Convolutional Neural Network for seismic phase classification, performance demonstration over a local seismic network.

Authors:

Jack Woollam[1*], Andreas Rietbrock[1,2], Angel Bueno[3], Silvio De Angelis[1].

[1]University of Liverpool, Liverpool, United Kingdom.

[2]Karsruhe Institute of Technology, Karlsruhe, Germany.

[3]University of Granada, Granada, Spain.

Abstract

Over the last two decades the amount of available seismic data has increased significantly fuelling the need for automatic processing to utilize the vast amount of information contained in such data sets. Detecting seismicity in temporary aftershock networks is one important example, which has become a huge challenge due to 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 post-seismic 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 optimised 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 out-performs the STA/LTA approach an 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 due to the required investment of time and resources; in addition, manual picking incorporates the subjectivity of different analysts which can bias pick accuracy. Further development of reliable automated picking methods are therefore essential to assist seismologists in their efforts to process large-scale datasets.

Historic Auto-pickers

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 STA/LTA approach (Allen, 1978; Allen, 1982; Earle & 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 & Kradolfer (1987) modified the STA/LTA incorporating an envelope function and a dynamic signal threshold into the characteristic function. There are numerous other approaches, including those based upon higher-order statistics (Saragiotis et al., 2002, 2004; Küperkock et al., 2010), autoregressive methods (Leonard & Kennet, 1999; Sleeman & Van Eck, 1999; Rastin et al., 2013), shallow neural networks (Wang & Teng, 1995; Dai & MacBeth, 1995, 1997; Zhao & Takano, 1999; Gentili & Michelini, 2006), methods which utilise wave polarisation (Ross & Ben-Zion, 2014; Baillard et al., 2014), and those which utilise pickers in tandem (Nippress et al., 2010). Whilst

there has been extensive development of auto-picker 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 auto-picking routines are typically inferior compared to human analysts

Traditional automated picking methods are manually optimized for individual networks and/or even on a station by station basis, fine tuning the 'characteristic functions' to distinguish body-wave phases from noise. E.g., 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 that S-wave phases are more difficult to pick as their onset is often masked by the coda of P-waves and manually extracted features will 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, where multiple layers of processing provide varying levels of abstraction (LeCun et al., 2015; Schmidhuber., 2015). Recent advancements in deep-learning techniques have yielded a suite of procedures which demonstrate 'super-human' 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 supervised machine learning that achieve exceptional results in classifying multi-dimensional inputs such as images, videos, and audio (Krizhevsky et al., 2012; Karpathy et al., 2014; LeCun & Bengio, 1995). CNNs apply repeated convolutional and pooling operations to the input data, resulting in a set of learnable filters which 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 optimisation 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 following an impulsive phase arrival. The major advantage of a CNN approach is how such features are then optimised. Shared weights throughout the network results in the systematic optimisation 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 labelled examples for training. Within the seismological community, large datasets of labelled data are readily available in the form of manually picked earthquake catalogues for many regions. We are now starting to see the adoption of deep-learning based methods to solve problems in seismological processing (e.g. Perol et al., 2018; Yoon et al., 2015, Ross et al., 2018; Zhu & Berozza, 2018; Titos et al., 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) catalogues of labelled examples collected over decades (Ross et al., 2018; Zhu & Berozza, 2018). We now investigate the dependency of the input data on classification performance by applying a CNN to classify seismic phases, where the network is trained over a relatively small catalogue of events (~11,000 P & S phase pairs). Can a relatively simple CNN architecture display similar performance improvements in the absence of an extensive training dataset? If a feature engineering approach demonstrates generalisation 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 catalogue of 411 events containing approximately 11,000 P-/S- phase pairs, located throughout the Iquique region of Northern Chile. The training catalogue has also been used to perform a minimum 1D velocity inversion (Woollam et al., 2019). Events occurred between March-May 2014 and 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 100Hz (Figure 1).

Manual picking of events was performed using Seismic Date eXplorer (SDX) software http://doree.esc.liv.ac.uk:8080/sdx/. We process the dataset applying a linear detrend. Whilst the CNN approach is shown to learn the characteristics of P-phases, S-phases, and noise (Zhu & Beroza, 2018), due to 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 2-25 Hz, a frequency range which lies in the passband of all instruments deployed.

Manual picks are represented probabilistically as a Gaussian function (σ = 1s, Figure 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' where the output probabilities were not impulsive enough to obtain a definitive phase-onset. Values lower than 1 second resulted in a high proportion of 'miss-picks' as 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 generalisation. 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, varying levels of Gaussian noise are then added to each batch[1], resulting in greater variations of signal vs. 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 catalogue of manually selected earthquakes for further studies. The input window size for the CNN is 6 seconds. To train the CNN, a given input batch is sequentially windowed with a timestep of 0.4 seconds. The windows are randomly shuffled before being used in training, preventing the CNN learning any unnecessary temporal order. A small time step is used to increase the total number of events during training; also, having the network learn to recognise the presence of phases at any point in the input window will help the network generalise 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 arrival.

[1] More information on parameters used to aiding generalisation provided in Data Appendix, section A.1

Methodology

Network architecture

The input to the network compromises three one-dimensional windows (x), where each window samples an individual component. For this given input, the network outputs the probability of either P-phase, S-phase, or Noise for each time sample within that window (Figure 4). Probabilities are output by applying the 'softmax' or normalised exponential function to the final layer

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

Where $\mathrm{j} = 1,2,3$ represents the P-phase, S-phase and Noise classes, $\alpha(\mathrm{x})$ contains the associated weights for the final layer. The input data are passed through repeated transformations; convolutional operations initially extract the appropriate features to characterise each class, the extracted features then go through repeated re-sampling stages, to output per-class probabilities. At each stage, a Rectified Linear Unit (ReLU) activation function is applied (Nair & Hinton, 2010). The cost function used to train the CNN is given by the negative log-likelihood $NLL(x, \theta)$. For a multi-class classification problem, where each class is characterised as a series of discrete probability distributions, $NLL(x, \theta)$ is also termed the cross-entropy loss function,

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

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). Eq. 2 is minimised using Adaptive Moment Estimation (ADAM, Kingma & Ba, 2014) along with batch training, the network weights are therefore updated at the end of each batch, over n training instances.

Hyperparameter optimisation (Bergstra & Bengio, 2012) is the derivation of the optimal network parameters and is a major challenge when designing neural network architectures. Parameters such as, number of layers, regularisation of layers, convolutional kernel shape, and the learning rate can all be optimised. 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. Once a robust network architecture is derived, a constrained search is performed for the best combination of hyper-parameters.

Our final network architecture consists of 3 convolutional layers, followed by 3 layers of up-sampling (Figure 4). Again, due to the limited nature of the training dataset, the focus for the network architecture is to limit the potential for overfitting. To localise the features corresponding to different classes, convolutional layers apply strided 1D convolutional filters along each component (Figure 3). The stride for the convolutional window is set to 4, this down-samples the time series 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 regularisation technique which randomly drops weights during training, reducing model complexity (Srivastava et al., 2014). One-dimensional 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 (Figure 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 $[p_{start}..p_{end}]$. The P-phase onset is set at the index of the maximum P-phase probability within this range. If the P-phase criterium is met, the corresponding S-phase is searched for within the searched window $[s_{start}..s_{end}]$, if $\sum_{i=s_{start}}^{s_{end}} s_i > s_{cut}$ 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:S phase picks using these criteria is 1:1. The parameters used in detecting phase onsets are provided in Table 1 – note that all index values are relative to the initial p_{cut} index:

Results

Predictions

The trained network takes a 6 second input window for 3-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-phases and S-phases, further confirming that deep-learning based classification methods engineer the appropriate features to accurately categorise P, S, and noise classes. This presents a major advantage over historic auto-picking methods which utilise manual feature extraction and often struggled to identify the S-phase. To obtain P/S phase onsets, we apply our autopicker function, with input parameters of Table 1, taking advantage of the simple temporal relationship between P and S phases to assign phase onsets (vertical lines on Figure 6). The phase onsets are then compared against the original manual picks and the residuals are plotted (Figure 7).

The residual distribution for the test dataset displays a good agreement in the centre of both the P- and S- residual distribution; however, the CNN has also picked extra events/phases in some waveforms. These extra phase picks may be accurate; however, any additional events are not represented in our classification vectors as 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 auto-picker performance. We perform an iterative inversion procedure, relocating both the original manual picks and the CNN picks for the initial dataset. The catalogues are relocated

using the VELEST routine (Kissling *et al*., 1994), which applies a minimum-1D velocity model along with station corrections to solve for hypocentre locations. Hypocentral parameters are solved for all events within the catalogue. 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 (+3s) 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 RMS residual larger than 3s to remove any picked events overlapping in time. Statistics of the residual distribution for the original manual picks compared against the CNN picks is 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 0.051s); however, S-phase picks of the CNN approach achieve similar performance to manual picking (σ decreased by 0.019s). 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 catalogue of events throughout Northern Chile, on the same temporary seismic network. Events were initially segmented using an iterative approach based on a STA/LTA trigger (Rietbrock et al., 2012) and provides 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 are provided in Table 3.

Figure 8 displays an event from the new catalogue 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 hypocentre distribution for both methods are 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 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 catalogues (minimum azimuthal gap < 220°) and for only the best-located events (minimum azimuthal gap < 160°). 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 when compared to the optimised STA/LTA approach.

The relocation residuals (Figure 10) are not just dependent upon accuracy of detected phases, but also on velocity variations not captured in the 1D model or station corrections affecting the residuals. As both catalogues were relocated with the same iterative re-location 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 catalogue. Investigating the residual distribution, the CNN approach has markedly improved both the overall relocation residual (Figure 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 0.230s for P-phases and 0.326s for S-phases when compared against the optimised 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 & Future work

Accurate and consistent catalogues of phase arrivals are of paramount importance to seismologists, as 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, as they only require the 3-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°), resulting in a decrease of 0.230s and 0.326s, respectively, when compared against an optimised STA/LTA picking approach (Rietbrock et al., 2012). The decreased variation in residual, indicates that a CNN based method is more consistent when auto picking, resulting in more accurate hypocentre relocations. We are close to reaching a point where supervised-learning based methods exhibit comparable or even increased performances when compared to manual picking by an expert seismologist (Zhu & Beroza, 2018; Ross et al., 2018). Until now, supervised learning-based methods have been trained using extensive training datasets (~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 generalisation, such methods can improve seismological processing workflows, not just for large catalogues, but for varying datasets. Future applications of deep-learning based methods in seismology include deploying such

pre-trained systems on poorly monitored areas/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 Institutes for Seismology (IRIS) data management centre for the temporary network data and also from the GEOFON data repository https://geofon.gfz-potsdam.de/waveform/archive/.

Acknowledgements

This work was partly supported by NERC grant: NE/M005879/1.

References

Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, *68*(5), pp.1521-1532.

Allen, R., 1982. Automatic phase pickers: their present use and future prospects. *Bulletin of the Seismological Society of America*, *72*(6B), pp.S225-S242.

Baer, M. and Kradolfer, U., 1987. An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, *77*(4), pp.1437-1445.

Baillard, C., Crawford, W.C., Ballu, V., Hibert, C. and Mangeney, A., 2013. An automatic kurtosis-based P-and S-phase picker designed for local seismic networks. *Bulletin of the Seismological Society of America*, *104*(1), pp.394-409.

Bergstra, J. and Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(Feb), pp.281-305.

Dai, H. and MacBeth, C., 1995. Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. *Geophysical journal international*, *120*(3), pp.758-774.

Dai, H. and MacBeth, C., 1997. The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings. *Journal of Geophysical Research: Solid Earth*, *102*(B7), pp.15105-15113.

Earle, P.S. and Shearer, P.M., 1994. Characterization of global seismograms using an automatic-picking algorithm. *Bulletin of the Seismological Society of America*, *84*(2), pp.366-376.

Gentili, S. and Michelini, A., 2006. Automatic picking of P and S phases using a neural tree. *Journal of Seismology*, *10*(1), pp.39-63.

Gomberg, J.S., Shedlock, K.M. and Roecker, S.W., 1990. The effect of S-wave arrival times on the accuracy of hypocenter estimation. *Bulletin of the Seismological Society of America*, *80*(6A), pp.1605-1628.

Hayes, G.P., Wald, D.J. and Johnson, R.L., 2012. Slab1. 0. A three-dimensional model of global subduction zone geometries. *Journal of Geophysical Research: Solid Earth*, *117*(B1).

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L., 2014. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 1725-1732).

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

Kissling, E., Ellsworth, W.L., Eberhart-Phillips, D. and Kradolfer, U., 1994. Initial reference models in local earthquake tomography. *Journal of Geophysical Research: Solid Earth*, *99*(B10), pp.19635-19646.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Küperkoch, L., Meier, T., Lee, J., Friederich, W. and EGELADOS Working Group, 2010. Automated determination of P-phase arrival times at regional and local distances using higher order statistics. *Geophysical Journal International*, *181*(2), pp.1159-1170.

LeCun, Y. and Bengio, Y., 1995. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), p.1995.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, *521*(7553), p.436.

347 Leonard, M. and Kennett, B.L.N., 1999. Multi-component autoregressive techniques for the
348 analysis of seismograms. *Physics of the Earth and Planetary Interiors*, *113*(1-4), pp.247-263.

349 Nair, V. and Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines.
350 In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807-
351 814).

352 Nippress, S.E.J., Rietbrock, A. and Heath, A.E., 2010. Optimized automatic pickers:
353 application to the ANCORP data set. *Geophysical Journal International*, *181*(2), pp.911-925.

354 Perol, T., Gharbi, M. and Denolle, M., 2018. Convolutional neural network for earthquake
355 detection and location. *Science Advances*, *4*(2), p.e1700578.

356 Rastin, S.J., Unsworth, C.P., Benites, R. and Gledhill, K.R., 2013. Using real and synthetic
357 waveforms of the Matata swarm to assess the performance of New Zealand GeoNet phase
358 pickers. *Bulletin of the Seismological Society of America*, *103*(4), pp.2173-2187.

359 Rietbrock, A., Ryder, I., Hayes, G., Haberland, C., Comte, D., Roecker, S. and Lyon-Caen,
360 H., 2012. Aftershock seismicity of the 2010 Maule Mw= 8.8, Chile, earthquake: Correlation
361 between co-seismic slip models and aftershock distribution?. *Geophysical Research*
362 *Letters*, *39*(8).

363 Ross, Z.E. and Ben-Zion, Y., 2014. Automatic picking of direct P, S seismic phases and fault
364 zone head waves. *Geophysical Journal International*, *199*(1), pp.368-381.

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

367 Saragiotis, C.D., Hadjileontiadis, L.J. and Panas, S.M., 2002. PAI-S/K: A robust automatic
368 seismic P phase arrival identification scheme. *IEEE Transactions on Geoscience and*
369 *Remote Sensing*, *40*(6), pp.1395-1404.

370 Saragiotis, C.D., Hadjileontiadis, L.J., Rekanos, I.T. and Panas, S.M., 2004. Automatic P
371 phase picking using maximum kurtosis and/spl kappa/-statistics criteria. *IEEE Geoscience*
372 *and Remote Sensing Letters*, *1*(3), pp.147-151.

373 Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural networks*, *61*,
374 pp.85-117.

375 Sleeman, R. and Van Eck, T., 1999. Robust automatic P-phase picking: an on-line
376 implementation in the analysis of broadband seismogram recordings. *Physics of the earth*
377 *and planetary interiors*, *113*(1-4), pp.265-275.

378 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014.
379 Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine*
380 *Learning Research*, *15*(1), pp.1929-1958.

381 Titos, M., Bueno, A., García, L. and Benítez, C., 2018. A Deep Neural Networks Approach to
382 Automatic Recognition Systems for Volcano-Seismic Events. *IEEE Journal of Selected*
383 *Topics in Applied Earth Observations and Remote Sensing*, *11*(5), pp.1533-1544.

384 Wang, J. and Teng, T.L., 1995. Artificial neural network-based seismic detector. *Bulletin of*
385 *the Seismological Society of America*, *85*(1), pp.308-319.

386 Woollam, J., Rietbrock, A., et al., Determination of a minimum 1D seismic velocity model for
387 the Iquique region of Northern Chile. *Geophysical Journal International,* submitted.

Yoon, C.E., O'Reilly, O., Bergen, K.J. and Beroza, G.C., 2015. Earthquake detection through computationally efficient similarity search. *Science advances*, *1*(11), p.e1501057.

Zhao, Y. and Takano, K., 1999. An artificial neural network approach for broadband seismic phase picking. *Bulletin of the Seismological Society of America*, *89*(3), pp.670-680.

Zhu, W. and Beroza, G.C., 2018. PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method. *arXiv preprint arXiv:1803.03211*.

Appendix

A.1 Hyper-parameters and data generalisation parameters.

Lognormal distribution used to scale individual event amplitudes is given by

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

We set μ = 0, and σ = 0.25, and sample the output probability distribution of eq. A.1, each sample is then used as a scale factor for event amplitudes.

Mailing addresses

Jack Woollam

Jack.Woollam@liverpool.ac.uk

Andreas Rietbrock

andreas.rietbrock@kit.edu

Angel Bueno

angelbueno@ugr.es

Silvio De Angelis

S.De-Angelis@liverpool.ac.uk

Figure captions

Figure 1 | Distribution of manually picked events throughout Northern Chile, stations are indicated by white triangles, event hypocentres are plotted as a function of depth.

Figure 2 | An example of input data (top) and classification data (bottom), inputs to the CNN are 3-component traces, linear-detrended, bandpass filtered between 2-25Hz. The associated classification vector for P-pick and S-pick are represented probabilistically as a Gaussian with σ = 1s.

Figure 3 | Schematic displaying how strided 1D convolutions quickly incorporate the long-term temporal dependencies of the input data into the convolution kernel.

Figure 4 | Overall CNN architecture, displaying the sequential convolution and re-sampling operations applied to the input window.

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 labelled above each vertical dashed line. Vertical dotted lines indicate the start or end of a P-/S-phase search window, where 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.

Figure 6 | Output CNN prediction probabilities when applied to identify phase onsets for the test dataset, phase onsets are indicated by vertical lines.

Figure 7 | Residual of CNN predicted phase onsets vs. original manual picks for the test dataset.

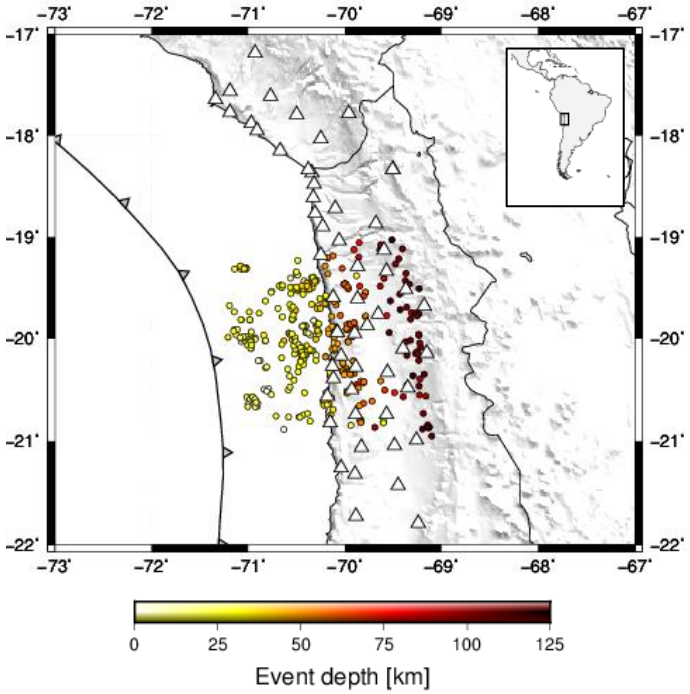
Figure 8 | Demonstrating the CNN auto picker performance on a new dataset for Northern Chile, where events were segmented using an STA/LTA trigger (Rietbrock et al., 2012). We only allow the auto picker to pick the presence of a single P-/S-phase per trace, to prevent relocation errors.

Figure 9 | Hypocentre relocation comparison for the STA/LTA auto picked catalogue (top) against the CNN auto picked catalogue (bottom) event relocations are plotted as a function of RMS residual, slab profile is provided by Hayes et al., (2012).

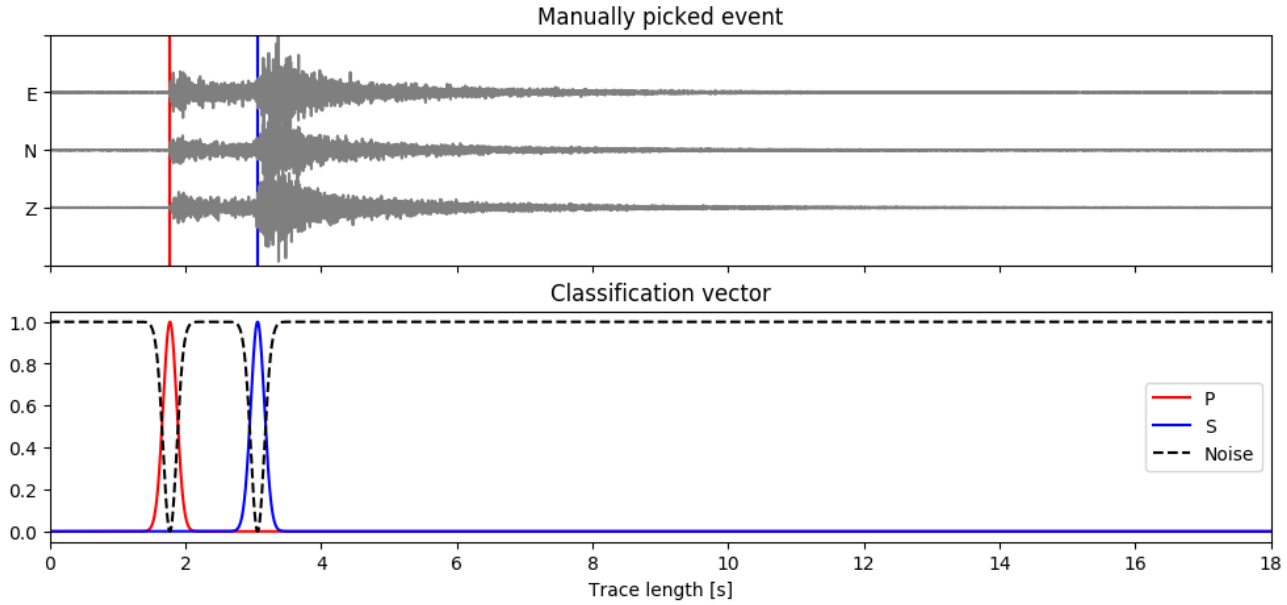
Figure 10 | Both auto picking methods phase residuals following hypocentral relocations, plotted for well-located events (minimum azimuthal gap < 160°) and the for entire relocated catalogues (minimum azimuthal gap < 220°).

Figures

Figure 1



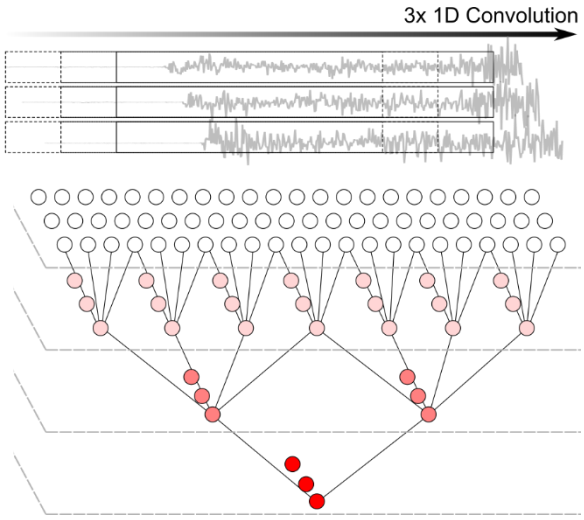
454 Figure 2



455

456

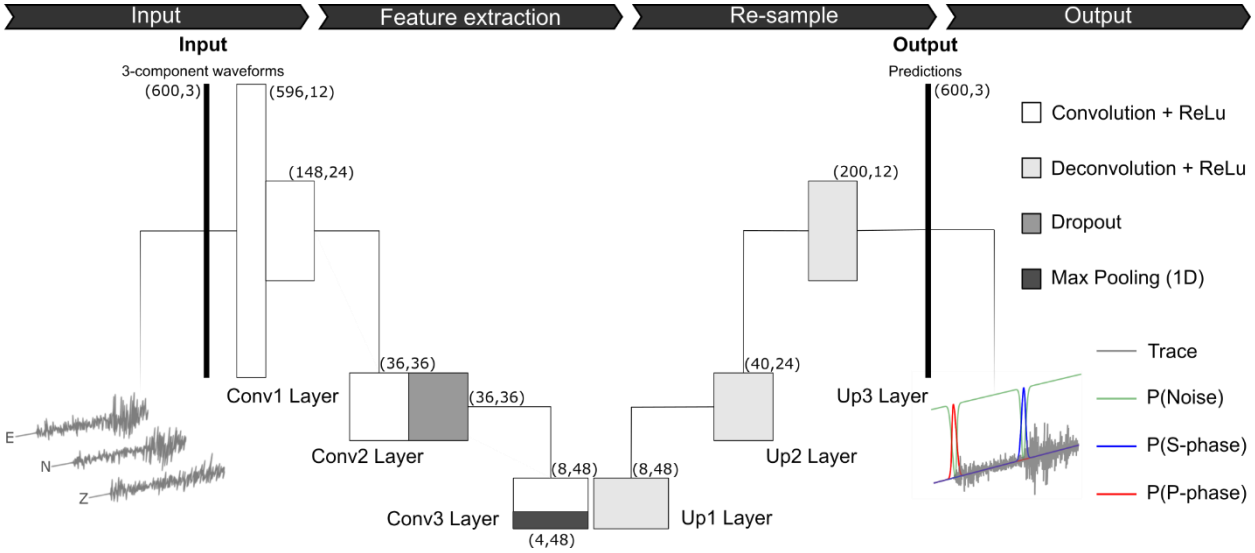
457 Figure 3



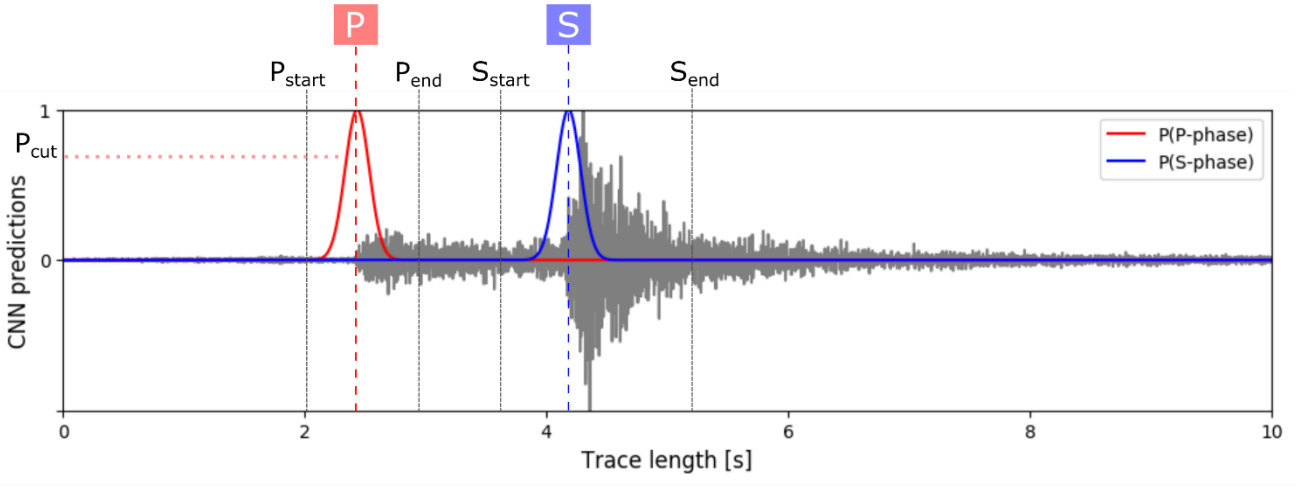
458

459

Figure 4


Input
Feature extraction
Re-sample
Output
Input
3-component waveforms
(600,3)
(596,12)
(148,24)
(36,36)
(36,36)
(8,48)
(4,48)
(8,48)
(40,24)
(200,12)
Output
Predictions
(600,3)
Conv1 Layer
Conv2 Layer
Conv3 Layer
Up1 Layer
Up2 Layer
Up3 Layer
E
N
Z
Convolution + ReLu
Deconvolution + ReLu
Dropout
Max Pooling (1D)
Trace
P(Noise)
P(S-phase)
P(P-phase)

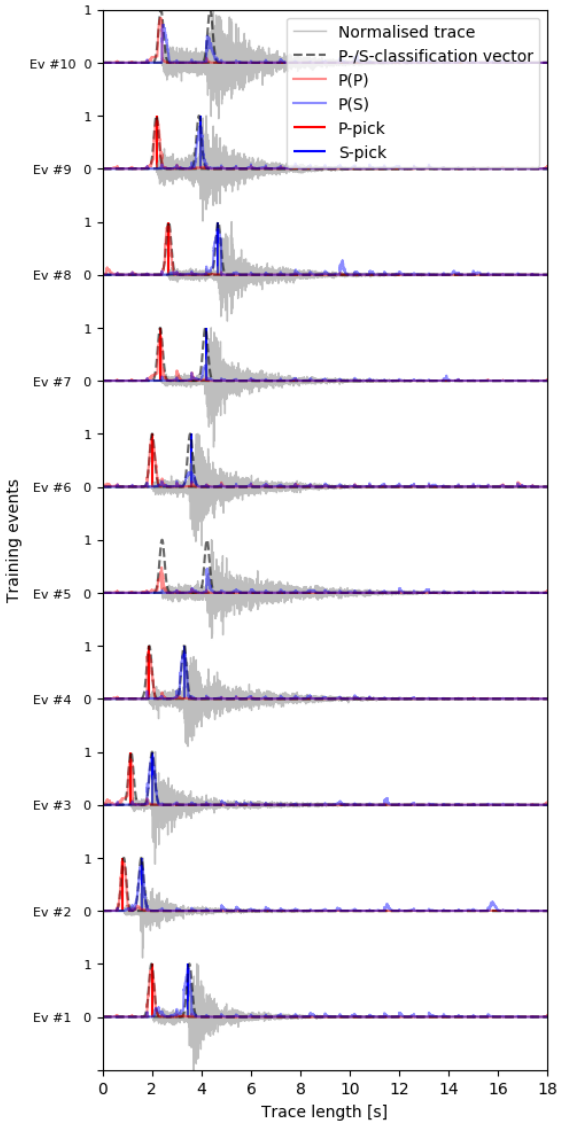
463 Figure 5



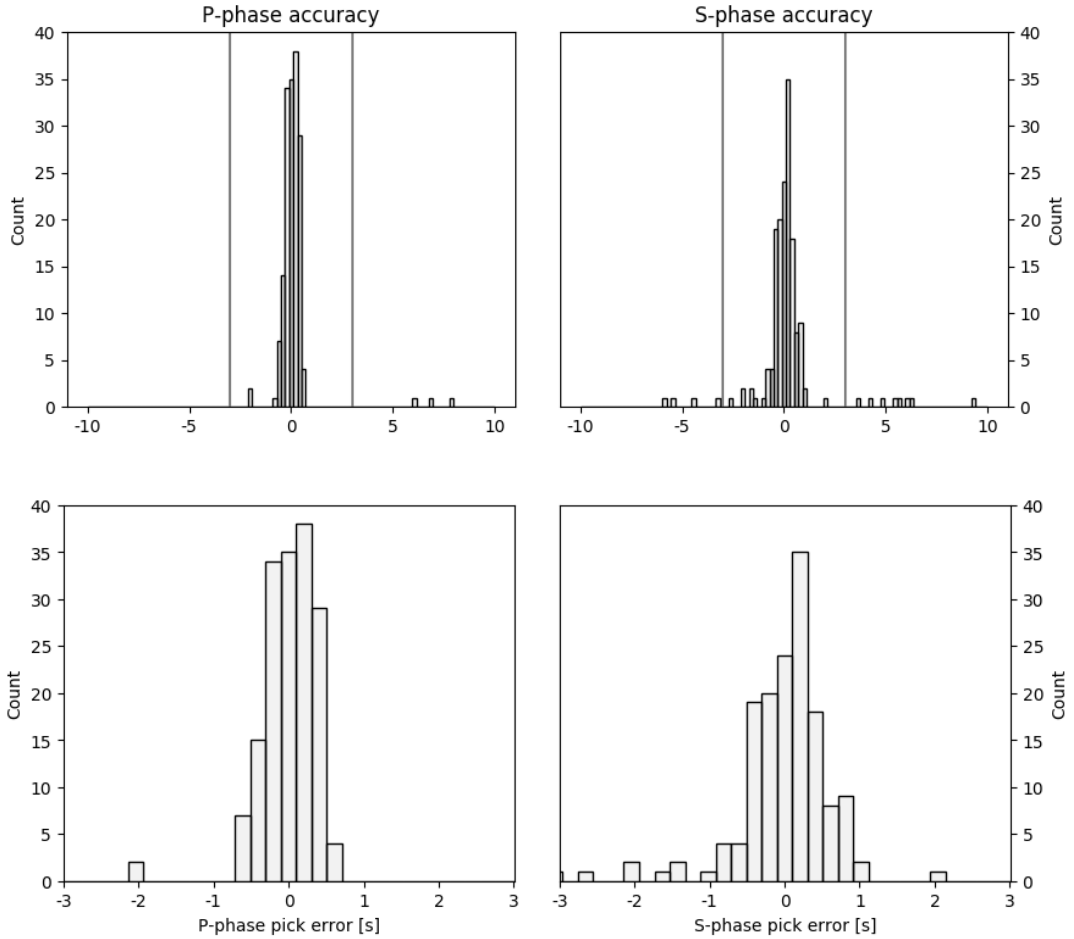
464

465

Figure 6


Normalised trace
P-/S-classification vector
P(P)
P(S)
P-pick
S-pick
Ev #10
Ev #9
Ev #8
Ev #7
Ev #6
Ev #5
Ev #4
Ev #3
Ev #2
Ev #1
Training events
0
2
4
6
8
10
12
14
16
18
Trace length [s]

469 Figure 7

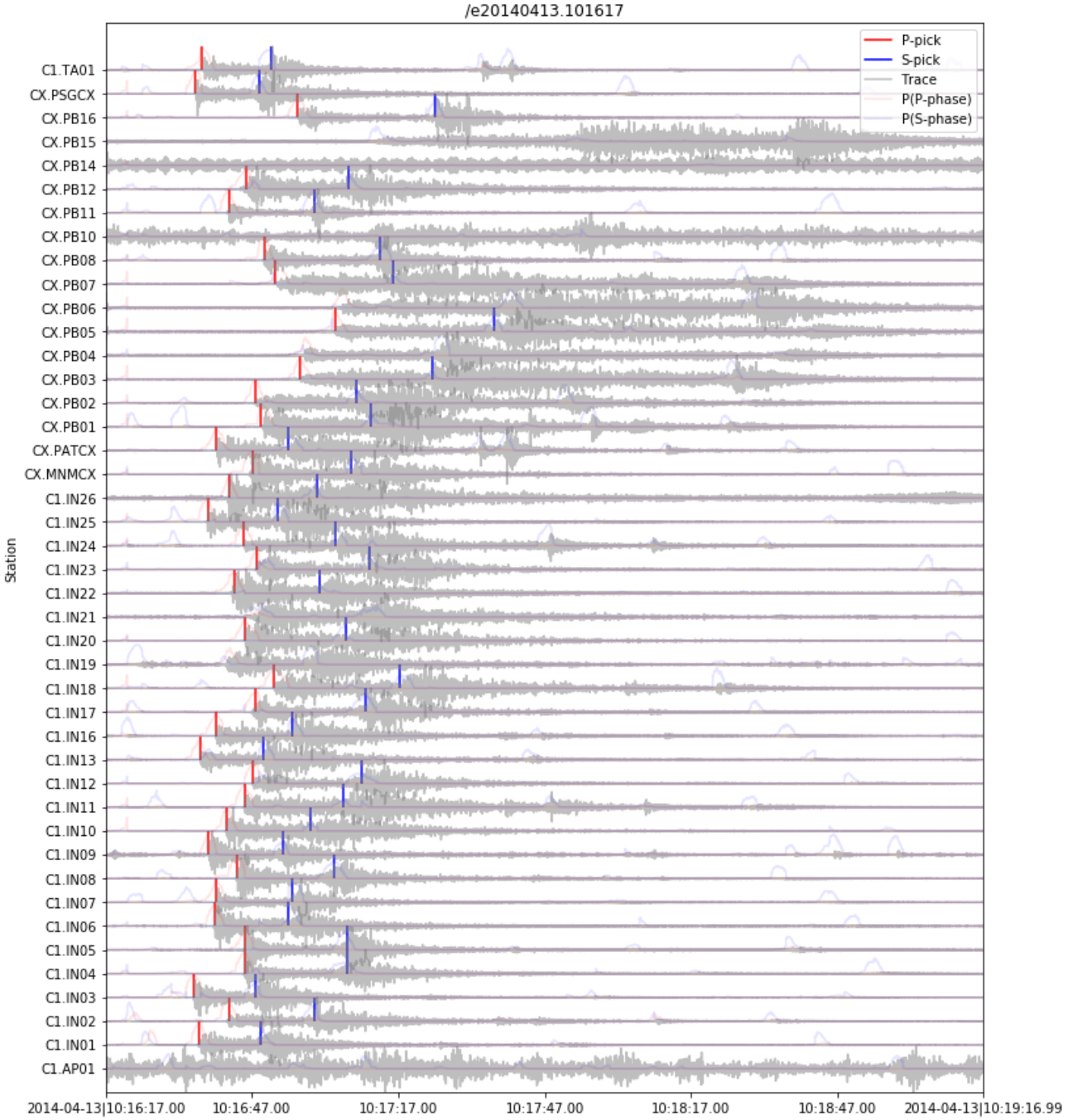


470

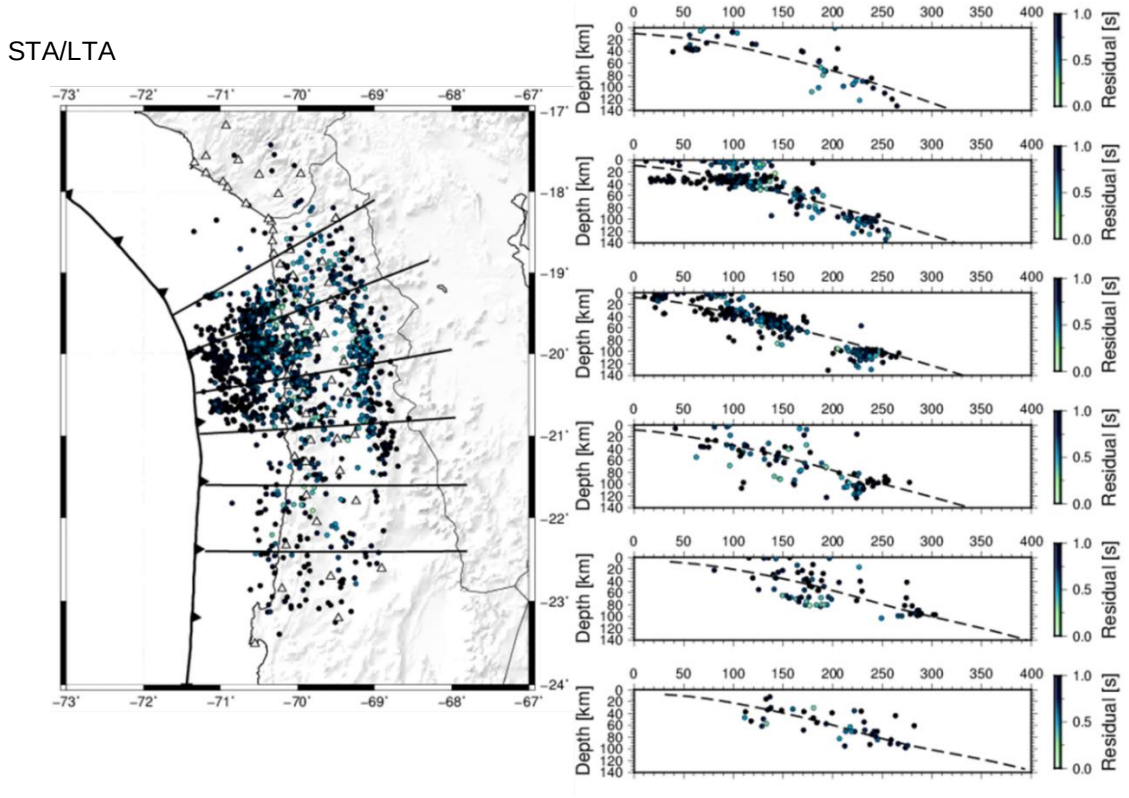
471

472

Figure 8


/e20140413.101617
P-pick
S-pick
Trace
P(P-phase)
P(S-phase)
C1.TA01
CX.PSGCX
CX.PB16
CX.PB15
CX.PB14
CX.PB12
CX.PB11
CX.PB10
CX.PB08
CX.PB07
CX.PB06
CX.PB05
CX.PB04
CX.PB03
CX.PB02
CX.PB01
CX.PATCX
CX.MNMCX
C1.IN26
C1.IN25
C1.IN24
C1.IN23
C1.IN22
C1.IN21
C1.IN20
C1.IN19
C1.IN18
C1.IN17
C1.IN16
C1.IN13
C1.IN12
C1.IN11
C1.IN10
C1.IN09
C1.IN08
C1.IN07
C1.IN06
C1.IN05
C1.IN04
C1.IN03
C1.IN02
C1.IN01
C1.AP01
Station
2014-04-13|10:16:17.00
10:16:47.00
10:17:17.00
10:17:47.00
10:18:17.00
10:18:47.00
2014-04-13|10:19:16.99

476 Figure 9



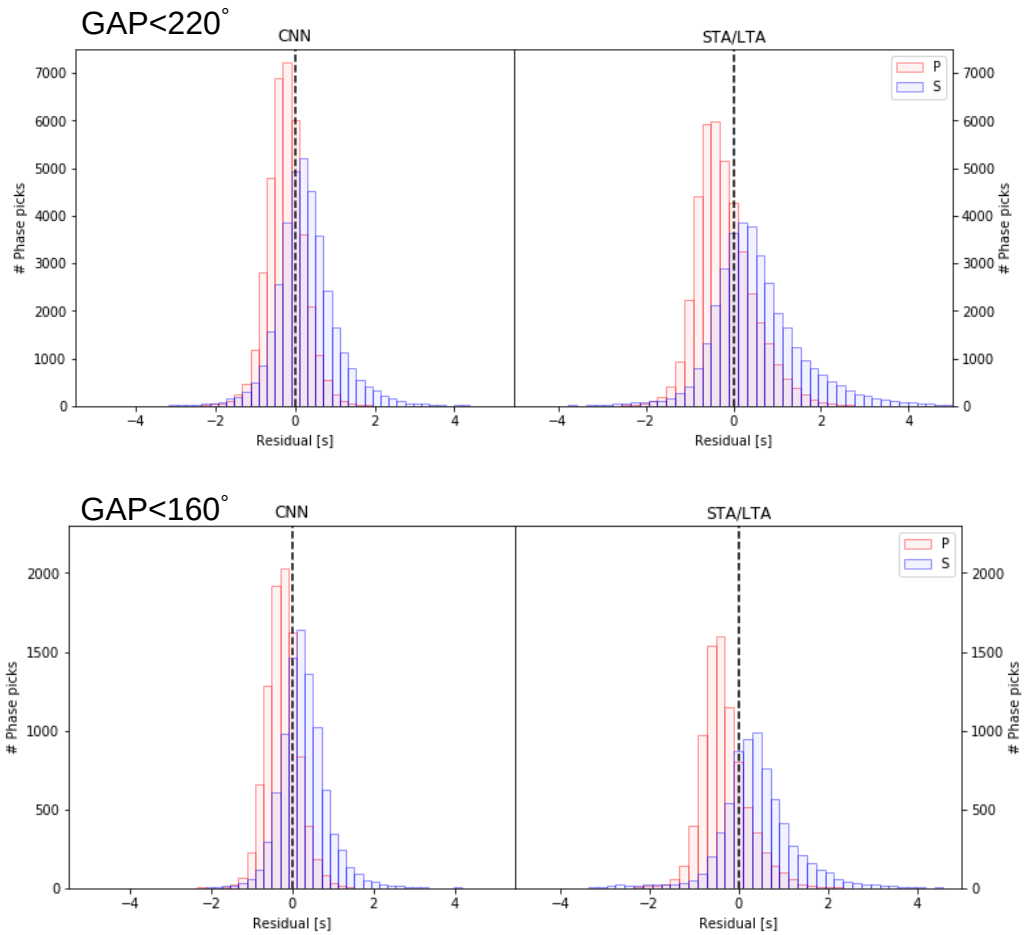
477

478

479

480 Figure 10

481



482

483

484

Tables

Table 1

Table 1 | Parameters applied to the autopicker function, which takes advantage of the temporal relationship between phases to identify phase onsets, all start/end indexes are given in samples (where sampling rate for all instruments = 100Hz).

p_{cut}	0.75	s_{sum}	5
p_{start}	-200	s_{start}	500
p_{end}	200	s_{end}	4000

Table 2

Table 2 | Statistics of residual distribution for original manually picked catalogue, both for the original manual picks and the CNN methodology picks.

GAP < 220°		CNN	Manual
P	μ	-0.261	**-0.124**
	σ	0.445	**0.394**
S	μ	**0.282**	0.390
	σ	0.749	**0.730**

Table 3

Table 3 | Overall auto picks on a separate catalogue of new events throughout Northern Chile.

	STA/LTA	CNN
P	72,655	77,623
S	63,353	77,623
Total	136,008	155,246

Table 4

Table 4 | Statistics of residual distribution for GAP < 220° and GAP < 160°.

		GAP < 220°		GAP < 160°	
		CNN	STA/LTA	CNN	STA/LTA
P	μ	-0.238	**-0.216**	**-0.247**	-0.333
	σ	**0.487**	0.696	**0.393**	0.623
S	μ	**0.277**	0.539	**0.270**	0.435
	σ	**0.780**	1.081	**0.596**	0.922