incorporation of bidirectionality into an LSTM network can improve seismic event detection procedures at low-mid S/Ns (< 2.7). Trained using only synthetic data and benchmarked against the commonly used STA/LTA event detection procedure, our BLSTM approach is shown to significantly increase the number of true detections whilst simultaneously reducing the number of false detections. The potential for

the BLSTM application for real time field detection is highlighted by accurate detection of a seismic events on two different monitoring arrays, as well as a compute time analysis highlighting the number of traces that can be processed in real time.

ACKNOWLEDGEMENTS

The authors would like to thank the Grane license partners Equinor Energy AS, Petoro AS, Vår Energi AS, and ConocoPhillips Skandinavia AS for allowing to present this work. The views and opinions expressed in this abstract are those of the Operator and are not necessarily shared by the license partners.

REFERENCES

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., 2016, Tensorflow: A system for large-scale machine learning: 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16), 265–283.

Allen, R. V., 1978, Automatic earthquake recognition and timing from single traces: Bulletin of the Seismological Society of America, **68**, 1521–1532.

Araya-Polo, M., T. Dahlke, C. Frogner, C. Zhang, T. Poggio, and D. Hohl, 2017, Automated fault detection without seismic processing: The Leading Edge, **36**, 208–214.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann, 2010, Obspy: A python toolbox for seismology: Seismological Research Letters, **81**, 530–533.

Binder, G., and D. Chakraborty, 2019, Detecting microseismic events in downhole distributed acoustic sensing data using convolutional neural networks, *in* SEG Technical Program Expanded Abstracts 2019: Society of Exploration Geophysicists, 4864–4868.

Birnie, C., K. Chambers, and D. Angus, 2017, Seismic arrival enhancement through the use of noise whitening: Physics of the Earth and Planetary Interiors, **262**, 80–89.

Birnie, C., K. Chambers, D. Angus, and A. L. Stork, 2016, Analysis and models of pre-injection surface seismic array noise recorded at the aquistore carbon storage site: Geophysical Journal International, **206**, 1246–1260.

———, 2020a, On the importance of benchmarking algorithms under realistic noise conditions: Geophysical Journal International, **221**, 504–520.

Birnie, C., and H. Jarraya, 2020, How to leverage advanced tensorflow and cloud computing for efficient deep learning on large seismic datasets: First EAGE Digitalization Conference and Exhibition, European Association of Geoscientists & Engineers, 1–5.

Birnie, C., H. Jarraya, F. Hansteen, and F. Schuchert, 2020b, Enhanced microseismic event detection using deep neural networks: 82nd EAGE Annual Conference & Exhibition, European Association of Geoscientists & Engineers, 1–5.

Bussat, S., M. Houbiers, and Z. Zarifi, 2018, Real-time microseismic overburden surveillance at the grane PRM field offshore norway: First Break, **36**, 63–70.

Caruana, R., S. Lawrence, and L. Giles, 2001, Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping: Advances in neural information processing systems, 402–408.

Chambers, K., J. Kendall, S. Brandsberg-Dahl, J. Rueda, et al., 2010, Testing the ability of surface arrays to monitor microseismic activity: Geophysical Prospecting, **58**, 821–830.

Chen, Y., G. Zhang, M. Bai, S. Zu, Z. Guan, and M. Zhang, 2019, Automatic waveform classification and arrival picking based on convolutional neural network: Earth and Space Science, **6**, 1244–1261.

Eaton, D. W., N. Igonin, A. Poulin, R. Weir, H. Zhang, S. Pellegrino, and G. Rodriguez, 2018, Induced seismicity characterization during hydraulic-fracture monitoring with a shallow-wellbore geophone array and broadband sensors: Seismological Research Letters, **89**, 1641–1651.

Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, 2013, Madagascar: Open-source software project for multidimensional data analysis and reproducible computational experiments: Journal of Open Research Software, **1**.

Gibbons, S. J., and F. Ringdal, 2006, The detection of low magnitude seismic events using array-based waveform correlation: Geophysical Journal International, **165**, 149–166.

Hochreiter, S., 1998, The vanishing gradient problem during learning recurrent neural nets

and problem solutions: International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, **6**, 107–116.

Hochreiter, S., and J. Schmidhuber, 1997, Long short-term memory: Neural computation, **9**, 1735–1780.

Joswig, M., 1993, Single-trace detection and array-wide coincidence association of local earthquakes and explosions: Computers & Geosciences, **19**, 207–221.

Kanamori, H., 2005, Real-time seismology and earthquake damage mitigation: Annu. Rev. Earth Planet. Sci., **33**, 195–214.

Kingma, D. P., and J. Ba, 2014, Adam: A method for stochastic optimization: arXiv preprint arXiv:1412.6980.

Kosko, B., 1988, Bidirectional associative memories: IEEE Transactions on Systems, man, and Cybernetics, **18**, 49–60.

Kumar, S., R. Vig, and P. Kapur, 2018, Development of earthquake event detection technique based on sta/lta algorithm for seismic alert system: Journal of the Geological Society of India, **92**, 679–686.

Linville, L., K. Pankow, and T. Draelos, 2019, Deep learning models augment analyst decisions for event discrimination: Geophysical Research Letters, **46**, 3643–3651.

Mousavi, S. M., W. Zhu, Y. Sheng, and G. C. Beroza, 2019, Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection: Scientific reports, **9**, 1–14.

Pearce, R., and B. Barley, 1977, The effect of noise on seismograms: Geophysical Journal International, **48**, 543–547.

Plenkers, K., J. R. Ritter, and M. Schindler, 2013, Low signal-to-noise event detection based on waveform stacking and cross-correlation: Application to a stimulation experiment: Journal of seismology, **17**, 27–49.

Poulin, A., R. Weir, D. Eaton, N. Igonin, Y. Chen, L. Lines, and D. Lawton, 2019, Focal-time analysis: A new method for stratigraphic depth control of microseismicity and induced seismic events: Geophysics, **84**, KS173–KS182.

Rodríguez-Pradilla, G., and D. W. Eaton, 2020, Automated microseismic processing and integrated interpretation of induced seismicity during a multistage hydraulic-fracturing stimulation, alberta, canada: Bulletin of the Seismological Society of America, **110**, 2018–2030.

Ronneberger, O., P. Fischer, and T. Brox, 2015, U-net: Convolutional networks for biomedical image segmentation: International Conference on Medical image computing and computer-assisted intervention, Springer, 234–241.

Ross, Z. E., Y. Yue, M.-A. Meier, E. Hauksson, and T. H. Heaton, 2019, Phaselink: A deep learning approach to seismic phase association: Journal of Geophysical Research: Solid Earth, **124**, 856–869.

Russell, B. H., 1988, Introduction to seismic inversion methods: SEG Books.

Schuster, M., and K. K. Paliwal, 1997, Bidirectional recurrent neural networks: IEEE transactions on Signal Processing, **45**, 2673–2681.

Skoumal, R. J., M. R. Brudzinski, and B. S. Currie, 2016, An efficient repeating signal detector to investigate earthquake swarms: Journal of Geophysical Research: Solid Earth, **121**, 5880–5897.

Song, F., H. S. Kuleli, M. N. Toksöz, E. Ay, and H. Zhang, 2010, An improved method for hydrofracture-induced microseismic event detection and phase picking: Geophysics, **75**, A47–A52.

Stork, A., C. Nixon, C. Hawkes, C. Birnie, D. White, D. Schmitt, and B. Roberts, 2018,

Is CO_2 injection at aquistore aseismic? A combined seismological and geomechanical study of early injection operations: International Journal of Greenhouse Gas Control, **75**, 107–124.

Stork, A. L., A. F. Baird, S. A. Horne, G. Naldrett, S. Lapins, J.-M. Kendall, J. Wookey, J. P. Verdon, A. Clarke, and A. Williams, 2020, Application of machine learning to microseismic event detection in distributed acoustic sensing data: Geophysics, **85**, KS149–KS160.

Thompson, M., M. Andersen, R. Elde, S. Roy, and S. Skogland, 2015, The startup of permanent reservoir monitoring for Snorre and Grane: 77th EAGE Annual Conference & Exhibition, European Association of Geoscientists & Engineers, 1–5.

Torrey, L., and J. Shavlik, 2010, Transfer learning, *in* Handbook of research on machine learning applications and trends: algorithms, methods, and techniques: IGI global, 242–264.

Trnkoczy, A., 2012, Understanding and parameter setting of STA/LTA trigger algorithm, *in* New Manual of Seismological Observatory Practice 2 (NMSOP-2): Deutsches GeoForschungsZentrum GFZ, 1–20.

Vaezi, Y., and M. Van der Baan, 2015, Comparison of the sta/lta and power spectral density methods for microseismic event detection: Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society, **203**, 1896–1908.

Waldeland, A. U., A. C. Jensen, L.-J. Gelius, and A. H. S. Solberg, 2018, Convolutional neural networks for automated seismic interpretation: The Leading Edge, **37**, 529–537.

Withers, M., R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo, 1998, A comparison of select trigger algorithms for automated global seismic phase and event detection: Bulletin of the Seismological Society of America, **88**, 95–106.

Wu, H., and B. Zhang, 2018, A deep convolutional encoder-decoder neural network in assisting seismic horizon tracking: arXiv preprint arXiv:1804.06814.

Yoon, C. E., O. O'Reilly, K. J. Bergen, and G. C. Beroza, 2015, Earthquake detection through computationally efficient similarity search: Science Advances, **1**, e1501057.

Zhao, Y., and K. Takano, 1999, An artificial neural network approach for broadband seismic phase picking: Bulletin of the Seismological Society of America, **89**, 670–680.

Zheng, J., J. Lu, S. Peng, and T. Jiang, 2018, An automatic microseismic or acoustic emission arrival identification scheme with deep recurrent neural networks: Geophysical Journal International, **212**, 1389–1397.

Zhu, W., and G. C. Beroza, 2019, PhaseNet: a deep-neural-network-based seismic arrival-time picking method: Geophysical Journal International, **216**, 261–273.

TABLES

Table 1: Parameter options for LSTM and BLSTM gridsearch procedure.

Parameter	**Options**
% of traces with event	25, 50, 75, 100
Window length (s)	0.5, 1, 2, 3
# of units	50, 100, 200
loss function	mean squared error, binary cross entropy

Table 2: Parameter ranges for exhaustive parameter sweep of STA/LTA trigger performance.

Parameter	**Minimum value**	**Maximum value**	**Step**
STA win. (s)	0.05	0.4	0.05
LTA win. (s)	1.0	4.0	0.5
Threshold	2	8	1

FIGURES

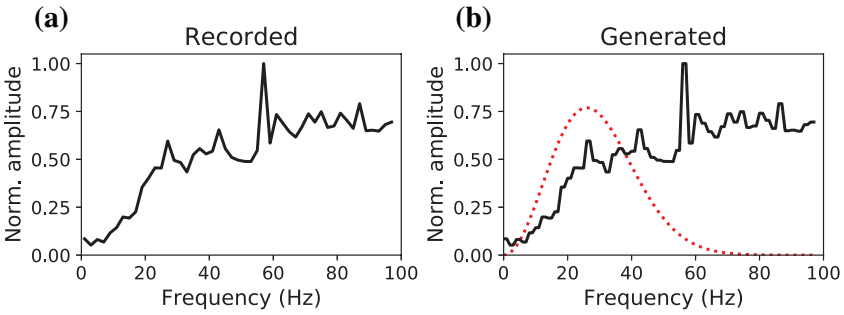

Figure 1: Frequency spectra of (a) recorded noise alongside (b) spectra of the noise and wavelet (dashed red line) from a sample synthetic seismic trace used for model training.

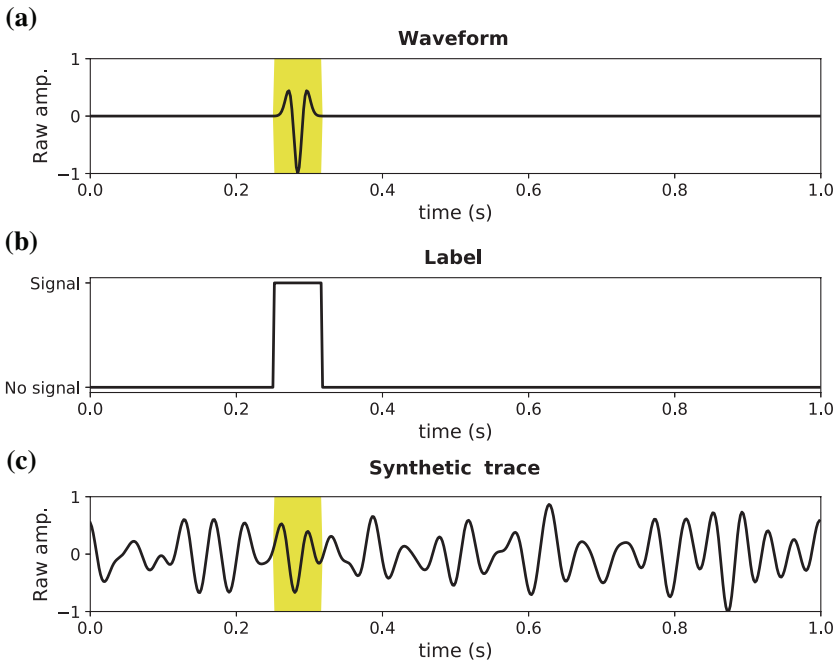

Figure 2: Labelling procedure of synthetic traces. The label in (b) is computed from the waveform data (a) prior to noise being added to generate the finalised synthetic trace (c). The highlighted area indicates where a signal is present in the normalised trace as defined by the labelling.

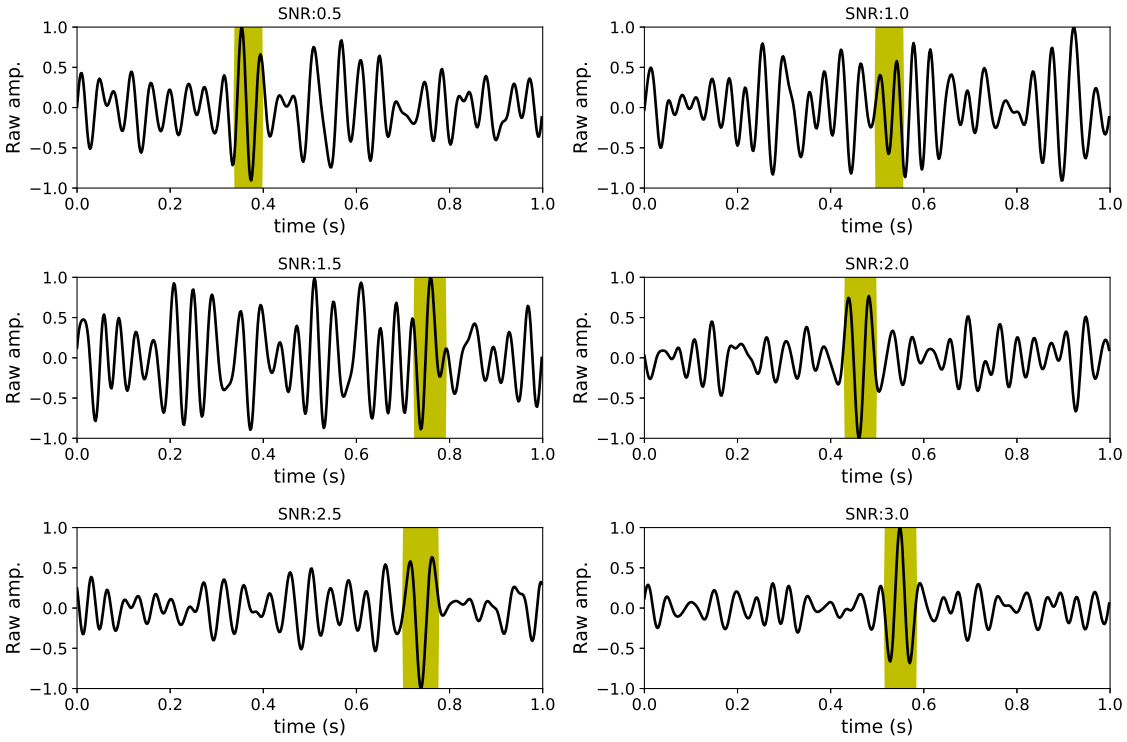

Figure 3: Examples of normalised synthetic seismic traces used for training and evaluation of event detection procedures. Highlighted area indicates where a signal is present in the trace.

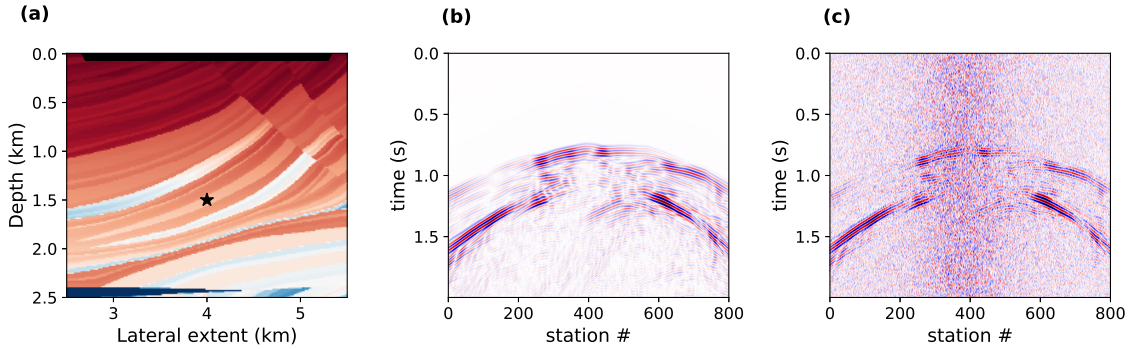

Figure 4: (a) P-wave Marmousi velocity model with the black star indicating the source location and black triangles indicating receiver positions. (b) Z-component wavefield data from elastic finite difference modelling of a source located as illustrated in (a) and (c) noisy synthetic data.

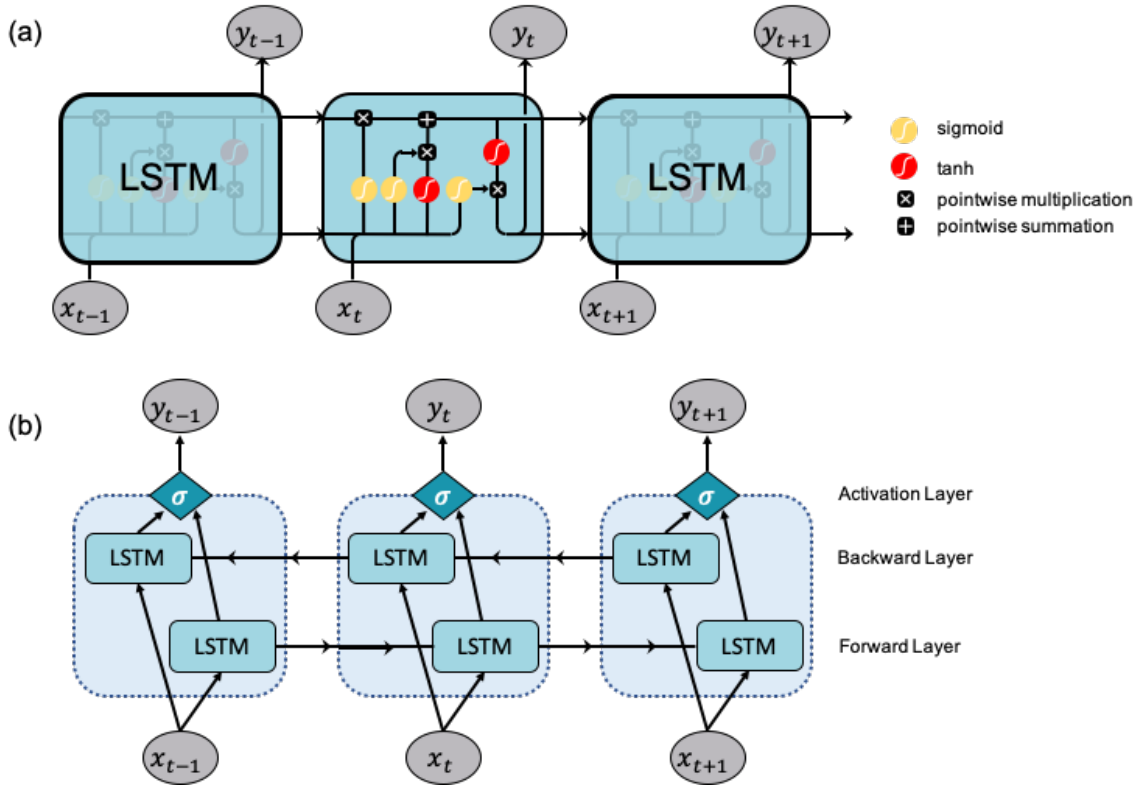

Figure 5: (a) Architecture of a single LSTM cell, and (b) architecture of a BLSTM cell, where x_t indicates inputs and y_t indicates outputs. The arrows indicate the direction of the information flow.

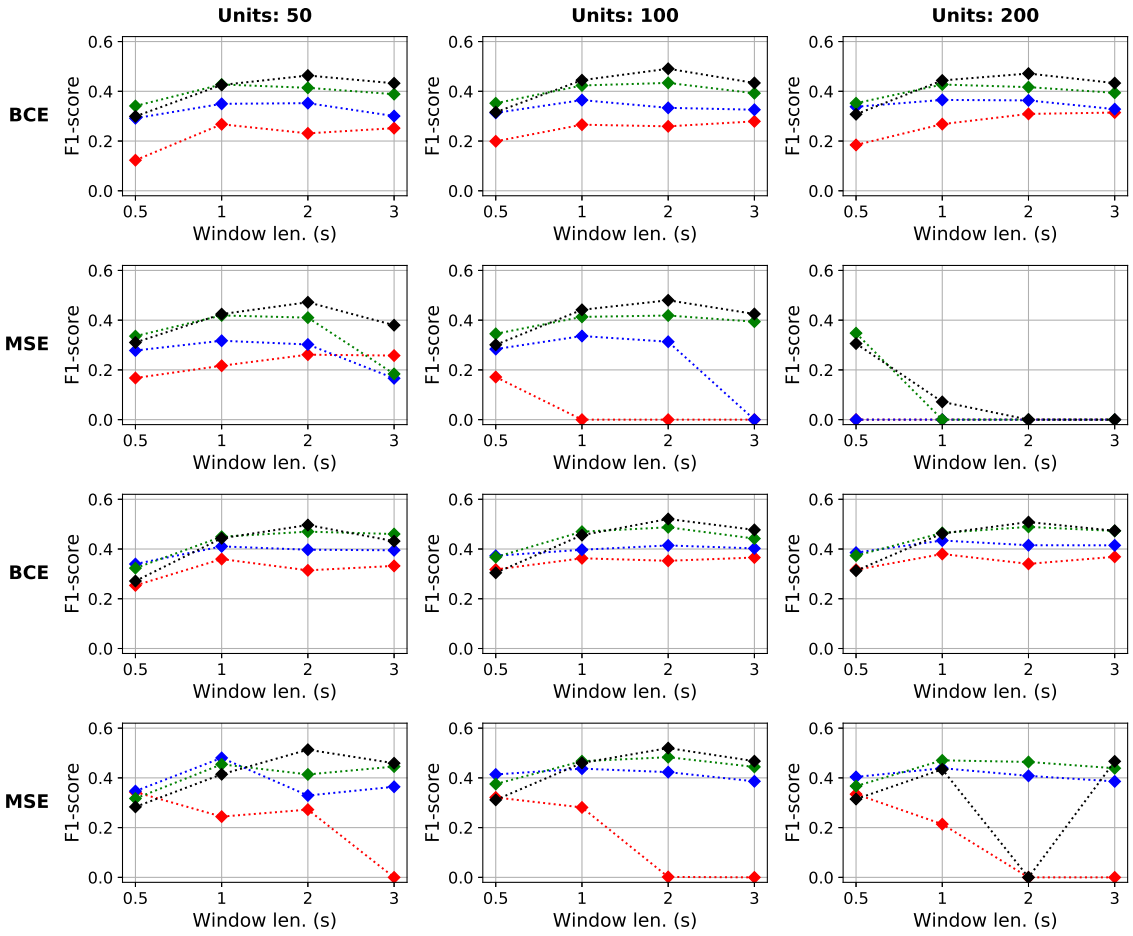

Figure 6: Grid search over window length, number of units and loss function for LSTM and BLSTM detection procedures, the top two and bottom two rows, respectively. The columns indicate the numbers of units in the models. The rows preceeded by "BCE" used a binary cross-entropy loss function whilst the rows preceeded by "MSE" used a mean-squared-error loss function. The colours indicate the percent of data samples that contained a seismic event. Red lines contained 25%, blue 50%, green 75% and black 100%.

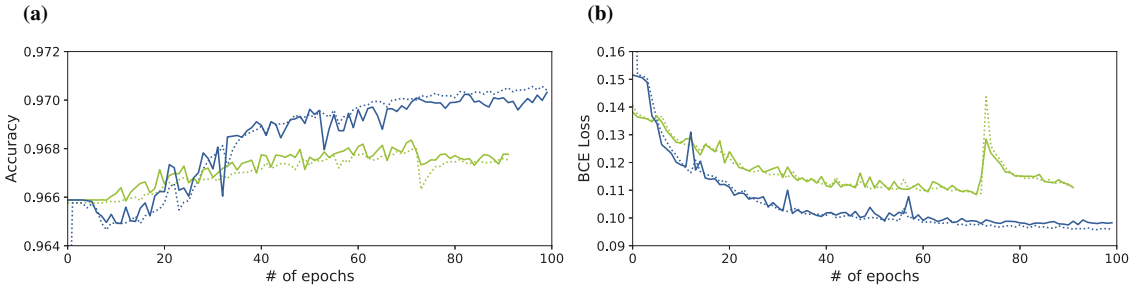

Figure 7: Evolution of (a) accuracy and (b) loss throughout the training of the LSTM (green) and BLSTM (blue) networks. Dashed lines indicate values computed on the training data whilst solid lines indicate values calculated on the validation set.

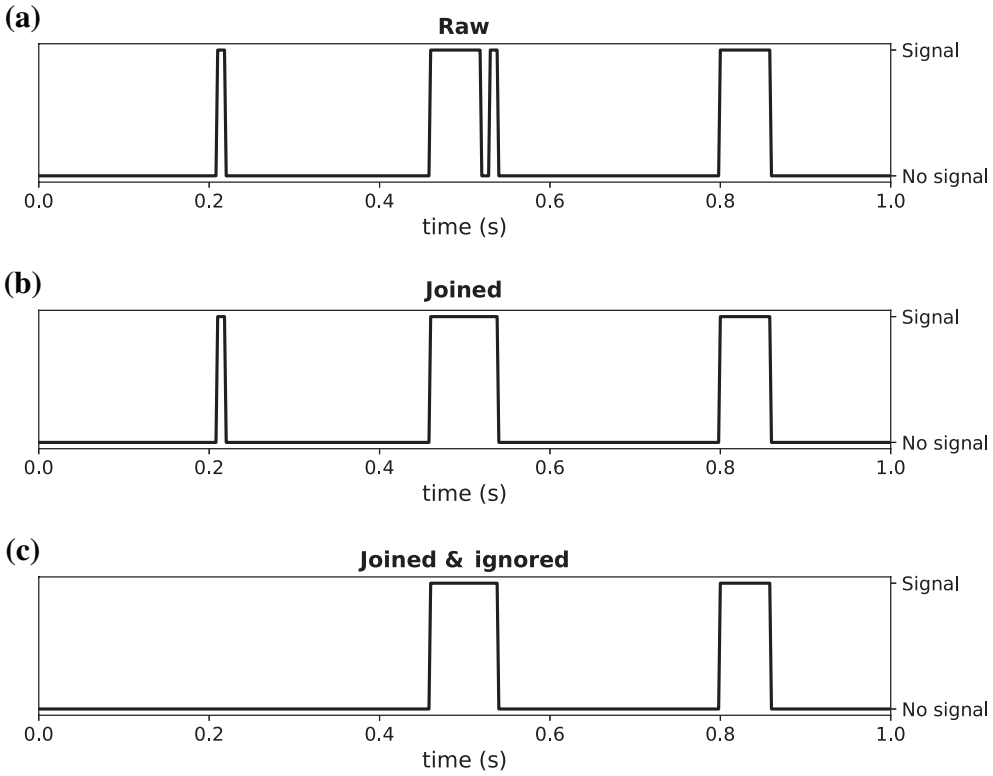

Figure 8: Post-processing of the output detection (a). First, nearby detected events are grouped together (b), followed by a removal of short-duration events (c).

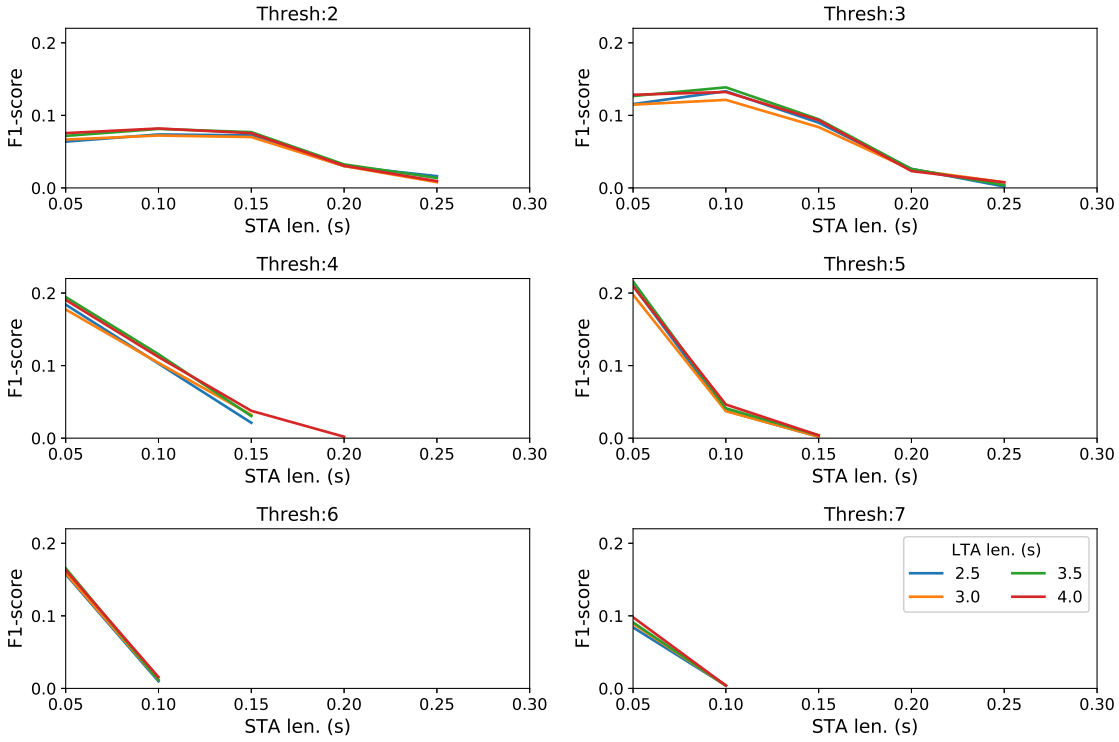

Figure 9: Grid search performed over synthetic traces to determine the optimum STA/LTA trigger parameters for this setting using the F1-score as the performance metric. The LTA window length is illustrated by the coloured lines while the STA window length varies along the x-axis. Each plot represents a different STA/LTA ratio threshold as annotated in the plots' titles.

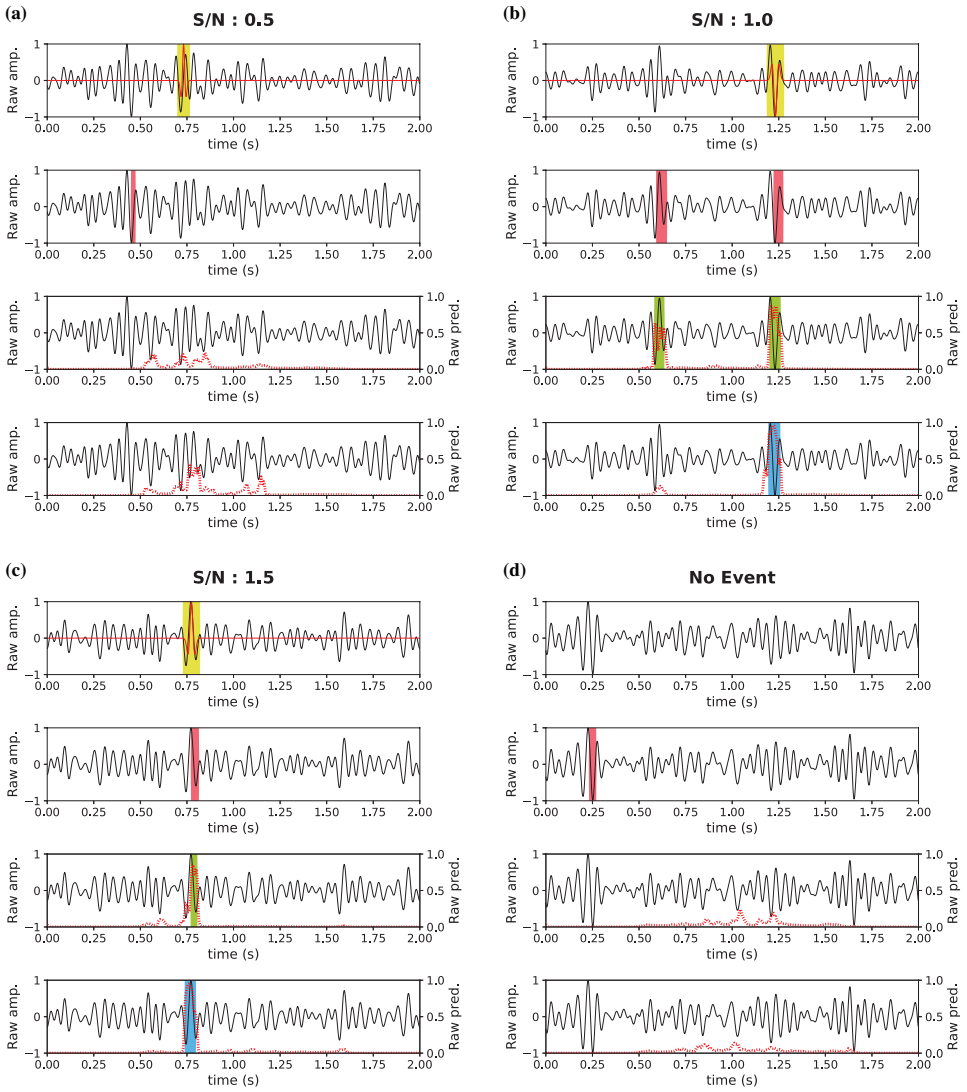

Figure 10: Event detection on synthetic seismic traces at S/N of (a) 0.5, (b) 1.0, (c) 1.5, and where no event is present in the trace (d). For each panel, the top row is the true label as defined when the synthetic trace was created with the wavelet overlaid (in red). The following rows from top to bottom respectively represent the STA/LTA detection, LSTM detection and the BLSTM detection. Shaded areas indicate where an event was detected whilst the red, dashed lines indicate the raw output of the LSTM and BLSTM predictions.

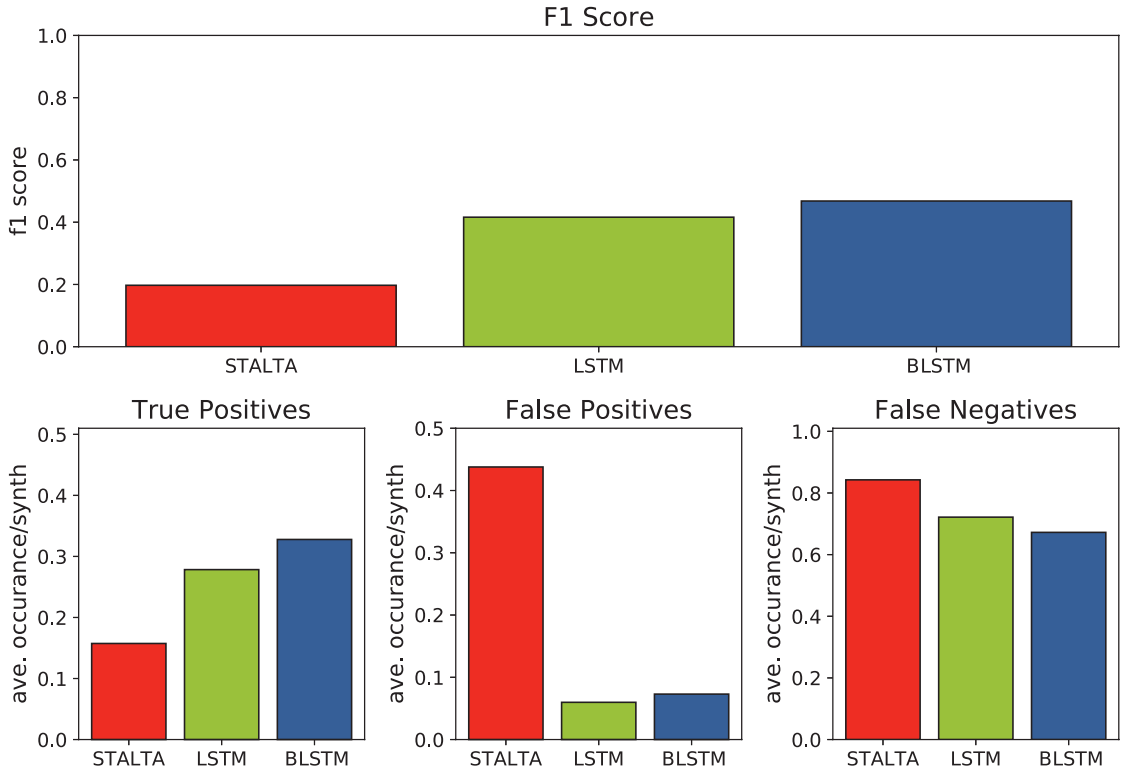

Figure 11: Evaluation of detection procedures tested on 5000 synthetic seismic traces, each containing a single seismic arrival ranging in S/N from 0.5 to 2.0. The y-axis highlights the average number of specific event type (e.g., True Positive) detected per synthetic, where each synthetic should have one true positive and zero false positives or negatives. False positives are synonymous with false triggers whilst false negatives illustrate the number of missed events.

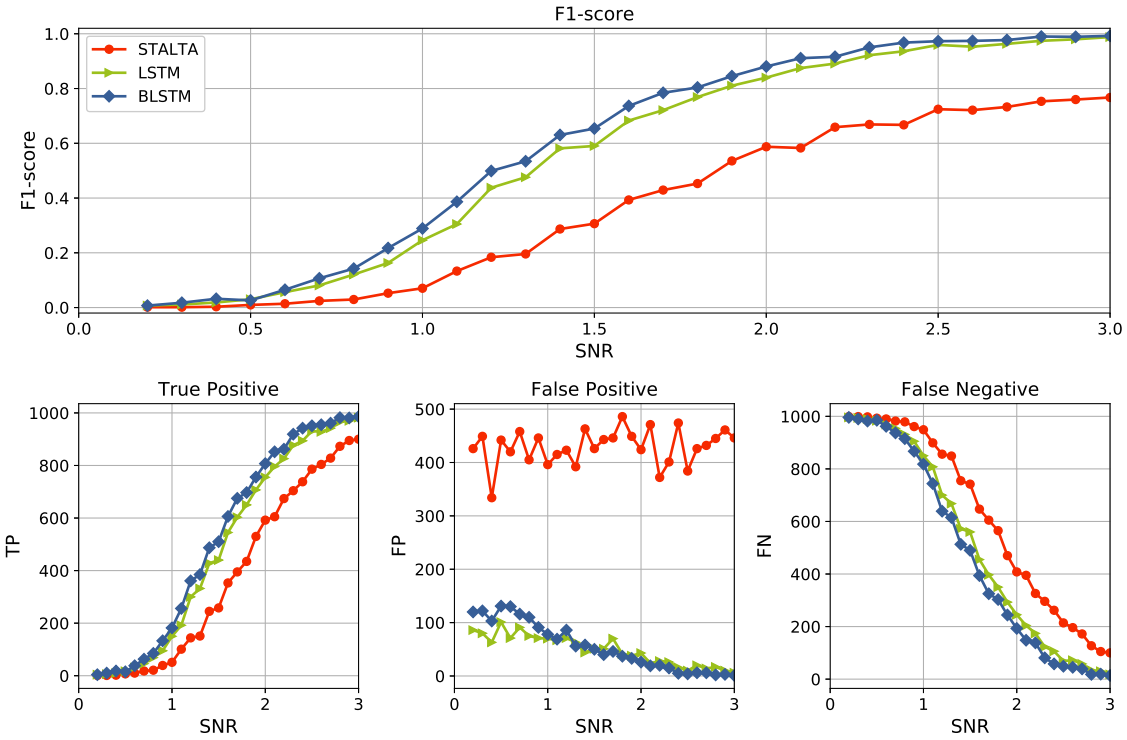

Figure 12: Analysis of the effect of noise level on the different detection procedures. The top plot illustrates the effect on the F1-score of the detections whilst the bottom three plots illustrate the number of true positives (TP), false positives (FP) and false negatives (FN), respectively. False positives are synonymous with false triggers whilst false negatives illustrate the number of missed events.

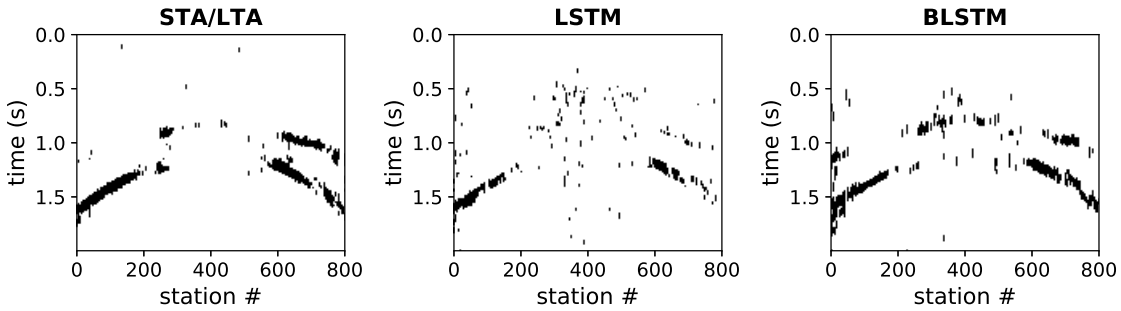

Figure 13: STA/LTA, LSTM and BLSTM detection of a source embedded in the Marmousi subsurface model, where the synthetic data has been generated via an elastic finite difference procedure with the focal mechanism $M_{11} = -1, M_{33} = 1$.

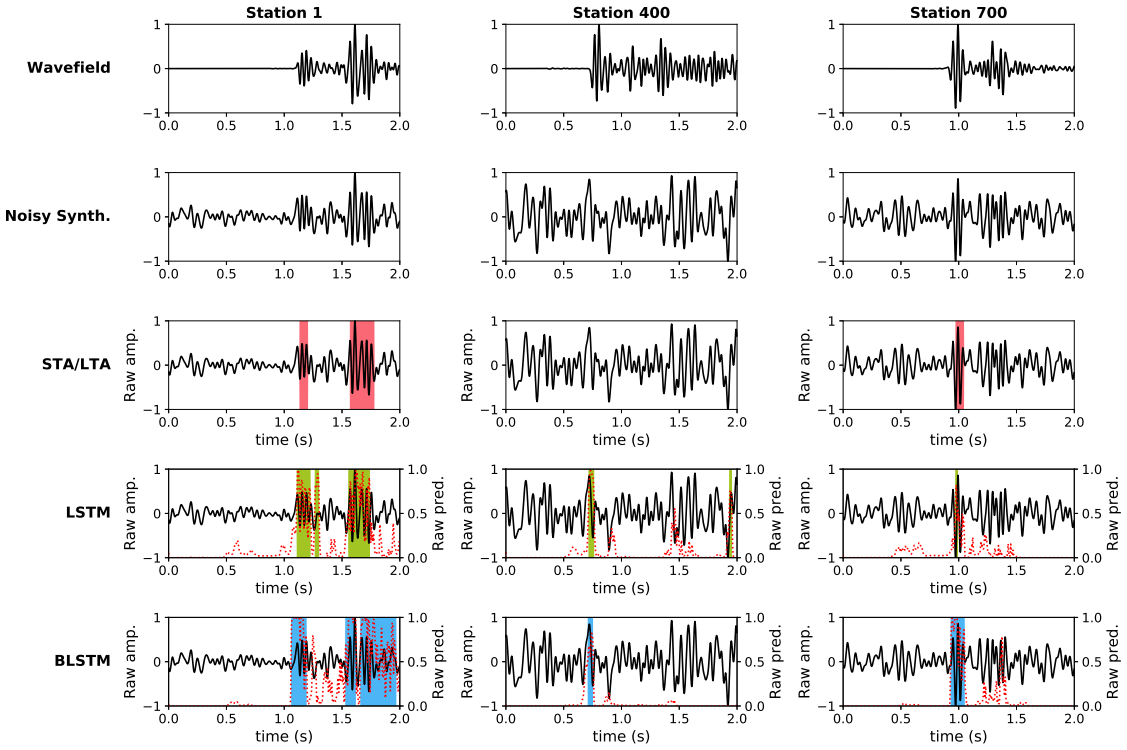

Figure 14: Event detection on three traces for a source embedded in the Marmousi subsurface model, where the synthetic data has been generated via an elastic finite difference procedure with the focal mechanism $M_{11} = -1, M_{33} = 1$. Shaded areas on the bottom three plots indicate where an event was detected. The red, dashed lines on the NN predictions (bottom two rows) illustrate the raw prediction values.

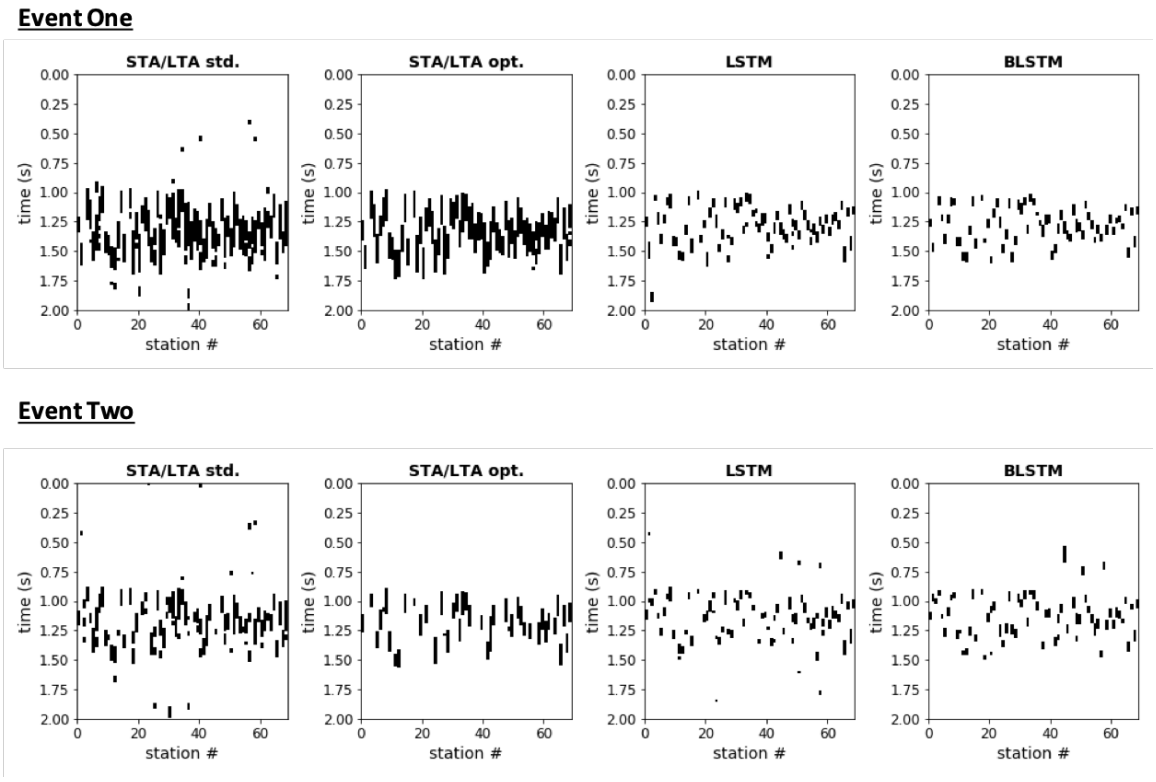

Figure 15: Event detections for two events observed on the ToC2ME array on November 1st 2016.

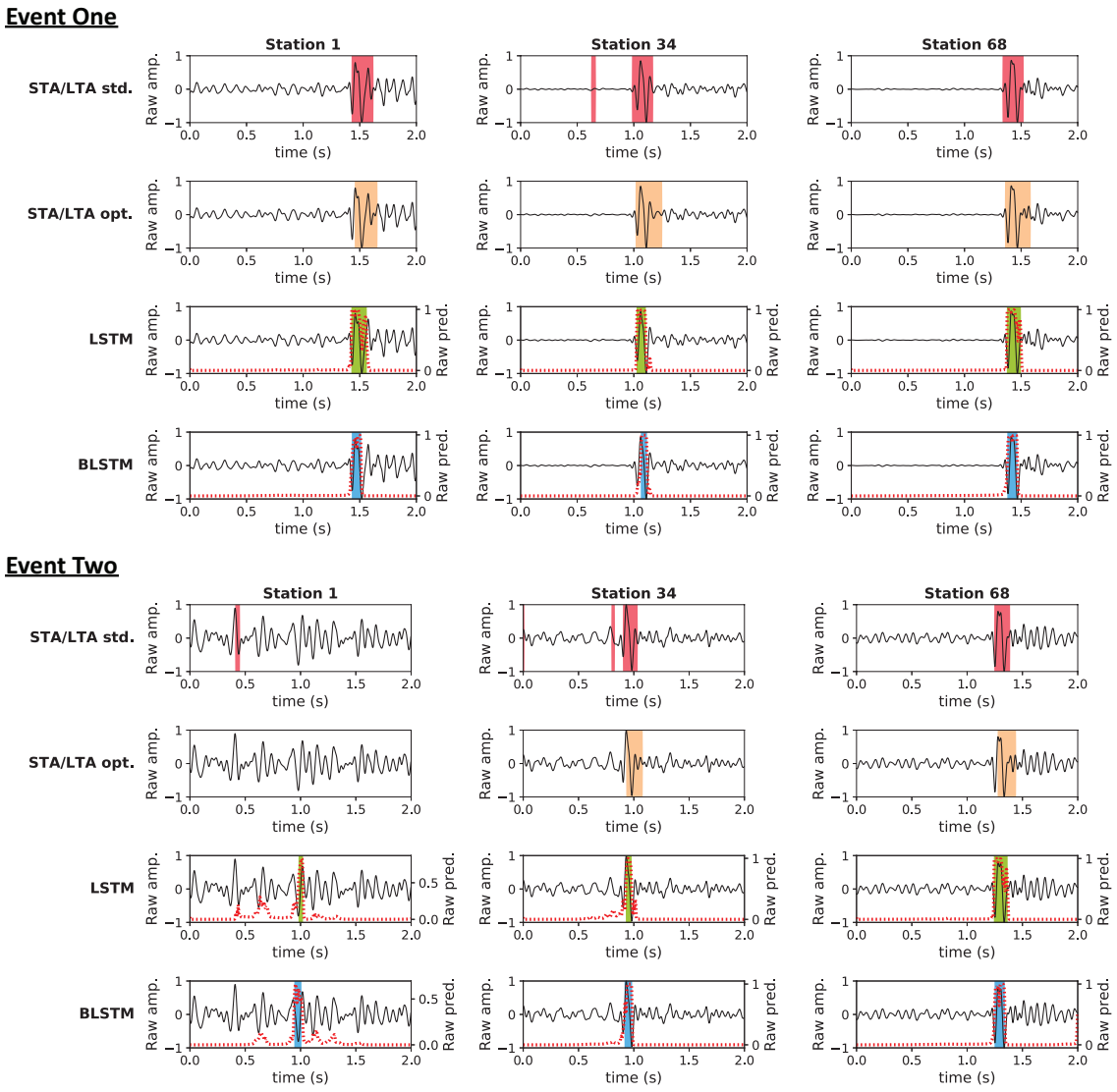

Figure 16: Event detection on single traces for two events observed on the ToC2ME array on November 1st 2016. Top row is the STA/LTA trigger with parameters determined via a parameter sweep on synthetic data, the second row is the optimum parameters for this recording, the third row is the LSTM detection and, the final row is the BLSTM detection. The red, dashed lines on the NN predictions (bottom two rows) illustrate the raw prediction values.

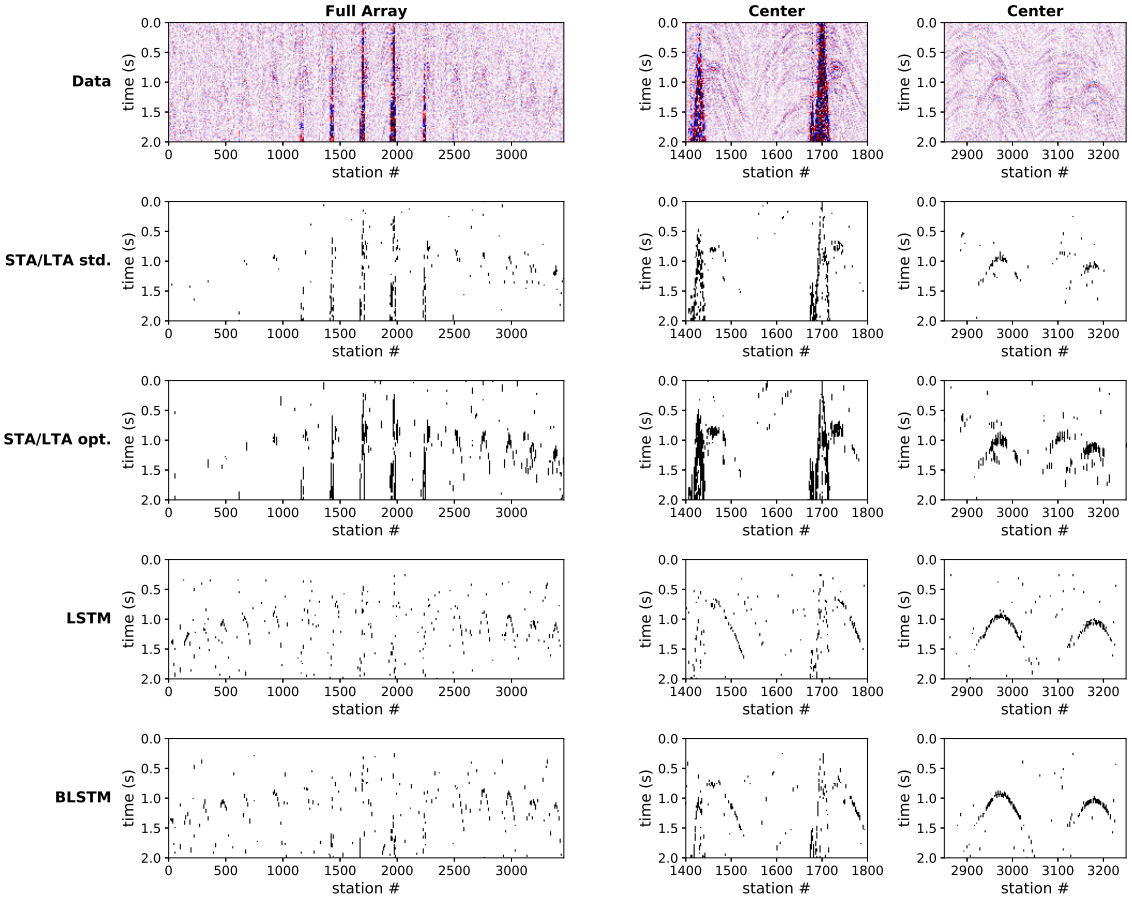

Figure 17: Field data (top) and detection results for a seismic event recorded on an offshore permanent sensor array. The center of the array, middle column, is contaminated by platform noise, while the edge of the array has a lower noise level but is further from the event source.

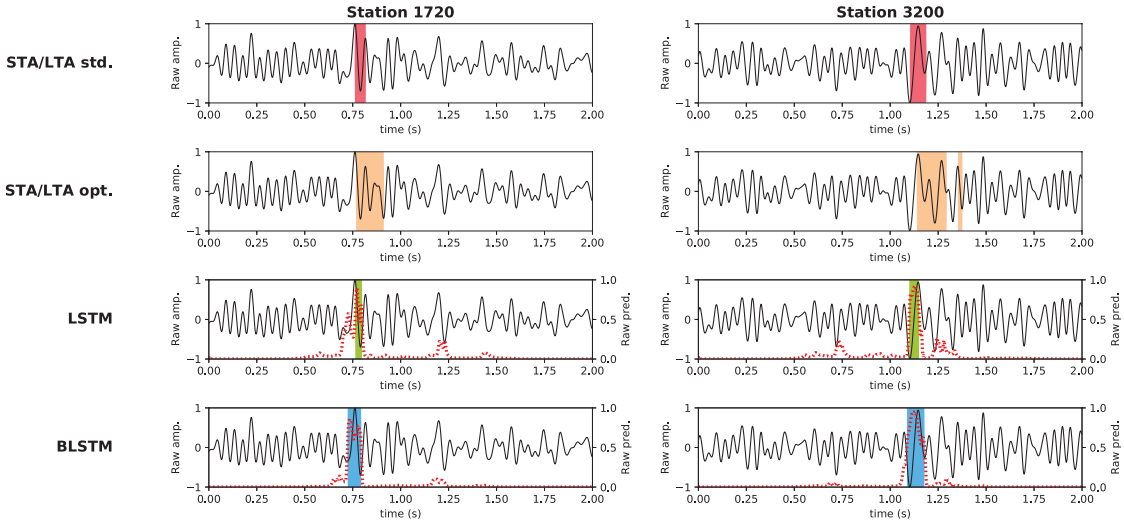

Figure 18: Detection on individual traces of G8 event from a station at the center of the array (left column) and from a station at the edge of the array (right column). Areas highlighted indicate the times when an event was detected by the STA/LTA autotrigger (top), the LSTM network (middle), and the BLSTM network (bottom). The red, dashed lines on the NN predictions (bottom two rows) illustrate the raw prediction values.

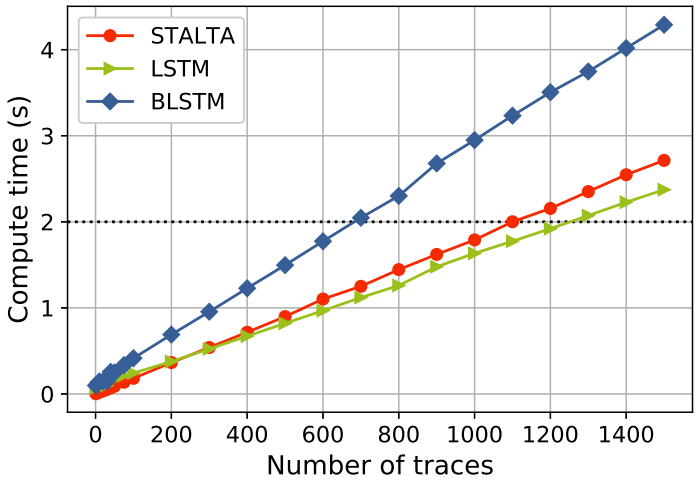

Figure 19: Compute time versus number of traces detection is performed on.