HHS Public Access

Author manuscript

Geophys J Int. Author manuscript; available in PMC 2022 January 01.

Published in final edited form as:

Geophys J Int. 2021 January 01; 224(1): 230–240.

Application of a convolutional neural network for seismic phase picking of mining-induced seismicity

Sean W. Johnson[1], **Derrick J. A. Chambers**[1], **Michael S. Boltz**[1], **Keith D. Koper**[2]

[1]National Institute for Occupational Safety and Health, Spokane Mining Research Division, 315 E Montgomery Av., Spokane WA, 99207, USA.

[2]University of Utah Seismograph Stations, 115 South 1460 East, Room 211 FASB SLC, UT 84112-0102, USA

SUMMARY

Monitoring mining-induced seismicity (MIS) can help engineers understand the rock mass response to resource extraction. With a thorough understanding of ongoing geomechanical processes, engineers can operate mines, especially those mines with the propensity for rock-bursting, more safely and efficiently. Unfortunately, processing MIS data usually requires significant effort from human analysts, which can result in substantial costs and time commitments. The problem is exacerbated for operations that produce copious amounts of MIS, such as mines with high-stress and/or extraction ratios. Recently, deep learning methods have shown the ability to significantly improve the quality of automated arrival-time picking on earthquake data recorded by regional seismic networks. However, relatively little has been published on applying these techniques to MIS. In this study, we compare the performance of a convolutional neural network (CNN) originally trained to pick arrival times on the Southern California Seismic Network (SCSN) to that of human analysts on coal-mine-related MIS. We perform comparisons on several coal-related MIS data sets recorded at various network scales, sampling rates and mines. We find that the Southern-California-trained CNN does not perform well on any of our data sets without retraining. However, applying the concept of transfer learning, we retrain the SCSN model with relatively little MIS data after which the CNN performs nearly as well as a human analyst. When retrained with data from a single analyst, the analyst-CNN pick time residual variance is lower than the variance observed between human analysts. We also compare the retrained CNN to a simpler, optimized picking algorithm, which falls short of the CNN's performance. We conclude that CNNs can achieve a significant improvement in automated phase picking although some data set-specific training will usually be required.

xik9@cdc.gov .

Data Availability

The de-contextualized event waveforms and phase picks used in this study were compiled by Chambers (2020) and can be found on Harvard's Dataverse at: https://doi.org/10.7910/DVN/5DGFJB. A simple python package for training and evaluating CNN and Baer models can be found at https://github.com/sjohnson5/CoalPick.

Moreover, initializing training with weights found from other, even very different, data sets can greatly reduce the amount of training data required to achieve a given performance threshold.

Keywords

Neural networks; fuzzy logic; Time-series analysis; Body waves; Induced seismicity

INTRODUCTION

Because seismic monitoring offers unique insight into the Earth's response to mining, it has become standard practice in deep underground hardrock mines, especially those experiencing rockbursting (Mendecki *et al.* 2010). However, it is less common, especially in the United States, for underground coal mines to adopt seismic monitoring, largely due to two factors: (1) the unique challenges and complexities associated with coal mining and shallow depositional geology have made it difficult to translate the well-developed technology and processing techniques used in hardrock mining, and (2) it is difficult for coal mines, which tend to be much larger and mined more quickly than hardrock mines, to justify the significant cost of a traditional in-mine seismic monitoring system, especially if they do not experience damaging seismic events. Despite this, seismic monitoring of underground coal mines is useful for a number of reasons: documenting seismicity, evaluating mine design performance and in some cases, detecting potentially dangerous ground stability issues (Swanson *et al.* 2016).

The standard product of seismic monitoring is an earthquake catalogue—a listing of information about each discrete seismic event including origin time, location, magnitude and often other source parameters such as radiated energy or the moment tensor. Event locations are particularly important for accurately interpreting the geomechanical significance of the seismicity because location errors propagate to other source parameter estimates. For coal mining environments, the greatest source of location error typically stems from inaccurately modelling the complex, time-dependent velocity structure in which the events occur (Collins *et al.* 2014; Boltz *et al.* 2016; Czarny *et al.* 2016). Another significant source of error is associated with determining body wave arrival times, which are used in standard location procedures. Generally, phase arrival times are determined automatically and then reviewed by a trained analyst to reduce errors and improve event locations. However, if automated phase picking can be improved, perhaps to the point of human levels of accuracy, it would reduce the cost and time investment required to process the data, resulting in better risk management and safer mines.

Over the past few decades, many automatic picking techniques have been developed for estimating body wave arrival times. Commonly used methods are based on detecting changes in observed energy, polarization, or other statistical properties of the recorded time-series (e.g. Baer & Kradolfer 1987; Withers *et al.* 1998; Baillard *et al.* 2013). Such picking methods are effective but require extensive network-specific tuning procedures (Nippress *et al.* 2010; Vassallo *et al.* 2012). Even when tuned, these picking methods typically fall short of human-level accuracy.

Several studies have also explored the use of relatively simplistic neural networks for phase-picking tasks (e.g. Dai & McBeth 1995; Dai & McBeth 1997; Gentili & Michelini 2006). However, recent advancements in computing capabilities, neural network architecture, training techniques and, most importantly, the availability of high-quality neural network software has enabled a new wave of advances. For example, recent studies have reported dramatic improvements in automatic *P*-arrival picking using convolutional neural networks (CNN, Kong *et al.* 2018; Bergen *et al.* 2019). CNNs are advantageous for image recognition and time-series tasks because they provide some degree of invariance to shifting and distortion (LeCun & Bendgio 1995). There are two types of phase-picking CNNs mentioned in the literature: regressors, which return a single scalar indicating pick time for each trace, and transformers, which return a characteristic function (CF) with a similar shape as the input. An example of a scalar CNN was developed by Ross *et al.* (2018a). The model operates on vertical-channel waveforms and returns the index of the picked *P* arrival. This model was trained on approximately 4.8 million *P* arrivals. The reported standard deviation of the difference between model and analyst picks (pick time residuals) was only 0.023 s or 2–3 samples after using the outer fence method to remove outliers. In contrast, Zhu & Beroza (2018) designed a transformer CNN which uses three-component traces and returns arrays indicating the probability of each sample belonging to three categories: *P*- and *S*-phase arrivals and noise. *P*- and *S*-arrival times are then selected as the maximum of a phase likelihood array if some threshold is exceeded. This model was trained on about 880 000 traces with picks for both *P*- and *S*-arrival times and the pick-time residuals had a standard deviation of 0.052 s 5–6 samples after excluding residuals higher than 0.5 s. Woollam *et al.* (2019) also created a CNN which outputs a similar CF indicating phase arrival probabilities. Although a fair comparison is difficult since Woollam *et al.* (2019) used residuals from a location procedure for assessing model performance rather than directly calculating analyst pick-time residuals, they demonstrated that reasonable results could be obtained by training the CNN on far less data than the other two studies (11 000 phase picks).

CNNs have also been developed for performing tasks other than arrival-time picking, for example determining first motions (Ross *et al.* 2018a), associating arrival-time picks into events (Ross *et al.* 2019), detecting and classifying body wave phase labels (Ross *et al.* 2018b), calculating magnitudes (Mousavi & Beroza 2019) and even detecting and locating events based only on waveforms (Perol *et al.* 2018). Although most studies focus on regional or global seismicity, several applications of CNNs have been applied to mining-induced seismicity (MIS) including:

1. Huang *et al.* (2018) trained a CNN to detect and locate events in an underground Chinese phosphate mine.
2. Wilkins *et al.* (2020) trained a CNN to identify induced events at an underground Australian coal while minimizing false detections related to operational noise.
3. Lin *et al.* (2019) used a CNN to classify signal types (blast, event and noise) originating in a Chinese copper mine.

However, little work has been published on applying CNN phase pickers to MIS. In this study, we assess the performance of a publicly accessible CNN *P*-phase picker (Ross *et*

al. 2018a) on several seismic networks of different scales and instrument compositions monitoring active longwall coal mines. We quantify the performance of the original CNN on the MIS data in order to determine how well the model transfers to data sets which are very different from the model's training data. In one sense, we expect the model to require data set specific tuning just like any other automatic phase picker. However, since deep-learning models are much more complex than traditional picking algorithms, we find it intriguing that perhaps the model has internalized some invariant phase picking concepts, just as a human analyst would. If this is the case, the model would be able to adequately pick *P*-arrival times on any seismic data set. After quantifying the original model's transferability, we then explore improving the model's performance through retraining the CNN for each data set utilizing the original weights as a starting point. Using the same training data set, we also optimize a simple *P*-arrival-time picker, a slightly modified version of the Baer and Kradolfer picking algorithm (Baer & Kradolfer 1987) included in the ObsPy Python package (Krischer *et al.* 2015), and compare its performance with the other models.

DATA

The data sets used in this study were collected by five different seismic networks monitoring underground longwall coal mines in the United States. Four of the networks (data sets A–D) were deployed and operated by the National Institute for Occupational Safety and Health (NIOSH). No geographic references are made to these operations as the mines wish to remain anonymous. The fifth network (data set E) is operated by the University of Utah.

Data sets A and B

Data sets A and B were derived from two separate temporary surface deployments at the same mine, consisting of 5-Hz, three-component MagSeis ZLand geophones sampling at 500 sps with a gain setting of 30 dB (Fig. 1). The geophones were centred over an active longwall for approximately one month in each deployment. The first deployment consisted of 11 stations covering approximately 0.15 km^2 and detected 12 499 seismic events with local magnitudes ranging from around −1.5 to 1.5. The second deployment consisted of 16 stations covering approximately 1.2 km^2 and detected 20 792 events with local magnitudes in a similar range. In both cases, we used ObsPy's sta/lta detector, coincidence filter, and Baer picker for event detection, association and initial phase picking.

Due to the large volume of data, only 1251 events (10 per cent of the total) from data set A were manually processed by a single analyst, resulting in 12 527 *P*-arrival picks that were used as a training data set. The test data set consisted of an additional 100 events (50 from each data set) that were each manually processed by four individuals, including the analyst which processed the training data. Multiple analysts processed the same events in order quantify analyst variability. The test sets featured 514–542 *P*-picks for data set A and 677–738 *P*-picks for data set B, depending on the analyst.

Data set C

Data set C was collected by a dense microseismic network of both in mine and surface sensors. The surface stations were 4.5-Hz geophones sampling at 1000 sps, and the

underground stations were 14-Hz, three-component geophones sampling at 5000 sps. The network operated for approximately 5 yr and recorded events associated with the mining of four longwall panels. Over 210 000 seismic triggers were recorded during the network lifetime, although many of the triggers were caused by noise associated with mine operations rather than induced seismicity.

For this study, we used analyst-processed events detected during an arbitrarily selected month. During this time period, the network consisted of 8 underground and 11 surface stations covering 5.5 km^2 (Fig. 2). There were 2345 events, located with 23 942 manual *P*-wave picks, that occurred during this time with local magnitudes ranging from −1 to 3. There were some quality issues identified with the manual processing, which was performed by a third party, but these were not significant enough to merit reprocessing. There were 5986 randomly selected traces (about 25 per cent of the total) that were used as the test set, and the rest were used for training.

Data set D

Data set D consists of two years of data collected by a local surface network surrounding a longwall coal mine (Fig. 3). The 130-km^2 network consists of 6 three-component Guralp 6TD broad-band seismometers sampling at 250 sps and three stations with a three-component EpiSensor accelerometer and a vertical L4 1-Hz geophone, each sampling at 100 sps. The Earthworm software suite (Friberg *et al.* 2010) was used to collect and store continuous data, detect seismic events and calculate preliminary locations and magnitudes. The data set includes 5808 MIS events, located with 31 987 manual *P*-wave picks, with moment magnitudes ranging from −1 to 2. Approximately 25 per cent (7997 randomly selected traces) of the picked traces were used for testing, and the rest were used for training.

Data set E

Data set E includes 1929 events located by the University of Utah Seismograph Stations (UUSS) that originated in the coal mining regions of Utah between 2012 October 01 and 2019 December 04 (Fig. 4). These events have local magnitudes ranging from 0.4 to 3.3. Mines have produced coal in these regions for over a century, and the associated seismicity has been extensively studied (e.g. Arabasz & Pechmann 2001; Arabasz *et al.* 2005). In order to follow the methodology of Ross *et al.* (2018a), we only used manually reviewed phase picks with source–receiver distances less than 120 km, resulting in 16 924 *P* arrivals on high-gain (EHZ and HHZ) channels. Approximately 25 per cent (4231 arrivals) of the data were used for testing, and the remaining 12 693 arrivals were used for training.

METHODOLOGY

We evaluated three models for automatic picking of *P*-wave arrival times for each data set: (1) base CNN: the CNN model and weights published by Ross *et al.* (2018a), (2) retrained CNN: the same CNN after retraining for each data set, and (3) trained Baer: the Baer picking algorithm whose parameters have been optimized for each data set.

Pre-processing

First, vertical-channel data were extracted from each station-event pair in the five data sets. Stations in data set D had two vertical channels, one from an L4 geophone and the other from an accelerometer so the channel on which the analyst made the *P*-arrival pick was selected. Next, all waveforms recorded by networks with heterogeneous sampling rates were downsampled to the network's lowest sampling rate. For data set C, the underground stations were downsampled from 5000 to 1000 Hz, and for data set D, the 250-Hz broad-band channels were downsampled to 100 Hz. The waveforms were downsampled rather than upsampled to accommodate the fixed 400 sample window required by the CNN. If the data were upsampled, too much of the initial impulse of the *P* arrival was often lost and the model performance would degrade sharply.

Due to data storage limitations, only triggered waveforms for data set C were archived. These triggered waveforms generally only had a very small number of samples available before the *P* trigger, especially after downsampling, which is problematic for many picking algorithms. For the waveforms that did not have at least 251 samples of pre-pick data, the available pre-pick data were repeated and prepended to the waveform until the 251 sample requirement was met. 15 traces were dropped because they did not contain at least 25 samples of pre-pick data. Alternately, we tried zero padding these waveforms, but it resulted in spurious picks on the first non-zero data point. This process was unique to data set C because there were no gaps or missing data in the other data sets.

Like Ross *et al.* (2018a), we applied a 1–20 Hz bandpass filter to data set E to remove low-frequency noise unrelated to MIS. A 0.5-Hz high-pass filter was applied on data set D rather than the bandpass filter because many of the events had significant energy above the 20-Hz band. The other data sets were not filtered as they consisted of high-frequency (5 Hz+) geophone data.

CNN training

For each data set, the base CNN model was retrained, using the Southern California Seismic Network (SCSN) model's weights as a starting point. The input arrays for training were created through the following procedure, based on that described by Ross *et al.* (2018a). First, the manually processed traces were sliced into 400-sample segments. In order to prevent the model from simply learning where the pick index begins, the data segments were randomly shifted such that the centre of the window was within ± 50 samples of the manual pick. This process was repeated for each trace five times in order to artificially increase the training set size. Next, each trace segment was normalized by the maximum of its absolute value. The input used to train the CNN is a 2-D array composed of the traces (the data) and a 1-D array of integers, which indicate the index of the manual pick (the target). All of the CNN's layers were allowed to update during training using the Huber loss function (Huber 1964) and Adam optimizer (Kingma & Ba 2014) as implemented by the TensorFlow library (Abadi *et al.* 2016). Training occurred for up to 25 epochs (complete passes over the training data) but was terminated early if the validation mean absolute error did not decrease for five consecutive epochs. The weights producing the lowest absolute mean validation error

throughout the entire training process were kept rather than the final weights, which may have started to overfit the training data.

Baer picker training

We used the same training data (including the same pre-processing) to optimize a Baer picker (Baer & Kradolfer 1987). The optimization used the differential evolution optimizer from the SciPy Python package (Storn & Price 1997) to minimize the inverse of the fitness function described by Vassallo *et al.* (2012). The Baer picker tended to pick later than the human analyst, which is common in automated picking algorithms, so we added an optimizable bias parameter to the Baer algorithm to allow all of the picks to shift by a constant value.

Model evaluation

The base and trained CNN, and trained Baer models were evaluated using the test data for each of the five data sets. For the test data, the traces were segmented into 400-sample windows and the analyst pick index was shifted using the same process described above. *P*-wave arrival time predictions from each model were compared to the picks made by the human analyst. Pick time residuals were calculated, and the same four descriptive statistics used by Ross *et al.* (2018a) were used to summarize the residual distributions: the 75th percentile of the absolute value of the pick time residuals ($Q_{|75|}$), the 90th percentile of the absolute value of the residuals ($Q_{|90|}$), the mean (μ) and the standard deviation (σ) of the residuals. Extreme outliers were included in the percentile calculations but removed using the outer fence method before determining the mean and standard deviation. Outlier removal was necessary because, in some cases, analyst and model picks were made on different events included in the input data. While associating phase picks with the correct events is critical for locating seismic sources, it falls outside the scope of this study.

RESULTS

Table 1 shows the number of phase picks used to train and test each data set, and the time required for each using the CPUs of an engineering workstation with 8 CPUs and 64 GBs of RAM. The Baer model optimization was performed on a single CPU thread with no attempt to implement concurrency and, consequently, the optimization time could probably be improved significantly with additional effort. Fig. 5 shows sample waveforms and phase picks from each data set.

Analyst variability

In order to estimate variability among human analysts, the test data from data sets A and B were processed by three analysts in addition to the standard analyst who originally processed the training data. Although the sample size is small, these can be used as an approximation of human-level performance for data sets A and B. Table 2 shows the various statistics for each analyst combination, and Fig. 6 shows the combined residual distribution.

Data set A

Fig. 7 shows histograms of the model-analyst pick residuals for data set A. The trained Baer model outperforms performs the base CNN model but falls short of the trained CNN model. The trained Baer and trained CNN evaluation statistics fall below the human variability measurements in all metrics. Effectively, this means uniformly applying the automated pickers results in lower pick time residuals than when multiple human analysts make *P* picks.

Data set B

The trained CNN model's weights and trained Baer model's parameters from data set A were used on data set B rather than retraining both models, because the networks were similar and located close together. Fig. 8 shows the residuals between each model and the analyst picks. For both the trained Baer and trained CNN models, the results are comparable to those of data set A, demonstrating that these models transfer well to this nearby network without retraining. Interestingly, although the trained CNN model was trained on data set A, it performs better on data set B for all metrics.

Data set C

Fig. 9 shows the residuals for data set C. The residuals in samples are higher, but due to the higher sampling rates in this data set (1000 sps), the temporal residuals are not as stark. Moreover, the waveforms recorded by this network, particularly for the underground stations, tended to be noisier and more difficult to pick. As mentioned previously, there were also some minor quality issues identified in the manually processed data. Around 15 per cent of the picks were 10 samples or greater away from the true *P*-wave arrival. The trained CNN clearly outperforms the Baer picker when presented with these challenges.

Data set D

Fig. 10 shows the pick residuals for data set D. In this case, the base CNN performs better than it did in the other data sets. This could be due to the sensors and network spacing being closer to those of the SCSN, although this was not the case for data set E. As with the other data sets, the trained CNN significantly outperformed the trained Baer model on this network.

Data set E

Fig. 11 shows the pick residuals for data set E. The trained CNN falls significantly short of the performance reported of the trained CNNs for the other data sets. We re-examined a subset of events and found no obvious quality issues (as were found for data set C). Possible reasons for both models' poor performance are covered in the Discussion section.

DISCUSSION

The trained CNN performed better than the trained Baer model on all data sets. Both models performed within levels of measured human variance for all but one of the evaluation metrics for data sets A and B. From an operational sense, both models probably perform

'well enough' to produce meaningful event locations. For example, when using picks from either model to locate events, 75 per cent of the events from data sets A and B locate within 16 m of the location resulting from manual *P*-arrival picking. It would be interesting to include many more optimized picking algorithms, trained on the same input data in the comparison.

Model transferability

The base CNN did not perform adequately for any of the data sets but was greatly improved through retraining. In order to quantify the benefits gleaned by transfer learning, and to determine how much data are required to adequately retrain the CNN starting with the SCSN weights, we explored several training restrictions using a variable number of seismic traces for training on data set A (Fig. 12). We found the following:

1. The improvements in training drop off sharply for both the Baer and the CNN around 200 traces and the improvements start to level off around 5000 traces, although minor improvements probably continue past the largest test data set of 10 000 traces.
2. We see no benefits to only allowing the outer (non-convolutional) layers to update during training.
3. A CNN with no starting weights achieves the same mean absolute error as the Baer picker once the test data set reaches about 5000 traces.

For an operator of a similar (local) network deployed in/around a coal mine, the first finding has practical significance; with only 5000 manually processed traces, a CNN model can be trained to pick P-arrival times with acceptable performance. Admittedly, there will still need to be human review of phase picks, particularly on events with high location residuals whose traces tend to contain multiple events, but the analyst workload would be greatly reduced compared to fully manual processing workflows.

The poor performance of the base model on all data sets is not surprising considering the significant differences between the MIS data sets and tectonic seismicity recorded by the SCSN, and certainly does not represent a deficiency of the original work. However, the CNNs failure to extrapolate to new types of seismicity and networks clearly demonstrates the CNN has not internalized the *general* task of phase picking as a human analyst might. Interestingly, the base model performed the worst for data set E (UUSS). This is unexpected considering that, of all the data sets examined in this study, UUSS is the most like SCSN in terms of instrumentation type, station spacing and source–receiver distance. However, coal mine seismicity from this area recorded at regional distances produces very distinct waveforms compared to typical tectonic earthquakes. For instance, coal seismicity tends to be much shallower and have smaller stress drops resulting in waveforms which are deficient in higher frequencies (Stein 2016). Comparing the effect of waveform frequency on model performance across all data sets is difficult given that the sampling frequencies are different between data sets. However, since the models are provided only with waveforms, they have no concept of absolute sampling rate so *a relative* sampling frequency can be assumed. Imposing a sampling frequency of 100 Hz on each data set does show a strong

relation between dominant relative frequency of a smoothed amplitude spectra and the absolute value of the base-model residuals (Fig. 13). This effect is likely due to two factors. First, the base CNN was trained mainly on waveforms with higher relative frequencies, so the model would expectedly perform worse when presented with lower relative frequency waveforms. Second, lower relative frequency waveforms are more emergent and thus harder to unambiguously pick. Frequency-related picking difficulty may explain why, even after training, the models performed worse on the data sets with lower dominant relative frequencies (data sets C and E) than those with higher dominant relative frequencies.

After retraining the base CNN, we attempted to quantify the performance degradation for picking on the original SCSN test data. If little or no degradation occurred, it would mean creating a general picker, one that would perform well on a wide variety of network and event types, could be possible with this CNN architecture. When using the CNN trained on data sets A and C, the pick time residuals for the SCSN data had standard deviations around 200 per cent and 120 per cent higher than the base CNN, while the standard deviation using the CNN trained on data set D was only around 30 per cent higher. Unfortunately, the degradations indicate that the CNN is somewhat network/training data dependent and cannot be generally applied without some retraining. However, as evidenced from the excellent transferability from data sets A to B, it may be possible to generate a small number of trained models to select from based on network and waveform characteristics.

Processing improvements

The original motivation for this research was to process the entirety of the data sets A and B deployments (36 012 events in total). We used less than 1 per cent of the data in order to train, test and evaluate the different models. We then processed the remaining events with the trained CNN from data set A in conjunction with a simple moving window scheme, which took around 5 d. Had a human analyst processed the same amount of data it would have taken approximately two years, assuming a 40-hr work week. Fig. 14 shows maps of the locations of all the events in data set A, using the picks made by an unoptimized Baer picker (the previous processing method) and using the picks made by the trained CNN. Clearly, the increase in pick quality had a significant impact on location accuracy and interpretability of seismicity.

Future work

Expanding this type of study to include additional picking algorithms, including CNN pickers which return CFs, and perhaps combining various pickers in concert, would be an interesting line of future research. A high-quality, open-source package which facilitates these types of studies through a unified Application Programming Interface would be a boon to both network operators and seismology researchers. A larger, more statistically rigorous effort to quantify the variability between human analysts accounting for network geometry and type, phase and experience levels would provide important benchmarks for assessing the performance of future automated phase-picking, detection and classification models.

In the near future, we expect neural network-based models will adequately perform the simpler tasks currently performed by seismic analysts. However, analysts will still be needed to provide over-sight, quality assurance and to process particularly unusual signals.

CONCLUSIONS

We have shown that a CNN trained on millions of regionally recorded earthquake traces to estimate *P*-wave arrival times does not adequately transfer to networks of different scales monitoring MIS in other geographical regions. However, the CNN can be retrained to effectively pick MIS recorded by local networks of varying sampling rates and instrument types using a surprisingly small amount of training data. Initializing training with the SCSN model weights greatly improves the resulting model's performance compared to training from scratch for a given training data set size. Applying the CNN on a regional MIS data set yielded limited success, likely due to differences in waveform frequency content.

We also demonstrated that the retrained CNN is superior to the Baer picking algorithm optimized on the same training data. Properly tuning any phase picker to a specific data set, however, remains an important consideration. Both the optimized Baer picker and the trained CNN model exhibit less variance in pick time residuals than the variance observed between human analysts. The application of improved phase picking models has the potential to reduce the time, cost and manual intervention required to extract actionable information from MIS. This will make it easier for ground-control experts at mines to understand the rock mass response to mining and more effectively detect and address certain types of stability issues.

ACKNOWLEDGEMENTS

This work would not be possible without the published models of Ross *et al.* (2018a), and we sincerely thank them for the extra effort expended to package and publish their models. We are grateful for the cooperation of several coal mines who wish to remain anonymous and the help from Relu Burlacu in extracting UUSS phase data. SWJ wrote the software to perform and evaluate phase picks. DJAC provided software development guidance and codes for integrating and analysing the various data sources. MSB located and plotted the seismic events. All three NIOSH authors collaboratively developed this manuscript and performed manual phase picking. KDK facilitated collaborations between NIOSH and UUSS and provided thoughtful discussions which helped direct the research.

REFERENCES

Abadi M et al. , 2016. Tensorflow: a system for large-scale machine learning, in 12th USENIX Symposium on Operating Systems Design and Implementation, Vol. 16, USENIX Association, pp. 265–283.

Arabasz WJ, Nava SJ, McCarter MK, Pankow KL, Pechmann JC, Ake J & McGarr A, 2005. Coal-mining seismicity and ground-shaking hazard: A case study in the Trail Mountain area, Emery County, Utah, Bull. seism. Soc. Am, 92(1), 18–30.

Arabasz WJ & Pechmann JC, 2001. Seismic Characterization of Coal-Mining Seismicity in Utah for CTBT Monitoring (No. UCRL-CR-143772), Lawrence Livermore National Laboratory, Livermore, CA (US).

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

Baillard C, Crawford WC, Ballu V, Hibert C & Mangeney A, 2013. An automatic kurtosis-based P-and S-phase picker designed for local seismic networks, Bull. seism. Soc. Am, 104(1), 394–409.

Bergen KJ, Johnson PA, de Hoop MV & Berzoa GC, 2019. Machine learning for data-driven discovery in solid Earth geoscience, Science, 363(6433), doi:10.1126/science.aau0323.

Boltz MS, Chambers DJA & Swanson PL, 2016. Effects of a three-dimensional velocity structure on the locations of coal mining-induced seismicity, in 50th US Rock Mechanics/Geomechanics Symposium. American Rock Mechanics Association.

Chambers D, 2020. Coal Mining Induced Seismicity Dataset, Harvard Dataverse, version 3, doi: 10.7910/DVN/5DGFJB.

Collins DS, Pinnock I, Toya Y, Shumila V & Trifu CI, 2014. Seismic event location and source mechanism accounting for complex geology and voids, in 48th US Rock Mechanics/Geomechanics Symposium. American Rock Mechanics Association.

Czarny R, Marcak H, Nakata N, Pilecki Z & Isakow Z, 2016. Monitoring velocity changes caused by underground coal mining using seismic noise, Pure appl. Geophys, 173(6), 1907–1916.

Dai H & MacBeth C, 1995. Automatic picking of seismic arrivals in local earthquake data using an artificial neural network, Geophys. J. Int, 120(3), 758–774.

Dai H & MacBeth C, 1997. The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings, J. geophys. Res.: Solid Earth, 102(B7), 15105–15113.

Friberg P, Lisowski S, Dricker I & Hellman S, 2010. Earthworm in the 21st century, in EGU General Assembly Conference, 12. Abstracts 12.12654.

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

Huang L, Li J, Hao H & Li X, 2018. Micro-seismic event detection and location in underground mines by using Convolutional Neural Networks (CNN) and deep learning, Tunnelling Underground Space Technol, 81, 265–276.

Huber PJ, 1964. Robust estimation of a location parameter, Ann. Math. Stats, 35(1), 73–101.

Kingma DP & Ba J, 2014. Adam: a method for stochastic optimization, a rXiv:1412.6980.

Kong Q, Trugman DT, Ross ZE, Bianco MJ, Meade BJ & Gerstoft P, 2018. Machine learning in seismology: turning data into insights, Seismol. Res. Lett, 90(1), 3–14.

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

LeCun Y & Bengio Y, 1995. Convolutional networks for images, speech, and time series, in The Handbook of Brain Theory and Neural Networks, Vol. 3361(10). MIT Press: Cambridge, MA.

Lin B, Wei X & Junjie Z 2019. Automatic recognition and classification of multi-channel microseismic waveform based on DCNN and SVM, Comp. Geosci, 123, 111–120.

Mendecki AJ, Lynch RA & Malovichko DA, 2010. Routine microseismic monitoring in mines, in Australian Earthquake Engineering Soc., Annual Conference, editor: Mccue K, Perth, Australia, pp. 1–33.

Mousavi SM & Beroza GC, 2019. A machine-learning approach for earthquake magnitude estimation, Geophys. Res. Lett, 47, doi:10.1029/2019GL085976.

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

Perol T, Gharbi M & Denolle M, 2018. Convolutional neural network for earthquake detection and location, Sci. Adv, 4(2), e1700578, doi:10.1126/sciadv.1700578. [PubMed: 29487899]

Ross ZE, Meier MA & Hauksson E, 2018a. P wave arrival picking and first-motion polarity determination with deep learning, J. geophys. Res.: Solid Earth, 123(6), 5120–5129.

Ross ZE, Meier MA, Hauksson E & Heaton TH, 2018b. Generalized seismic phase detection with deep learning, Bull. seism. Soc. Am, 108(5A), 2894–2901.

Ross ZE, Yue Y, Meier MA, Hauksson E & Heaton TH, 2019. PhaseLink: a deep learning approach to seismic phase association, J. geophys. Res.: Solid Earth, 124(1), 856–869.

Stein JR, 2016. Seismic source discrimination in the Wasatch Plateau Region of central Utah, MS thesis, The University of Utah.

Storn R & Price K, 1997. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, J. Glob. Optim, 11(4), 341–359.

Swanson P, Boltz MS & Chambers D, 2016. Seismic Monitoring Strategies for Deep Longwall Coal Mines, National Institute for Occupational Safety and Health, Washington, DC (US).

Vassallo M, Satriano C & Lomax A, 2012. Automatic picker developments and optimization: a strategy for improving the performances of automatic phase pickers, Seismol. Res. Lett, 83(3), 541–554.

Wilkins AH, Strange A, Duan Y & Luo X, 2020. Identifying microseismic events in a mining scenario using a convolutional neural network, Comp. Geosci, 137, 104418. doi: 10.1016/j.cageo.2020.104418.

Withers M, Aster R, Young C, Beiriger J, Harris M, Moore S & Trujillo J, 1998. A comparison of select trigger algorithms for automated global seismic phase and event detection, Bull. seism. Soc. Am, 88(1), 95–106.

Woollam J, Rietbrock A, Bueno A & De Angelis S, 2019. Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network, Seismol. Res. Lett, 90(2A), 491–502.

Zhu W & Beroza GC, 2018. PhaseNet: a deep-neural-network-based seismic arrival-time picking method, Geophys. J. Int, 216(1), 261–273.

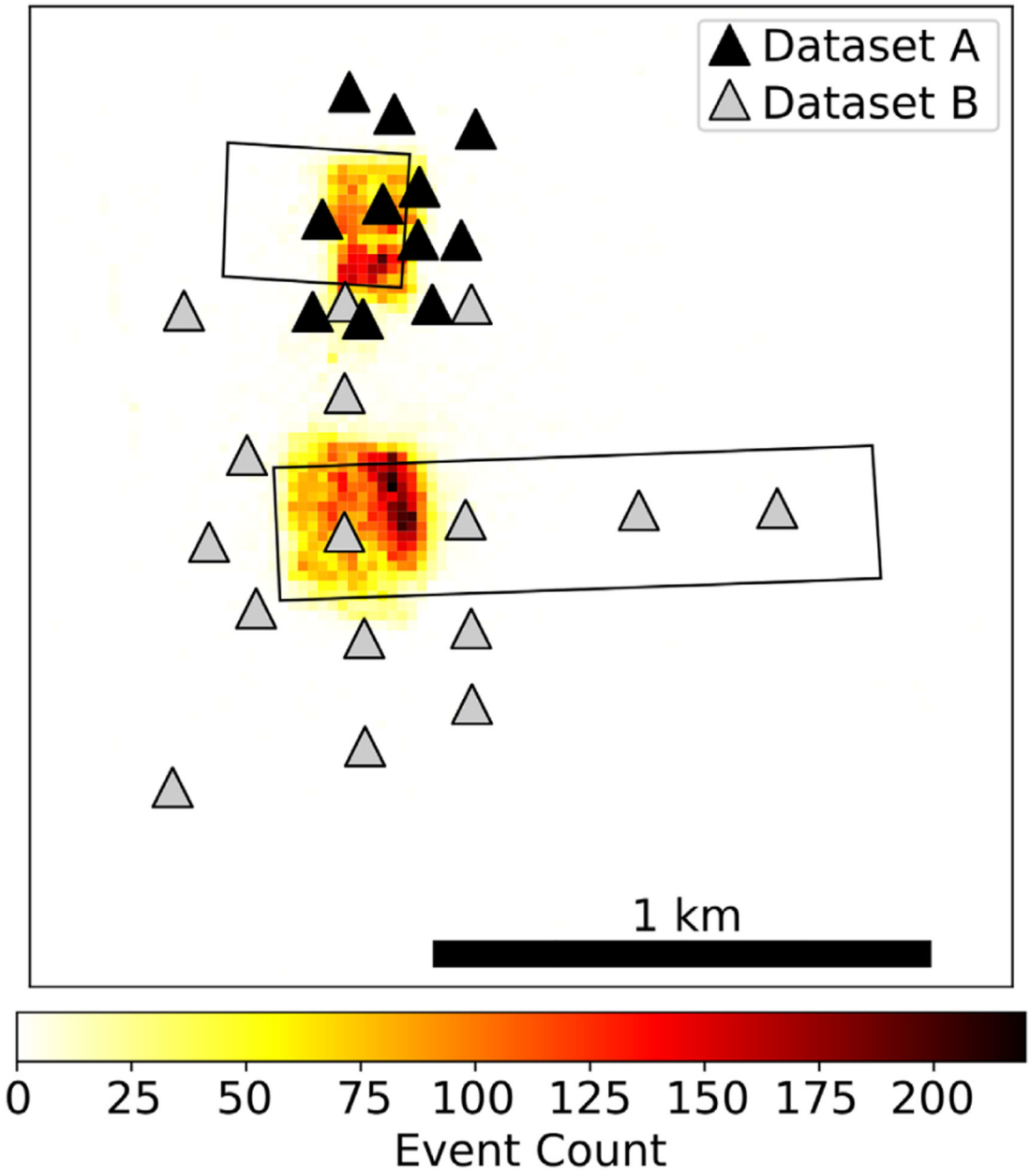

Figure 1.
Plan-view plot of data sets A and B. Triangles represent the locations of nodes, and rectangles outline longwall panels. The shading denotes the number of events that occurred within a 400-m^2 area.

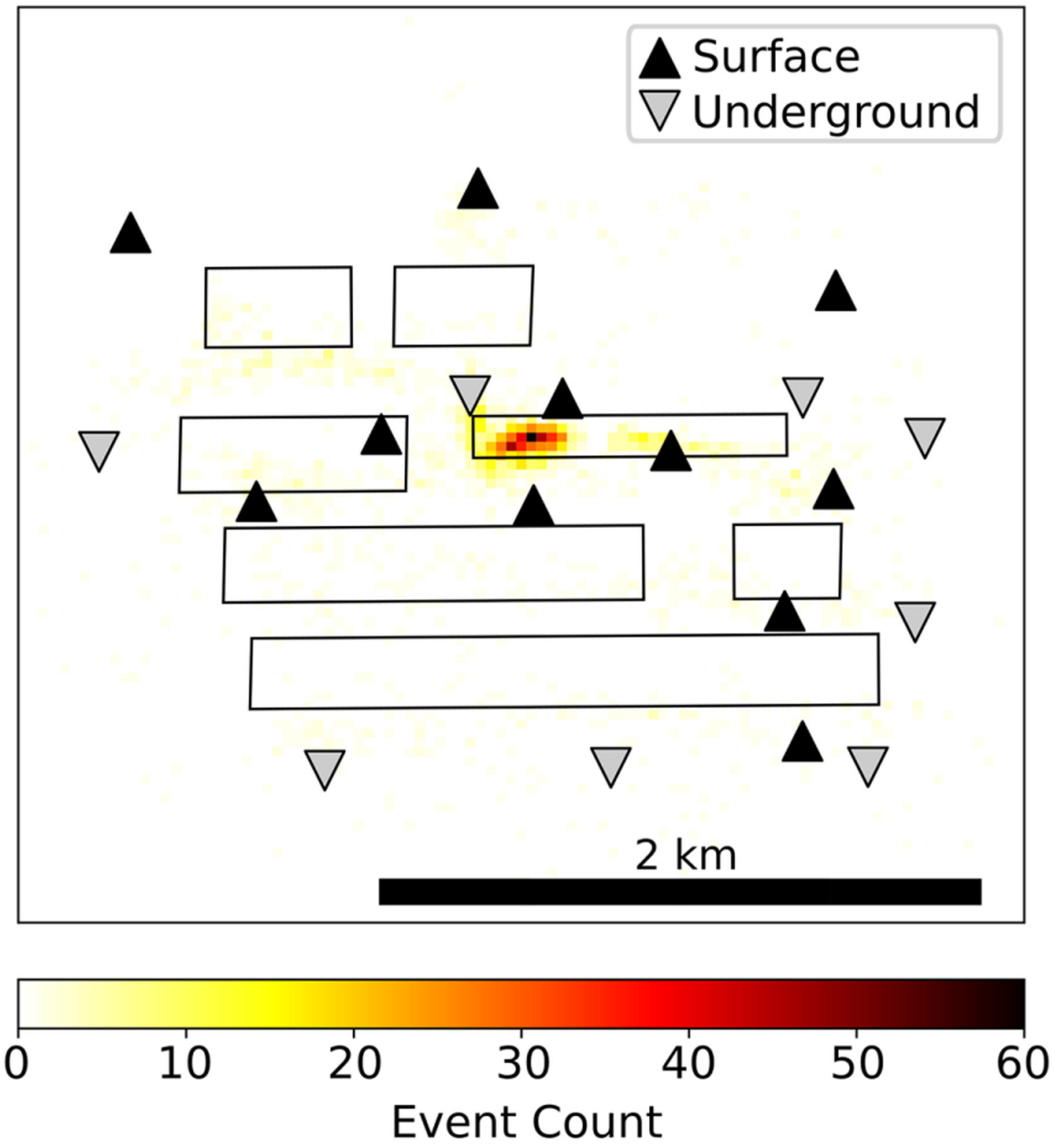

Figure 2.

Plan-view plot of data set C. Triangles indicate the location of surface sensors, while inverted triangles demarcate underground sensors. The rectangles outline the longwall panels, and shading denotes the number of events that occurred within a 1050-m^2 area.

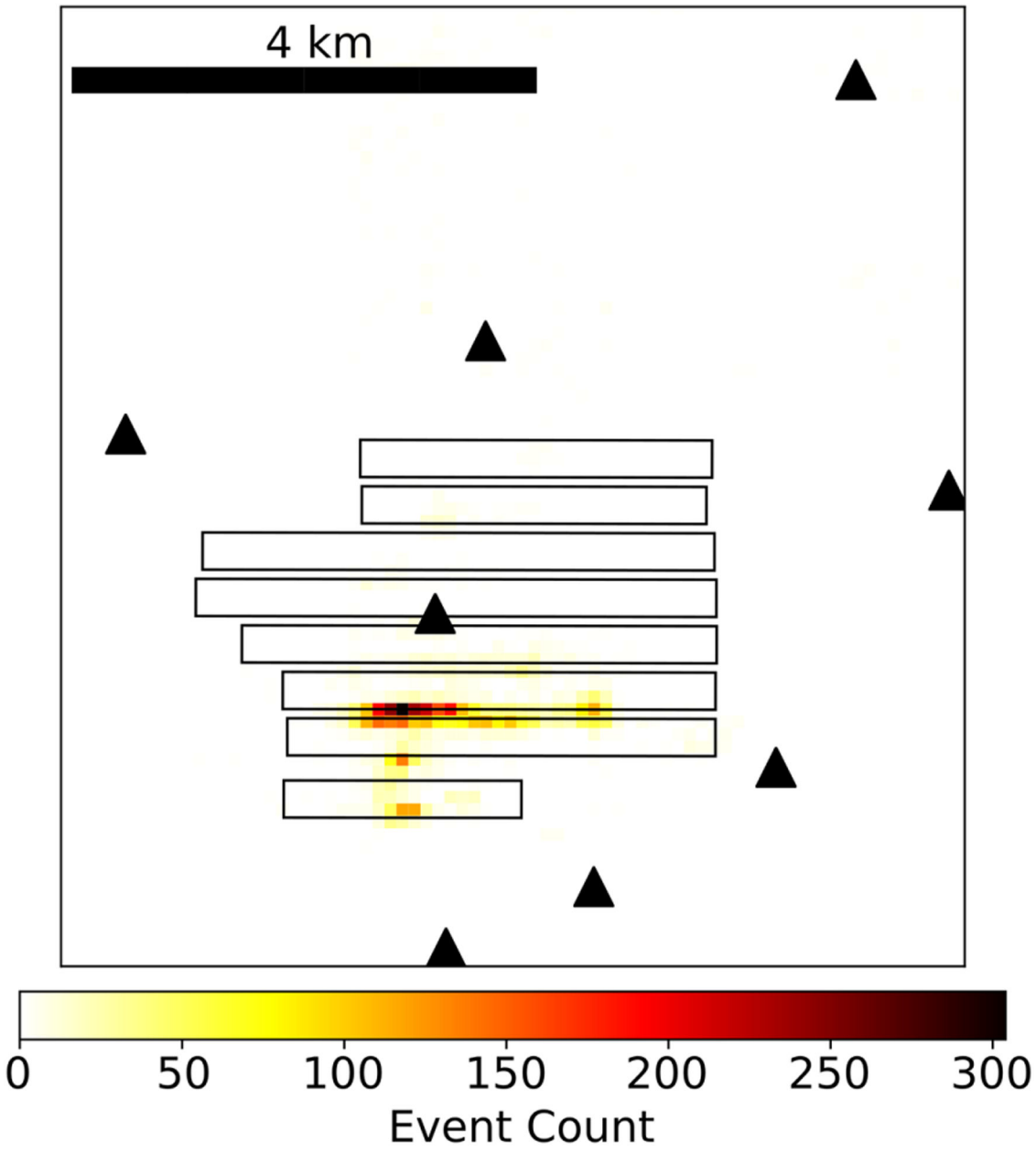

Figure 3.
Plan-view plot of data set D. Black triangles represent the locations of sensors, and rectangles outline the longwall panels. The shading denotes the number of events that occurred within a 11 500-m^2 area.

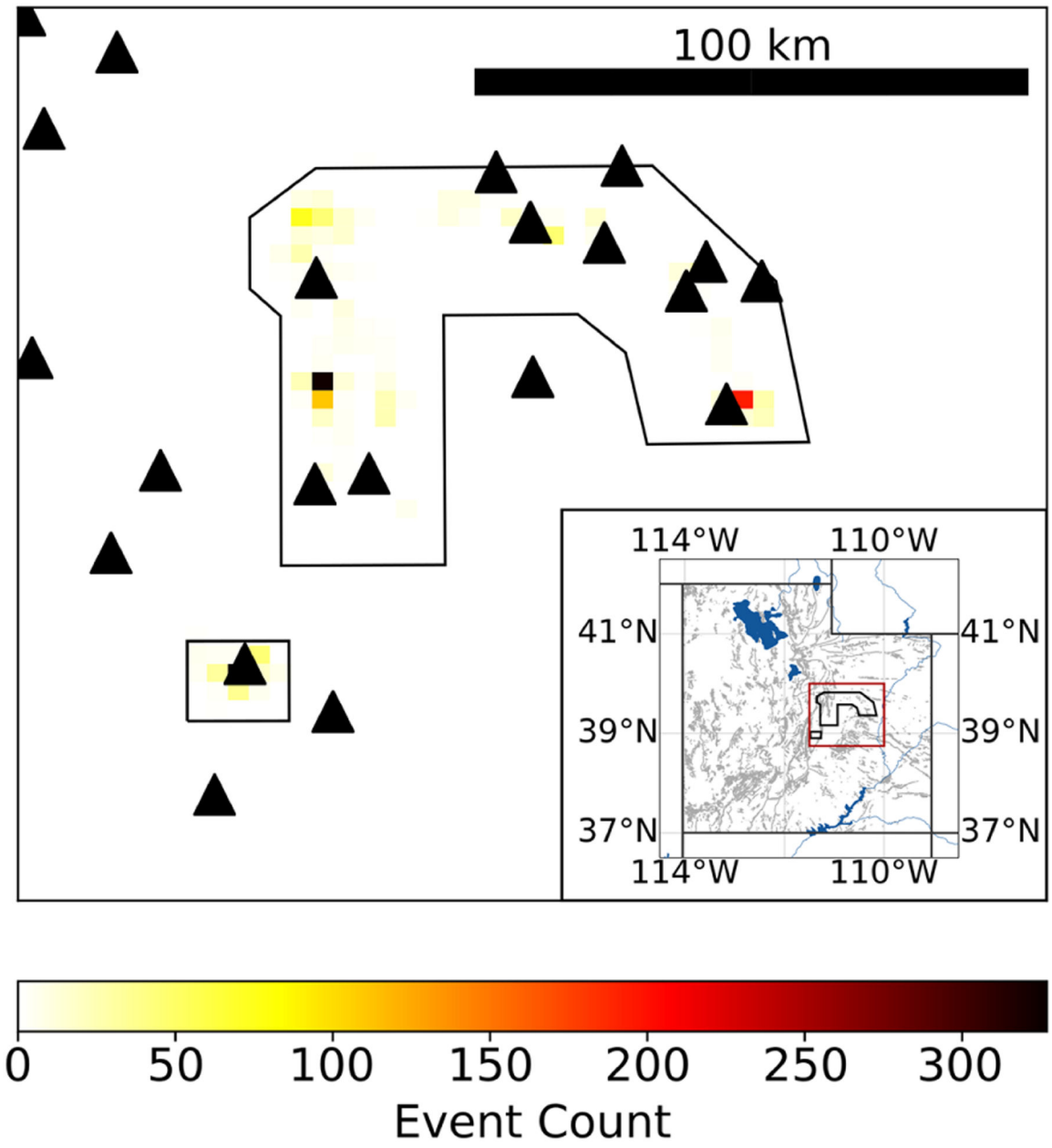

Figure 4.
Plan-view plot of data set E. Black triangles mark the locations of UUSS sensors. The dark lines delineate the coal mining regions. The inset shows the region's location within a map of the state of Utah, USA.

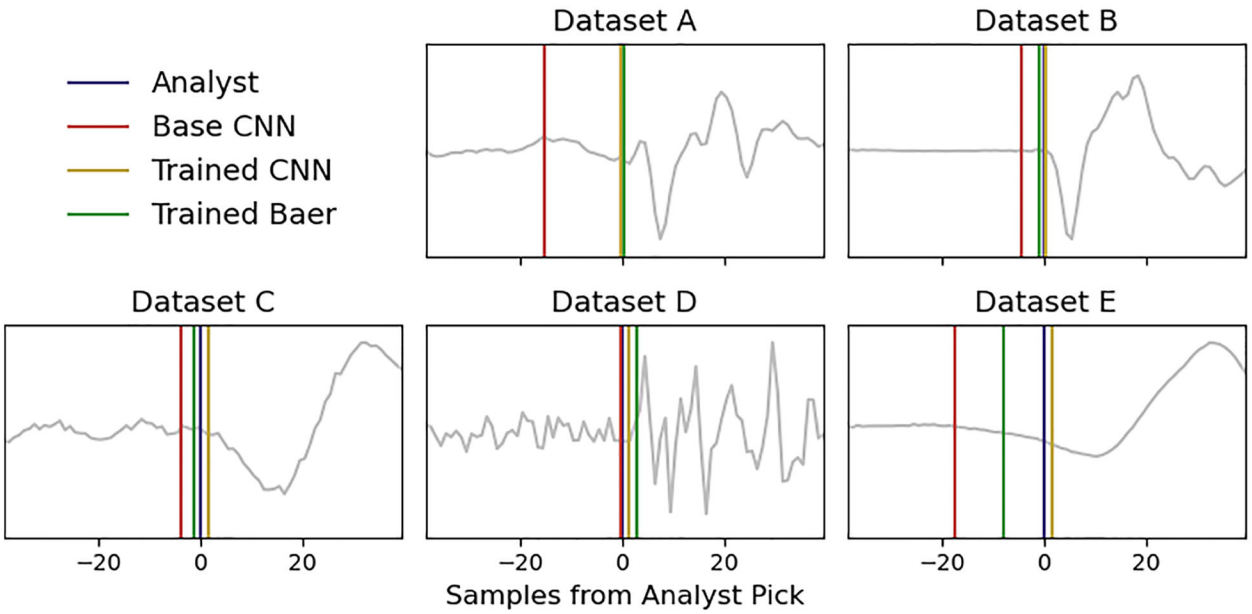

Figure 5.
Zoomed-in sample traces from each data set showing picks from each model/analyst.

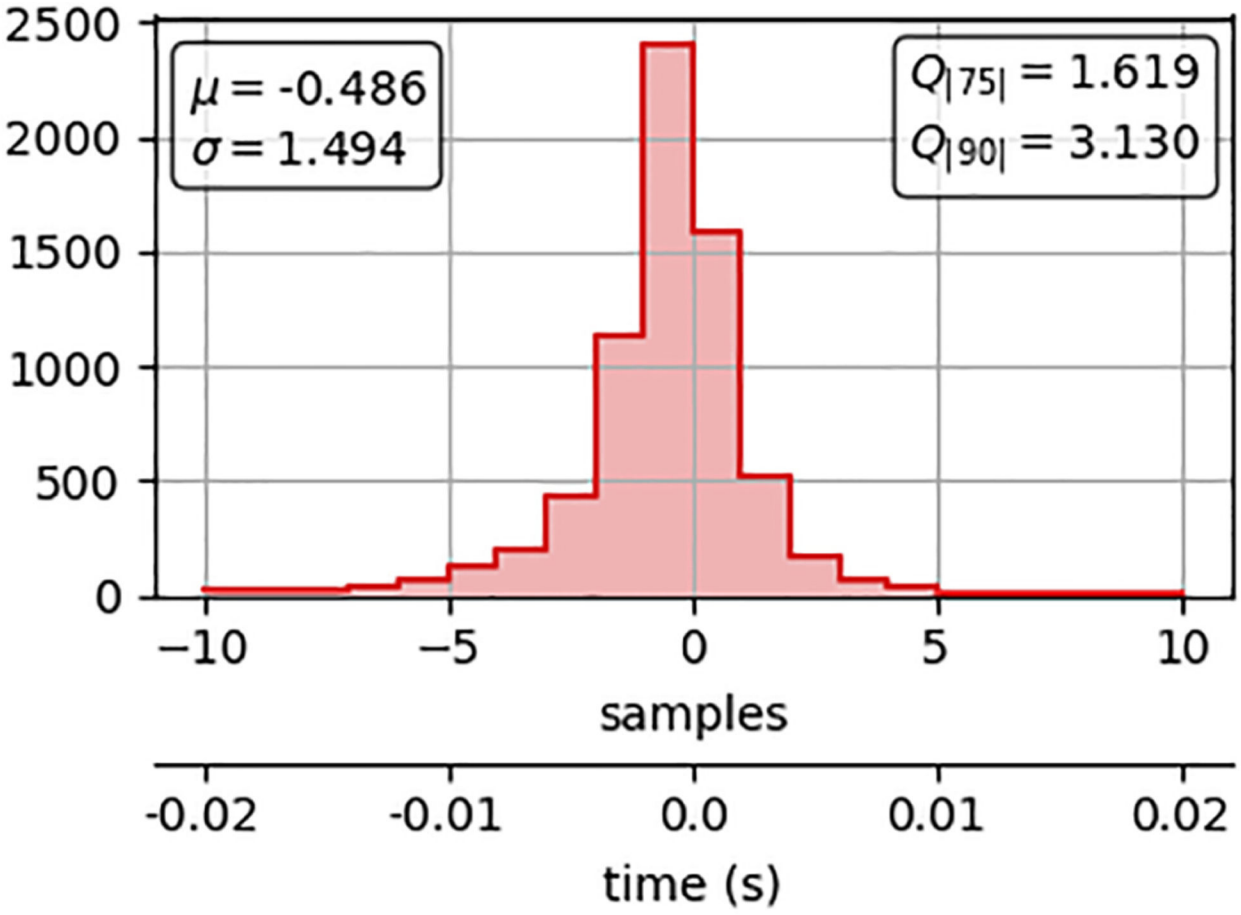

Figure 6.

Summation of residuals between each analyst pair for the test sets of data sets A and B. Residual statistics of mean (μ), standard deviation (σ) and 75th and 90th percentiles of the absolute value of the residuals ($Q_{|75|}$ and $Q_{|90|}$), are shown in samples.

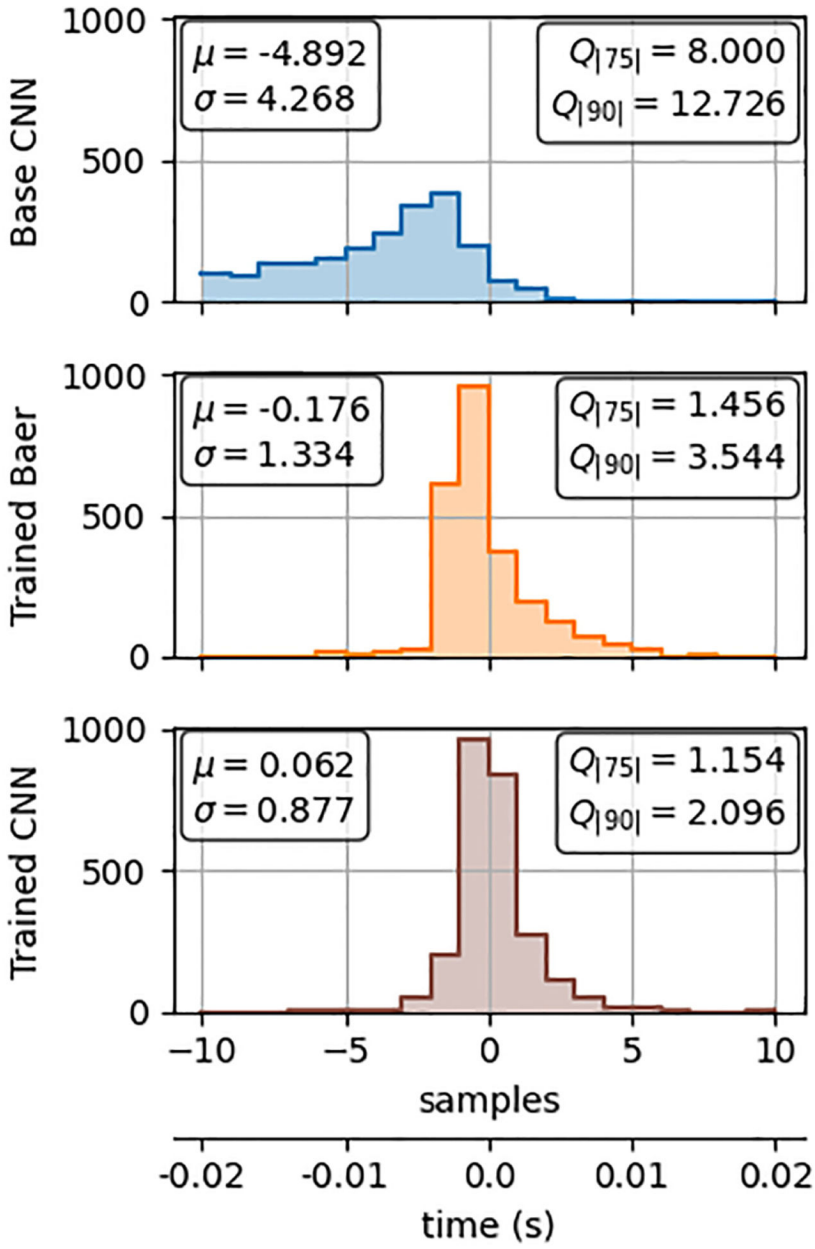

Figure 7.

Comparison of each model pick to the analyst's picks for the data set A test set. Statistics are shown in samples. Residual statistics of mean (μ), standard deviation (σ) and 75th and 90th percentiles of the absolute value of the residuals ($Q_{|75|}$ and $Q_{|90|}$), are shown in samples.

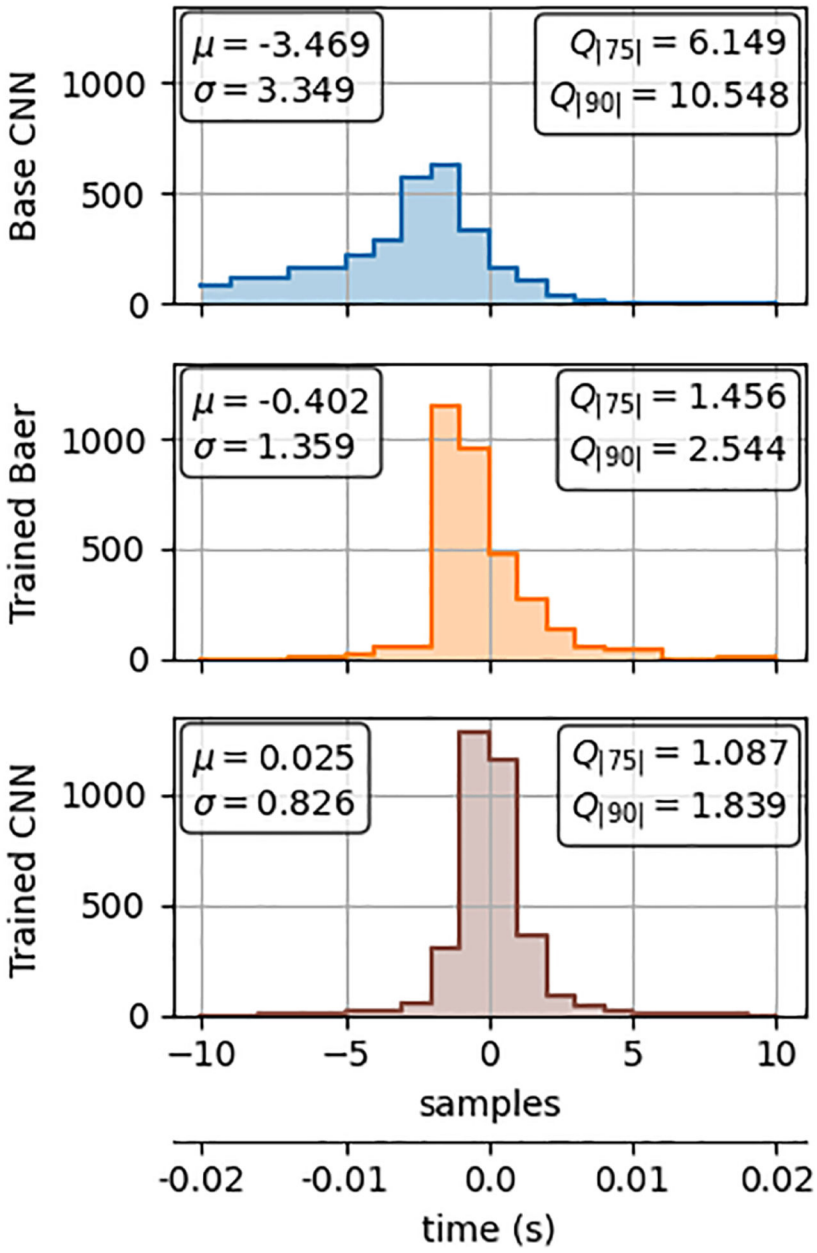

Figure 8.

Comparison of each model pick to the analyst's picks for the data set B test set. Statistics are shown in samples. Residual statistics of mean (μ), standard deviation (σ) and 75th and 90th percentiles of the absolute value of the residuals ($Q_{|75|}$ and $Q_{|90|}$), are shown in samples.

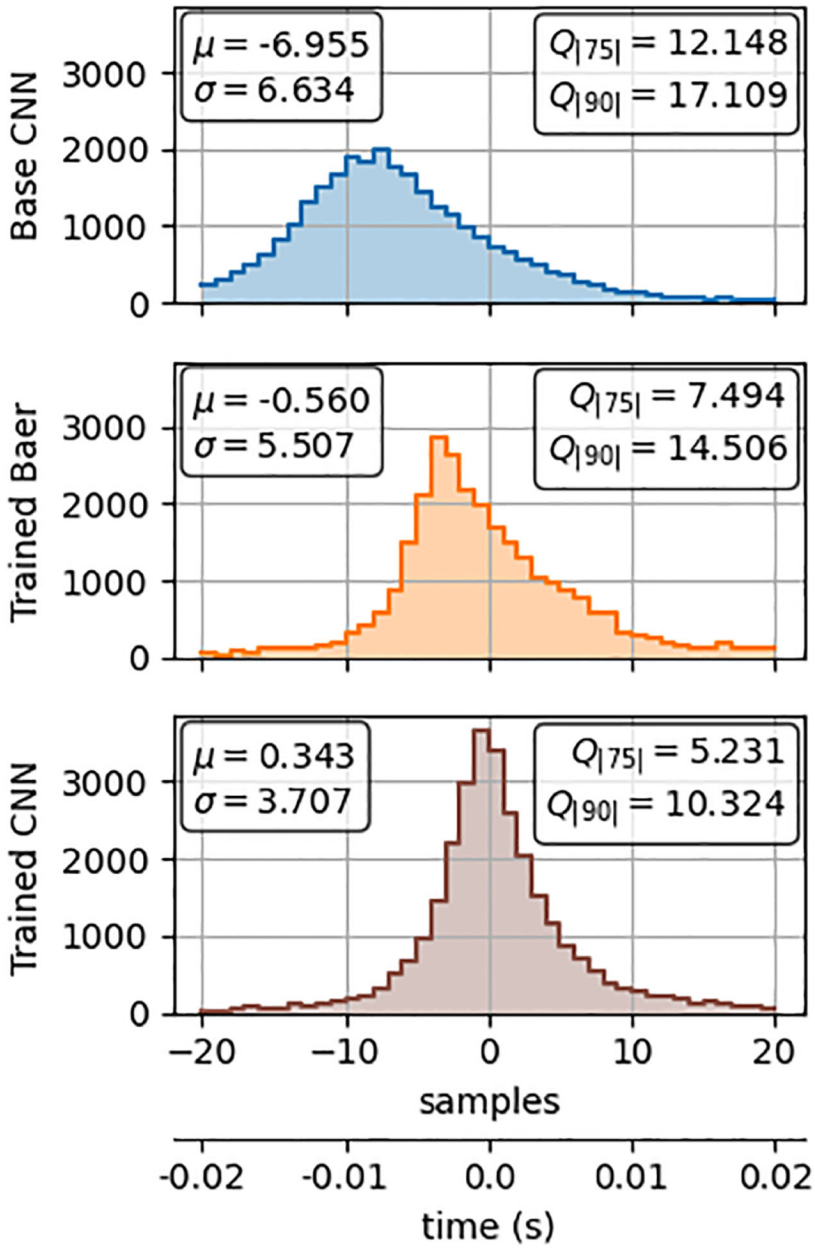

Figure 9.

Comparison of each model pick to the analyst's picks for the data set C test set. Statistics are shown in samples. Residual statistics of mean (μ), standard deviation (σ) and 75th and 90th percentiles of the absolute value of the residuals ($Q_{|75|}$ and $Q_{|90|}$), are shown in samples.

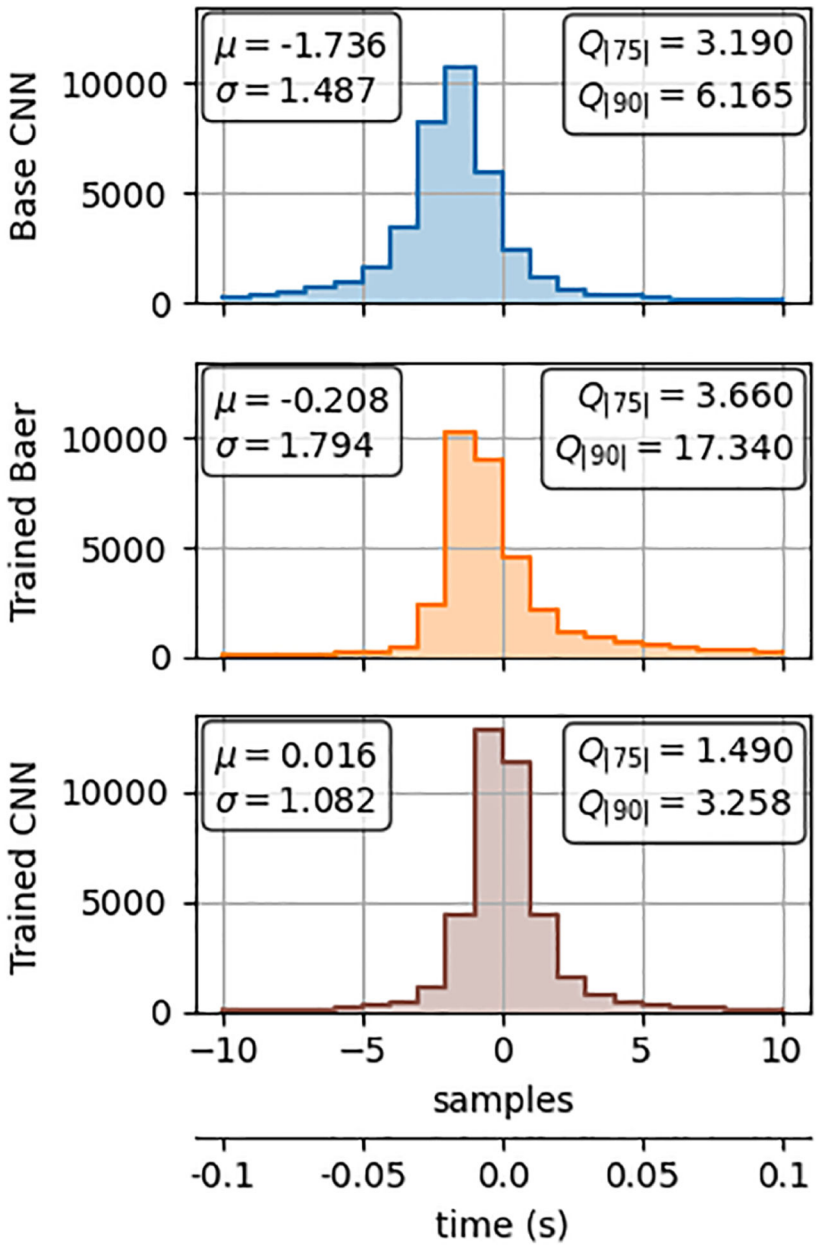

Figure 10.

Comparison of each model's picks to the analyst's picks for the data set D test set. Statistics are shown in samples. Residual statistics of mean (μ), standard deviation (σ) and 75th and 90th percentiles of the absolute value of the residuals ($Q_{|75|}$ and $Q_{|90|}$), are shown in samples.

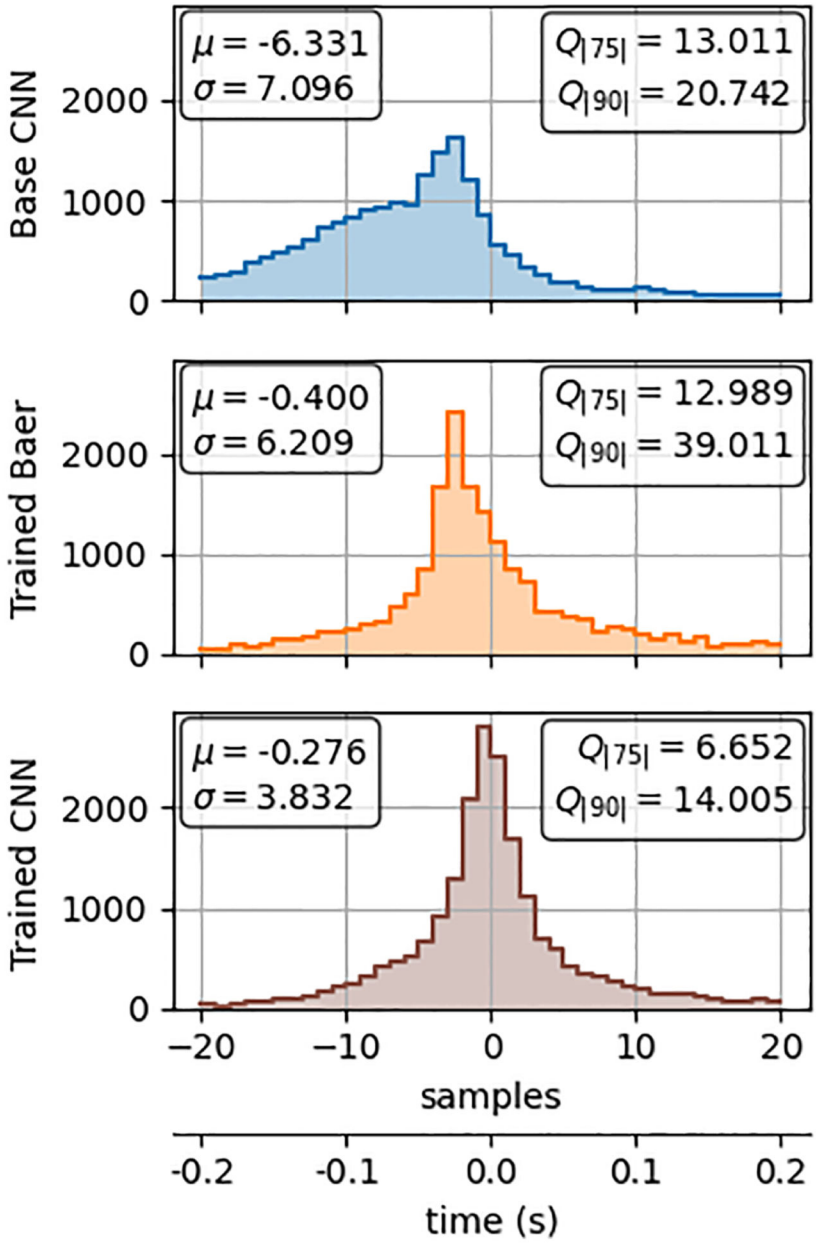

Figure 11.

Comparison of each model's picks to the analyst's picks for the data set E test set. Statistics are shown in samples. Residual statistics of mean (μ), standard deviation (σ) and 75th and 90th percentiles of the absolute value of the residuals ($Q_{|75|}$ and $Q_{|90|}$), are shown in samples.

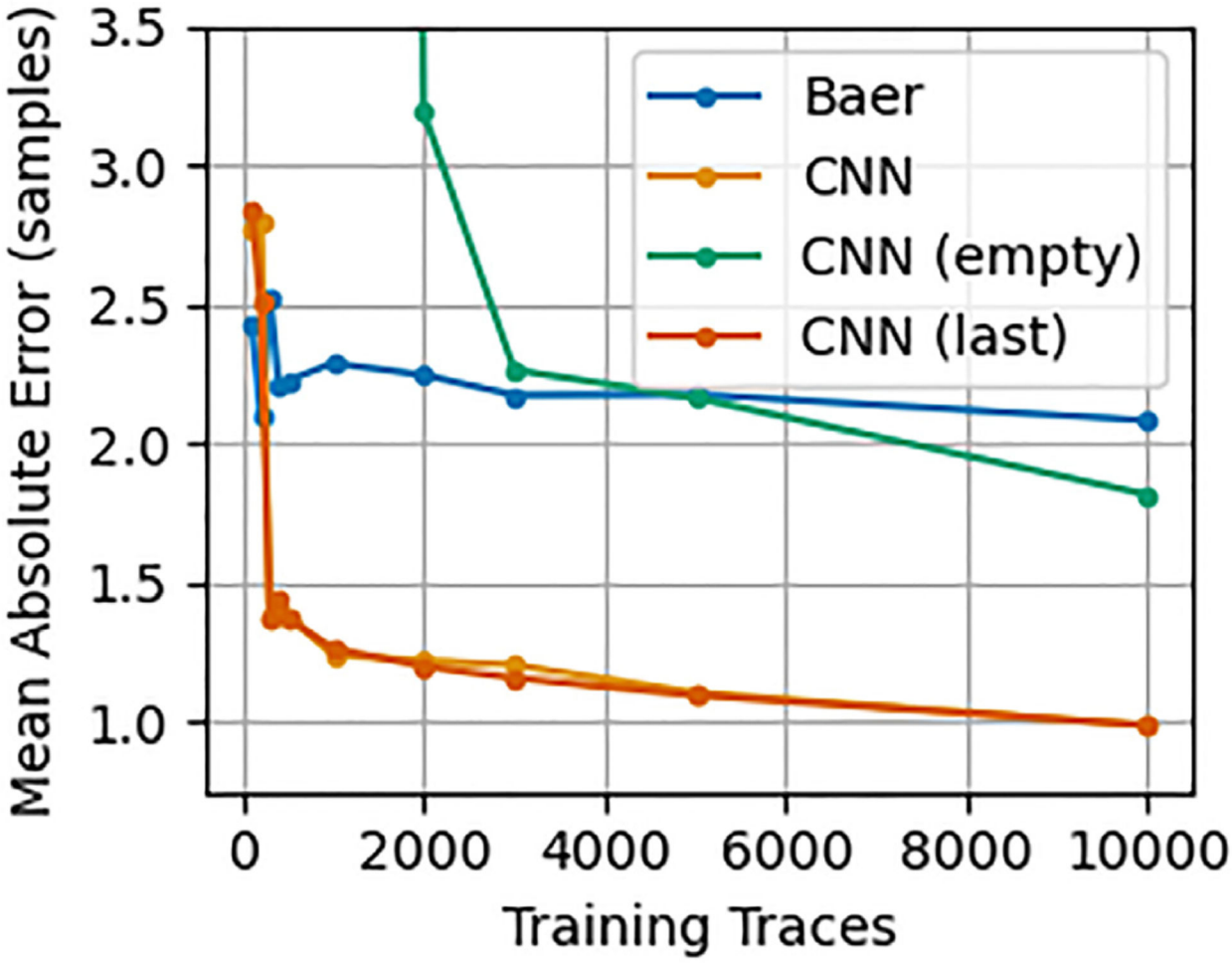

Figure 12.

Mean absolute error of pick time residuals for various models trained on differing numbers of traces for data set A. Baer is the trained Baer picker, CNN is the base CNN (which starts with the SCSN weights), CNN (empty) starts with random weights, and CNN (last) is the base CNN but only the last three layers of the network are allowed to update during training.

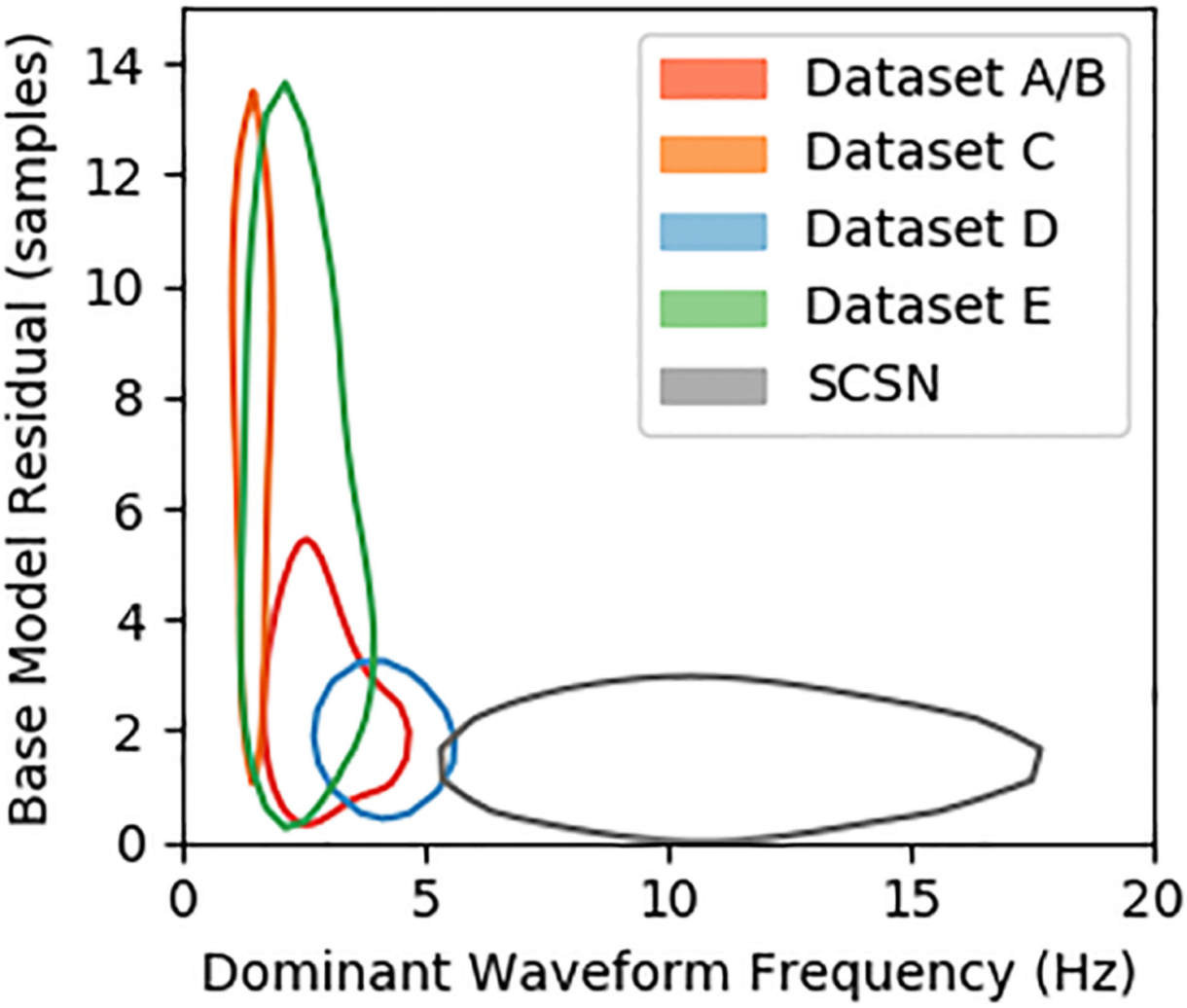

Figure 13.

Dominant frequency, assuming a nominal sampling rate of 100 Hz versus the absolute value of the base model residuals. A one-bin Kernel Density Estimate is shown for each data set.

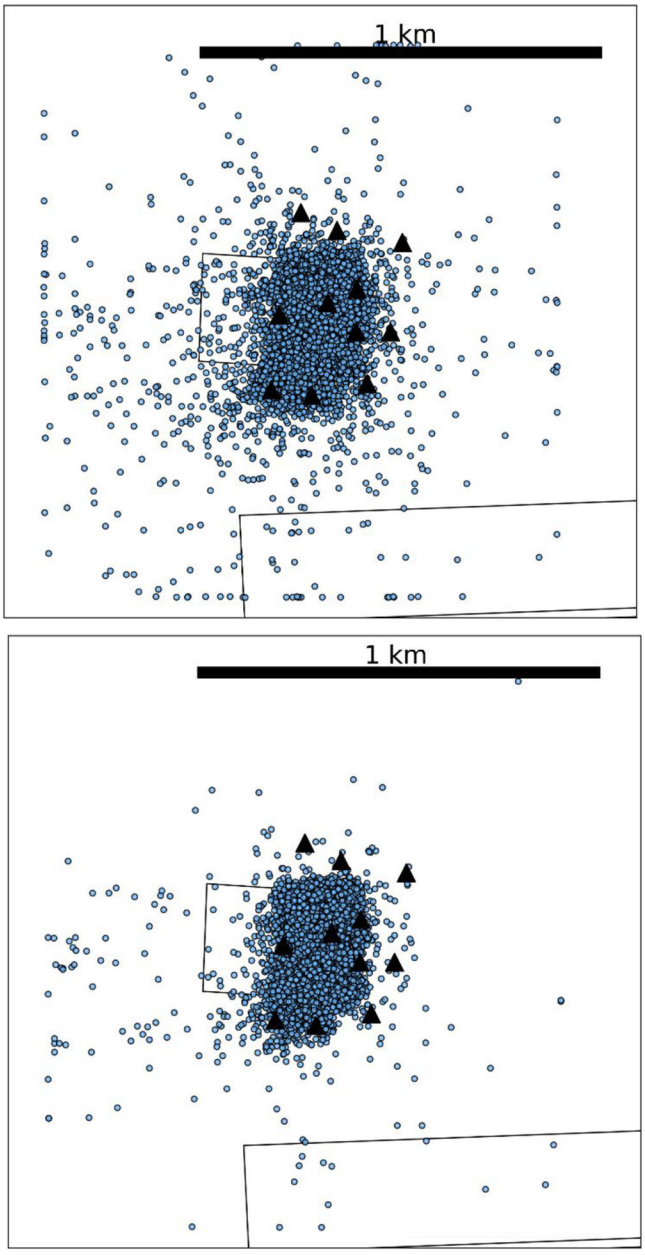

Figure 14.
The entirety of data set A: (top) shows the events located using picks made by the original, unoptimized Baer picker, and (bottom) shows the events located using picks made by the trained model.

Table 1.

The training and testing time of different models. Data set B has no training values because we simply used the models trained on data set A.

		Training				Testing	
Data set	**Number of training traces**	**CNN time (min)**	**CNN epochs**	**Baer time (min)**	**Number of test traces**	**CNN time (s)**	**Baer time (s)**
A	12527	17.52	14	39.52	514	0.87	0.14
B	–	–	–	–	677	1.11	0.10
C	17937	14.43	5	65.87	5981	9.10	0.94
D	23990	25.10	8	92.72	7997	12.20	1.28
E	12693	10.36	5	47.67	4231	6.48	0.63

Table 2.

Analyst comparisons residuals. All statistics are in samples.

Analyst 1	Analyst 2	μ	σ	$Q_{\|75\|}$	$Q_{\|90\|}$
Standard analyst	Analyst A	− 1.120	1.235	1.785	3.169
Standard analyst	Analyst B	− 1.055	1.332	1.937	3.119
Standard analyst	Analyst C	− 1.020	1.468	1.826	3.904
Analyst A	Analyst B	0.143	1.266	1.330	2.594
Analyst A	Analyst C	0.186	1.186	1.372	2.629
Analyst B	Analyst C	− 0.053	1.515	1.616	3.293
Mean statistics		− 0.487	1.645	1.334	3.118