

Convolutional Neural Network for Seismic Phase Classification, Performance Demonstration over a Local Seismic Network

by Jack Woollam, Andreas Rietbrock, Angel Bueno, and Silvio De Angelis

ABSTRACT

Over the past two decades, the amount of available seismic data has increased significantly, fueling the need for automatic processing to use the vast amount of information contained in such data sets. Detecting seismicity in temporary aftershock networks is one important example that has become a huge challenge because of the high seismicity rate and dense station coverage. Additionally, the need for highly accurate earthquake locations to distinguish between different competing physical processes during the postseismic period demands even more accurate arrival-time estimates of seismic phase. Here, we present a convolutional neural network (CNN) for classifying seismic phase onsets for local seismic networks. The CNN is trained on a small dataset for deep-learning purposes (411 events) detected throughout northern Chile, typical for a temporary aftershock network. In the absence of extensive training data, we demonstrate that a CNN-based automatic phase picker can still improve performance in classifying seismic phases, which matches or exceeds that of historic methods. The trained network is tested against an optimized short-term average/long-term average (STA/LTA) based method (Rietbrock *et al.*, 2012) in classifying phase onsets for a separate dataset of 3878 events throughout the same region. Based on station travel-time residuals, the CNN outperforms the STA/LTA approach and achieves location residual distribution close to the ones obtained by manual inspection.

INTRODUCTION

Accurate detection of earthquake signals generated within the Earth is a fundamental and challenging task in seismology. Traditionally, the optimal method of identifying seismic phases involves a trained analyst manually inspecting seismograms and determining individual phase arrival times. Continuous developments in data acquisition and storage have resulted in vast, unprecedented increases in the volume of available seismic data. For such large-scale datasets, traditional manual picking methods are rendered unfeasible because of the required investment of time and resources; in addition, manual picking

incorporates the subjectivity of different analysts that can bias pick accuracy. Further development of reliable automated picking methods is therefore essential to assist seismologists in their efforts to process large-scale datasets.

Historic Autopickers

The pressing need for a reliable automatic phase picker is not new, and numerous methods have been proposed to detect *P*- and *S*-wave onsets automatically. The most commonly used method for automatic phase picking is still the short-term average/long-term average (STA/LTA) approach (Allen, 1978, 1982; Earle and Shearer, 1994), which measures the ratio between the energy of the seismic signal over a short-term and a long-term window; any values of the STA/LTA ratio above a defined cut-off threshold represent a phase arrival. Baer and Kradolfer (1987) modified the STA/LTA by incorporating an envelope function and a dynamic signal threshold into the characteristic function. There are numerous other approaches, including those based on higher-order statistics (Saragiotis *et al.*, 2002, 2004; Küperkoch *et al.*, 2010), autoregressive methods (Leonard and Kennett, 1999; Sleeman and Van Eck, 1999; Rastin *et al.*, 2013), shallow neural networks (Dai and MacBeth, 1995, 1997; Wang and Teng, 1995; Zhao and Takano, 1999; Gentili and Michelini, 2006), methods that use wave polarization (Baillard *et al.*, 2013; Ross and Ben-Zion, 2014), and methods that use pickers in tandem (Nippres *et al.*, 2010). Although there has been extensive development of autopicker routines, automated picking algorithms cannot currently match the accuracy of an experienced analyst. This is attributed to the complex nature of earthquake source and propagation, with multiple physical processes affecting the wavefield; variations in attenuation, noise interference, source mechanism, and energy partitioning at interfaces all affect the observed waveform.

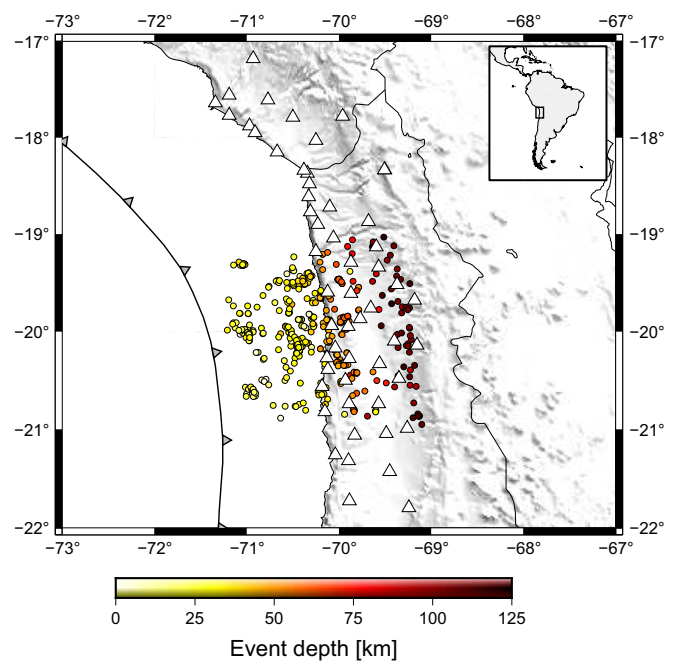
Why Historic Autopicking Routines are Typically Inferior Compared with Human Analysts

Traditional automated picking methods are manually optimized for individual networks or even on a station by station

basis, fine-tuning the characteristic functions to distinguish body-wave phases from noise. For example, triggers can be based on the frequency content of a trace, kurtosis, or some other combination of manually extracted features. One common problem is that *S*-wave phases are more difficult to pick because their onset is often masked by the coda of *P*-wave, and manually extracted features often struggle to identify the *S*-wave in such instances (Gomberg *et al.*, 1990).

Advancements in Deep Learning

Rather than extracting individual features, deep learning-based algorithms focus on learning representations of data, in which multiple layers of processing provide varying levels of abstraction (LeCun *et al.*, 2015; Schmidhuber, 2015). Recent advancements in deep learning techniques yielded a suite of procedures that demonstrate superhuman performance when applied to solve problems in fields ranging from computer vision (Krizhevsky *et al.*, 2012) to speech recognition (Hinton *et al.*, 2012). Convolutional neural networks (CNNs) are a form of supervised machine learning that achieves exceptional results in classifying multidimensional inputs such as images, videos, and audios (LeCun and Bengio, 1995; Krizhevsky *et al.*, 2012; Karpathy *et al.*, 2014). CNNs apply repeated convolutional and pooling operations to the input data, resulting in a set of learnable filters that automatically engineer the appropriate features for classification. The appropriate features are extracted by fine-tuning of the network's internal parameters (or weights) via a computer-based optimization process. The intrinsic properties of CNNs make them an ideal method for natural signal classification (LeCun *et al.*, 2015). Natural signals often demonstrate local connections between samples, an example being the higher amplitudes observed immediately after an impulsive phase arrival. The major advantage of a CNN approach is how such features are then optimized. Shared weights throughout the network result in the systematic optimization of decision boundaries to find the best-weighted combination of local features to classify phase onsets. Another major factor behind the success of deep learning methods is that the only required input is a large dataset of labeled examples for training. Within the seismological community, large datasets of labeled data are readily available in the form of manually picked earthquake catalogs for many regions. We are now starting to see the adoption of deep learning-based methods to solve problems in seismological processing (e.g., Yoon *et al.*, 2015; Perol *et al.*, 2018; Ross *et al.*, 2018; Titos *et al.*, 2018; Zhu and Beroza, 2018). Preliminary results indicate such methods can match or even surpass human levels of performance in seismic phase classification. So far, CNN approaches have been trained over extensive (~million) catalogs of labeled examples collected over decades (Ross *et al.*, 2018; Zhu and Beroza, 2018). We now investigate the dependency of the input data on classification performance by applying a CNN to classify seismic phases, in which the network is trained over a relatively small catalog of events (~11,000 *P*- and *S*-phase pairs). Can a relatively simple CNN architecture display similar performance improvements in the



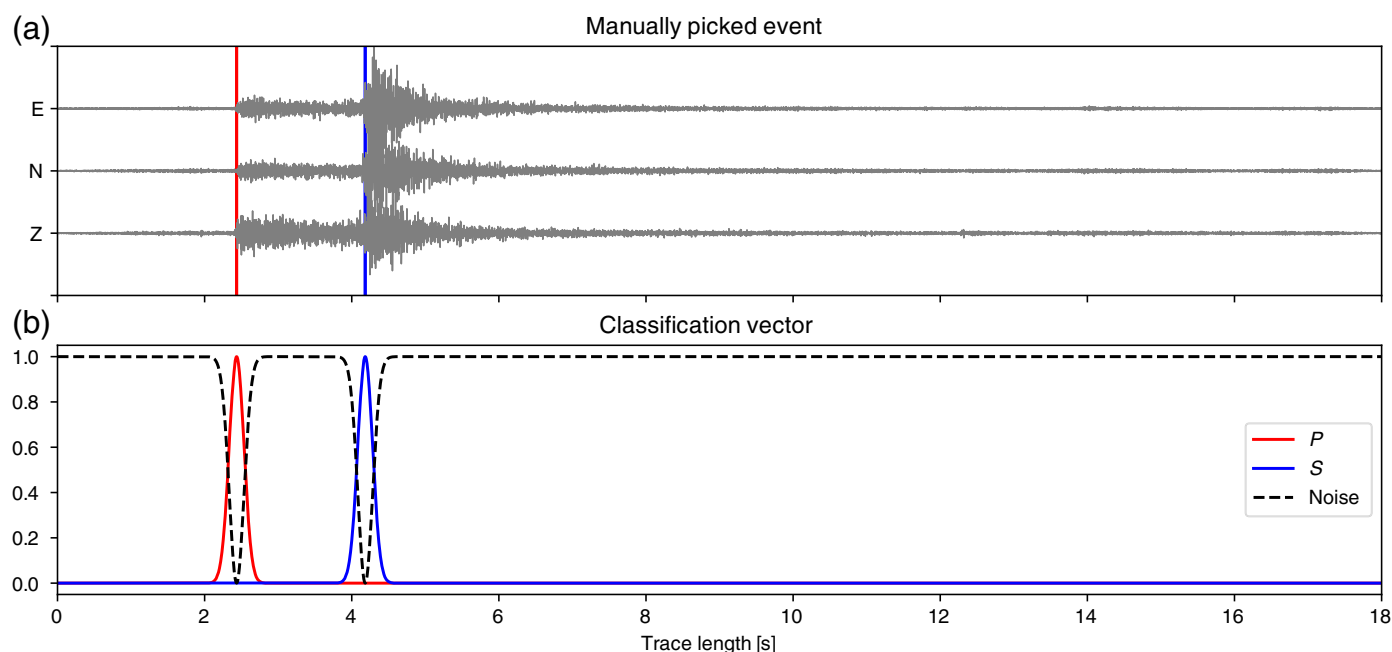
▲ **Figure 1.** Distribution of manually picked events throughout northern Chile. Stations are indicated by white triangles, and event hypocenters are plotted as a function of depth. The color version of this figure is available only in the electronic edition.

absence of an extensive training dataset? If a feature engineering approach demonstrates generalization capabilities when trained over a small local dataset with inherent biases, this will further validate the potential of deep learning-based methods over traditional techniques for seismic phase classification.

DATA

The dataset used in training the CNN is a manually picked catalog of 411 events containing approximately 11,000 *P*-/*S*-phase pairs located throughout the Iquique region of northern Chile. The training catalog has also been used to perform a minimum 1D velocity inversion. Events occurred between March and May 2014 are recorded over a network of 65 broadband and short-period stations distributed throughout northern Chile and southern Peru; all stations use a sampling frequency of 100 Hz (Fig. 1).

Manual picking of events was performed using seismic data explorer software (see [Data and Resources](#)). We process the dataset applying a linear detrend. Although the CNN approach is shown to learn the characteristics of *P* phases, *S* phases, and noise (Zhu and Beroza, 2018), because of our limited training dataset, the CNN network will only be presented with a small portion of noise examples. To limit the potential for the CNN to erroneously identify noise it has not been trained on as phases, and to homogenize the data set due to different instrumentation; we bandpass filter the data between 2 and 25 Hz, a frequency range that lies in the passband of all instruments deployed.



▲ **Figure 2.** An example of (a) input data and (b) classification data; inputs to the convolutional neural network (CNN) are three-component traces, linear detrended, and bandpass filtered between 2 and 25 Hz. The associated classification vector for *P* pick and *S* pick are represented probabilistically as a Gaussian with $\sigma = 1$ s. The color version of this figure is available only in the electronic edition.

Manual picks are represented probabilistically as a Gaussian function ($\sigma = 1$ s, Fig. 2), reducing the bias associated with erroneous picks. The σ parameter was determined through manual parameter testing. Larger σ values resulted in the network acting more as an event detector in which the output probabilities were not impulsive enough to obtain a definitive phase onset. Values lower than 1 s resulted in a high proportion of miss-picks because manual pick errors not captured by the classification vector had a detrimental effect on engineering the appropriate features for phase classification. The dataset is split into training, validation, and test batches (with ratios of 80:10:10, respectively).

Dataset Augmentation and Training

Deep learning-based classifiers contain a significant number of trainable parameters in the solution space; therefore, an extremely large number of examples are needed to prevent overfitting of the training dataset and to enhance generalization. Our dataset is relatively small for deep learning purposes. To overcome the limitations associated with a small training dataset, we perform several additional processing steps. Events are scaled by multiplication of a value drawn from a lognormal distribution, the ends of the segmented event are tapered to limit impulsive amplitude spikes generated by processing, and varying levels of Gaussian noise are then added to each batch,¹ resulting in greater variations of signal versus background noise. The training events are therefore modified to show a range of arrival types rather than the high-magnitude,

well-recorded events that are typically seen in a small catalog of manually selected earthquakes for further studies. The input window size for the CNN is 6 s. To train the CNN, a given input batch is sequentially windowed with a timestep of 0.4 s. The windows are randomly shuffled before being used in training, preventing the CNN learning any unnecessary temporal order. A small timestep is used to increase the total number of events during training; also, having the network learn to recognize the presence of phases at any point in the input window will help the network generalize beyond the training dataset. Formatting the input data in such a way reduces the biases associated with our small dataset and enhances the capability of the network to pick varying types of arrivals.

METHODOLOGY

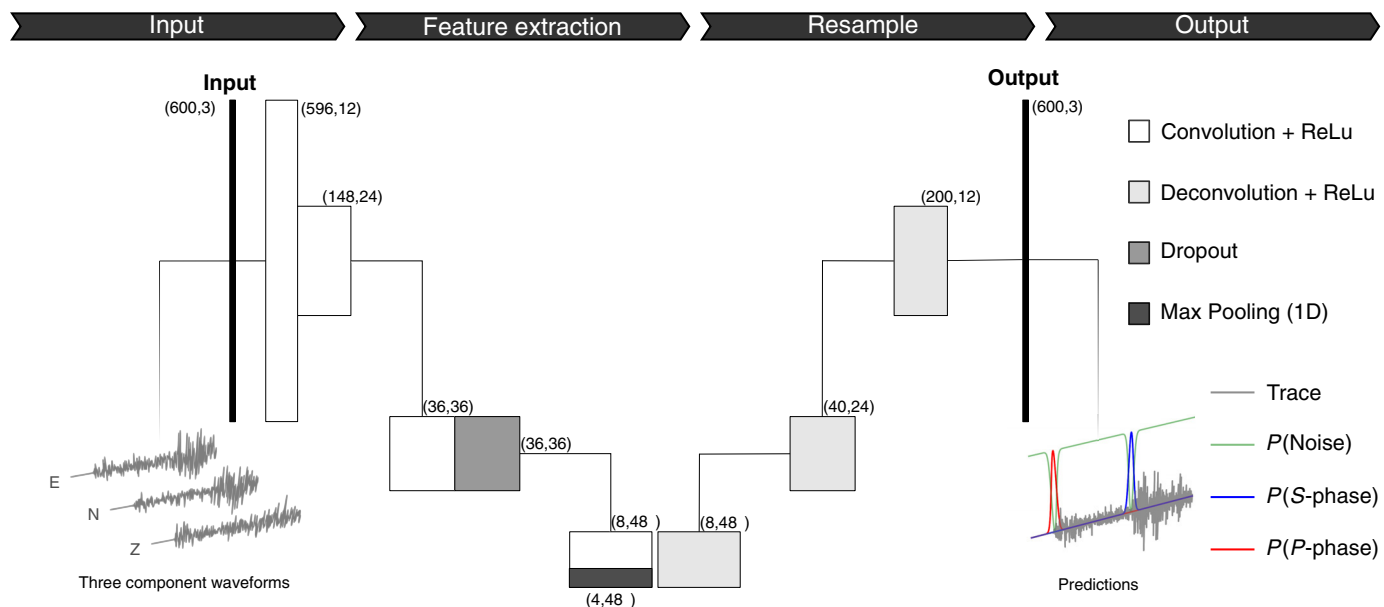
Network Architecture

The input to the network comprises three 1D windows (x), in which each window samples an individual component. For this given input, the network outputs the probability of *P*-phase, *S*-phase, or noise for each time sample within that window (Fig. 3). Probabilities are output by applying the softmax or normalized exponential function to the final layer

$$p(Y = i|x) = \frac{e^{\alpha_i(x)}}{\sum_{j=1}^3 e^{\alpha_j(x)}}, \quad (1)$$

in which $j = 1, 2, 3$ represents the *P*-phase, *S*-phase, and noise classes and $\alpha(x)$ contains the associated weights for the final layer. The input data are passed through repeated transformations; convolutional operations initially extract the appropriate

¹More information on parameters used in aiding generalization is provided in the [Appendix](#).



▲ **Figure 3.** Overall CNN architecture displaying the sequential convolution and resampling operations applied to the input window. The color version of this figure is available only in the electronic edition.

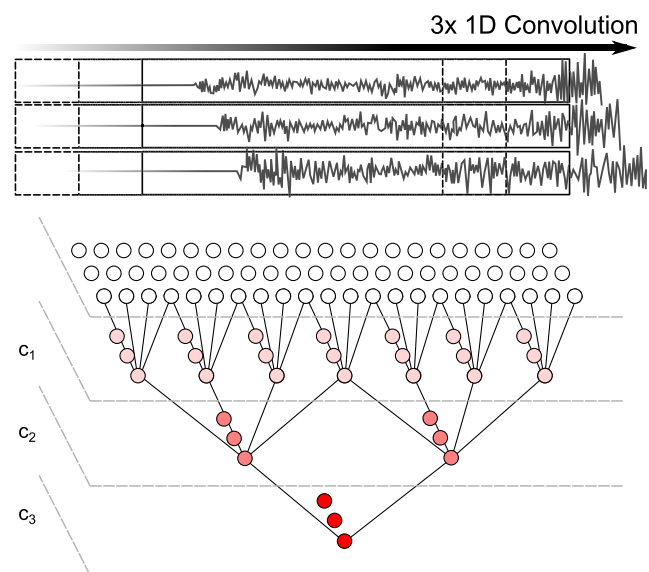
features to characterize each class, the extracted features then go through repeated resampling stages, to output per-class probabilities. At each stage, a rectified linear unit activation function is applied (Nair and Hinton, 2010). The cost function used to train the CNN is given by the negative log likelihood $NLL(x, \theta)$. For a multiclass classification problem, in which each class is characterized as a series of discrete probability distributions, $NLL(x, \theta)$ is also termed the cross-entropy loss function,

$$NLL(x, \theta) = - \sum_{k=1}^3 \sum_{n=0}^{n-N} \log(p(c_k | x_n, \theta)), \quad (2)$$

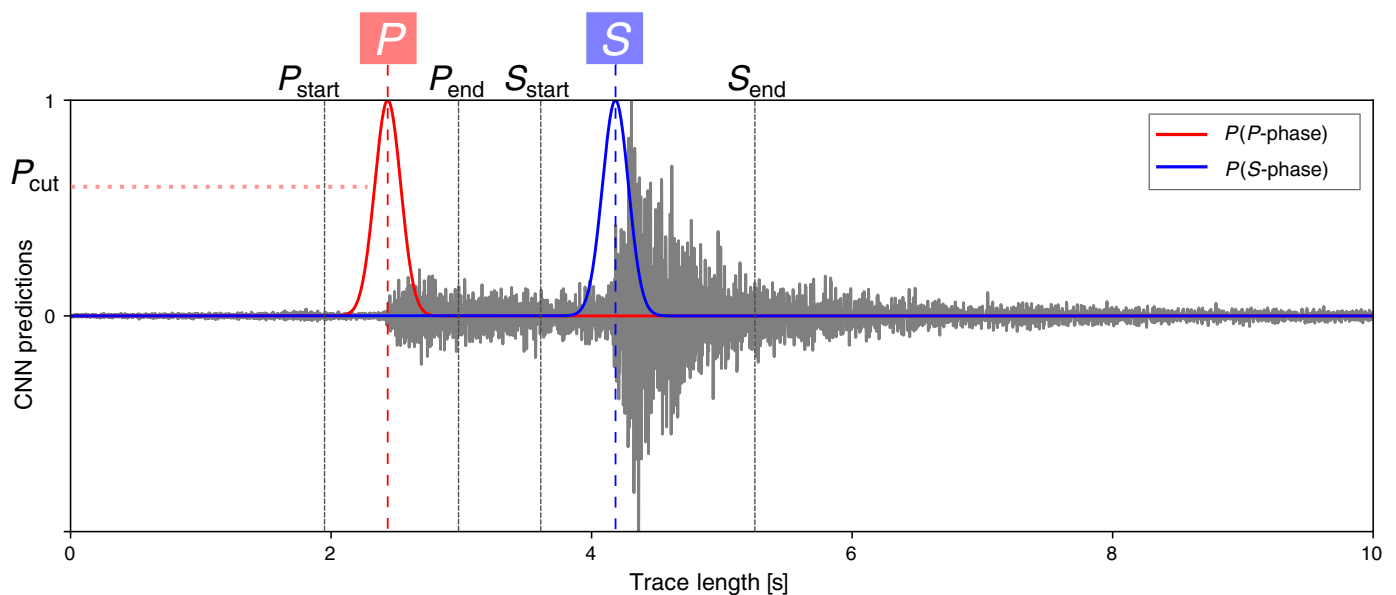
in which N represents the total number of training instances, c_k corresponds to the class label assigned to the input (x_n), and the network weights (θ). Equation (2) is minimized using adaptive moment estimation (Kingma and Ba, 2014) along with batch training, the network weights are therefore updated at the end of each batch, over n training instances.

Hyperparameter optimization (Bergstra and Bengio, 2012) is the derivation of the optimal network parameters and is a major challenge when designing neural network architectures. Parameters such as the number of layers, regularization of layers, convolutional kernel shape, and learning rate can all be optimized. Methods to solve this problem consist of grid search, random search, and manual estimation. As our study aims to demonstrate that a robust CNN can be trained on small datasets, the focus is on efficient implementation over more time-consuming systematic search methods. After a robust network architecture is derived, a constrained search is performed for the best combination of hyperparameters.

Our final network architecture consists of three convolutional layers followed by three layers of upsampling (Fig. 3). Again, because of the limited nature of the training dataset, the focus for the network architecture is to limit the potential for overfitting. To localize the features corresponding to different classes, convolutional layers apply strided 1D convolutional filters along each component (Fig. 4). The stride for the convolutional window is set to 4, this downsamples the time series



▲ **Figure 4.** Schematic displaying how strided 1D convolutions quickly incorporate the long-term temporal dependencies of the input data into the convolution kernel. The color version of this figure is available only in the electronic edition.



▲ **Figure 5.** Displaying how the temporal relationship between P and S phases is used to identify phase onsets from the output CNN probabilities. Solid lines correspond to the output P -/ S -phase probabilities, vertical dashed lines indicate phase onsets, and the phase type is labeled above each vertical dashed line. Vertical dotted lines indicate the start or end of a P -/ S -phase search window, in which the corresponding labels are again presented at the top of each line. The horizontal dotted line represents the p_{cut} parameter used in determining phase onsets. The color version of this figure is available only in the electronic edition.

by a factor of 4 for each layer, reducing the overall number of free parameters and allowing for quicker incorporation of long-term temporal dependencies into the convolution kernel. A dropout parameter is added to the second convolutional layer. Dropout is a regularization technique that randomly drops weights during training, reducing model complexity (Srivastava *et al.*, 2014). 1D max-pooling is applied to the final convolutional layer, further reducing the overall number of networks weights.

Picking Phases

To obtain P - and S -phase onsets from the CNN output probabilities, we use knowledge of the simple temporal relationships between P and S phases to determine onset times (Fig. 5).

For the P -phase probability distribution $p = \{p_1, p_2, p_3, \dots, p_n\}$ and the S -phase probability distribution $s = \{s_1, s_2, s_3, \dots, s_n\}$, if P -phase probabilities exceed a defined cut-off threshold p_{cut} , the P -phase onset is searched for within the window $[p_{\text{start}}, p_{\text{end}}]$. The P -phase onset is set at the index of the maximum P -phase probability within this range. If the P -phase criterion is met, the corresponding S -phase is searched for within the searched window $[s_{\text{start}}, s_{\text{end}}]$ if $\sum_{i=s_{\text{start}}}^{s_{\text{end}}} s_i > s_{\text{cut}}$ then the S -phase is set at the index of the maximum S -phase probability within search window. Both conditions must be satisfied for an event to be picked; consequently, the ratio of P - to S -phase picks using these criteria is 1:1. The parameters used in detecting phase onsets are provided in Table 1; all index values are relative to the initial p_{cut} index.

RESULTS

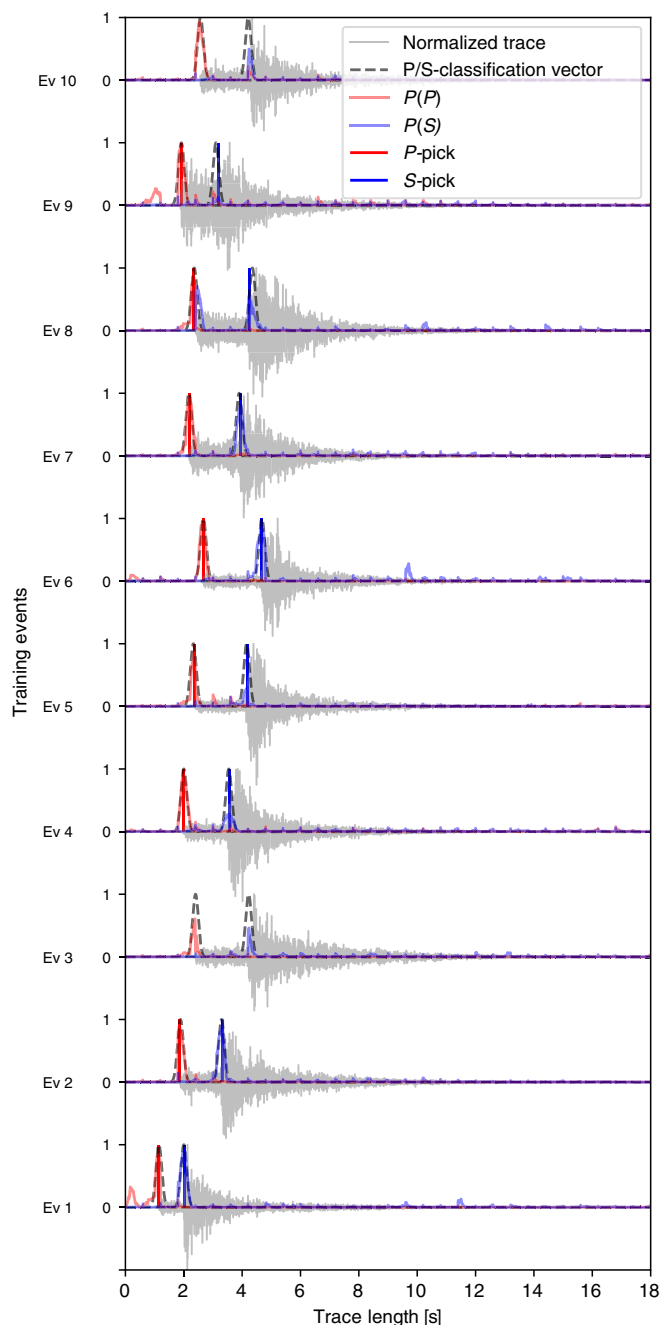
Predictions

The trained network takes a 6-s input window for three-component data and makes phase predictions for each time sample within the window. Figure 6 displays a sample of the output phase probabilities for events in the test dataset. The predictions display a clear distinction between P and S phases, further confirming that deep learning-based classification methods engineer the appropriate features to accurately categorize P , S , and noise classes. This presents a major advantage over historic autopicking methods that use manual feature extraction and often struggled to identify the S -phase. To obtain P -/ S -phase onsets, we apply our autopicker function with the input parameters of Table 1, taking advantage of the simple temporal relationship between P and S phases to assign phase onsets (vertical

Table 1
Parameters Applied to the Autopicker Function

P -Pick	Parameter Value	S -Pick	Parameter Value
p_{cut}	0.75	s_{sum}	5
p_{start}	-200	s_{start}	500
p_{end}	200	s_{end}	4000

Autopicker function takes advantage of the temporal relationship between phases to identify phase onsets. All start and end indexes are given in samples (sampling rate for all instruments = 100 Hz).



▲ **Figure 6.** Output CNN prediction probabilities when applied to identify phase onsets for the test dataset; phase onsets are indicated by vertical lines. The color version of this figure is available only in the electronic edition.

lines in Fig. 6). The phase onsets are then compared with the original manual picks, and the residuals are plotted (Fig. 7).

The residual distribution for the test dataset displays a good agreement in the center of both the P - and S -residual distribution; however, the CNN has also picked phases corresponding to extra arriving events in some waveforms. These extra phase picks may be accurate, but any additional events are not represented in our classification vectors because a detailed association of individual phases to specific events

arriving simultaneously is beyond the scope of this work. This negatively affects the residual distribution and is responsible for several of the large outliers observed.

Relocation Testing

To overcome the issue of extra picked arrival times from simultaneously occurring events biasing our residual comparison, we perform an additional test to remove arrival times from any events overlapping in time. This additional test provides a more consistent assessment of autopicker performance. We perform an iterative inversion procedure, relocating both the original manual picks and the CNN picks for the initial dataset. The catalogs are relocated using the VELEST routine (Kissling *et al.*, 1994), which applies a minimum 1D velocity model along with station corrections to solve for hypocenter locations. Hypocentral parameters are solved for all events within the catalog. When using VELEST, all phase picks within a segmented trace are assigned to a single event during relocation. The large outlier residuals a significant distance (+3 s) from the trend are attributed to multiple picked events in the same segmented trace being erroneously classified as a single event in VELEST. We therefore reject events with root mean square (rms) residual larger than 3 s to remove any picked events overlapping in time. Statistics of the residual distribution for the original manual picks compared with the CNN picks are provided in Table 2.

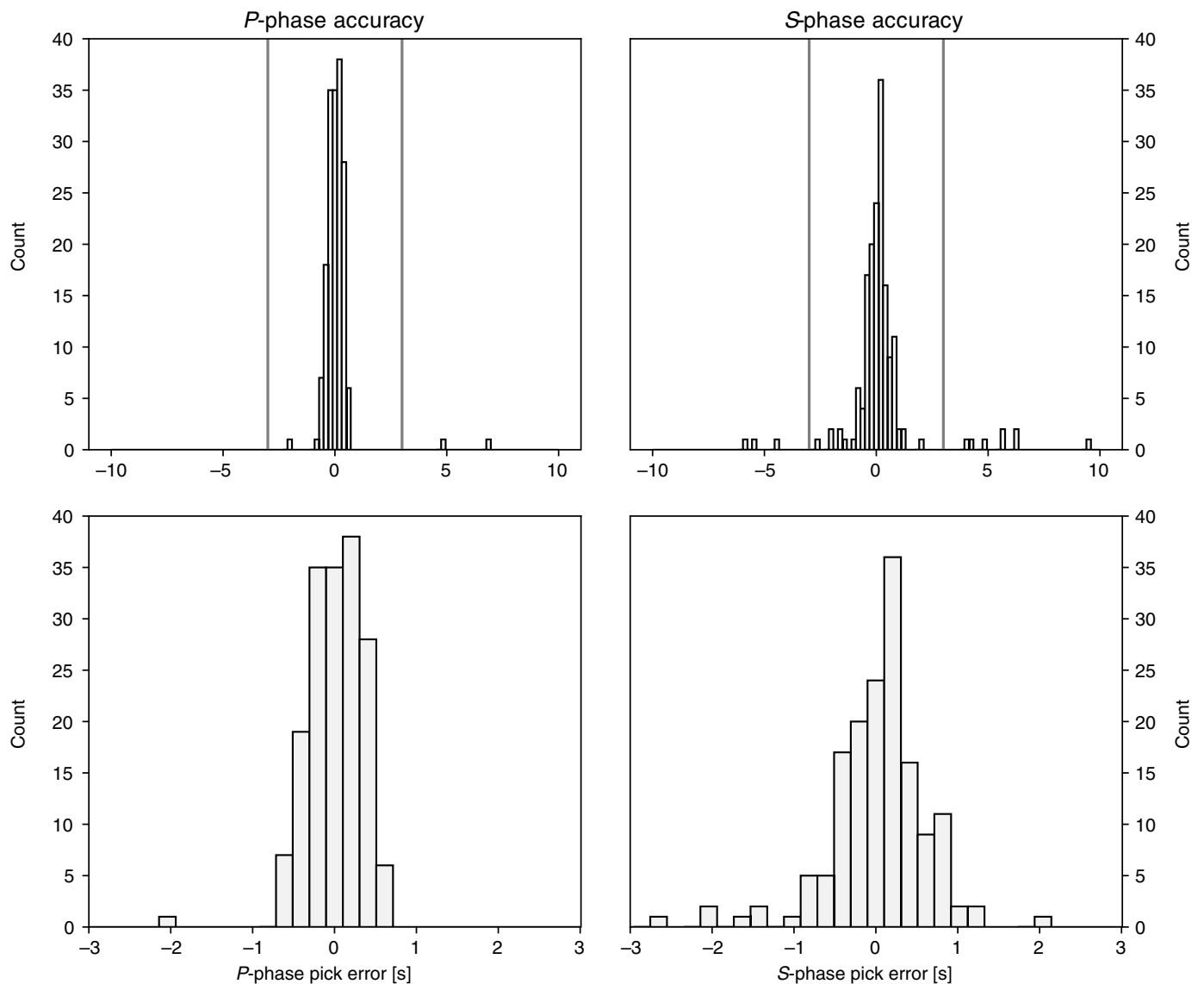
The residual distribution indicates that manually picked P phases are slightly more accurate than CNN P -phase picks (σ decreased by 0.051 s); however, S -phase picks of the CNN approach achieve similar performance to manual picking (σ decreased by 0.019 s). We recognize that our training and test data set used for the earthquake location data set are not independent; however, the residual distribution obtained from the CNN methodology is similar to that of the manual picks of an expert seismologist.

Autopicker Comparison

To further test the CNN picker, we apply the CNN methodology in predicting phase onsets for a separate catalog of events throughout northern Chile on the same temporary seismic network. Events were initially segmented using an iterative approach based on an STA/LTA trigger (Rietbrock *et al.*, 2012) and provide a useful test case for the CNN method. The relocation procedure is again applied to compare performance. The initial number of phase picks for both methods is provided in Table 3.

Figure 8 displays an event from the new catalog picked using the CNN method, multiple event arrivals are again present in the traces. To limit the effect of this issue on our residual comparison, we set both the STA/LTA and CNN method to only pick a single P -/ S -phase pair per trace and again use the iterative relocation procedure to assess residual.

The relocated hypocenter distribution for both methods is displayed in Figure 9. It can be clearly observed that locations are more clustered in the CNN approach and are better concentrated along the plate interface, indicating the greater



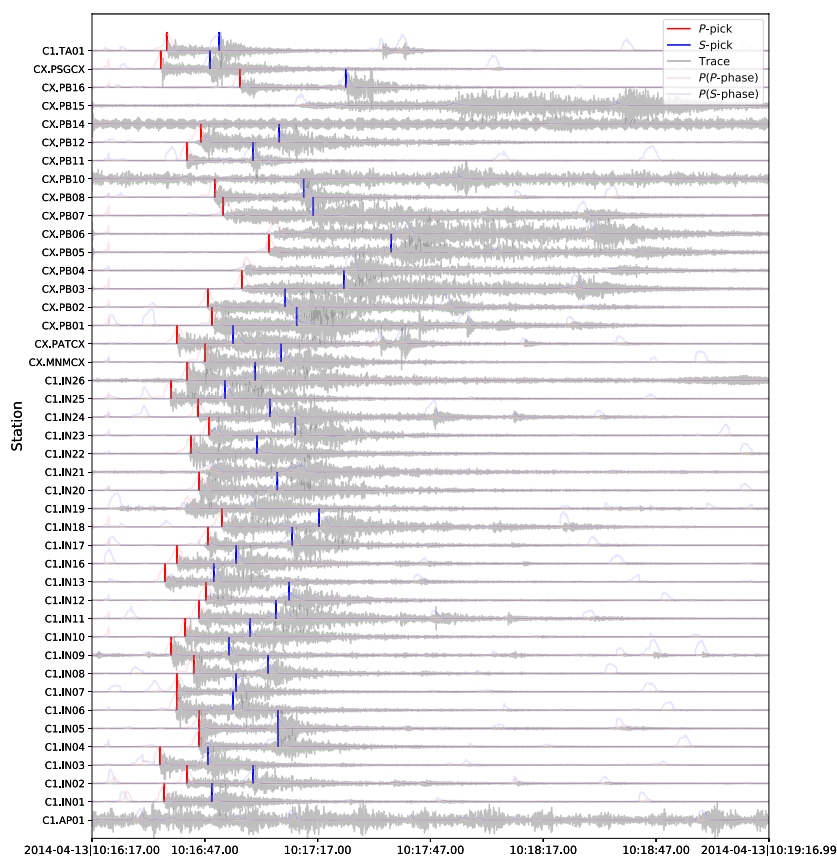
▲ **Figure 7.** Residual of CNN predicted phase onsets versus original manual picks for the test dataset.

consistency in phase picks for the CNN approach. Phase residuals for the relocated events are displayed in Figure 10; we show residuals for both the final catalogs (minimum azimuthal

gap $< 220^\circ$) and for only the best-located events (minimum azimuthal gap $< 160^\circ$). Statistics for the residual distributions are displayed in Table 4. Assuming a normal distribution, the CNN method exhibits decreased variance in phase residual for both P and S phases compared with the optimized STA/LTA approach.

Table 2 Statistics of Residual Distribution for Original Manually Picked Catalog, Both for the Original Manual Picks and the CNN Methodology Picks			
Gap $< 220^\circ$		CNN	Manual
P	μ	-0.261	-0.124
	σ	0.445	0.394
S	μ	0.282	0.390
	σ	0.749	0.730
Bold values indicate optimum parameter value. CNN, convolutional neural network.			

Table 3 Overall Autopicks on a Separate Catalog of New Events Throughout Northern Chile		
	STA/LTA	CNN
P	72,655	77,623
S	63,353	77,623
Total	136,008	155,246
STA/LTA, short-term average/long-term average.		



▲ **Figure 8.** Demonstrating the CNN autopicker performance on a new dataset for northern Chile, where events were segmented using short-term average/long-term average (STA/LTA) trigger (Rietbrock *et al.*, 2012). We only allow the autopicker to pick the presence of a single *P*/*S*-phase per trace to prevent relocation errors. The color version of this figure is available only in the electronic edition.

The relocation residuals (Fig. 10) are not just dependent on the accuracy of detected phases but also on velocity variations not captured in the 1D model or station corrections affecting the residuals. Because both catalogs were relocated with the same iterative relocation procedure using the same 1D velocity model and station delay terms, discrepancies in residual distributions should directly reflect the relative consistency of picks in each catalog. Investigating the residual

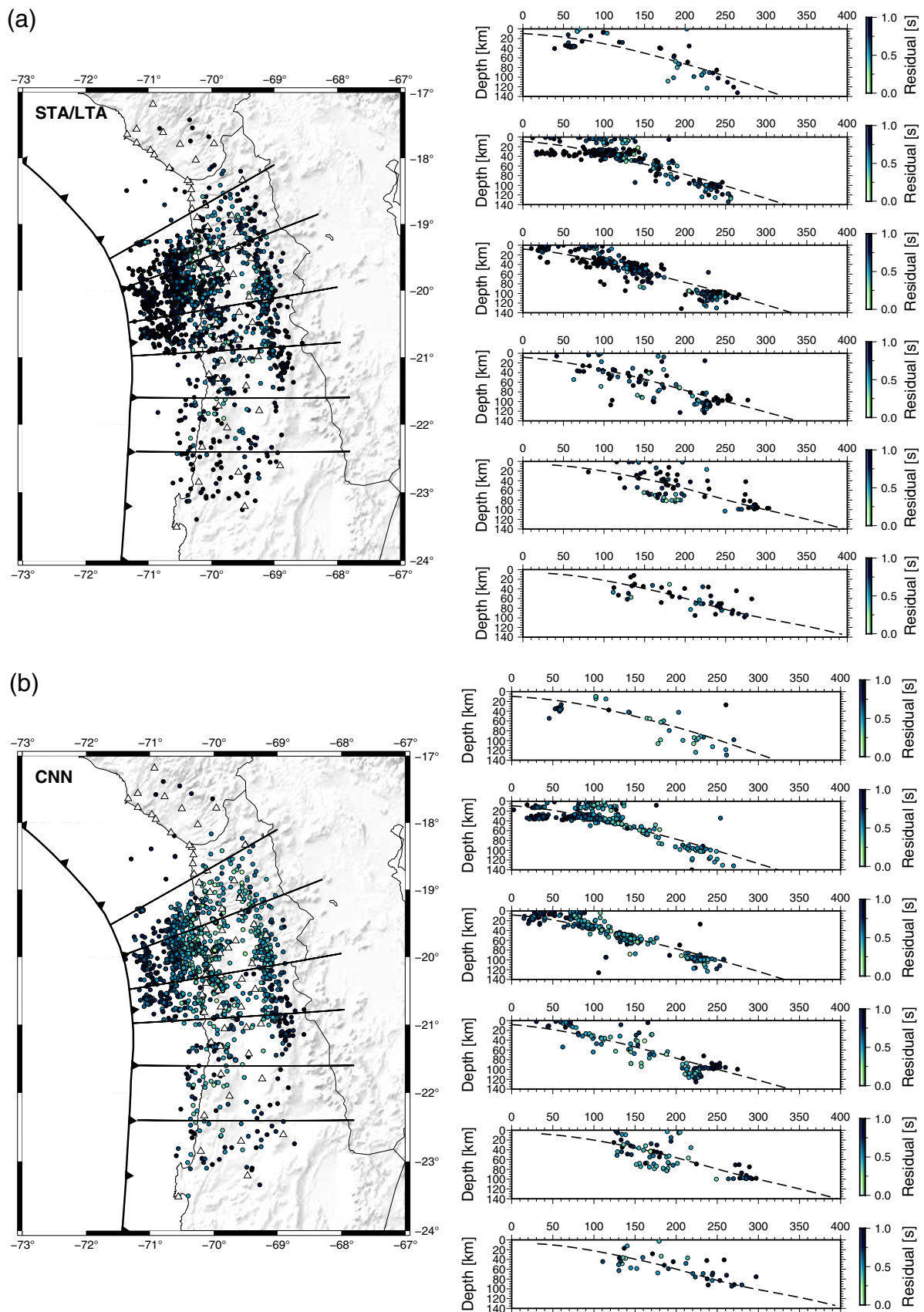
distribution, the CNN approach has markedly improved both the overall relocation residual (Fig. 9) and the variation in residual for both *P* and *S* phases. In addition to this, the difference in σ for the well-located events is shown to be more accurate for the CNN approach, with σ improving by 0.230 s for *P* phases and 0.326 s for *S* phases compared with the optimized STA/LTA-picking approach. The statistics of the residual distribution are also in a similar range to that of the manual picks (see Table 2).

DISCUSSION AND FUTURE WORK

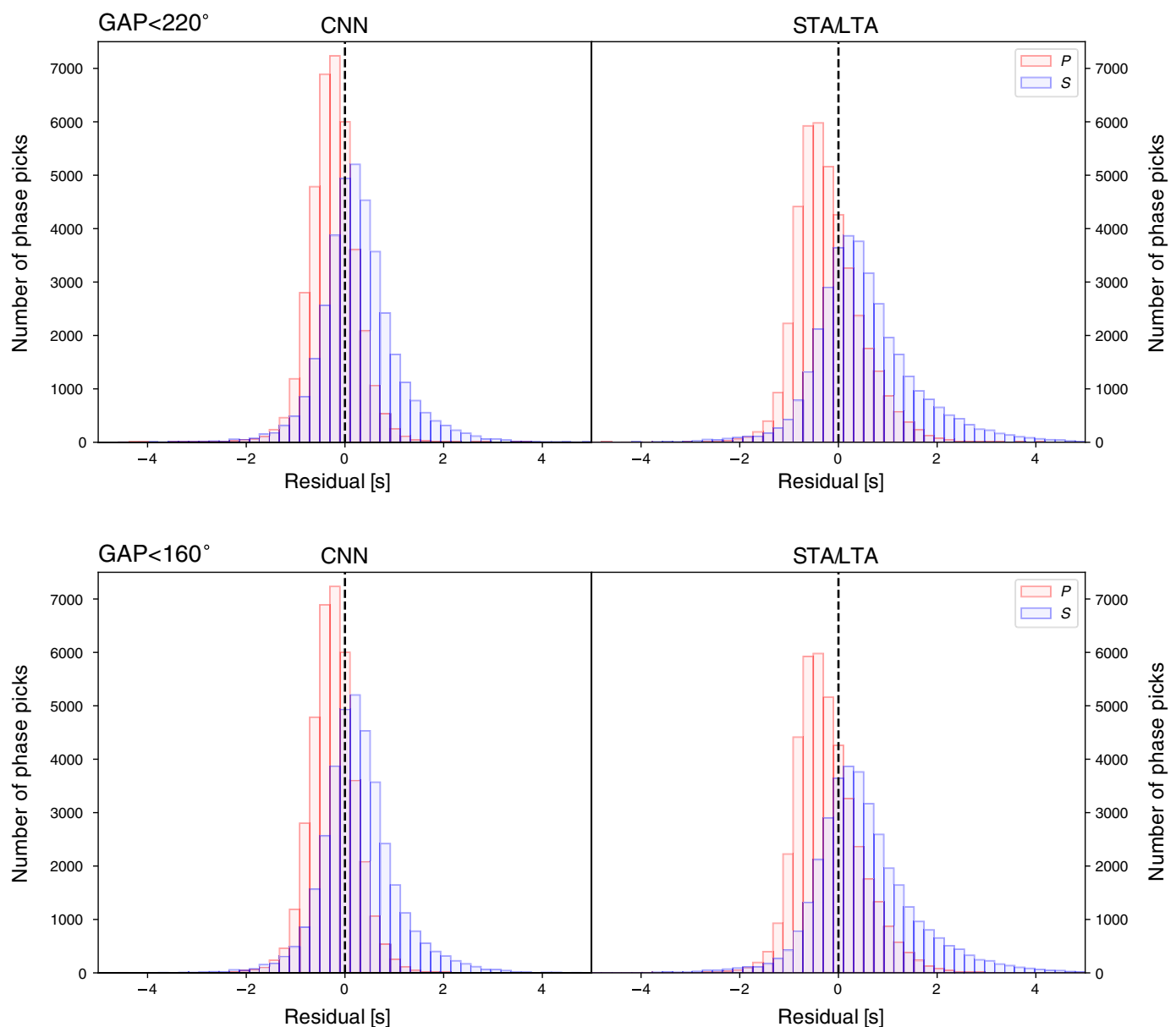
Accurate and consistent catalogs of phase arrivals are of paramount importance to seismologists because they typically form the starting point for further seismological investigations. The rapidly increasing amount of seismic data available, along with constant developments in computational capabilities, have resulted in the seismological community now increasingly turning to machine learning-based methods to improve the efficiency of seismic processing. As shown, automatic feature engineering approaches such as CNNs hold promise for seismic phase classification because they only require the three-component data as an input, and the features engineered from the data combine to detect the general characteristics of *P* phases, *S* phases, and noise. Our experimental results show that even when data are scarce, a simple CNN architecture significantly improves the σ of *P*- and *S*-pick

residuals, especially for well-locatable events (minimum azimuthal gap $< 160^\circ$), resulting in a decrease of 0.230 and 0.326 s, respectively, compared with an optimized STA/LTA-picking approach (Rietbrock *et al.*, 2012). The decreased variation in residual indicates that a CNN-based method is more consistent when autopicking, resulting in more accurate hypocenter relocations. We are close to reaching a point at which supervised learning-based methods exhibit comparable or even increased performances compared with manual picking by an expert seismologist (Ross *et al.*, 2018; Zhu and Beroza, 2018). Until now, supervised learning-based methods have been trained using extensive training datasets (\sim millions of examples). The results from our work add to the literature of supervised learning-based methods for seismic phase classification and demonstrate that with appropriate considerations regarding overfitting and generalization, such methods can improve seismological processing workflows, not just for large catalogs but for varying datasets. Future applications of deep learning-based methods in seismology include deploying such pretrained systems on poorly

Table 4 Statistics of Residual Distribution for Gap $< 220^\circ$ and 160°					
		Gap $< 220^\circ$		Gap $< 160^\circ$	
		CNN	STA/LTA	CNN	STA/LTA
<i>P</i>	μ	−0.238	−0.216	−0.247	−0.333
	σ	0.487	0.696	0.393	0.623
<i>S</i>	μ	0.277	0.539	0.270	0.435
	σ	0.780	1.081	0.596	0.922
Bold values indicate optimum parameter value.					



▲ **Figure 9.** Hypocenter relocation comparison for (a) the STA/LTA autopicked catalog against (b) the CNN autopicked catalog event relocations are plotted as a function of root mean square (rms) residual. Slab profile is provided by [Hayes et al. \(2012\)](#). The color version of this figure is available only in the electronic edition.



▲ **Figure 10.** Both autopicking methods phase residuals following hypocentral relocations, plotted for well-located events (minimum azimuthal gap < 160°) and for the entire relocated catalogs (minimum azimuthal gap < 220°). The color version of this figure is available only in the electronic edition.

monitored areas or areas of interest, resulting in improved data recovery, and efficient automation of seismic workflows.

DATA AND RESOURCES

All data used in this study can be downloaded from the Incorporated Research Institutions for Seismology (IRIS) Data Management Center for the temporary network data and also from the GEOForschungsNetz (GEOFON) data repository (<https://geofon.gfz-potsdam.de/waveform/archive/>, last accessed September 2018). Seismic data explorer are available at <http://doree.esc.liv.ac.uk:8080/sdx> (last accessed September 2018). ☒

ACKNOWLEDGMENTS

This work was partly supported by National Environment Research Council (NERC) Grant Numbers NE/M005879/1, NE/R010412/1, NE/P00105X/1, and NE/P007708/1.

REFERENCES

- Allen, R. (1982). Automatic phase pickers: Their present use and future prospects, *Bull. Seismol. Soc. Am.* **72**, no. 6B, S225–S242.
- Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.* **68**, no. 5, 1521–1532.
- Baer, M., and U. Kradolfer (1987). An automatic phase picker for local and teleseismic events, *Bull. Seismol. Soc. Am.* **77**, no. 4, 1437–1445.

- Baillard, C., W. C. Crawford, V. Ballu, C. Hibert, and A. Mangeney (2013). An automatic kurtosis-based P-and S-phase picker designed for local seismic networks, *Bull. Seismol. Soc. Am.* **104**, no. 1, 394–409.
- Bergstra, J., and Y. Bengio (2012). Random search for hyper-parameter optimization, *J. Machine Learn. Res.* **13**, 281–305.
- Dai, H., and C. MacBeth (1995). Automatic picking of seismic arrivals in local earthquake data using an artificial neural network, *Geophys. J. Int.* **120**, no. 3, 758–774.
- Dai, H., and C. MacBeth (1997). The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings, *J. Geophys. Res.* **102**, no. B7, 15,105–15,113.
- Earle, P. S., and P. M. Shearer (1994). Characterization of global seismograms using an automatic-picking algorithm, *Bull. Seismol. Soc. Am.* **84**, no. 2, 366–376.
- Gentili, S., and A. Michelini (2006). Automatic picking of P and S phases using a neural tree, *J. Seismol.* **10**, no. 1, 39–63.
- Gomberg, J. S., K. M. Shedlock, and S. W. Roecker (1990). The effect of S-wave arrival times on the accuracy of hypocenter estimation, *Bull. Seismol. Soc. Am.* **80**, no. 6A, 1605–1628.
- Hayes, G. P., D. J. Wald, and R. L. Johnson (2012). Slab1.0: A three-dimensional model of global subduction zone geometries, *J. Geophys. Res.* **117**, no. B1, doi: [10.1029/2011JB008524](https://doi.org/10.1029/2011JB008524).
- Hinton, G., L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, *IEEE Signal Process. Mag.* **29**, no. 6, 82–97, doi: [10.1109/MSP.2012.2205597](https://doi.org/10.1109/MSP.2012.2205597).
- Karpathy, A., G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei (2014). Large-scale video classification with convolutional neural networks, *Proc. of the IEEE conference on Computer Vision and Pattern Recognition*, 1725–1732.
- Kingma, D. P., and J. Ba (2014). *Adam: A Method for Stochastic Optimization*, available at <http://arxiv.org/abs/1412.6980> (last accessed September 2018).
- Kissling, E., W. L. Ellsworth, D. Eberhart-Phillips, and U. Kradolfer (1994). Initial reference models in local earthquake tomography, *J. Geophys. Res.* **99**, no. B10, 19,635–19,646.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, 1097–1105.
- Küperkoch, L., T. Meier, J. Lee, W. Friederich, and EGELADOS Working Group (2010). Automated determination of P-phase arrival times at regional and local distances using higher order statistics, *Geophys. J. Int.* **181**, no. 2, 1159–1170.
- LeCun, Y., and Y. Bengio (1995). Convolutional networks for images, speech, and time series, in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib (Editor), Vol. 3361, no. 10, MIT Press, Cambridge, Massachusetts.
- LeCun, Y., Y. Bengio, and G. Hinton (2015). Deep learning, *Nature* **521**, no. 7553, 436.
- Leonard, M., and B. L. N. Kennett (1999). Multi-component autoregressive techniques for the analysis of seismograms, *Phys. Earth Planet. In.* **113**, nos. 1/4, 247–263.
- Nair, V., and G. E. Hinton (2010). Rectified linear units improve restricted boltzmann machines, *Proc. of the 27th International Conference on Machine Learning (ICML-10)*, 807–814.
- Nippres, S. E. J., A. Rietbrock, and A. E. Heath (2010). Optimized automatic pickers: Application to the ANCORP data set, *Geophys. J. Int.* **181**, no. 2, 911–925.
- Perol, T., M. Gharbi, and M. Denolle (2018). Convolutional neural network for earthquake detection and location, *Sci. Adv.* **4**, no. 2, e1700578, doi: [10.1126/sciadv.1700578](https://doi.org/10.1126/sciadv.1700578).
- Rastin, S. J., C. P. Unsworth, R. Benites, and K. R. Gledhill (2013). Using real and synthetic waveforms of the Matata swarm to assess the performance of New Zealand GeoNet phase pickers, *Bull. Seismol. Soc. Am.* **103**, no. 4, 2173–2187.
- Rietbrock, A., I. Ryder, G. Hayes, C. Haberland, D. Comte, S. Roecker, and H. Lyon-Caen (2012). Aftershock seismicity of the 2010 Maule $M_w = 8.8$, Chile, earthquake: Correlation between co-seismic slip models and aftershock distribution?, *Geophys. Res. Lett.* **39**, no. 8, doi: [10.1029/2012GL051308](https://doi.org/10.1029/2012GL051308).
- Ross, Z. E., and Y. Ben-Zion (2014). Automatic picking of direct P, S seismic phases and fault zone head waves, *Geophys. J. Int.* **199**, no. 1, 368–381.
- Ross, Z. E., M.-A. Meier, and E. Hauksson (2018). P-wave arrival picking and first-motion polarity determination with deep learning, *J. Geophys. Res.* **123**, no. 6, 5120–5129, doi: [10.1029/2017JB015251](https://doi.org/10.1029/2017JB015251).
- Sarigiots, C. D., L. J. Hadjileontiadis, and S. M. Panas (2002). PAI-S/K: A robust automatic seismic P phase arrival identification scheme, *IEEE Trans. Geosci. Remote Sens.* **40**, no. 6, 1395–1404.
- Sarigiots, C. D., L. J. Hadjileontiadis, I. T. Rekanos, and S. M. Panas (2004). Automatic P phase picking using maximum kurtosis and /spl kappa/-statistics criteria, *IEEE Geosci. Remote Sens. Lett.* **1**, no. 3, 147–151.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview, *Neural Netw.* **61**, 85–117.
- Sleeman, R., and T. van Eck (1999). Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings, *Phys. Earth Planet. In.* **113**, nos. 1/4, 265–275.
- Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: A simple way to prevent neural networks from overfitting, *J. Machine Learn. Res.* **15**, no. 1, 1929–1958.
- Titos, M., A. Bueno, L. García, and C. Benítez (2018). A deep neural networks approach to automatic recognition systems for volcano-seismic events, *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **11**, no. 5, 1533–1544.
- Wang, J., and T.-L. Teng (1995). Artificial neural network-based seismic detector, *Bull. Seismol. Soc. Am.* **85**, no. 1, 308–319.
- Yoon, C. E., O. O'Reilly, K. J. Bergen, and G. C. Beroza (2015). Earthquake detection through computationally efficient similarity search, *Sci. Adv.* **1**, no. 11, e1501057, doi: [10.1126/sciadv.1501057](https://doi.org/10.1126/sciadv.1501057).
- Zhao, Y., and K. Takano (1999). An artificial neural network approach for broadband seismic phase picking, *Bull. Seismol. Soc. Am.* **89**, no. 3, 670–680.
- Zhu, W., and G. C. Beroza (2018). *PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method*, available at <http://arxiv.org/abs/1803.03211> (last accessed September 2018).

APPENDIX

HYPERPARAMETERS AND DATA GENERALIZATION PARAMETERS

Lognormal distribution used to scale individual event amplitudes is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{\left(\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)}, \quad x > 0. \quad (\text{A1})$$

We set $\mu = 0$ and $\sigma = 0.25$, and sample the output probability distribution of equation (A1). Each sample is then used as a scale factor for event amplitudes.

Jack Woollam
Silvio De Angelis
University of Liverpool
Jane Herdman Building

*4 Brownlow Street
Liverpool L69 3GP, United Kingdom
Jack. Woollam@liverpool.ac.uk
S.De-Angelis@liverpool.ac.uk*

*Andreas Rietbrock
Geophysical Institute (GPI)
Karlsruhe Institute of Technology
Hertzstraße 16
76187 Karlsruhe
Germany
andreas.rietbrock@kit.edu*

*Angel Bueno
Department of Signal Theory, Telematics and
Communications
University of Granada
Calle Periodista Daniel Saucedo Aranda
18014 Granada
Spain
angelbueno@ugr.es*

Published Online 16 January 2019