doi:10.26443/seismica.v4i2.1579

Picking Induced Seismicity with Deep Learning (piSDL)

Janis Heuel * [1], **Vincent Maurer** [2], **Michael Frietsch** [1], **Andreas Rietbrock** [1]

[1]Geophysical Institute, Karlsruhe Institute of Technology, Karlsruhe, Germany, [2]ÉS-Géothermie, Strasbourg, France

Author contributions: *Conceptualization*: J. Heuel, A. Rietbrock. *Software*: J. Heuel. *Validation*: V. Maurer. *Writing - Original draft*: J. Heuel. *Writing - Review & Editing*: J. Heuel, V. Maurer. *Visualization*: J. Heuel. *Project administration*: M. Frietsch. *Funding acquisition*: A. Rietbrock, M. Frietsch.

Abstract Training deep-learning picking models with several published data sets can be easily done through the Python toolbox SeisBench. Most of the data sets contain earthquakes recorded at local, regional and teleseismic distances, with only limited data in the low magnitude, close distance region. Applying current published PhaseNet models to induced seismicity data leads to only a few events being detected and trained PhaseNet models are not able to outperform well-established workflows in seismology. Here we present a new seismological data set and trained PhaseNet models for picking induced seismicity with deep-learning (piSDL). PhaseNet was trained with 171,182 three component waveforms from 40,576 events. Noise samples were added in the training data set to reduce the number of false picks. In this study, we noticed that a good earthquake training data set and noise samples from the analysed area are both important to detect more seismic events with a newly trained PhaseNet model. We validated our new PhaseNet models at a geothermal site in Rittershoffen (France). The models trained with the new data set and noise samples clearly outperform PhaseNet's original published model and traditional methods in seismology by detecting up to 62% more events compared to a seismicity catalogue published by an agency.

Production Editor: Andrea Llenos
Handling Editor: Hongyu Sun
Copy & Layout Editor: Hannah F. Mark

Signed reviewer(s): Cindy Lim Shin Yee

Received: January 30, 2025
Accepted: August 1, 2025
Published: August 19, 2025

1 Introduction

Detecting earthquakes and determining seismic phase arrivals are among the most important processing steps in seismology. For example, precise onset times of different seismic phases are essential for accurate source locations and travel time tomography. Since manual seismic phase picking is time-intensive, and the number of seismological stations worldwide continues to grow almost exponentially, automated earthquake detection and phase picking algorithms have been developed over the last decades (e.g. Allen, 1982; Baer and Kradolfer, 1987; Diehl et al., 2009; Goforth and Herrin, 1981; Küperkoch et al., 2010; Leonard and Kennett, 1999; Sleeman and van Eck, 1999). In addition, due to the increase in induced seismicity caused by the development of geothermal plants, automatic algorithms are also needed to accurately pick the phase arrivals.

In recent years, deep-learning pickers, such as Generalized phase detection (GPD, Ross et al., 2018), PhaseNet (Zhu and Beroza, 2018), EQTransformer (Mousavi et al., 2020) or PhaseNet-TF (Xi et al., 2024) have been developed. Instead of calculating explicit features from three component seismic waveforms, deep-learning algorithms learn from large training data sets to implicitly determine these features from millions of different seismic waveforms and labeled phase onsets. As shown, for example, by Mousavi et al. (2020) or Münchmeyer et al. (2022), deep-learning pickers outperform traditional automated phase picking algorithms. They can detect both P and S arrivals in one processing step, and approach the accuracy of manual phase picking. In recent years, many data sets for different regions have been published for training such deep-learning pickers (e.g. STEAD (Mousavi et al., 2019), INSTANCE (Michelini et al., 2021), TXED (Chen et al., 2024), CREW (Suarez and Beroza, 2024)). Most of the published data sets contain earthquakes recorded at local and regional distances. Among these published data sets, several picking models were trained to test which of the above pickers performs best (Münchmeyer et al., 2022) and to apply these pickers to different regions, such as China (Zhu et al., 2023) or to ocean bottom seismometers (Bornstein et al., 2024; Niksejel and Zhang, 2024). However, Chai et al. (2020) and Jiang et al. (2021) noticed that both the original published models of PhaseNet and EQTransformer do not perform well when they are applied directly to different geographical regions. Furthermore, the models might fail when applied to data sets with small distances between source and receiver (i.e. 10-20 km) and low signal-to-noise ratio (SNR) events (Dai et al., 2023), as they have not been trained on this type of data. These type of events are, for example, typical for induced seismicity.

This study presents a new data set and trained PhaseNet models for picking induced seismicity with deep-learning (piSDL). In the following, the new data set is presented to train PhaseNet using earthquake waveforms with low SNR, small distances between source and seismometer, and multiple events within a time window. Most of the recorded waveforms in the data set are from induced seismic events. We trained PhaseNet from scratch with the presented new data set and

*Corresponding author: janis.heuel@kit.edu

also applied transfer learning by fine tuning previously trained and published models. After presenting details of the used data, we show how we trained PhaseNet with our data set. To artificially increase the size of the data set, we developed two new augmentation methods, which we included into the Python framework SeisBench (Woollam et al., 2022) (see Data and Code Availability). Furthermore, we show how the number of noise waveforms affects the picking performance of trained PhaseNet models and how the number of false picks can be mitigated by this approach. In the last part of this study, different PhaseNet models are applied to continuous data recorded at a geothermal field in Rittershoffen, France (Baujard et al., 2017; Maurer et al., 2020), to evaluate the performance of the new PhaseNet models.

2 Data Set

To train PhaseNet, we gathered three-component seismic waveforms from several earthquake catalogues, including both induced seismicity events and natural low-magnitude events ($M_L \leq 2$). A total of 321,946 waveform samples were collected from various agencies and locations. However, after initial training of the PhaseNet model and testing it with our dataset (see Methods), the model's performance was unsatisfactory. Following several iterations of training and testing, we removed all waveform samples with only a single phase arrival in a time window (Suarez and Beroza, 2024) (i.e. each seismic event in the presented data set is labeled by a P- and S-arrival). Additionally, all samples that are likely mislabelled by analysts, such as extremely weak earthquake signals or earthquakes that are only visible when taking information from neighbouring stations, were excluded (Fig. 1), since our temporary trained PhaseNet model was not able to detect any phase arrival. After these preprocessing steps, the final data set consists of 171,182 waveforms from 40,576 different seismic events that have been recorded at 455 seismological stations. Figure 2 summarises the workflow for how we derived the final data set, including initial training of PhaseNet models. In addition to the recorded earthquake waveforms, we added noise waveform samples to the data set. Details of how we selected the noise samples are described in Methods. The data were recorded with sampling rates between 100 and 300 Hz. One part of the data set includes three component waveforms with multiple events within a time window (Fig. 3). The earthquake magnitudes range from -1 to 4.5, but no distinction was made between different magnitude types across the datasets from various regions and agencies. Figure 4 illustrates the distribution of earthquake magnitudes (a), SNR (b), epicentral distance (c), and the relationship between SNR and epicentral distance (d) in our dataset compared to the Stanford Earthquake Dataset (STEAD, Mousavi et al., 2019). Following Chen et al. (2024), the SNR was estimated using the formula:

$$\mathrm{SNR} = 20 \cdot \log_{10} \frac{S}{N}\,, \tag{1}$$

where S and N are the root-mean-square of the demeaned signal and noise window, respectively. The signal is determined from 1 s after the S-arrival and noise from a 1 s window before the P-arrival (Chen et al., 2024). For the visualisation in Fig. 4, the largest value of the SNR from each of the three components (Z, N, E) was chosen. Negative SNR values result from phase onsets which are, for example, in the coda of a previous event. Compared to STEAD, our dataset for induced seismicity contains significantly more events with low SNR and small source-station distances (Fig. 4b). Our dataset includes induced seismicity events caused by hydraulic fracturing-based fluid injection operations in the Dawson-Septimus area, Canada (Roth et al., 2020); coal mine flooding in the Ruhr area, Germany (Floodrisk, Rische et al., 2022); and geothermal operations in Insheim (Germany, Küperkoch et al., 2010), Rittershoffen (France, Maurer et al., 2020), St. Gallen (Switzerland, Diehl et al., 2017), Soultz-sous-Forêts (France, e.g. Genter et al., 2010; Schill et al., 2017), and Vendenheim (France, Schmittbuhl et al., 2021). The Rittershoffen dataset includes waveforms with multiple phase onsets within a single trace (Fig. 3). Additionally, natural low-frequency earthquakes from the Eifel region (Germany, e.g., Hensch et al., 2019; Koushesh and Ritter, 2024; Ritter et al., 2024) were included, as this catalogue provides seismic picks with low SNR events. All available events between 2009 and 2023 with $M_L \leq 2$ from the Swiss Seismological Service (Swiss Seismological Service (SED) At ETH Zurich, 1983) were also included, as this catalogue features high-quality, manually picked phase onsets. Table 1 summarises the different data sets and Figure S1 shows maps with source locations and station distributions.

3 Methods

3.1 Model training

To train a deep-learning phase-picking model for induced seismicity, we used the implementation of PhaseNet (Zhu and Beroza, 2018) in SeisBench (Woollam et al., 2022). As shown by Münchmeyer et al. (2022), PhaseNet is one of the best-performing deep-learning phase-picking approaches and has 30% fewer training parameters (i.e. 268,499) than EQTransformer (Mousavi et al., 2020). PhaseNet is a modified U-Net (Ronneberger et al., 2015) consisting of a down-sampling branch using 1D convolutions and an up-sampling branch employing 1D deconvolutions. PhaseNet is a sequence to sequence model. The input is an unfiltered, min-max normalised 30 s three-component seismogram with a sampling rate of 100 Hz (i.e. data with different sampling rates are resampled in SeisBench), so the input layer has a dimension of 3×3001 samples. The output layer has the same shape as the input layer and represents three probability distributions for P wave, S wave and noise (i.e. neither P nor S). Since the manually picked phase arrivals are included in the data set, the probability distributions are automatically determined during the training process and act as the ground truth for the output layer. The output probability distributions for P and S

Data set	Mag. of completeness	Agency	Picking method	Num. of events	Num. of waveforms
Dawson-Septimus area	1.1	RUB	automatic & manual	7777	55415
Floodrisk	-0.1	RUB	automatic & manual	2303	8926
Insheim	0.3	KIT and BGR	automatic & manual	805	2012
Rittershoffen	0.4	ES Géothermie	automatic & manual	7201	24152
St. Gallen	0.3	SED	manual	350	2219
Soultz-sous-Forêts	0.1	Unknown	manual	6035	7851
Vendenheim	1.1	BCSF	manual	190	985
Eifel	0.0	KIT and LED	manual	1046	5020
SED	0.9	SED	manual	14869	64602
Induced data set	0.8			40576	171182

Table 1 Further information about all gathered data sets. Automatic picks have been double checked by analysts. Figure S1 shows maps with earthquake locations and station distributions for each data set. Abbreviations for agencies: RUB: Ruhr-University Bochum; KIT: Karlsruhe Institute of Technology; BGR: Federal Institute for Geosciences and Natural Resources; SED: Swiss Seismological Service; LED: Landeserdbebendienst Baden-Württemberg, BCSF: Le Bureau Central Sismologique Français.

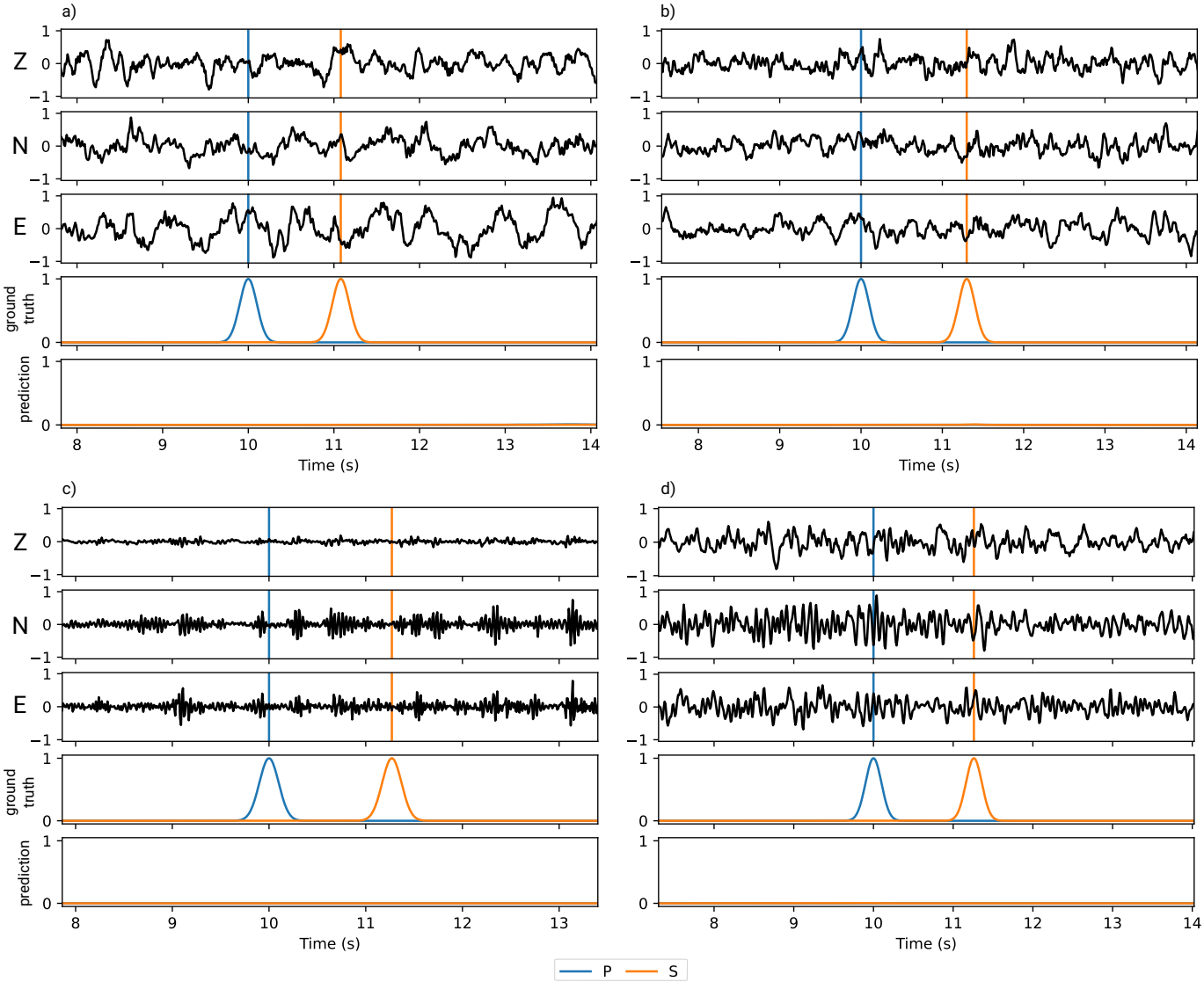

Figure 1 Examples of removed waveform samples from the data set, comparing the ground truth probabilistic function for P- and S-arrivals with the predicted values. From top to bottom in each subfigure: normalised, unfiltered vertical (Z) and horizontal (N, E) components with manually picked P- and S-arrivals, ground truth probabilistic functions for P and S, and predicted probabilistic functions by our initially trained PhaseNet model. These events were removed as the trained PhaseNet model failed to detect P- or S-arrivals.

arrivals is a Gaussian distribution with a standard deviation of 0.1 s (i.e. 10 samples at 100 Hz sampling frequency) and with a maximum of one at the labeled arrival time (Zhu and Beroza, 2018). During the training of

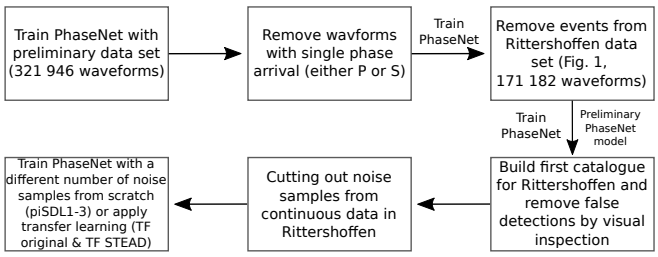

Figure 2 Workflow how the final induced seismicity dataset was built through various pre-processing steps and testing of fully trained PhaseNet models. After optimising the data set, a preliminary PhaseNet model was trained and false event detections were removed by visual inspection. Using phase arrivals from this preliminary catalogue was needed to cut out noise samples from continuous data in Rittershoffen.

the neural network, the cross-entropy loss between the ground truth and the predicted probability distributions is calculated. The data were augmented by randomly shifting events within the trace and rotating the horizontal components of the three-component input seismogram by an arbitrary angle. Two training strategies were tested:

1. **Training from scratch:** Models were trained from randomly initialised weights. The training dataset was artificially increased by duplicating metadata entries and applying data augmentation, exposing the model to a greater variety of data.

2. **Transfer Learning:** Training continued from pre-trained models (e.g., Chai et al., 2020; Bornstein et al., 2024; Zhu et al., 2024). Pre-trained weights from the original PhaseNet model (Zhu and Beroza, 2018) and a model trained on STEAD (Mousavi et al., 2019; Münchmeyer et al., 2022) were used. Transfer learning requires smaller problem-specific datasets since the pre-trained model has already learned general features from a large-scale dataset (Pouyanfar et al., 2018). For these cases, no artificial data augmentation was applied. We chose these pre-trained models because the original PhaseNet model is widely used to create new earthquake catalogues (e.g. Chai et al., 2020; Jiang et al., 2021; Becker et al., 2024; Castillo et al., 2024), was also successfully applied to cases of induced seismicity (Wong et al., 2021; Lim et al., 2024), and STEAD contains a high-quality global earthquake dataset from regional seismic events. Note that the input data of the original published PhaseNet model is standardised by its mean and standard deviation instead of using min-max normalisation as described above.

The models were trained with a batch size of 256, a learning rate of 0.001, and up to 250 epochs on an NVIDIA A100-40 GPU with 40 GB of memory. The Adam optimiser (Kingma and Ba, 2014) was used for optimisation. The training data set was split into 70% for training, 20% for validation, and 10% for testing (each single data set in Tab. 1 was split by the given ratio). During the training, the validation loss was monitored and if the validation loss did not change over 15 consecutive epochs, training was stopped (early stopping). To ensure that the best model was saved, we only saved the model if the validation loss continued to decrease. After the training was completed, the model was evaluated with the 10% test data set.

3.2 Noise samples

As described by Jiang et al. (2021), PhaseNet models trained without noise samples can produce numerous false picks. To address the false picks, we added noise samples from Rittershoffen and STEAD to our dataset and trained multiple models with varying numbers of noise samples (Fig. 5). To obtain noise samples from continuous data recorded at Rittershoffen, we first built a seismicity catalogue for January 2024 using one catalogue generated with SeisComp (Helmholtz-Centre Potsdam - GFZ German Research Centre for Geosciences and gempa GmbH, 2008) and another derived from a preliminary PhaseNet model for induced seismicity (details about building a catalogue with PhaseNet are given in Building an induced seismicity catalogue). The preliminary PhaseNet model was only trained with our data set for induced seismicity (Fig. 2), and therefore the derived catalogue contains many false detections. After manually inspecting the catalogue and removing false detections, 30 s three-component noise waveforms were randomly extracted from all stations at the Rittershoffen site (Fig. 2). After gathering noise waveform samples, we then trained 13 models with the induced data set and different numbers of noise samples ranging from 0 to 60,000. Half of the noise samples were from STEAD and the other half from Rittershoffen. To overcome small perturbations of a single model, we trained in total 130 models with different noise samples and each model was tested on six hours of continuous data from Rittershoffen, where all events were well-documented.

3.3 Model evaluation

Each model was evaluated on the 10% test dataset. Precision, recall, and the harmonic mean (F1-score) were calculated using

$$\text{Precision: } P = \frac{TP}{TP + FP}\,, \tag{2}$$

$$\text{Recall: } R = \frac{TP}{TP + FN}\,, \tag{3}$$

$$\text{F1-score: } F1 = 2 \cdot \frac{P \cdot R}{P + R}\,, \tag{4}$$

where TP represents the true positives, FP are the false positives, and FN denotes false negatives. A pick is counted as a true positive if the probability distribution exceeds a certain decision threshold (pick threshold in Fig. 7) within ± 0.25 s of the true phase arrival. The selection uncertainty of ± 0.25 s was chosen in order to have a strict evaluation criterion for the test data set. A false positive is a predicted phase arrival that

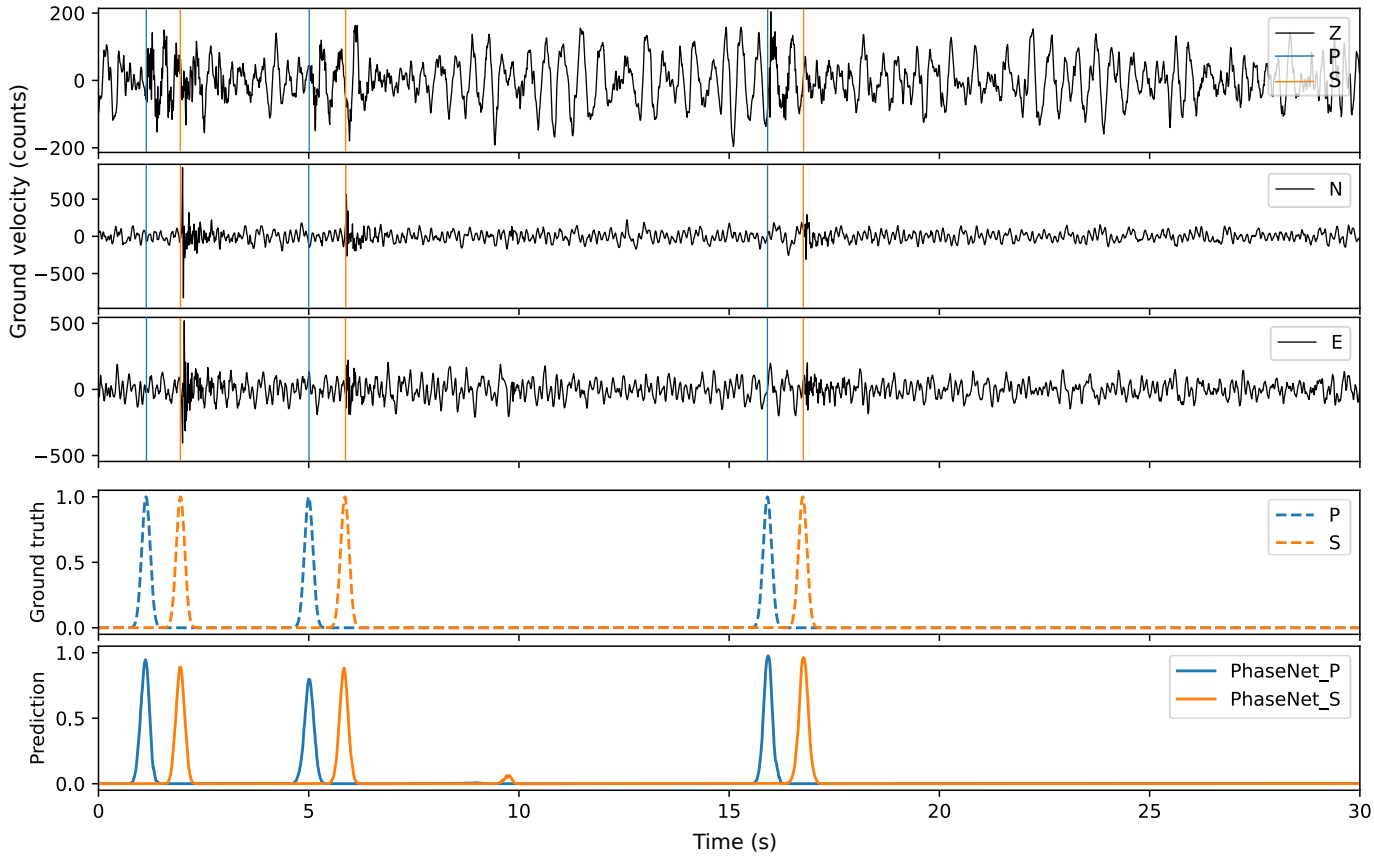

Figure 3 Example of an input waveform sample containing multiple events within a time window, recorded at a station near Rittershoffen (France). From top to bottom: unfiltered vertical component (Z), first horizontal (N), and second horizontal (E) component with manually picked P- and S-arrivals, ground truth probabilistic functions for P and S, and the predicted probabilistic functions by the trained PhaseNet model.

does not match any true arrival within a larger window (typically ± 2 s) around the true phase arrival, i.e. the predicted pick uncertainty is > 0.25 s. False negatives are the number of true picks without a peak above the pick threshold within the large window of the predicted output (e.g. Myklebust and Köhler, 2024). Models were tested across various pick thresholds to identify optimal performance on the test dataset and to compare model performance.

3.4 Building an induced seismicity catalogue

Once we found models that perform well on our test data set, we applied these models to one month of continuous data from Rittershoffen. These models are trained with our new data set for induced seismicity and three different noise data sets (Tab. 2). Obtaining the picks is done by the PhaseNet implementation in SeisBench (Woollam et al., 2022) by dividing the continuous data stream into overlapping 30 s windows. For each window the output probabilities of P, S and noise are predicted by the PhaseNet model and they are merged to produce a continuous probability stream for the entire day by selecting the maximum value within overlapping sections. We used a constant picking threshold for P and S of 0.2, as well as the default overlapping of 1500 samples between neighbouring windows. As we trained three different models from scratch for induced seismicity with different numbers of noise samples (Tab. 2), we were interested whether a semblance-based ensembler (Yuan et al., 2023) is able to reduce the number of false picks, when applying our trained models. As described by Yuan et al. (2023), the output probability distribution is given by

$$C(t_i) = W(t_i) C^0(t_i)^\nu \,, \tag{5}$$

where $W(t_i)$ is a weighting function, defined as

$$W(t_i) = \max\left(\mathrm{Pb}_1(t_i), \mathrm{Pb}_2(t_i), \ldots, \mathrm{Pb}_m(t_i)\right) \tag{6}$$

and $C^0(t_i)$ denotes the original semblance for either P or S, which is defined as

$$C^0(t_i) = \frac{\sum_{t_i-\delta_t/2}^{t_i+\delta_t/2} \left(\sum_{m=1}^{M} \mathrm{Pb}_m\right)^2}{M \sum_{t_i-\delta_t/2}^{t_i+\delta_t/2} \left(\sum_{m=1}^{M} (\mathrm{Pb}_m)^2\right)} \,. \tag{7}$$

The predicted probability distributions of each model m are denoted by Pb_m, where t_i represents the i$^{\text{th}}$ time sample, δ_t is the time window length in seconds (here 0.5 s), M is the total number of models used for predictions, and the exponent ν in Equation 5 balances noise suppression and signal coherence, typically set to 2. After picking the seismic phases, a seismic phase associator (PyOcto, Münchmeyer, 2024) is used to identify all seismic events. The associated seismic events were relocated using NLLoc (Lomax et al., 2001, 2009). The velocity model which is necessary for the association and relocation and parameters for PyOcto are given in Tab. 3 and Tab. 4, respectively.

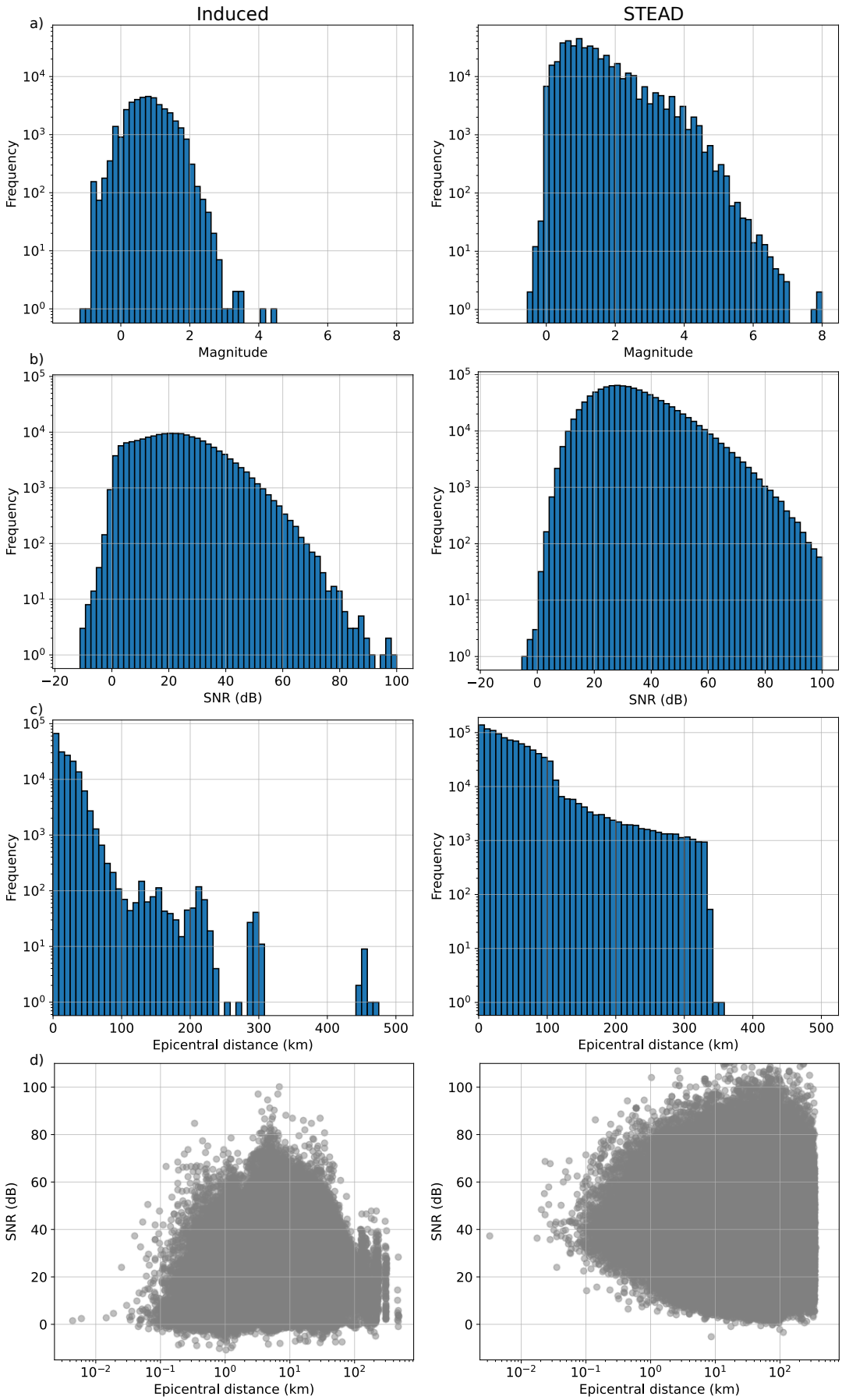

Figure 4 Magnitude frequency distribution of the induced data set and STEAD (a), SNR histogram for both data sets (b), epicentral distance distribution (c), and scatter plot of SNR versus epicentral distance (d). Note that the induced data set includes more events with low SNR and shorter source-station distances.

Model name	Rittershoffen	STEAD
PN piSDL1	15 000	10 000
PN piSDL2	20 000	5 000
PN piSDL3	20 000	15 000

Table 2 Number of noise samples for different models trained with the induced data set. The noise samples are either from Rittershoffen or from STEAD.

depth (km)	v_P (km/s)	v_S (km/s)
0.0	1.85	0.86
0.8	2.87	1.34
1.6	5.80	3.31
2.6	5.82	3.32
3.6	5.85	3.34
4.6	5.87	3.35
5.6	5.90	3.37
6.6	5.92	3.38
7.6	5.95	3.40

Table 3 Velocity model (Cuenot et al., 2008), which is used for the association with PyOcto (Münchmeyer, 2024) and the relocation of seismic events with NLLoc (Lomax et al., 2001, 2009). Note the velocity model was derived for the region of Soultz-sous-Forêts, which is close to the analysed site in Rittershoffen.

4 Results

4.1 Noise sample tests

After training an initial PhaseNet model with all available induced data, we observed poor performance of the model, when applied to continuous seismic data from Rittershoffen (Fig. 6b). In addition, the original PhaseNet model, published by Zhu and Beroza (2018) (in the following PN original), performed much better than our model (Fig. 6a) but missed some events (see Application to continuous data). Figure 6 shows the detected picks of PN original (a) and our trained model for induced seismicity without (b) and with (c) noise samples, respectively, by analysing half an hour of continuous data from Rittershoffen. PN original detected 118 P and 97 S arrivals for all twelve stations analysed, while the induced seismicity model trained without noise samples detected 524 P and 478 S arrivals. Many of these picked phase onsets are false picks, since the time window contains two events with a total of 14 P and 14 S arrivals. After training PhaseNet with our induced seismicity data set and noise samples, the model detected only 33 P and 36 S arrivals. Both piSDL models (Fig. 6a & b) picked all of the ground truth phase arrivals; however, the original PhaseNet model only picked seven P but all 14 S arrivals. The reduced number of picks when training with noise samples could imply that adding noise samples from the analysed stations in Rittershoffen enables PhaseNet to learn characteristics from noise as well as from earthquake waveforms, improving its ability to distinguish between both. However, when training PhaseNet with an earthquake data set and noise samples the number of picks is reduced the more noise samples are added to the data set.

To investigate if there is an optimum size for the noise data set, we gathered up to 60,000 noise samples from STEAD and Rittershoffen and trained 13 models with the induced data set and different numbers of noise samples, i.e. the first model was trained with zero noise samples, a second with 5000 and the 13th model with 60,000 noise samples. Since each model works slightly differently, we trained ten models for each single noise data set. This results in a total number of 130 models. Each model was then tested on six hours of continuous data from Rittershoffen. We manually picked 42 P-picks and 49 S-picks in these six hours, which are part of associated events. However, it is possible that low magnitude events are only visible at single stations that are not included in the seismic catalogue. Figure 5 summarises the noise tests of all trained models and shows the average number of P and S picks in relation to the number of noise samples and the percentage of detected picks that match known picks from the catalogue. However, the number of matching picks never reached 100%. Picks are missed either because of too noisy conditions or due to the windowing to predict the probability distributions for 30 s windows (i.e. overlapping of 1500 samples for neighbouring windows). This inconsistency has already been pointed out by Park et al. (2024).

However, by adding 5 - 15% of noise samples to the training data set, the number of false picks can be reduced. Although some picks may be missed at individual stations, seismic phase associators are still able to associate the corresponding events.

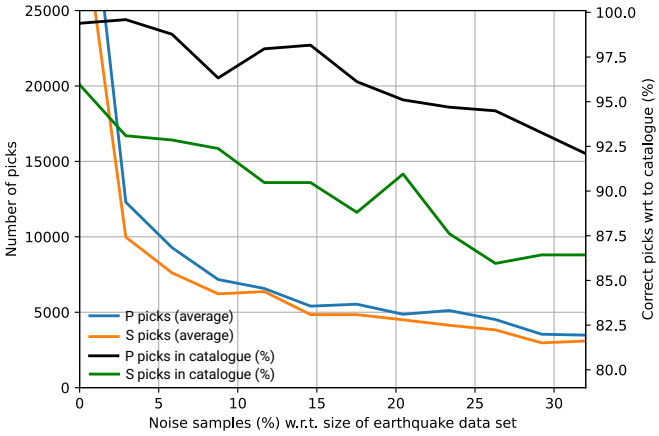

Figure 5 Finding the optimal number of noise samples to reduce the number of false picks when applying trained PhaseNet models to six hours of continuous data from Rittershoffen. The six hours contain 42 P- and 49 S-picks which are part of associated events. The number of noise samples are given in percent with respect to the size of waveforms in the training data set (i.e. 171,182 waveforms). In total 130 PhaseNet models were trained with different numbers of noise samples. The blue and orange curves are the average number of P- and S-picks over all trained models and the black and green curves represent the average of detected picks within the six hours of continuous data.

4.2 Model performance

To evaluate the model performance we tested each model on the 10% test data set and determined preci-

Parameter	Value	Description
zlim	[0, 30]	Depth limits in km
n_picks	10	Minimum required number of picks per event
n_p_picks	3	Minimum required number of P picks per events
n_s_picks	4	Minimum required number of S picks per events
n_p_and_s_picks	3	Minimum required number of stations that have both P and S picks
time_before	10	Overlap between consecutive time slices (should match travel time for seismic waves through the network)

Table 4 Settings for PyOcto (Münchmeyer, 2024) to associate events from all available picks. The default values are used for the parameters not mentioned.

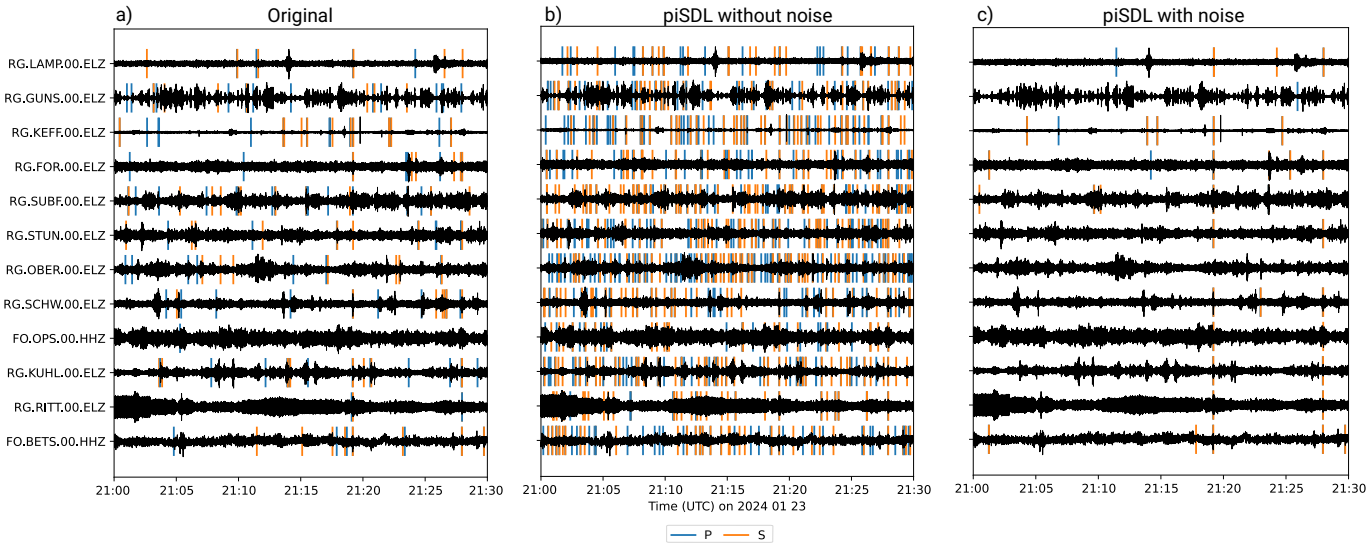

Figure 6 Comparison of picking performance on 30 min of continuous data from Rittershoffen. (a) The original published model by Zhu and Beroza (2018), (b) our model only trained with induced seismicity data and (c) our model trained with 20 000 noise samples from Rittershoffen and 15,000 from STEAD (Mousavi et al., 2019) and the induced seismicity data set. The original model detected 118 P and 97 S arrivals, the model without noise 524 P and 478 S arrivals and the model trained with noise samples detected 33 P and 36 S picks. Blue and orange vertical lines represent P and S picks, respectively. Two events are reported by the agency in the given time window. After phase association, the original PhaseNet model (a) detects a single event, piSDL without noise (b) detects twelve events and piSDL with noise (c) is able to detect only the two events.

sion, recall, and F1-score. We tested each model with 20 equally distributed pick thresholds from the range $[10^{-3}, 1]$. An optimal pick threshold can be derived from the precision-recall curve by finding a precision-recall value pair that is closest to the point $(1, 1)$, since a model with a precision and recall of 1 would represent a perfect model. However, there is a trade-off between precision and recall. A higher pick threshold reduces the number of false positives, i.e. an increase in precision, but also increases the number of false negatives, which leads to a lower recall. Figure 7 illustrates the precision-recall curve and the precision, recall, and F1 score as a function of the pick threshold for different PhaseNet models tested for P and S phases. The PN original (Fig. 7c, d) and the PhaseNet model trained on STEAD (Fig. 7g, h; PN STEAD) show similar performance when applied to the test data set. The optimum P and S thresholds determined are 0.04 and 0.02 for PN original and 0 for PN STEAD, i.e. PN STEAD does not work successfully on our test data set. Also the F1-score for the P phase in both models is below 0.8 for pick thresholds in the range 0.2 to 0.6. For the S phases, the F1 score is slightly higher, but overall both models perform poorly for the induced seismicity data. Therefore, retraining of both models with the induced seismicity data set or training a new model from scratch is necessary. Figure 7a and b demonstrates the performance of a model trained from scratch, which is very similar with the performance of the transfer-learned STEAD model (Fig. 7i, j; TF STEAD). Both models show F1-scores of above 0.95 for P and S up to pick thresholds of about 0.6. Also the optimal thresholds for P and S derived from the precision-recall curves are ≥ 0.2. However, PN original still performs poorly even after applying transfer learning using the induced data set (Fig. 7e, f).

Figure 8 shows some successful predictions from the test data set using PhaseNet trained with the induced data set (in the following PN piSDL). The SNR is between -8.2 dB and 24.8 dB. The examples also demonstrate that the new model is able to pick multiple phases, i.e. two P and S arrivals, in a single time frame (Fig. 8e). The peaks of the predicted probabilities align with the manually labelled phases; however, the maximum high of one for the Gaussian distribution is not reached for most examples. Figure 9 shows some failed cases. The noise is different from the examples in Fig. 8 and most likely the PhaseNet model has learnt from the noise data set to predict these types of phase arrivals as noise.

However, for very low pick thresholds (i.e. pick threshold ≤ 0.05), the phase arrivals in Fig. 9a, b, and d could be predicted. Only the events in Fig. 9c are too noisy. However, the second event in Fig. 9a is successfully predicted by the model.

4.3 Application to continuous data

As shown in Figure 6, the trained model can easily be applied to continuous recorded data using SeisBench (Woollam et al., 2022). To test our new and established PhaseNet models (models in Tab. 5), we analysed one month of continuous data from twelve stations close to Rittershoffen (France). Further, we associated all predicted seismic phases to derive a seismicity catalogue with PyOcto and relocated the associated events with NLLoc. Each catalogue was build using the same thresholds for P and S (0.2) and 50% overlapping, i.e. 1500 samples. The number of picked P and S arrivals and detected events for each model are shown in Table 5. In Figure 10 we compare the different derived catalogues against each other and Figure S5 shows the earthquake locations and stations. For example, applying PN original to the continuous data leads to 20 associated events in January 2024; however, the transfer-learned original PhaseNet model (TF original) detected 31 events. Both catalogues have 19 events in common, meaning that the catalogue derived from the original model found one event that was not detected by the transfer-learned model and the catalogue derived from the TF original contains twelve events that have not been detected using the PN original. To find common events between different catalogues, we first compared the origin times. Since we only have a very limited number of events, we then compared the predicted picks and waveforms for each event and station manually. If both the origin time and the picks at the stations match, the events from different catalogues were counted as common events. Most events have been detected using the transfer-learned STEAD model (TF STEAD) and PN piSDL1 (both 39 events). Further, we compare three different models (PN piSDL1-3) trained with our induced seismicity data set and different numbers of noise samples. Table 2 shows the different numbers of noise samples for each trained model with the induced data set. Both transfer-learned models (TF original and TF STEAD) have been trained with 20,000 noise samples from Rittershoffen and 25,000 randomly selected noise samples from STEAD. All catalogues are also compared with the automatically derived catalogue from SeisComp (24 events), one catalogue that contains all 32 induced seismic events at the site in Rittershoffen in January 2024 (ML + manually (induced) in Fig. 10), and one catalogue that contains all events (37) that have been labelled as induced as well as natural events (ML + manually (all) in Fig. 10). The last two catalogues have been derived using PhaseNet and a manual inspection afterwards of the detected events. To test whether the noise samples affect the number of associated events, we trained a model without event data from Rittershoffen (PN piSDL (no Ritt.) in Fig. 10). However, using this model results in only 22 events, and the model misses many small events that were successfully associated by PN piSDL3, for example. Furthermore, this model has only 16 events in common with PN original. Combining our three different models trained from scratch (i.e. PN piSDL1-3) using a semblance-based ensembler results in 31 detections. In comparison to TF STEAD, both catalogues have 29 events in common; however, the number of picks is reduced by more than 50% (Tab. 5). When comparing the catalogues from PN piSDL3 and SeisComp, one event was not detected by our new PhaseNet model. After manual analysis of all events in both catalogues (i.e. PN piSDL3 and SeisComp), we noticed an overlapping event. This means a second event starts in the coda of a previous event (Fig. 11). The first event was successfully detected by the associator; however, the second event in Figure 11a has not been associated, although all required phases have been picked correctly by our model. We also tested a second phase association algorithm (GaMMA, Zhu et al., 2022) which also failed. However, both events have not been detected when applying the original model (Fig. 11b), since the original model misses several phases and also labeled, for example, at station KUHL an S arrival as a P wave. Figure 11c and d demonstrates the differences in picking noisy phase arrivals between PN piSDL3 and PN original. The original PhaseNet model (Fig. 11d) is able to detect most of the S-phases; however, it does not pick the noisy P-phases. On the contrary, model PN piSDL3 (Fig. 11c) also picks these noisy P-arrivals. More comparisons between picking performances of PN piSDL3 and PN original are given the the supplementary material (Figs. S6 - S9).

Model name	P-picks	S-picks	Events
PN original	127,761	105,496	20
PN STEAD	9,567	17,508	6
TF original	97,373	83,178	31
TF STEAD	149,394	120,031	39
PN piSDL1	127,113	126,459	39
PN piSDL2	163,150	94,049	35
PN piSDL3	118,226	89,068	37
PN piSDL3 (no Ritt.)	87,849	55,252	22
Semblance	71,872	55,266	31

Table 5 Number of predicted P and S picks and associated events for different models. Note PN piSDL1-3 have been trained with varying numbers of noise samples from Rittershoffen and STEAD. The model PN piSDL3 (no Ritt.) was trained without event data from Rittershoffen. The supplementary material shows the number of picked phases at each station for model PN piSDL1.

5 Discussion

This study introduces an approach to seismic phase picking for induced seismicity using deep learning, focusing on optimising PhaseNet. The results underline the importance of new tailored data sets for induced seismicity and the inclusion of noise samples during the training process to mitigate the number of false picks and false detections after phase association. Fur-

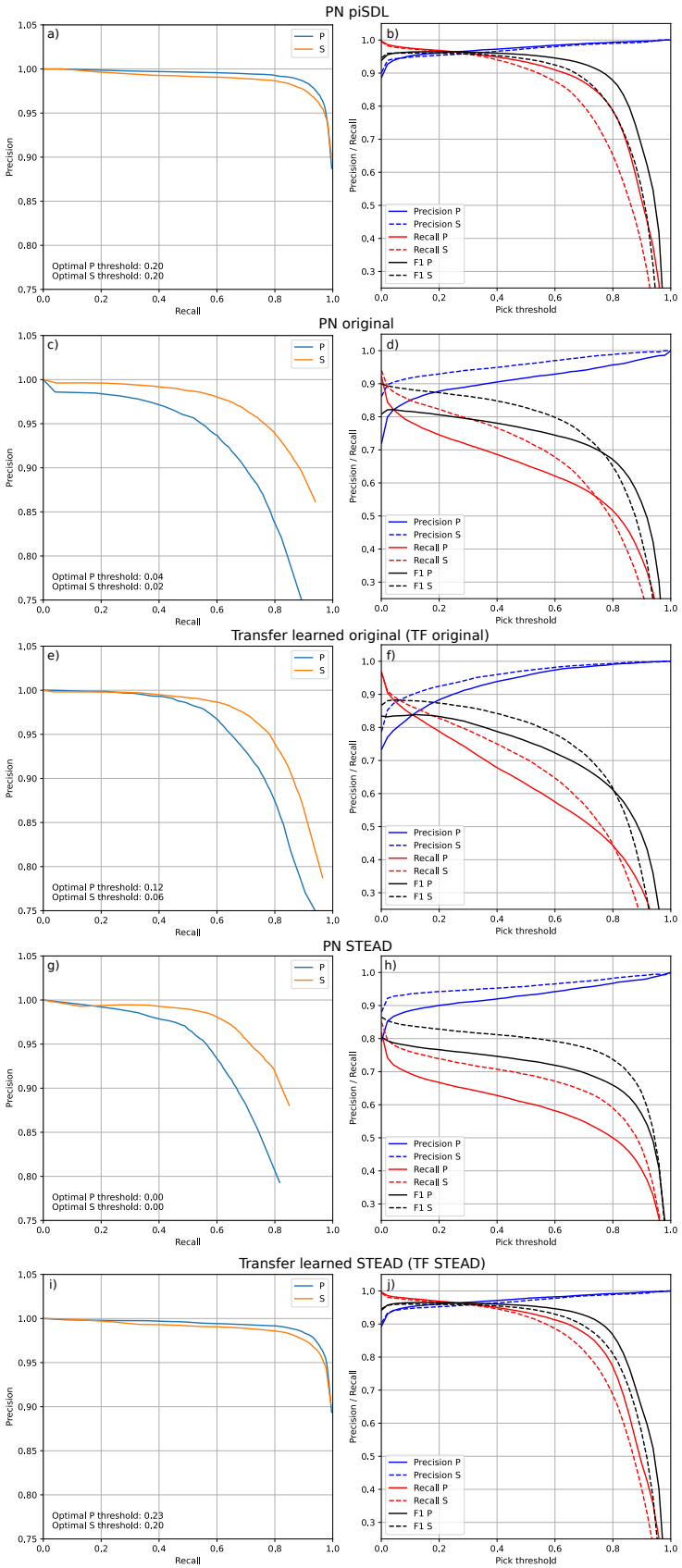

Figure 7 Evaluation of different PhaseNet models on the 10% test data set. Recall-precision curve (left column) and precision, recall, and F1-score against different pick thresholds (right column) for P and S wave. The optimal pick threshold is determined from each precision-recall curve by finding the point on that curve that is closest to the point $(1, 1)$. This point would represent the perfect model. Our model trained from scratch on the induced data set (a, b) and the transfer-learned STEAD model (i, j) perform best. Both pre-trained models (original (c, d) and STEAD (g, h)) do not perform well. However, the original model is slightly better than STEAD, although the original model was only trained on data from southern California seismic network.

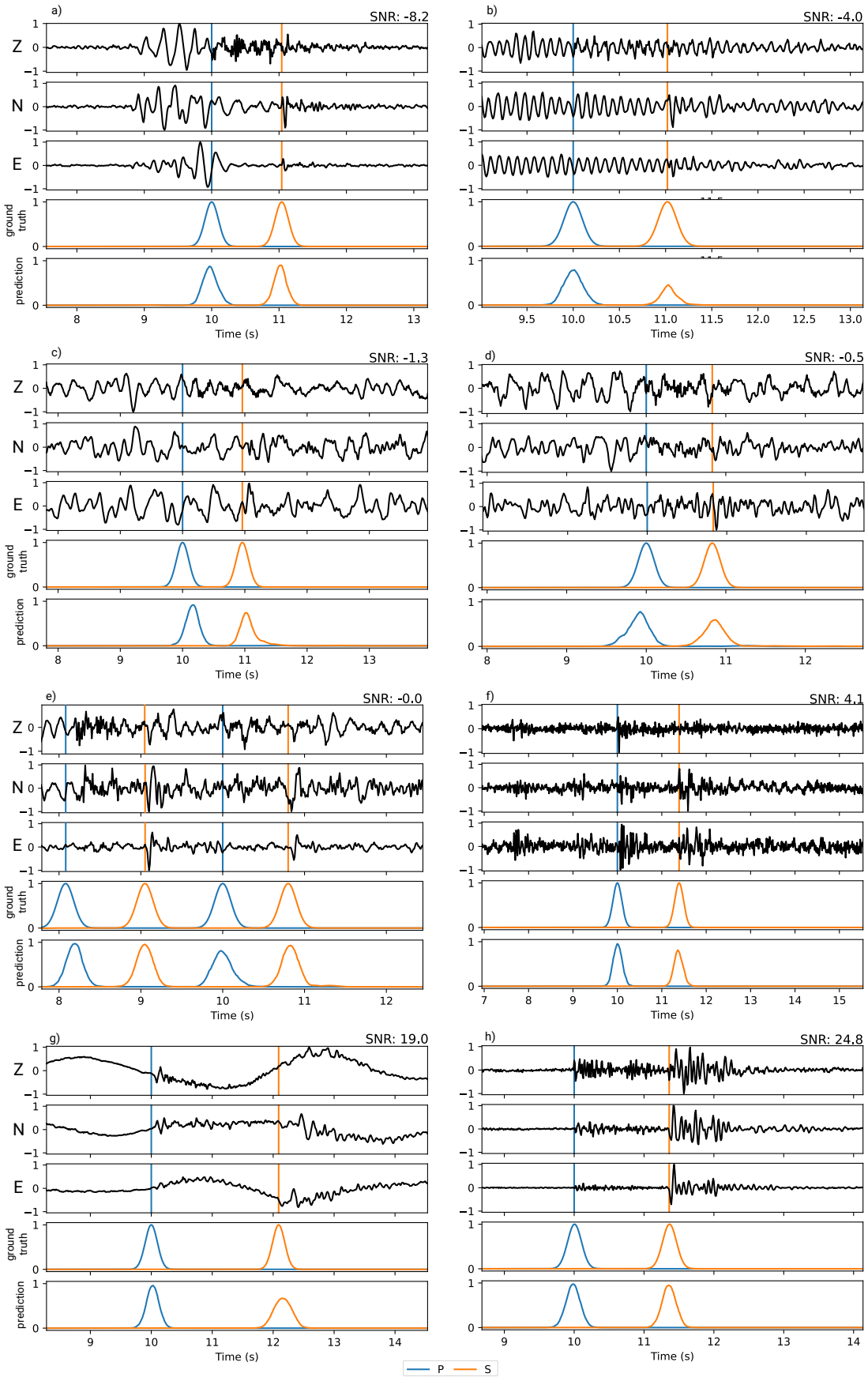

Figure 8 Examples of waveforms from the test data set and successful predictions by the piSDL model for different noise conditions, i.e. different SNR. The upper subfigures in each plot are the three-component waveforms and the lower subfigures are the ground truth probability distributions for P and S and the predicted one. The blue and orange vertical lines are the manually labeled P and S phase arrivals, respectively. The SNR in (e) is determined from the second event.

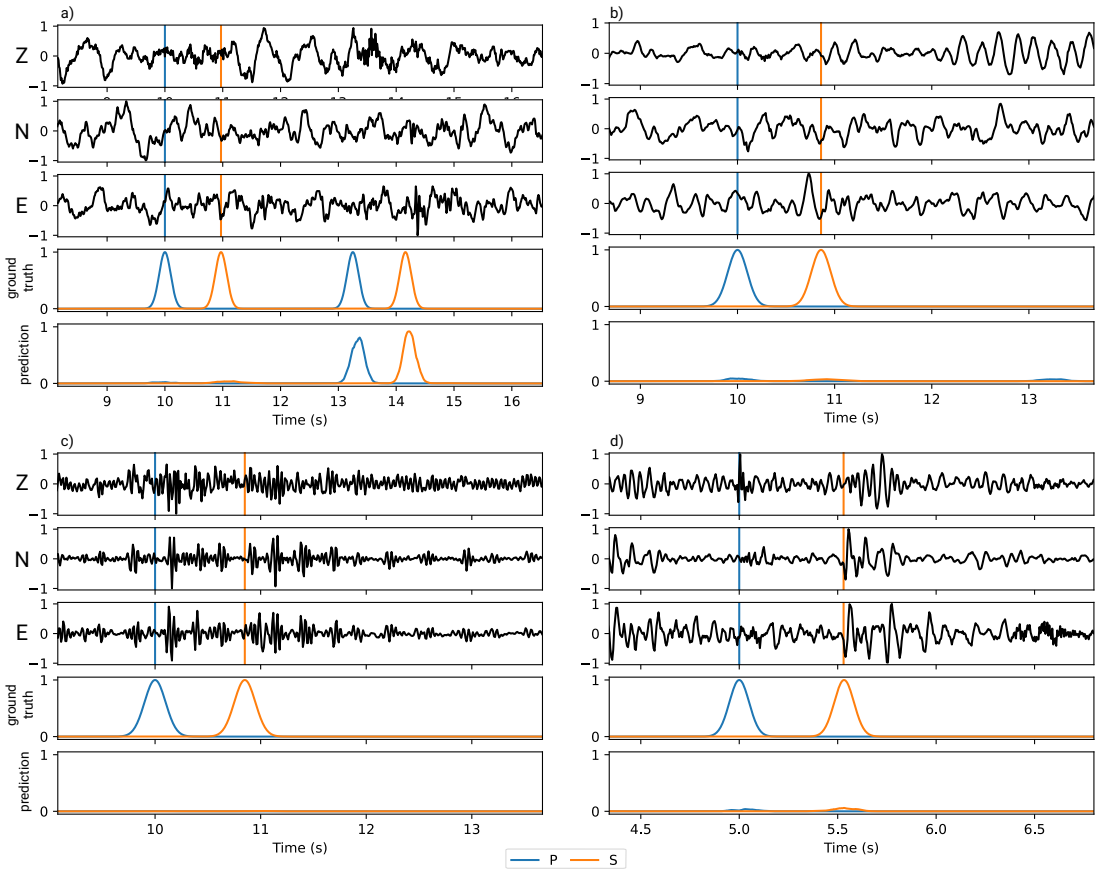

Figure 9 Examples of waveforms from the test data set and failed predictions by the piSDL model. Note in (a) the first event is counted as a failed example, even though the second event was successfully predicted.

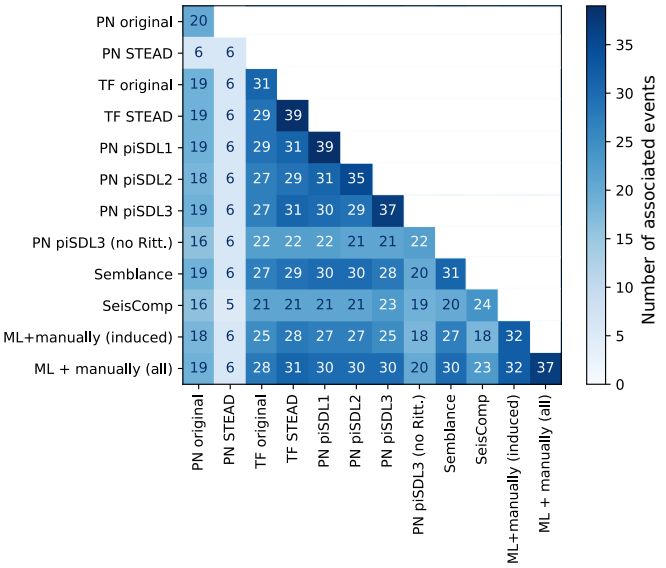

Figure 10 Comparison of different event catalogues against each other. The diagonal represents the number of associated events in January 2024 for each single model. Other values denote the number of common events between different catalogues. We compared origin times and associated waveforms to find common events between the different catalogues. Note PN piSDL1-3, TF original, and TF STEAD have been trained with varying numbers of noise samples from Rittershoffen and STEAD (Details about the number of noise samples are given in the text).

ther, applying different PhaseNet models to continuous data shows that the new models are able to detect more events at a geothermal site close to Rittershoffen (France).

The induced data set was designed to address specific challenges, including low SNR, small source-receiver distances, and complex waveforms with multiple events in a single time window. Figure 4 highlights the distinct characteristics of the induced data set compared to STEAD, showcasing its higher proportion of low-SNR events. This data set enhances the model's capability to identify phase arrivals under challenging conditions. The removal of mislabeled waveforms (Fig. 1) and the addition of noise samples further refined the training process, ensuring more reliable model predictions and mitigating the number of false picks. Our proposed data set contains induced as well as natural low magnitude earthquakes. Excluding these natural low magnitude earthquakes when training a PhaseNet model does not change the model performance, since the earthquakes waveforms of induced and natural ones are indistinguishable (e.g. Dahm et al., 2015; Schoenball et al., 2015). However, we have not yet tested the new PhaseNet model trained with our derived induced data set only on natural earthquakes to evaluate whether the new presented PhaseNet model is able to detect more low magnitude events than existing models.

Adding noise samples from both STEAD and Rittershoffen was pivotal in reducing false picks, as illustrated in Figure 5. Reducing the number of false picks is one

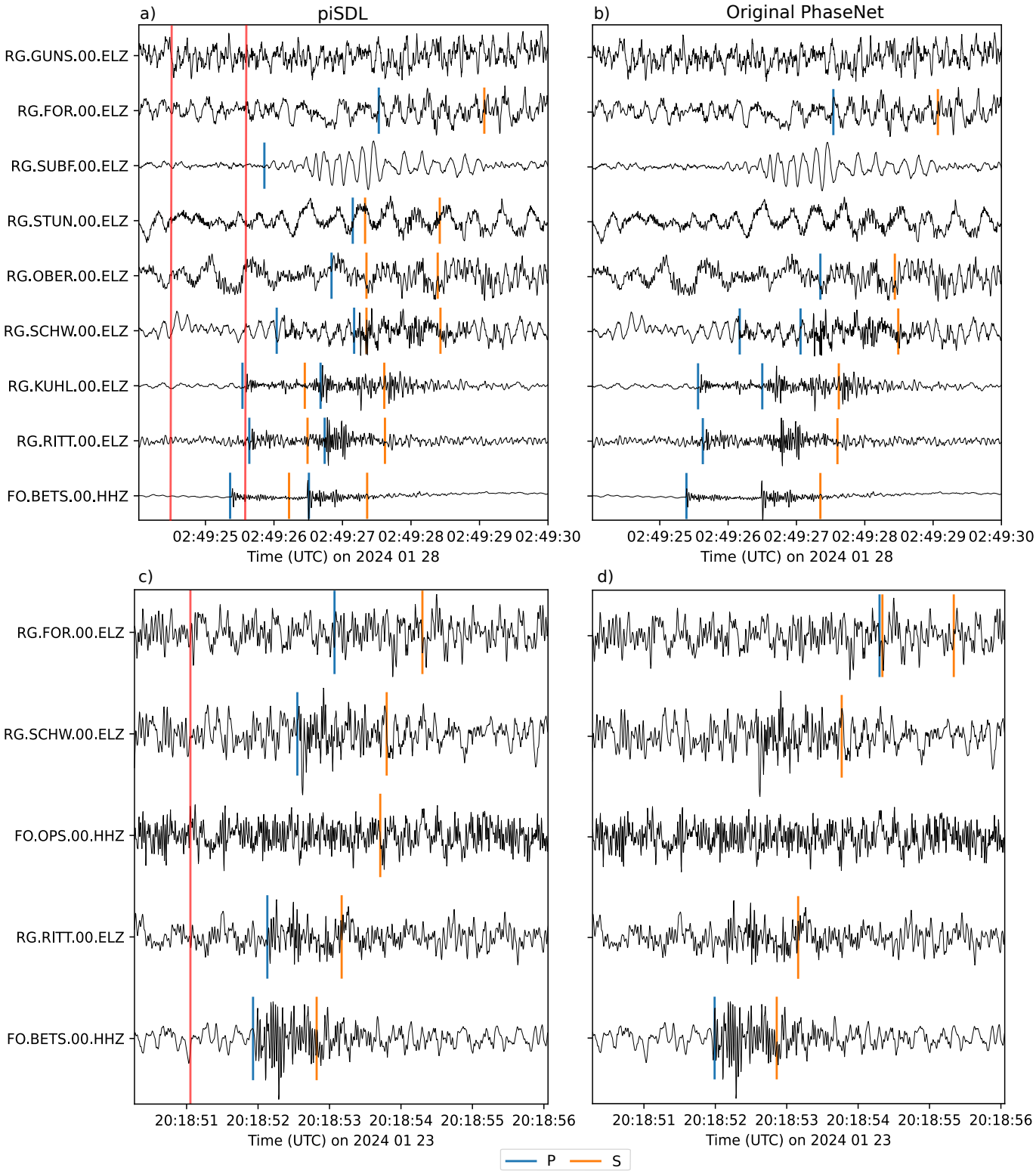

Figure 11 Two events within a few seconds at the Rittershoffen site and predicted picks by PN piSDL3 (a) and PN original (b). In (c) and (d) a third event is shown for which PN piSDL3 predicted several noisy picks and PN original only P and S arrivals at station BETS and S phases at RITT, SCHW and FOR. The waveforms are unfiltered and only the vertical component is shown. Since the original PhaseNet model predicted incorrect phase onsets (b), the associated event is also incorrect. Vertical red lines show the origin time of each seismic event. More events are shown in the supplementary material (Fig. S6 - S9).

of the most important tasks to avoid false detections after phase association, since the tested phase association algorithms do not take the waveform shape into account. Models trained without noise data showed significantly higher numbers of P and S picks (Fig. 6b). Conversely, adding 10–15% noise samples optimised the model's ability to differentiate between noise and seismic phases (i.e. P and S arrivals). In this study, we used noise samples from STEAD as well as Rittershoffen, since our new trained PhaseNet models are applied at the geothermal site in Rittershoffen. Thus, the model learns features from typical noise samples from Rit-

tershoffen as well as general features from worldwide recorded noise samples. However, we have not tested whether a PhaseNet model trained using only noise samples from STEAD performs differently than when only noise samples from Rittershoffen are used. To build a noise data set, we previously compiled an earthquake catalogue in Rittershoffen, using a PhaseNet model that was trained without noise samples. Afterwards, we selected random 30 s time windows from the continuous data where no seismic event was known in this 30 s. This selected window was then added to the noise data set. Since training of PhaseNet requires a large number of waveform samples, and therefore manual inspection of the noise samples is not possible, the proposed method does not ensure that the randomly selected noise samples do not contain low magnitude earthquakes or earthquakes that were only visible at a single station and are not part of the previously derived seismicity catalogue.

The performance metrics, including precision, recall, and F1-score (Fig. 7), reveal that models trained on the induced data set achieved superior results compared to pre-existing models. Notably, the transfer-learned STEAD model and the PN piSDL models (PhaseNet models trained from scratch with the induced seismicity data set) consistently outperformed the original published PhaseNet model. The transfer-learned original model performs worse on the test data set in comparison to the transfer-learned STEAD model (Fig. 7). The different performances are caused by the different training data sets of both models. PhaseNet STEAD was trained with approximately 1 million earthquake waveforms from $\sim$ 450,000 earthquakes, recorded at 2613 worldwide deployed seismological stations (Mousavi et al., 2019) and approximately 200,000 noise samples from these stations. The original published PhaseNet model, however, was trained with 779,514 earthquake waveforms from 889 stations of the southern California seismic network and 234,117 different events in that region (Zhu and Beroza, 2018). No noise samples were added during training of the original model. This means the STEAD model has learned more features from a higher variability of data. Nevertheless, the original model predicts more picks on continuous data than the model trained on STEAD, which results in more event detections.

The application of the trained models to continuous seismic data from Rittershoffen (Fig. 6c) validated their practical utility. Table 5 shows the number of associated events and picks for each model. The use of a semblance-based ensembler reduced the number of P and S picks while maintaining a high number of detected events. The transfer-learned STEAD and PN piSDL models outperformed PN original in detecting induced seismic events. Furthermore, as demonstrated in Figure 11a and b, PN original predicts at station KUHL an incoming S wave as a P wave. As demonstrated in Figure 4, the induced data set is built to have more events with low SNR. Figure 11c and d clearly show how PN piSDL3 picks noisy P-arrivals, while PN original only successfully detects most of the S-arrivals. Figures S6 - S9 also highlight that the newly trained PhaseNet models picks more noisy P-arrivals. However, as shown in Figures 1 and 9, PN piSDL3 might fail in some examples. Since manual seismic phase picking allows the integration of information from neighbouring stations, analysts might see small wiggles and in the waveforms and and pick these wiggles as seismic phase arrivals. PhaseNet only uses information from the input three-component waveform and has no spatial information from neighbouring stations. For these cases, seismic phase picking approaches such as Phase Neural Operator (PhaseNo Sun et al., 2023) will probably help to detect even more noisy phase arrivals.

Retraining PhaseNet models with our proposed induced seismicity data set does result in a much higher detection rate. As already hypothesised by Bornstein et al. (2024), PhaseNet does not have such a high impact on transfer learning due to the low number of training parameters. Comparing the models trained from scratch and the transfer-learned models, all models perform very similarly but the number of picks differs. It is noteworthy that the models trained with more noise samples have fewer picks, i.e. fewer false picks. However, variability in the number of associated events across models indicates potential areas for further refinement, such as more careful selection criteria for the noise data set. Further improvements in the workflow for building seismic catalogues might also reduce the number of false detections: our workflow did not exclude events with source locations outside of the analysed network, and challenges in associating overlapping seismic events (Fig. 11) still remain. In addition, the performance of the model during an earthquake sequence in Rittershoffen was not tested in this study and needs to be verified in the future. Applying a model trained from scratch without earthquake waveforms from Rittershoffen results in a similar number of detections as for the original model, even though the model was trained with noise samples from the site. Both derived catalogues contain all large events; however, low magnitude events are missed by both models. These findings coincide with the results by Chai et al. (2020) and Jiang et al. (2021) that PhaseNet models do not generalise well and transfer learning with small data sets for the region of interest is a mandatory task to detect all earthquakes.

6 Conclusions

The induced data set is a new data set specifically designed to address low SNR events and small source-receiver distances to train the deep neural network PhaseNet to pick arrivals of P and S waves. The data set mostly contains induced seismicity events from geothermal sites, waste water disposal, and coal mine flooding, as well as natural low magnitude events from the Swiss Seismological Service and the Eifel region in Germany. PhaseNet models trained on this specific data set demonstrated significant performance advantages over models trained on general-purpose data sets such as STEAD or PhaseNet's original published model. Including noise samples during model training mitigates the number of false picks and enhances the

model's ability to differentiate between noise and seismic signals. The PhaseNet models trained and transfer-learned with the induced data set achieved higher precision and recall rates compared to pre-existing models, showcasing the effectiveness of specialized training strategies. However, challenges remain, particularly in detecting and associating overlapping seismic events, which highlights the need for more robust phase association algorithms. The results underscore the importance of specific designed training data sets to train or fine tune existing PhaseNet models for different tasks, and training PhaseNet with noise samples is the most important step to reduce the number of false picks. However, some challenges still remain when training and applying PhaseNet models: conducting broader hyperparameter searches (Park and Shelly, 2024) could yield even more robust models. Adding an antialiasing layer (Park et al., 2024) might help to avoid missing picks, because of the overlapping of neighbouring windows when working with continuous data. Exploring novel loss functions such as focal or dice loss (Park and Beroza, 2025) might help to overcome the selection of the pick threshold for practitioners to pick P and S waves. Furthermore, since induced seismic events mostly have low magnitudes, the earthquake waveforms are only visible at stations which are close the hypocentre. At other stations, the recorded earthquake waveforms might be noisy. Adding spatial information to seismic phase picking approaches, as done by Sun et al. (2023), might help to pick more events in noisy environments.

Acknowledgements

This project was funded by the Bundesministerium für Bildung und Forschung under the name "AI-based monitoring of geothermal Seismicity" with the grant number 01|S23030A. The authors thank the handling Editor Hongyu Sun, Cindy Lim Shin Yee, and an anonymous reviewer for reviewing our manuscript and providing constructive feedback to improve our manuscript.

Data and Code Availability

The compiled data set and trained model will be accessible via SeisBench (https://github.com/seisbench/seisbench) but the published data set does not contain data from Rittershoffen due to the ownership of ÉS Géothermie and waveform data from the Eifel region, due to an embargo till 2026. The provided data set through SeisBench contains 142,001 waveforms. The data set has a size of 35 GB. The Python scripts to train PhaseNet is available under https://github.com/JanisHe/train_phasenet (last accessed: 2025-08-11) and to build catalogues from continuous seismic data under https://github.com/JanisHe/seisbench_catalogues (last accessed: 2025-08-11). All plots are created with Matplotlib (Hunter, 2007)

References

Allen, R. Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72 (6B):S225–S242, 12 1982. doi: 10.1785/BSSA07206B0225.

Baer, M. and Kradolfer, U. An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, 77(4):1437–1445, 08 1987. doi: 10.1785/BSSA0770041437.

Baujard, C., Genter, A., Dalmais, E., Maurer, V., Hehn, R., Rosillette, R., Vidal, J., and Schmittbuhl, J. Hydrothermal characterization of wells GRT-1 and GRT-2 in Rittershoffen, France: Implications on the understanding of natural flow systems in the rhine graben. *Geothermics*, 65:255–268, 2017. doi: 10.1016/j.geothermics.2016.11.001.

Becker, D., McBrearty, I. W., Beroza, G. C., and Martínez-Garzón, P. Performance of AI-Based Phase Picking and Event Association Methods after the Large 2023 Mw 7.8 and 7.6 Türkiye Doublet. *Bulletin of the Seismological Society of AmericaSearch Dropdown Menu*, 2024. doi: 10.1785/0120240017.

Bornstein, T., Lange, D., Münchmeyer, J., Woollam, J., Rietbrock, A., Barcheck, G., Grevemeyer, I., and Tilmann, F. PickBlue: Seismic phase picking for ocean bottom seismometers with deep learning. *Earth and Space Science*, 11(1):e2023EA003332, 2024. doi: 10.1029/2023EA003332.

Castillo, E., Siervo, D., and Prieto, G. A. Colombian Seismic Monitoring Using Advanced Machine-Learning Algorithms. *Seismological Research Letters*, 2024. doi: 10.1785/0220240036.

Chai, C., Maceira, M., Santos-Villalobos, H. J., Venkatakrishnan, S. V., Schoenball, M., Zhu, W., Beroza, G. C., Thurber, C., and Team, E. C. Using a deep neural network and transfer learning to bridge scales for seismic phase picking. *Geophysical Research Letters*, 47(16):e2020GL088651, 2020. doi: 10.1029/2020GL088651.

Chen, Y., Savvaidis, A., Saad, O. M., Dino Huang, G.-C., Siervo, D., O'Sullivan, V., McCabe, C., Uku, B., Fleck, P., Burke, G., et al. TXED: The Texas earthquake dataset for AI. *Seismological Research Letters*, 95(3):2013–2022, 2024. doi: 10.1785/0220230327.

Cuenot, N., Dorbath, C., and Dorbath, L. Analysis of the microseismicity induced by fluid injections at the EGS site of Soultz-sous-Forêts (Alsace, France): implications for the characterization of the geothermal reservoir properties. *Pure and applied geophysics*, 165:797–828, 2008. doi: 10.1007/s00024-008-0335-7.

Dahm, T., Cesca, S., Hainzl, S., Braun, T., and Krüger, F. Discrimination between induced, triggered, and natural earthquakes close to hydrocarbon reservoirs: A probabilistic approach based on the modeling of depletion-induced stress changes and seismological source parameters. *Journal of Geophysical Research: Solid Earth*, 120(4):2491–2509, 2015. doi: 10.1002/2014JB011778.

Dai, Z., Zhou, L., Hu, X., Qu, J., and Li, X. Generalization of PhaseNet in Shandong and its application to the Changqing M4.1 earthquake sequence. *Earthquake Science*, 36(3):212–227, 2023. doi: 10.1016/j.eqs.2023.04.003.

Diehl, T., Kissling, E., Husen, S., and Aldersons, F. Consistent phase picking for regional tomography models: application to the greater Alpine region. *Geophysical Journal International*, 176(2): 542–554, 02 2009. doi: 10.1111/j.1365-246X.2008.03985.x.

Diehl, T., Kraft, T., Kissling, E., and Wiemer, S. The induced earthquake sequence related to the St. Gallen deep geothermal project (Switzerland): Fault reactivation and fluid interactions imaged by microseismicity. *Journal of Geophysical Research: Solid Earth*, 122(9):7272–7290, 2017. doi:

10.1002/2017JB014473.

Genter, A., Goerke, X., Graff, J.-J., Cuenot, N., Krall, G., Schindler, M., and Ravier, G. Current status of the EGS Soultz geothermal project (France). In *world geothermal congress, WGC2010, Bali, Indonesia*, pages 25–29, 2010.

Goforth, T. and Herrin, E. An automatic seismic signal detection algorithm based on the Walsh transform. *Bulletin of the Seismological Society of America*, 71(4):1351–1360, 08 1981. doi: 10.1785/BSSA0710041351.

Helmholtz-Centre Potsdam - GFZ German Research Centre for Geosciences and gempa GmbH. The SeisComP seismological software package. GFZ Data Services., 2008. doi: 10.5880/GFZ.2.4.2020.003.

Hensch, M., Dahm, T., Ritter, J., Heimann, S., Schmidt, B., Stange, S., and Lehmann, K. Deep low-frequency earthquakes reveal ongoing magmatic recharge beneath Laacher See Volcano (Eifel, Germany). *Geophysical Journal International*, 216(3): 2025–2036, 01 2019. doi: 10.1093/gji/ggy532.

Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

Jiang, C., Fang, L., Fan, L., and Li, B. Comparison of the earthquake detection abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes. *Earthquake Science*, 34(5): 425–435, 2021. doi: 10.29382/eqs-2021-0038.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014. https://arxiv.org/abs/1412.6980.

Koushesh, K. and Ritter, J. R. An adaptive 6-dimensional floating-search multi-station seismic-event detector (A6-DFMSD) and its application to low-frequency earthquakes in the East Eifel Volcanic Field, Germany. *Journal of Applied Volcanology*, 13(1):9, 2024. doi: 10.1186/s13617-024-00147-8.

Küperkoch, L., Meier, T., Lee, J., Friederich, W., and Group, E. W. Automated determination of P-phase arrival times at regional and local distances using higher order statistics. *Geophysical Journal International*, 181(2):1159–1170, 05 2010. doi: 10.1111/j.1365-246X.2010.04570.x.

Leonard, M. and Kennett, B. Multi-component autoregressive techniques for the analysis of seismograms. *Physics of the Earth and Planetary Interiors*, 113(1):247–263, 1999. doi: 10.1016/S0031-9201(99)00054-0.

Lim, C. S. Y., Lapins, S., Segou, M., and Werner, M. J. Deep learning phase pickers: how well can existing models detect hydraulic-fracturing induced microseismicity from a borehole array? *Geophysical Journal International*, 240(1):535–549, 10 2024. doi: 10.1093/gji/ggae386.

Lomax, A., Zollo, A., Capuano, P., and Virieux, J. Precise, absolute earthquake location under Somma–Vesuvius volcano using a new three-dimensional velocity model. *Geophysical Journal International*, 146(2):313–331, 2001. doi: 10.1046/j.0956-540x.2001.01444.x.

Lomax, A., Michelini, A., Curtis, A., and Meyers, R. Earthquake location, direct, global-search methods. *Encyclopedia of complexity and systems science*, 5:2449–2473, 2009.

Maurer, V., Gaucher, E., Grunberg, M., Koepke, R., Pestourie, R., and Cuenot, N. Seismicity induced during the development of the Rittershoffen geothermal field, France. *Geothermal Energy*, 8(1): 5, 2020. doi: 10.1186/s40517-020-0155-2.

Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., and Lauciani, V. INSTANCE–the Italian seismic dataset for machine learning. *Earth System Science Data*, 13(12):5509–5544, 2021. doi: 10.5194/essd-13-5509-2021.

Mousavi, S. M., Sheng, Y., Zhu, W., and Beroza, G. C. STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. *IEEE Access*, 7:179464–179476, 2019. doi: 10.1109/ACCESS.2019.2947848.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., and Beroza, G. C. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, 11(1):1–12, 2020. doi: 10.1038/s41467-020-17591-w.

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., et al. Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *Journal of Geophysical Research: Solid Earth*, 127(1):e2021JB023499, 2022. doi: 10.1029/2021JB023499.

Myklebust, E. B. and Köhler, A. Deep learning models for regional phase detection on seismic stations in Northern Europe and the European Arctic. *Geophysical Journal International*, 239(2): 862–881, 08 2024. doi: 10.1093/gji/ggae298.

Münchmeyer, J. PyOcto: A high-throughput seismic phase associator. *Seismica*, 3(1), Jan. 2024. doi: 10.26443/seismica.v3i1.1130.

Niksejel, A. and Zhang, M. OBSTransformer: a deep-learning seismic phase picker for OBS data using automated labelling and transfer learning. *Geophysical Journal International*, page ggae049, 2024. doi: 10.1093/gji/ggae049.

Park, Y. and Beroza, G. C. Reducing the Parameter Dependency of Phase-Picking Neural Networks with Dice Loss. *The Seismic Record*, 5(1):55–63, 01 2025. doi: 10.1785/0320240028.

Park, Y. and Shelly, D. R. The Value of Hyperparameter Optimization in Phase-Picking Neural Networks. *The Seismic Record*, 4 (3):231–239, 09 2024. doi: 10.1785/0320240025.

Park, Y., Delbridge, B. G., and Shelly, D. R. Making Phase-Picking Neural Networks More Consistent and Interpretable. *The Seismic Record*, 4(1):72–80, 2024. doi: 10.1785/0320230054.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., Shyu, M.-L., Chen, S.-C., and Iyengar, S. S. A survey on deep learning: Algorithms, techniques, and applications. *ACM computing surveys (CSUR)*, 51(5):1–36, 2018. doi: 10.1145/3234150.

Rische, M., Fischer, K. D., and Friederich, W. FloodRisk–Induced seismicity by mine flooding–Observation, characterisation and relation to mine water rise in the eastern Ruhr area (Germany). *Journal of Applied & Regional Geology/Zeitschrift der Deutschen Gesellschaft für Geowissenschaften (ZDGG)*, 173(4), 2022. doi: 10.1127/zdgg/2023/0346.

Ritter, J. R., Koushesh, K., Schmidt, B., Föst, J.-P., Bühler, J., Hensch, M., and Mader, S. M. Seismological monitoring of magmatic and tectonic earthquakes in the East Eifel Volcanic Field, Germany. *Journal of Seismology*, pages 1–26, 2024. doi: 10.1007/s10950-024-10257-w.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. doi: 10.1007/978-3-319-24574-4_28.

Ross, Z. E., Meier, M.-A., and Hauksson, E. P-wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, 2018. doi: 10.1029/2017JB015251.

Roth, M. P., Verdecchia, A., Harrington, R. M., and Liu, Y. High-resolution imaging of hydraulic-fracturing-induced earthquake clusters in the Dawson-Septimus area, Northeast British Columbia, Canada. *Seismological Research Letters*, 91(5): 2744–2756, 2020. doi: 10.1785/0220200086.

Schill, E., Genter, A., Cuenot, N., and Kohl, T. Hydraulic performance history at the Soultz EGS reservoirs from stimulation and long-term circulation tests. *Geothermics*, 70:110–124, 2017. doi: 10.1016/j.geothermics.2017.06.003.

Schmittbuhl, J., Lambotte, S., Lengliné, O., Grunberg, M., Jund, H., Vergne, J., Cornet, F., Doubre, C., and Masson, F. Induced and triggered seismicity below the city of Strasbourg, France from November 2019 to January 2021. *Comptes Rendus. Géoscience*, 353(S1):561–584, 2021. doi: 10.5802/crgeos.71.

Schoenball, M., Davatzes, N. C., and Glen, J. M. Differentiating induced and natural seismicity using space-time-magnitude statistics applied to the Coso Geothermal field. *Geophysical Research Letters*, 42(15):6221–6228, 2015. doi: 10.1002/2015GL064772.

Sleeman, R. and van Eck, T. Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Physics of the Earth and Planetary Interiors*, 113(1):265–275, 1999. doi: 10.1016/S0031-9201(99)00007-2.

Suarez, A. L. A. and Beroza, G. Curated Regional Earthquake Waveforms (CREW) Dataset. *Seismica*, 3(1), 2024. doi: 10.26443/seismica.v3i1.1049.

Sun, H., Ross, Z. E., Zhu, W., and Azizzadenesheli, K. Phase neural operator for multi-station picking of seismic arrivals. *Geophysical Research Letters*, 50(24):e2023GL106434, 2023. doi: 10.1029/2023GL106434.

Swiss Seismological Service (SED) At ETH Zurich. National Seismic Networks of Switzerland, 1983. doi: 10.12686/SED/NETWORKS/CH.

Wong, W. C. J., Zi, J., Yang, H., and Su, J. Spatial-temporal evolution of injection-induced earthquakes in the Weiyuan Area determined by machine-learning phase picker and waveform cross-correlation. *Earth and Planetary Physics*, 5(6):520–531, 2021. doi: 10.26464/epp2021055.

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., et al. SeisBench—A toolbox for machine learning in seismology. *Seismological Society of America*, 93(3):1695–1709, 2022. doi: 10.48550/arXiv.2111.00786.

Xi, Z., Wei, S. S., Zhu, W., Beroza, G. C., Jie, Y., and Saloor, N. Deep learning for deep earthquakes: insights from OBS observations of the Tonga subduction zone. *Geophysical Journal International*, 238(2):1073–1088, 06 2024. doi: 10.1093/gji/ggae200.

Yuan, C., Ni, Y., Lin, Y., and Denolle, M. Better Together: Ensemble Learning for Earthquake Detection and Phase Picking. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–17, 2023. doi: 10.1109/TGRS.2023.3320148.

Zhu, J., Li, Z., and Fang, L. USTC-Pickers: a unified set of seismic phase pickers transfer learned for China. *Earthquake Science*, 36(2):95–112, 2023. doi: 10.1016/j.eqs.2023.03.001.

Zhu, J., Fang, L., Miao, F., Fan, L., Zhang, J., and Li, Z. Deep learning and transfer learning of earthquake and quarry-blast discrimination: applications to southern California and eastern Kentucky. *Geophysical Journal International*, 236(2):979–993, 2024. doi: 10.1093/gji/ggad463.

Zhu, W. and Beroza, G. C. PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273, 2018. doi: 10.1093/gji/ggy423.

Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L., and Beroza, G. C. Earthquake phase association using a Bayesian Gaussian mixture model. *Journal of Geophysical Research: Solid Earth*, 127(5):e2021JB023249, 2022. doi: 10.1029/2021JB023249.