

PhaseNet: a deep-neural-network-based seismic arrival-time picking method

Weiqliang Zhu and Gregory C. Beroza

Department of Geophysics, Stanford University, CA 94035-2215, USA. E-mail: zhuwq@stanford.edu

Accepted 2018 October 11. Received 2018 October 9; in original form 2018 March 08

SUMMARY

As the number of seismic sensors grows, it is becoming increasingly difficult for analysts to pick seismic phases manually and comprehensively, yet such efforts are fundamental to earthquake monitoring. Despite years of improvements in automatic phase picking, it is difficult to match the performance of experienced analysts. A more subtle issue is that different seismic analysts may pick phases differently, which can introduce bias into earthquake locations. We present a deep-neural-network-based arrival-time picking method called “PhaseNet” that picks the arrival times of both *P* and *S* waves. Deep neural networks have recently made rapid progress in feature learning, and with sufficient training, have achieved super-human performance in many applications. PhaseNet uses three-component seismic waveforms as input and generates probability distributions of *P* arrivals, *S* arrivals and noise as output. We engineer PhaseNet such that peaks in the probability distributions provide accurate arrival times for both *P* and *S* waves. PhaseNet is trained on the prodigious available data set provided by analyst-labelled *P* and *S* arrival times from the Northern California Earthquake Data Center. The data set we use contains more than 700 000 waveform samples extracted from over 30 yr of earthquake recordings. We demonstrate that PhaseNet achieves much higher picking accuracy and recall rate than existing methods when applied to the waveforms of known earthquakes, which has the potential to increase the number of *S*-wave observations dramatically over what is currently available. This will enable both improved locations and improved shear wave velocity models.

Key words: Neural networks, fuzzy logic; Time-series analysis; Body waves; Computational seismology; Earthquake monitoring and test-ban treaty verification.

1 INTRODUCTION

Earthquake detection and location are fundamental to seismology. The quality of earthquake catalogues depends critically on both the number and accuracy of arrival-time measurements. Earthquake arrival-time measurement, or phase picking, is often carried out by network analysts who base their phase picks on expert judgment and years of experience. As the rate of seismometer deployment continues to accelerate; however, it is becoming increasingly difficult to keep up with the data flow. This is particularly true for dense networks in areas of particular interest or concern that now may contain over 1000s of sensors. Phase pickers are particularly challenged by *S* waves, because they are not the first arriving waves, and they emerge from the scattered waves of the *P* coda. *S*-wave arrival times are particularly useful because they can be used to reduce the depth-origin trade-off that can afflict earthquake locations based on *P* waves alone, and because *S*-wave structure is important for strong ground motion prediction.

Decades of research has been devoted to automatic phase picking, including methods based on amplitude, standard deviation or

energy; statistical methods and shallow neural networks. The short-term average/long-term average (STA/LTA) method (Allen 1978) is commonly used and tracks the ratio of energy in a short-term window with that in a long-term window. Peaks above a threshold mark impulsive *P* or *S* wave arrivals. This method is efficient, often effective, but susceptible to noise and has low accuracy for arrival times, particularly for shear waves. Baer & Kradolfer (1987) improved the STA/LTA method using the envelope as the characteristic function. Sleeman & Van Eck (1999) applied joint autoregressive (AR) modelling of the noise and seismic signal and used the Akaike Information Criterion to determine the onset of a seismic signal. Approaches based on higher-order statistics, including kurtosis and skewness, were developed to identify the transition from Gaussianity to non-Gaussianity, which coincides with the onset of the seismic event, even in the presence of noise (Saragiotis *et al.* 2002; Küperkoch *et al.* 2010). Traditional shallow neural networks were tested by Gentili & Michelini (2006) to pick *P* and *S* phases, based on four manually defined features: variance, absolute value of skewness, kurtosis and a combination of skewness and kurtosis predicted based on sliding windows. While most phase picking

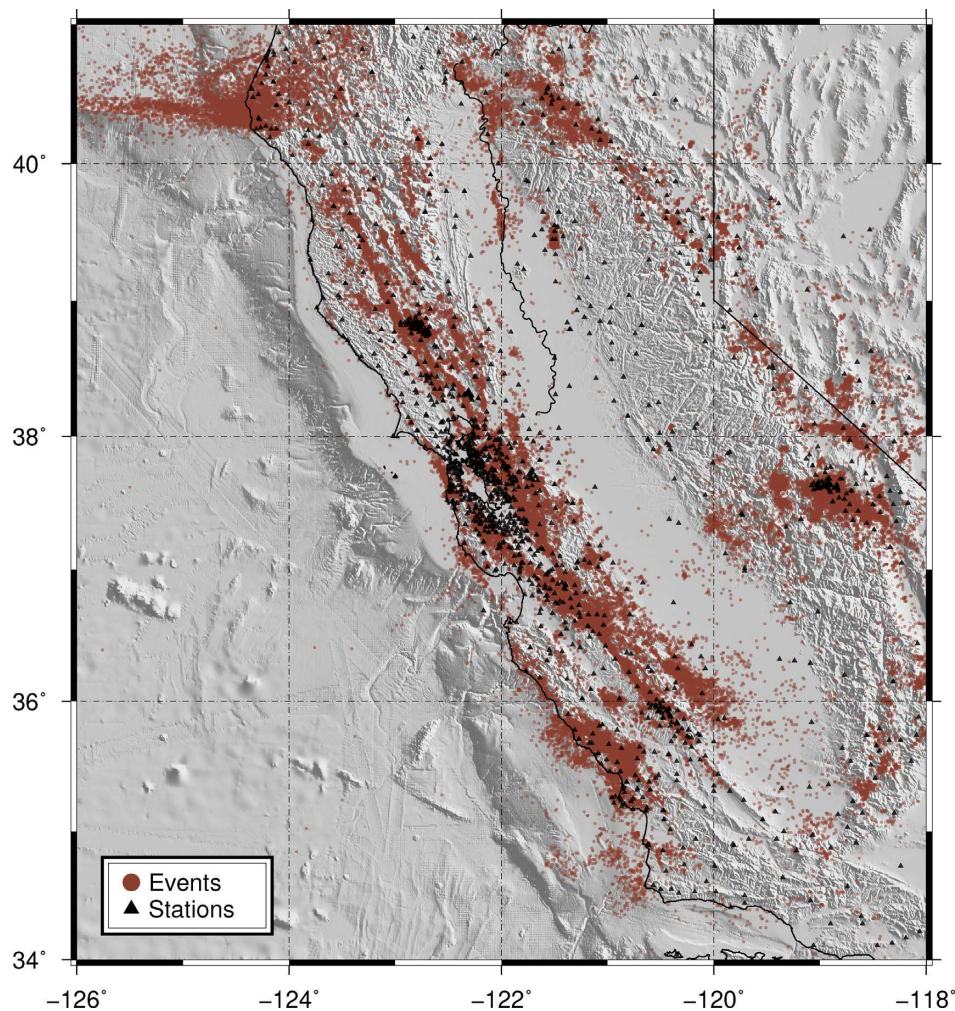


Figure 1. The locations of 234 117 earthquakes (red points) and 889 seismic stations (black triangles) in the Northern California Earthquake Catalog.

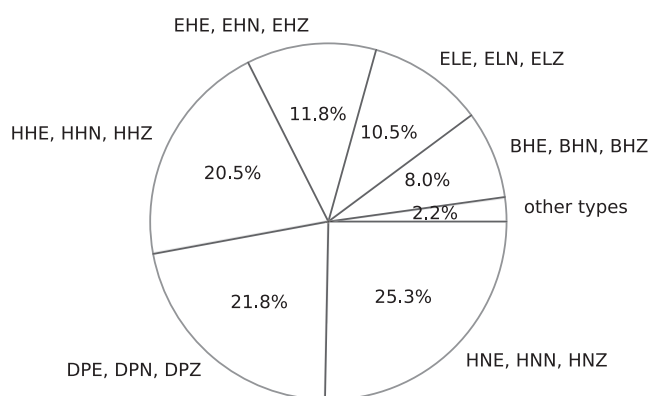


Figure 2. The proportion of different instrument types in the data set. The first letter is the band code: H: high broad band, D: very very short period, E: short period. The second letter is the instrument code: N: accelerometer, P: very short-period seismometer, H: high-gain seismometer, L: low-gain seismometer. The third letter is the orientation code: E: east–west direction, N: north–south direction, Z: vertical direction.

algorithms focus on *P* waves, Ross & Ben-Zion (2014) utilized polarization analysis to distinguish between *P* and *S* waves primarily to improve *S*-wave arrival time measurements. Despite the substantial efforts outlined above, the accuracy of automated phase picking

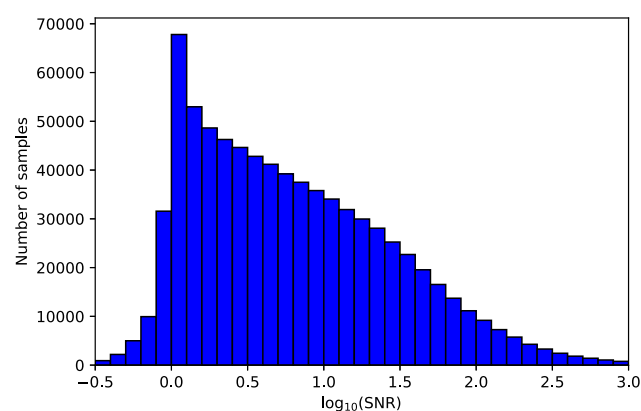


Figure 3. Signal-to-noise ratio (SNR) distribution. The SNR is calculated by the ratio of standard deviations of two 5 s windows following and preceding the *P* arrival.

algorithms lags that of experienced analysts. This is attributable to the fact that earthquake waveforms are highly complex due to multiple effects including source mechanism, stress drop, scattering, site-effects, phase conversions and interference from a multitude of noise sources. Traditional automated methods use manually defined features that require careful data processing, like bandpass filtering and setting an activation threshold.

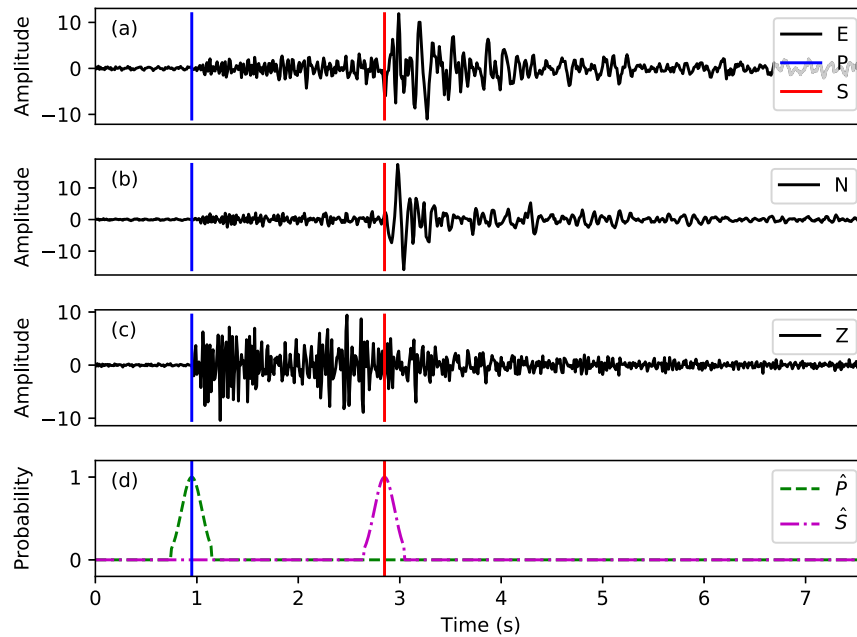


Figure 4. A sample from the data set. (a) – (c) Seismograms of the “ENZ” (East, North, Vertical) components. The blue and red vertical lines are the manually picked P and S arrival times. (d) The converted probability masks for P and S picks. The shape is a truncated Gaussian distribution with a mean (μ) of the arrival time and a standard deviation (σ) of 0.1 s.

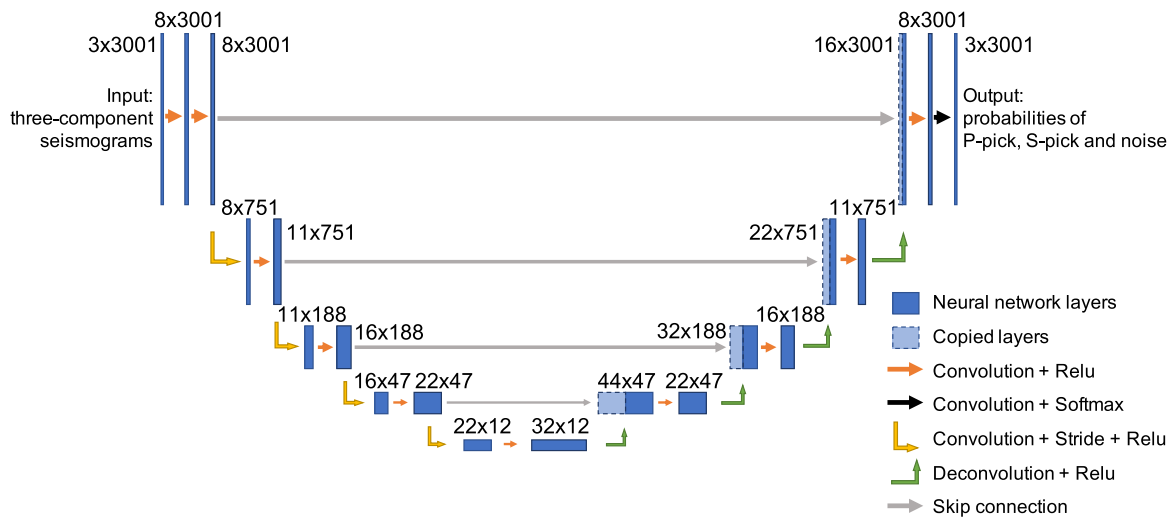


Figure 5. The network architecture. The input is the 30-s three-component seismograms sampled at 100 Hz, so the input has a dimension of 3×3001 . The output is three probabilities with the same length as input for P pick, S pick and noise. The blue rectangles represent layers inside the neural network. The numbers near them are the dimensions of each layer, which follow a format of “number of channels \times length of each channel”. The arrows are operations applied between layers, whose meanings are noted in the low right corner. The input seismic data go through four down-sampling stages and four up-sampling stages. The down-sampling is done by 1-D convolution and stride. We have set the length of convolution kernel to seven data points and the stride step to four data points. The up-sampling is done by deconvolution, which recovers the input length of the previous stage. A skip connection at each stage directly concatenates the left output to the right layer without going through the deeper layers, which improves convergence during training. The blue rectangles with dashed boundaries are the layers copied directly by the skip connection. The softmax normalized exponential function is used to set probabilities in the last layer.

In this paper, we present a deep neural network algorithm, PhaseNet, for seismic phase picking. Instead of using manually defined features, deep neural networks learn the features from labelled data, both noise and signal, which proves a powerful advantage for complex seismic waveforms. The network is trained on the substantial catalogue of available P and S arrival times picked by experienced analysts. Unfiltered three-component seismic waveforms

are the input to PhaseNet, which is trained to output three probability distributions: P wave, S wave and noise. The neural network is trained on the target probability distributions of known earthquake waveforms. Peaks in the P wave and S wave probability distributions are designed to correspond to the predicted P and S arrival times. We demonstrate that PhaseNet provides high accuracy and recall rate for both P and S picks, and achieves significant improve-

Table 1. Evaluation metrics on the test data set. Pickers with residuals ($\Delta t < 0.1$ s) are counted as true positive picks. The mean ($\mu(\Delta t)$) and standard deviation ($\sigma(\Delta t)$) are calculated on residuals ($\Delta t < 0.5$ s) whose distributions are shown in Fig. 6.

Evaluation indicator	Phase	PhaseNet	AR picker
Precision	<i>P</i>	0.939	0.558
	<i>S</i>	0.853	0.195
Recall	<i>P</i>	0.857	0.558
	<i>S</i>	0.755	0.144
F1 score	<i>P</i>	0.896	0.558
	<i>S</i>	0.801	0.165
$\mu(\Delta t)$ (ms)	<i>P</i>	2.068	11.647
	<i>S</i>	3.311	27.496
$\sigma(\Delta t)$ (ms)	<i>P</i>	51.530	83.991
	<i>S</i>	82.858	181.027

ment compared with a traditional STA/LTA method. PhaseNet has the potential to provide comprehensive, superior performance for standard earthquake monitoring.

2 DATA

Seismological archives include tremendous numbers of manually picked *P* and *S* wave arrivals, which represent an exceptionally rich training set of labelled data that is ideal for deep learning (Fig. 1). In this paper, we gathered available digital seismic waveform data based on the Northern California Earthquake Data Center Catalog (NCEDC 2014). We use three-component data that have both *P* and *S* arrival times. This leaves us 779 514 recordings in the data set. We use stratified sampling based on stations to divide this data set into training, validation and test data sets, with 623 054, 77 866 and 78 592 samples, respectively. The training and validation sets are used during training, fine-tuning parameters and model selection. The test set is only used to evaluate the final performance and

results of PhaseNet. This data set has a diversity of waveform characteristics. It includes different types of instruments in Northern California Seismic Network and covers a wide range of signal-to-noise ratio (SNR). The proportion of each instrument in the data set is shown in Fig. 2. The distribution of SNR is shown in Fig. 3. The SNR is calculated by the ratio of standard deviations of the 5 s following and the 5 s preceding the *P* arrival. The complexity of this data set makes it challenging for automatic phase picking, but it provides a more comprehensive performance evaluation.

We apply minimal data pre-processing to the training data. We randomly select a 30-s time window that includes the *P* and *S* arrival times as the input of PhaseNet. The position of the arrivals within the window are varied to ensure that the algorithm does not just learn the windowing scheme. All data are sampled at 100 Hz, which is the most common sampling rate in the raw data set, so that the 30-s input waveforms have 3001 data points for each component. We normalize each component waveform by removing its mean and dividing it by the standard deviation (Figs 4a–c). The manually picked time points in the data set may not be the true *P/S* arrivals,

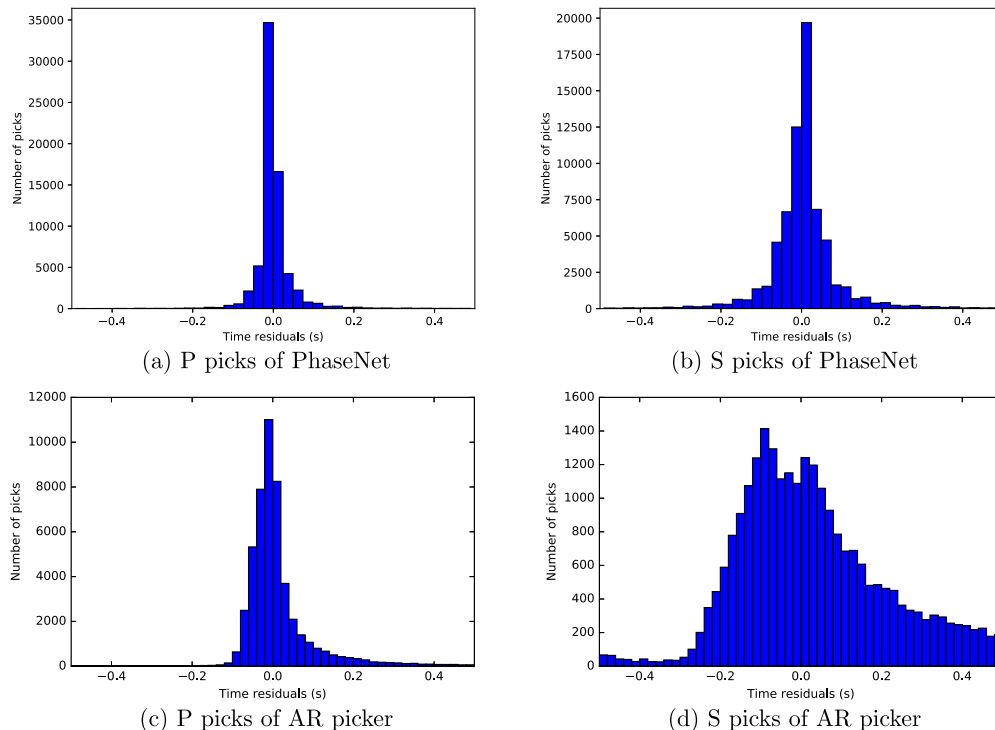


Figure 6. The distribution of residuals (Δt) of PhaseNet (upper panels) and AR picker (lower panels) on the test data set.

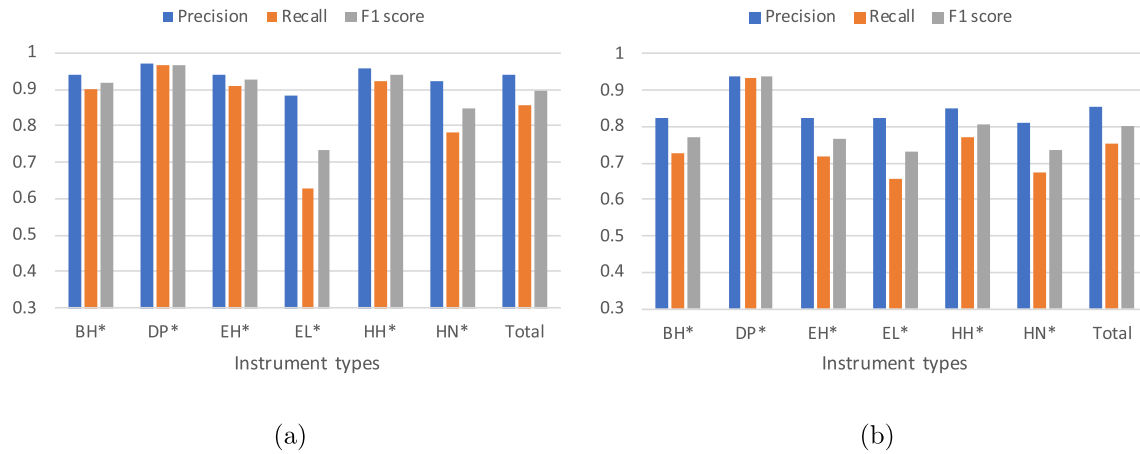


Figure 7. Performances on different instrument types. (a) P picks. (b) S picks. The meaning of x-axis labels is the same as in Fig. 2. The “total” data set is the same test data set used in Table 1.

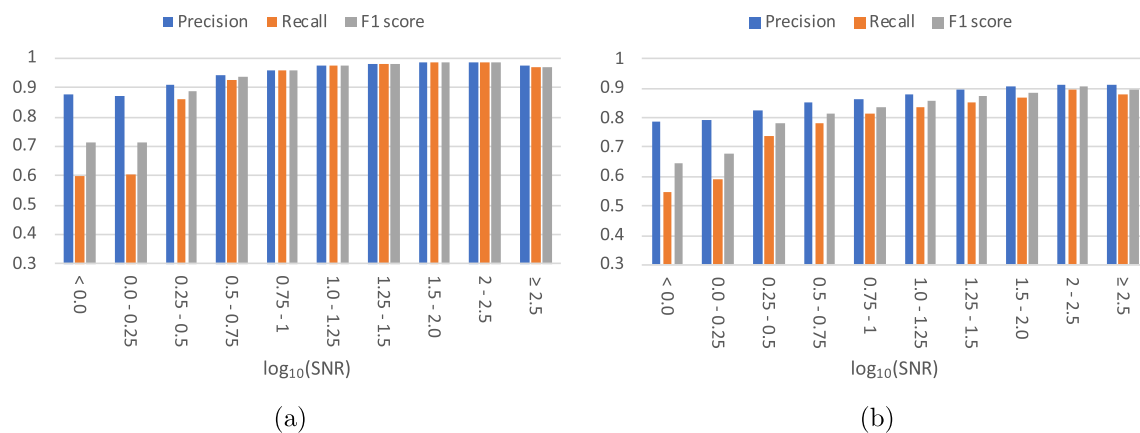


Figure 8. Performances of different signal-to-noise ratios (SNR). (a) P picks. (b) S picks.

but we expect the ground-truth arrival times will be centred on the manual picks with some uncertainty. For this reason we apply a mask with the shape of a Gaussian distribution around the manual picks. Therefore, the time point picked by analysts has the highest probability, while the nearby data points have reduced probabilities (Fig. 4d). The standard deviation of the Gaussian distribution is set to 0.1 s. Representing manual picks probabilistically allows the algorithm to reduce the influence of picking errors in the data set. Because we have considered the probabilities of the nearby data points, the mask increases the amount of information in P and S picks relative to noise, and helps accelerate convergence. Here the noise includes all data points that are not first arrivals of P or S waves. The probability distribution of noise is calculated by

$$\text{Prob}(\text{noise}) = 1 - \text{Prob}(P) - \text{Prob}(S),$$

where “Prob” is the probability of each class. After the conversion using a Gaussian distribution mask, we can extract accurate arrival times from the peaks of probability distributions predicted by PhaseNet.

3 METHOD

The architecture of PhaseNet (Fig. 5) is modified from U-Net (Ronneberger *et al.* 2015) to deal with 1-D time-series data. U-net is a deep neural network approach used in biomedical image processing that seeks to localize properties in an image. The mapping to our

problem is to localize the properties of our time-series into three classes: P pick, S pick and noise. The inputs are three-component seismograms of known earthquakes. The outputs are probability distributions of P wave, S wave and noise. In our experiments, the input and output sequences contain 3001 data points for each component (30 s long, sampled at 100 Hz). The input seismic data go through four down-sampling stages and four up-sampling stages. Inside each stage, we apply 1-D convolutions and rectified linear unit (ReLU) activations. The down-sampling process is designed to extract and shrink the useful information from raw seismic data to a few neurons, so each neuron in the last layer makes up a broadly receptive window. The up-sampling process expands and converts this information into probability distributions of P wave, S wave and noise for each time point. A skip connection at each depth directly concatenates the left output to the right layer without going through the deeper layer. This helps improve convergence during training (Ronneberger *et al.* 2015; Li *et al.* 2017). The 1-D convolution size is set to seven data points. The stride step for down-sampling is set to four data points, so after each stride the channel length is condensed into one-fourth of its original dimension, while the deconvolution operation (Noh *et al.* 2015) for up-sampling expands the condensed layers by a factor of four to recover its previous length. We have added padding at the front and the back of each layer during convolutions to make the input and output sequences have the same length. Fig. 5 shows the size of each layer and the operations of convolution and deconvolution. The softmax normalized

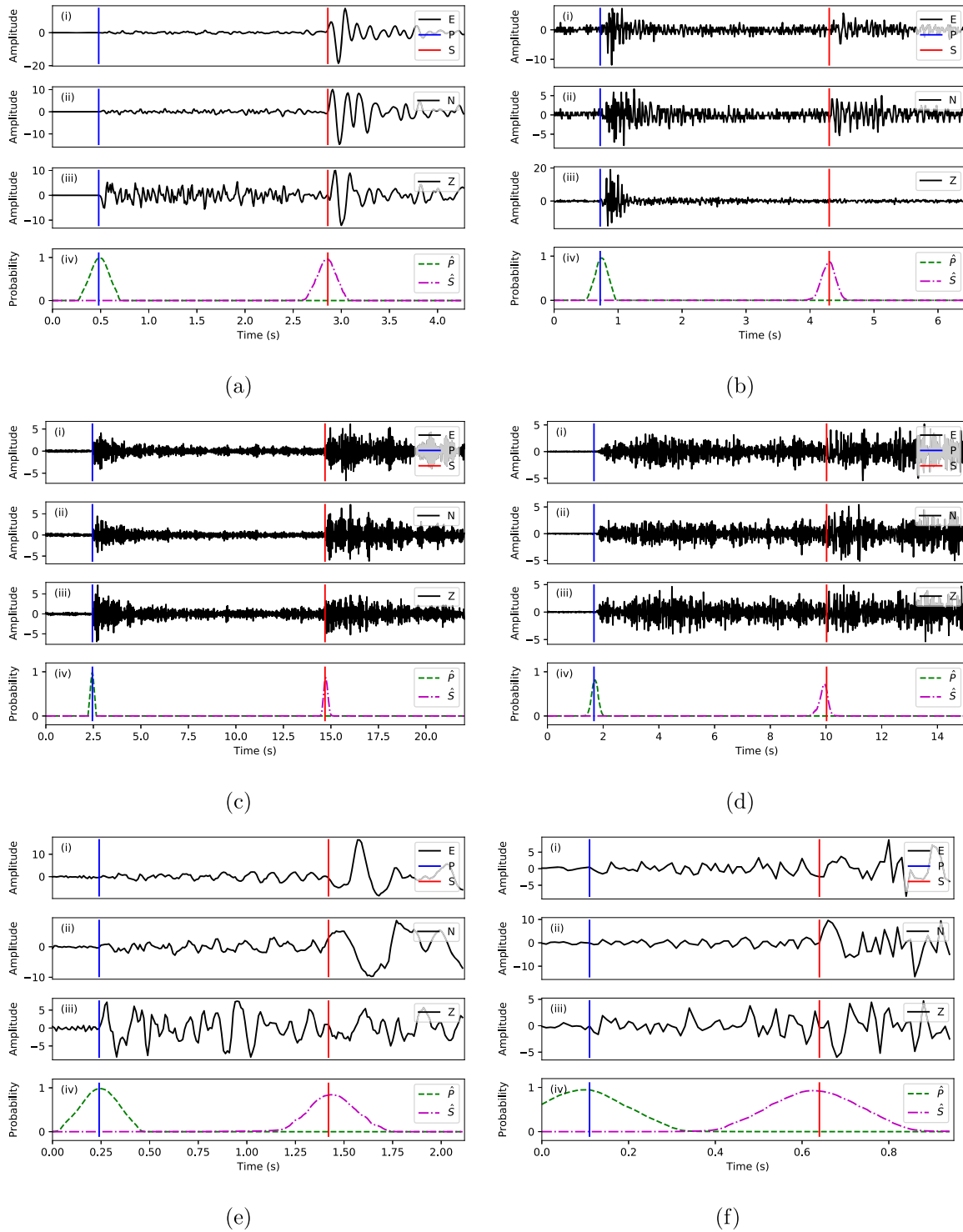


Figure 9. Examples of good picks ($\Delta t < 0.1$ s) in the test data set. The upper sub-figures (i–iii) are the “ENZ” components of seismograms. The lower sub-figures (iv) are the predicted probability distributions of P wave (\hat{P}) and S wave (\hat{S}). The blue and red vertical lines are the P and S arrival times picked by analysts.

exponential function is used to set probabilities in the last layer:

$$q_i(x) = \frac{e^{z_i(x)}}{\sum_{k=1}^3 e^{z_k(x)}},$$

where $i = 1, 2, 3$ represents noise, P and S classes. $z(x)$ are the unscaled values of the last layer. The loss function is defined using

cross-entropy between the true probability distribution ($p(x)$) and predicted distribution ($q(x)$):

$$H(p, q) = - \sum_{i=1}^3 \sum_x p_i(x) \log q_i(x),$$

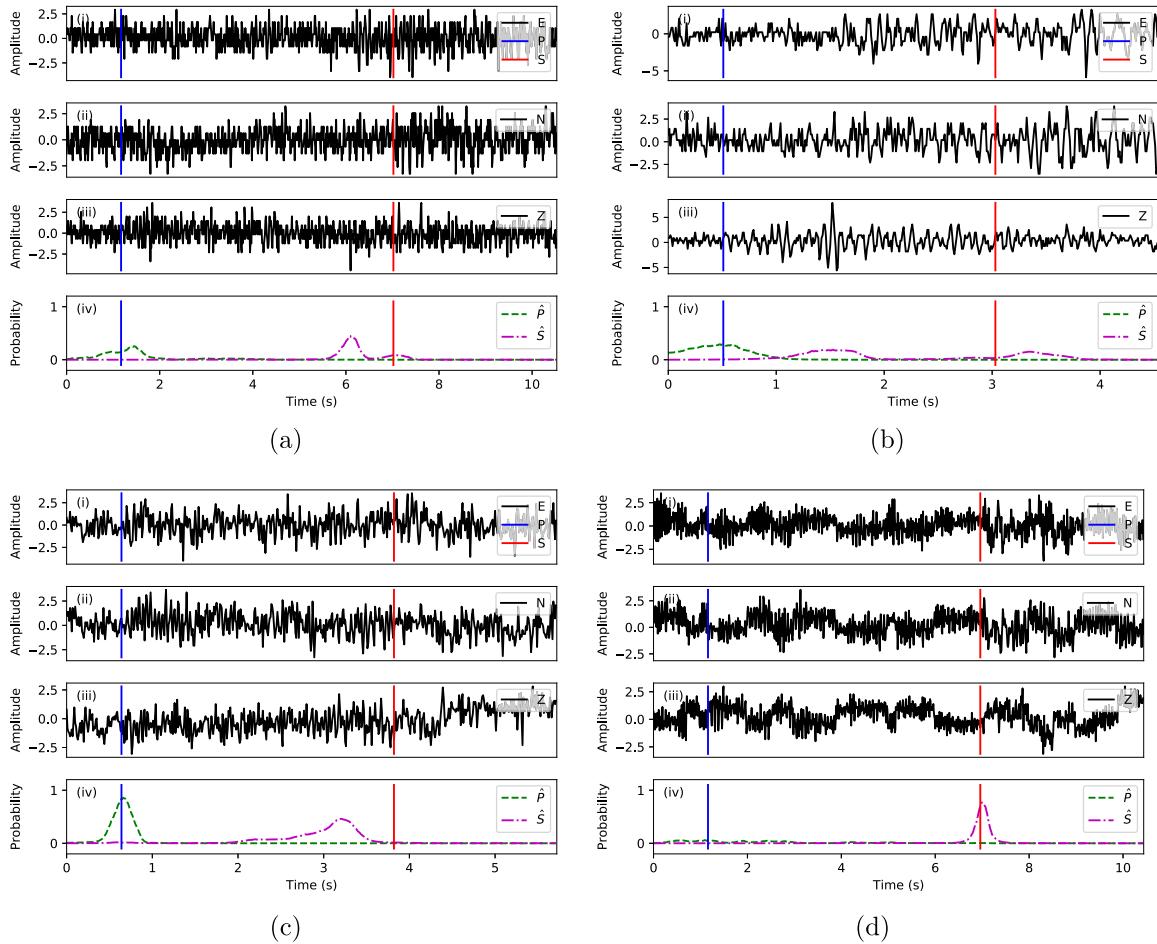


Figure 10. Examples of bad picks in the test data set. (a) and (b) are examples of no P or S picks predicted. (c) is an example of bad S picks. (d) is an example of bad P picks.

which measures the divergence between the two probability distributions. The P and S arrival times are extracted from the peaks of output probability distributions (Duarte 2015).

4 EXPERIMENTS

We have chosen the evaluation metrics: precision, recall, F1 score, mean (μ) and standard deviation (σ) of time residuals (Δt) between picks of PhaseNet and analysts to test the performance of PhaseNet (Powers 2011). Precision, recall and F1 are standard measures of performance defined as

$$\text{Precision : } P = \frac{T_p}{T_p + F_p},$$

$$\text{Recall : } R = \frac{T_p}{T_p + F_n},$$

$$\text{F1 score : } F1 = 2 \frac{P \times R}{P + R},$$

where T_p is the number of true positives, F_p is the number of false positives and F_n is the number of false negatives. Peak probabilities above 0.5 are counted as positive picks. Arrival-time residuals that are less than 0.1 s ($\Delta t < 0.1$ s) are counted as true positives. Picks with larger residuals are counted as false positives. F1 score is a

balanced criterion between precision and recall. For example, if we set a very high threshold for positive picks, only the best picks are reported positive which is only a small portion of total true picks, so that we can get a very high precision but a very low recall. The opposite case of a very low threshold will lead to low precision but high recall. For both cases, F1 score will be low, which is less sensitive to the selection of threshold and could provide more accurate evaluation of algorithms' performance. We compare our results with those obtained by the open-source "AR picker" (Akazawa 2004) implemented in Obspy (Beyreuther *et al.* 2010). The results of both PhaseNet and AR picker are shown in Table 1. For our data set, our method achieved significant improvements, particularly for the S waves. Because S waves emerge from the scattered waves of the P coda, picking S arrivals is more challenging for automatic methods.

Fig. 6 shows the distribution of time residuals between the automated and human-labelled P and S picks. The residual distributions of the P picks are much narrower than for the S picks, which is consistent with the fact that P wave arrivals are expected to be clearer and hence easier to pick. The residual distributions of both P and S picks for PhaseNet are distinctly narrower and do not have obvious biases compared with the results from the AR picker.

Fig. 7 shows performances of PhaseNet on different instrument types. The same model, which is trained on all instrument types, is used for testing, while the test data set is divided based on each instrument type. Without changing any parameters or thresholds, the performance of PhaseNet is robust on different instruments. Despite

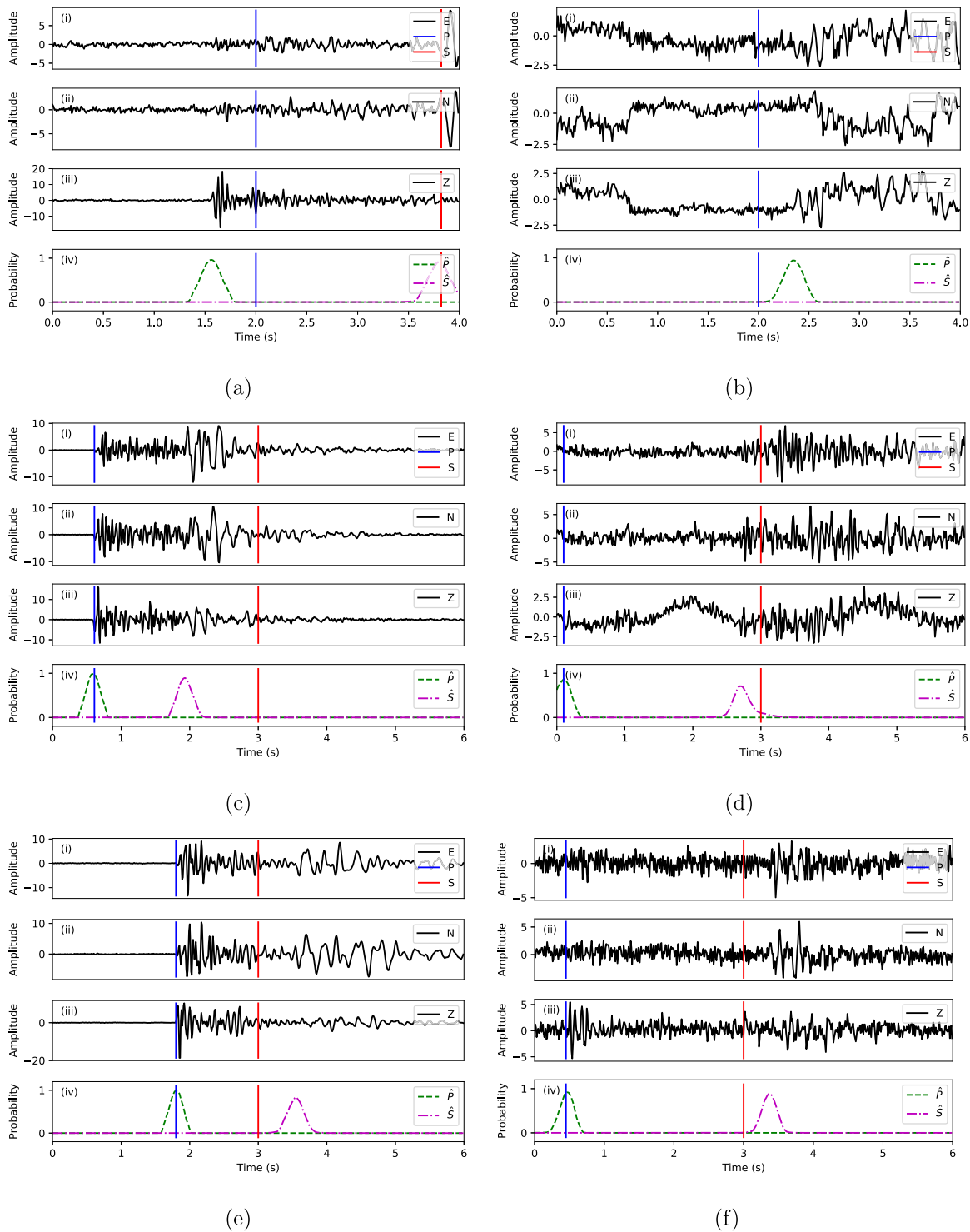


Figure 11. Examples where manual picks may not be accurate in the test data set. (a) and (b) are ambiguous *P* picks. (c)–(f) are ambiguous *S* picks.

the waveform differences between short period and broad band, high gain and low gain, accelerometer and seismometer, PhaseNet learns the common features needed to detect *P* and *S* phases and picks the correct arrival times.

Fig. 8 shows the change of performances of PhaseNet with different SNRs. The test set is divided into 10 different categories based on the value of $\log_{10}(\text{SNR})$. Precision, recall and F1 score are calculated for each category. All the three evaluation metrics

increase as the improvement of SNR. The F1 score exceeds 0.9 for *P* and 0.8 for *S* when the value of $\log_{10}(\text{SNR})$ is over 0.5. The precision of PhaseNet is high even for low SNR data while the recall rate becomes relatively smaller. This reflects the fact that for noisy data the seismic signals may be overwhelmed by noise and become harder to detect.

It is instructive to look at a handful of representative results. Fig. 9 shows good examples from the test data set. For different

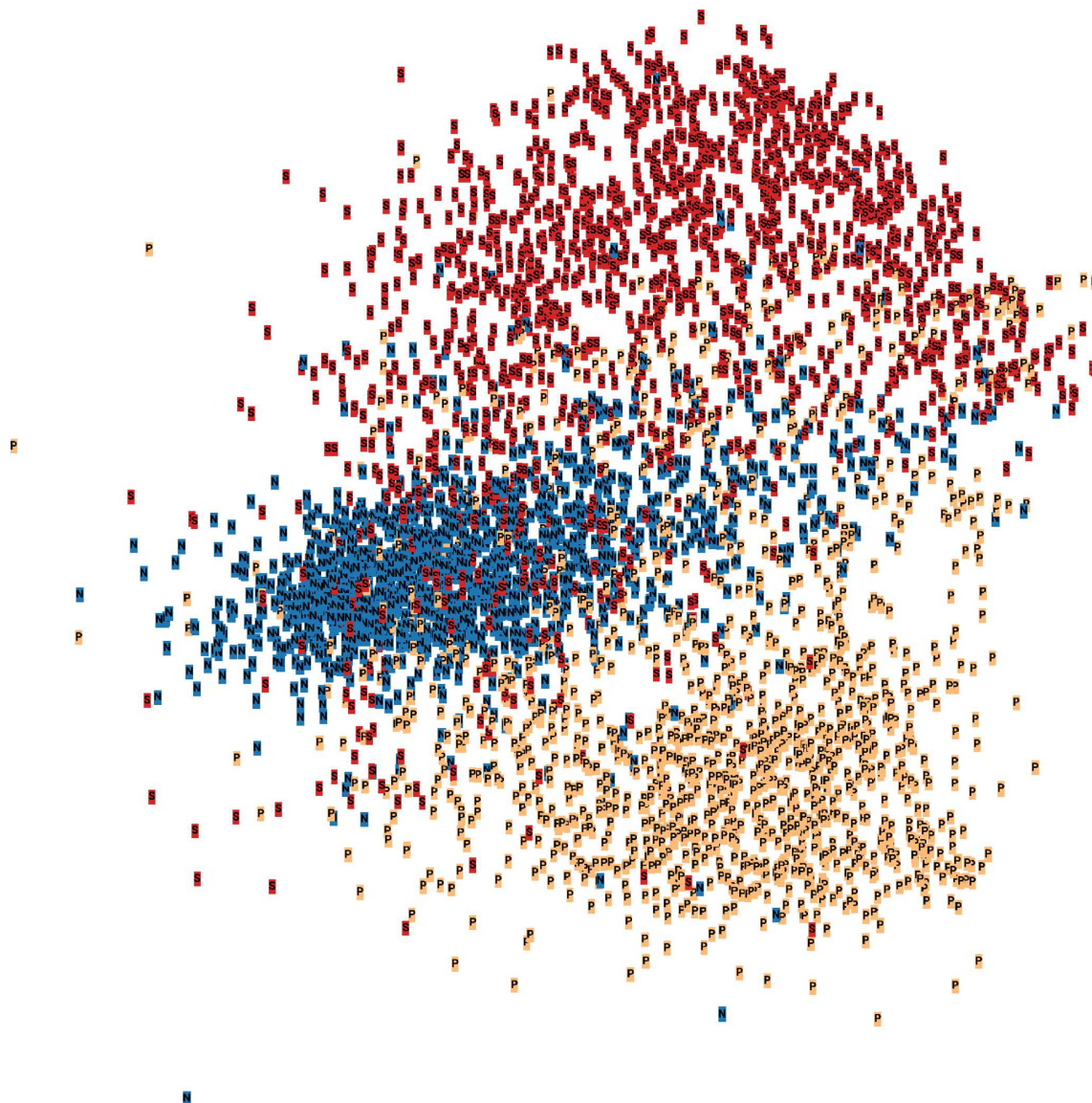


Figure 12. PCA visualization of weights in the deepest layer. The red, yellow and blue dots represent input data with P arrivals, S arrivals or noise only.

waveform characteristics and P – S time intervals, PhaseNet successfully picks the P and S arrivals. The peaks of the predicted distributions accurately align with the analyst-labelled P and S picks. Fig. 10 shows some apparently failed cases. The P and S first arrivals are harder to distinguish and the waveforms are more noisy and complex than those in Fig. 9. Fig. 11 shows some interesting cases where the P or S arrival times picked by analysts may be incorrect. The predictions of the neural networks appear more reasonable and consistent. Because there are subjective factors in seismic-phase picking, analysts may use different criteria to pick arrivals. Picks by the same analysts may also differ at different times.

To analyse the representations that PhaseNet has learned, we train another model without the skip connection which forces all the information to go through the deepest layer, and apply PCA (Principal Component Analysis) to the neural weights of the deepest layer (Fig. 5). The neural network condenses the knowledge from high dimensional raw waveforms into a few parameters in the deepest layer, which means that these low dimensional neural

weights should contain the information needed to determine P versus S arrivals. We feed in seismic data with P arrivals, S arrivals or only noise, and record the corresponding vectors in the deepest layer. The PCA visualization (Fig. 12) shows that these condensed vectors group to different regions for P pick, S pick and noise. This demonstrates that the neural network has learned to extract the characteristic features to differentiate between P pick, S pick and noise from the raw data and capture them in the condensed neural weights in the deepest layer.

PhaseNet predicts the probability distributions of P and S picks for every data point in the time-series, so it may be applied to continuous data for earthquake detection. We have created continuous seismic data by stacking waveforms of eight different events (Fig. 13). These events are shifted to make the arrival-time interval between adjacent events equal to 6 s. We have applied both basic STA/LTA in Obspy and our PhaseNet method to this sequence. The short and long sliding windows of the STA/LTA method are set to 0.2 and 2 s, respectively. PhaseNet does not need a sliding window, but takes the whole 60-s waveforms as input and outputs of three

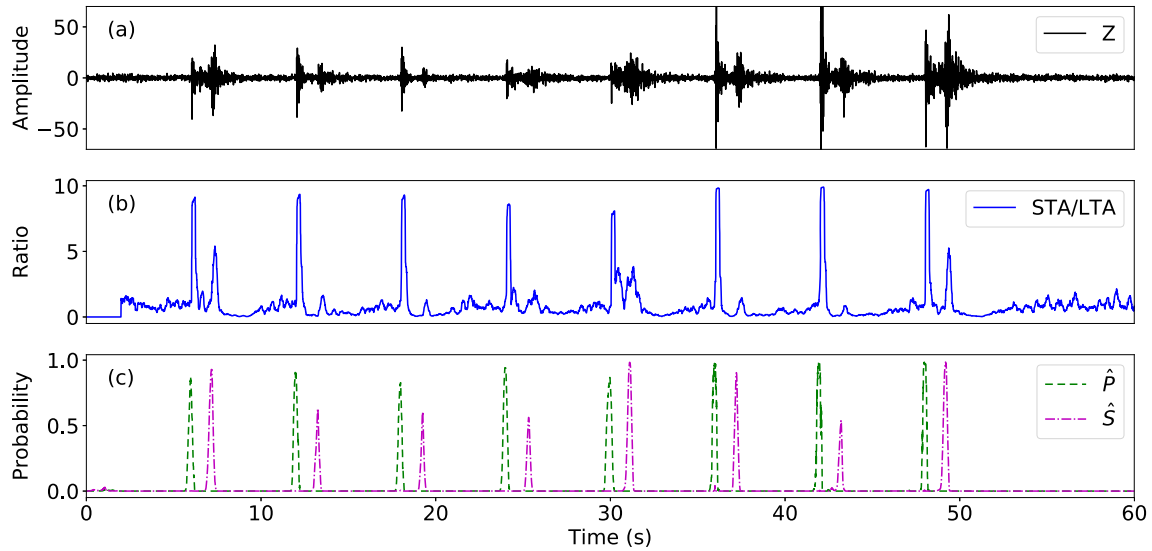


Figure 13. Synthetic continuous seismic waveforms. (a) Waveform of vertical component. (b) Output of basic STA/LTA in Obspy. (c) Output of PhaseNet. The continuous data are created by stacking waveforms of eight events. The first-arrival-time interval between adjacent events is 6 s. The STA/LTA method runs on the vertical component. The PhaseNet runs on three components.

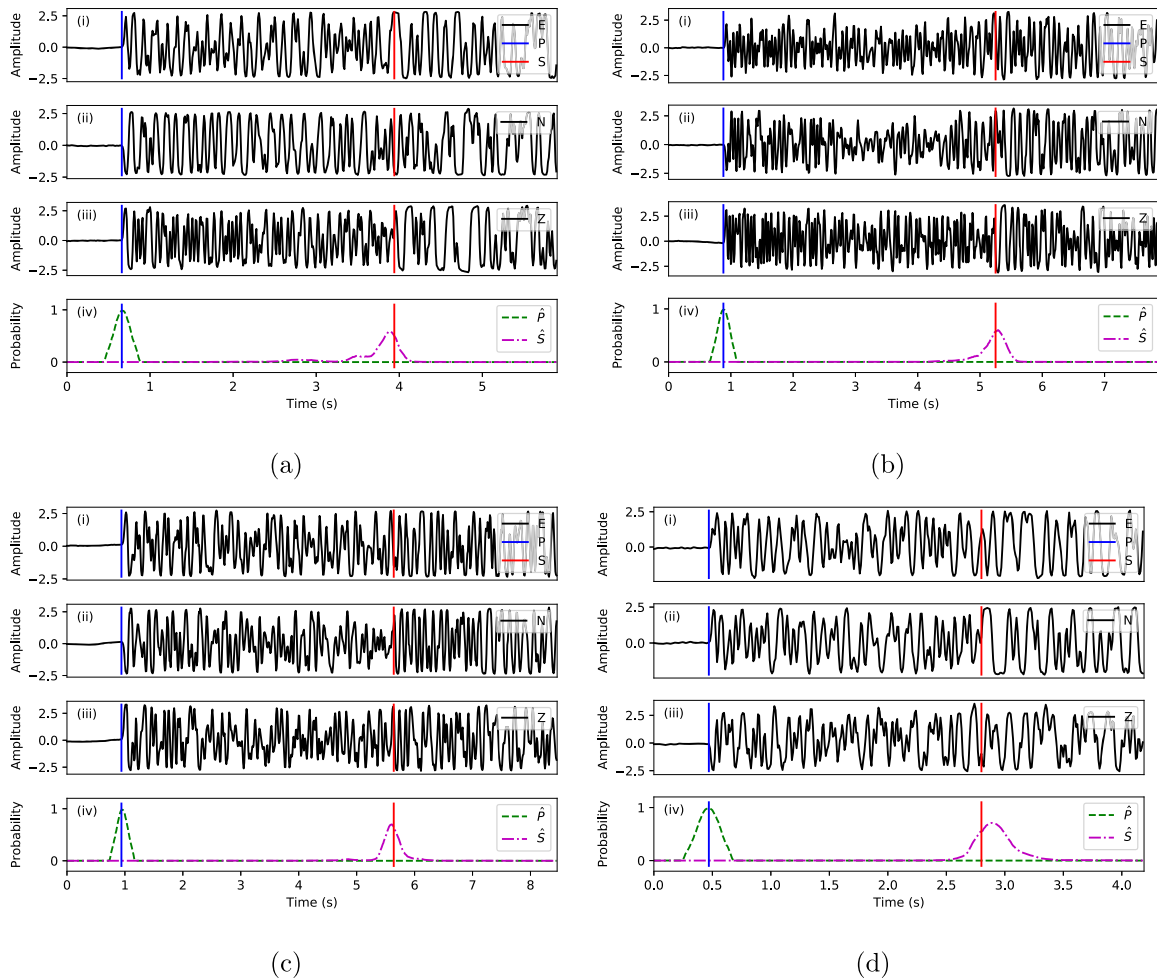


Figure 14. Examples of amplitude clipped waveforms.

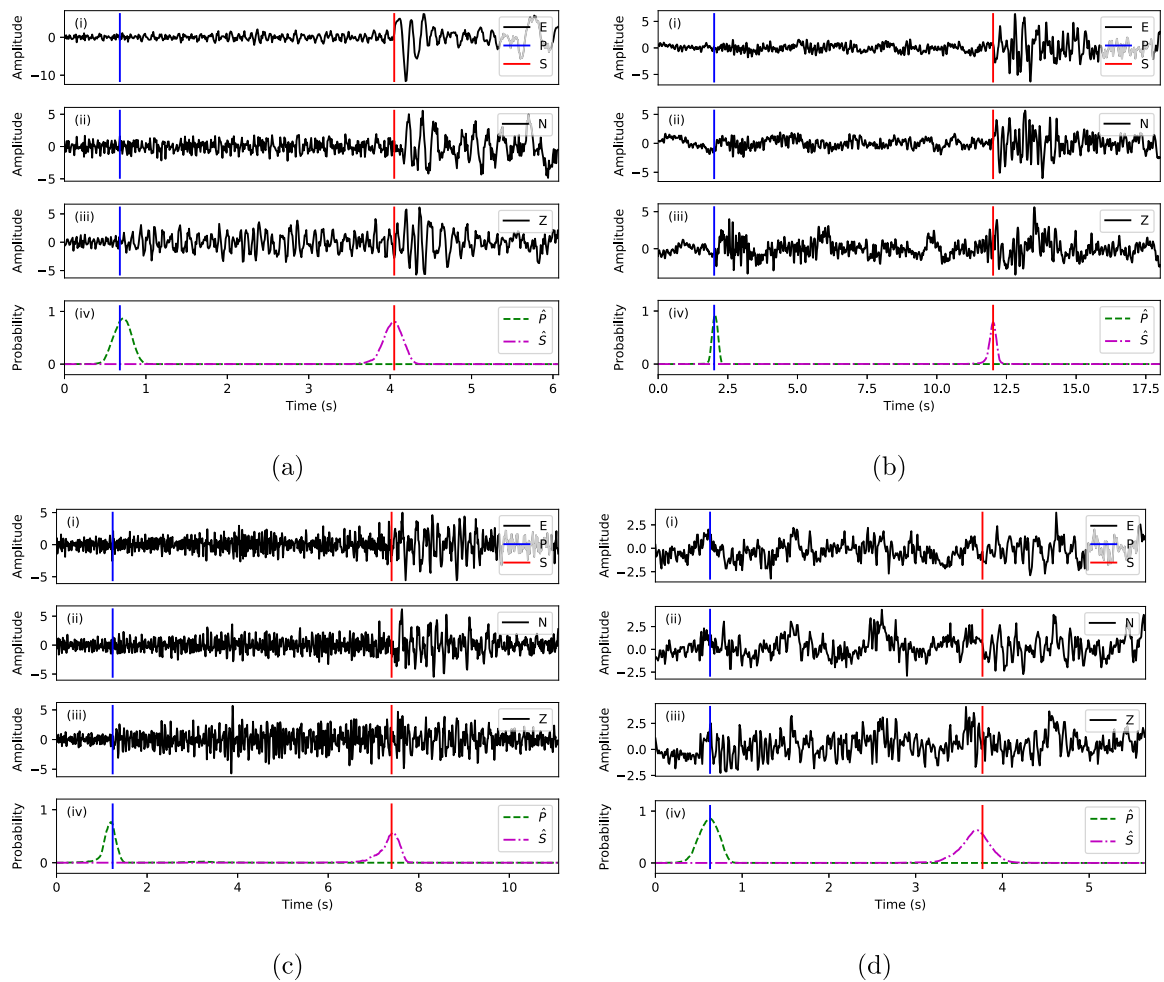


Figure 15. Examples of low SNR data.

probability sequences for P pick, S pick and noise. The output sequences in Fig. 13 show that PhaseNet produces similar spikes as STA/LTA methods, which are commonly used for earthquake detection; however, PhaseNet can also differentiate between P and S arrivals. This information could also be used to reduce false detections, because events with both P and S picks are more likely to be true earthquakes compared with the undifferentiated spikes reported by STA/LTA. The PhaseNet model in this paper is trained on a data set of detected earthquakes. It is designed to pick seismic phases accurately. In order to apply PhaseNet for detection on continuous data, a new data set that includes more non-seismic signals should be used for training to let it learn the features to differentiate the noise spikes that look similar to seismic phases.

5 DISCUSSION

We have shown that PhaseNet can detect and pick P and S arrivals effectively within known earthquake waveforms. The F1 score provides a balanced assessment of algorithm performance in both precision and recall. PhaseNet achieves an F1 score of 0.896 for P arrivals and 0.801 for S arrivals, which is substantially better than the AR picker (0.558 for P arrivals and 0.165 for S arrivals). We have chosen a strict threshold for true positive ($\Delta t < 0.1$ s) during evaluation. If we were to relax this standard, the F1 score would be

even higher. Our method differs from that proposed by Ross & Ben-Zion (2014), because PhaseNet does not explicitly use polarization analysis to separate P from S waves. PhaseNet automatically learns features, which might implicitly include polarization, to distinguish P from S waves. We find that the improvement in S picks is more significant than the improvement to P picks, which suggests that the features learned from data are more effective than manually defined features.

The STA/LTA method is based on detecting a sudden change in waveform amplitude. But the S phase is always contaminated by the P coda, which degrades the STA/LTA ratio to select an accurate S pick. PhaseNet has an advantage here in that it can learn features other than amplitude both to detect S waves and to differentiate between P and S waves. Fig. 14 shows examples of PhaseNet applied to clipped waveforms. Although the amplitude is strongly clipped, PhaseNet is still able to pick S arrivals successfully.

We have not pre-processed the data with denoising techniques such as bandpass filtering. As a result, our data set contains a number of low SNR data. We apply the AR picker after pre-processing the data with a bandpass filter of 0.1–30 Hz. Without filtering, its performance would be substantially degraded. PhaseNet does not require this pre-processing because it not only learns the characteristics of P and S waves, but it also learns what kind of data is noise. Fig. 15 shows several prediction results on low SNR data, for which it would be difficult for analysts to pick P and S arrivals.

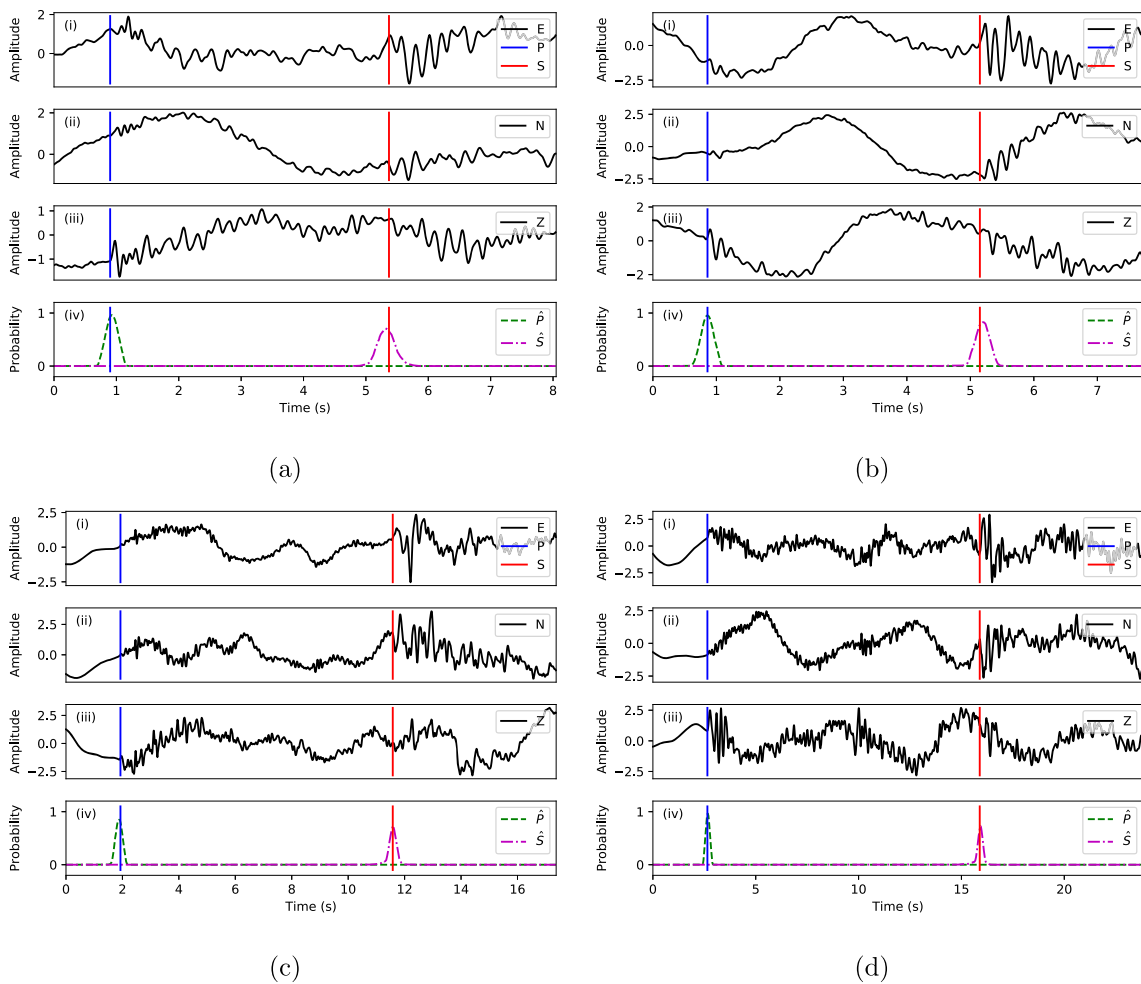


Figure 16. Examples with background variation.

Despite these challenges, PhaseNet predicts accurate arrival times at high probabilities. Fig. 16 shows examples with strong low frequency background noise. PhaseNet can accurately pick both P and S phases without the need for filtering.

The STA/LTA method is sensitive to the threshold selected to determine P or S arrivals, and there is an inevitable trade-off between too high and too low thresholds. Moreover, it is prone to a delayed arrival time if the threshold is set too high. Instead of an unbounded STA/LTA ratio, PhaseNet estimates probability distributions between $[0, 1]$. We have set the threshold of probability to 0.5 for both P and S picks. Tuning this threshold can further improve the performance, but the effect is not significant. Unlike STA/LTA, this threshold will not systematically bias arrival times, because this threshold is only used to decide if it is a positive pick. The accurate arrival time is measured from the peak of the probability distribution and does not depend directly on this threshold.

PhaseNet is not constrained by the number of earthquakes inside one time window. Because it predicts probability sequences of the same length as input waveforms, there could be several peaks or no peaks in P or S probability distributions inside the window. As shown in Fig. 13, PhaseNet converts the 60-s waveforms into probability distributions with several spikes of P and S arrivals. We can apply PhaseNet to continuous data to generate running probability distributions of P and S arrivals, which can be used as the basis of earthquake detector when paired with an association algorithm.

Accurate phase arrival times can be used to get absolute earthquake locations and to develop seismic velocity models. PhaseNet provides an improved method to get accurate S arrivals, which will be useful for developing better S -wave velocity models and improving earthquake locations.

6 CONCLUSION

Deep learning methods are improving rapidly. An important ingredient for improving them is the existence of large labelled data sets. In seismology, we are fortunate to have such large data sets ready at hand in the form of decades of arrival times with accompanying waveforms. We are on the verge of, or perhaps have already arrived at, a threshold where neural networks are “superhuman” in the sense that they can outperform human analysts. In this paper, we have built a training data set using manually picked P and S arrival times from the Northern California Seismic Network catalog. We have developed PhaseNet, a deep neural network algorithm that uses three component waveform data to predict the probability distributions of P pick, S pick and noise. We extract arrival times from the peaks of these distributions. Test results show that our method achieves significant improvements compared with existing methods, particularly for S waves. PCA visualization shows that the condensed neural weights contain characteristics that allow the separation of P wave, S wave and noise. While further testing

against existing methods is required, we are not far from making such a capability operational. An increase in accurate *P* and *S* arrival times will help us to continue to extract as much information as possible from rapidly growing waveform data sets for earthquake monitoring, and the ability to extract reliable *S* arrivals will allow us to improve shear wave velocity models substantially, which will be especially useful for prediction of path effects in strong ground motion prediction. Finally, we note that PhaseNet can also be used for other phases for which manually labelled training data sets are available.

ACKNOWLEDGEMENTS

We thank Lind S. Gee and Stephane Zuzlewski for their help in downloading and processing the catalogue and waveform data from NCEDC. We thank Seyed Mostafa Mousavi, Yixiao Sheng, Clara Yoon, Kaiwen Wang and William Ellsworth for helpful discussions. Waveform data, metadata or data products for this study were accessed through the Northern California Earthquake Data Center (NCEDC). This research is supported by the National Science Foundation (NSF) grant number EAR-1818579.

REFERENCES

- Akazawa, T., 2004. Technique for automatic detection of onset time of *P*- and *S*-phases in strong motion records in *13th World Conference on Earthquake Engineering*, Vancouver, Vol. **786**, pp. 786.
- Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces, *Bull. seism. Soc. Am.*, **68**(5), 1521–1532.
- Baer, M. & Kradolfer, U., 1987. An automatic phase picker for local and teleseismic events, *Bull. seism. Soc. Am.*, **77**(4), 1437–1445.
- Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.*, **81**(3), 530–533.
- Duarte, M., 2015. Notes on scientific computing for biomechanics and motor control. Available at: <https://github.com/demotu/BMC>.
- Gentili, S. & Michelini, A., 2006. Automatic picking of *P* and *S* phases using a neural tree, *J. Seismol.*, **10**(1), 39–63.
- Küperkoch, L., Meier, T., Lee, J. & Friederich, W., 2010. Automated determination of *P*-phase arrival times at regional and local distances using higher order statistics, *J. geophys. Int.*, **181**(2), 1159–1170.
- Li, H., Xu, Z., Taylor, G., Studer, C. & Goldstein, T., 2017. Visualizing the loss landscape of neural nets, arXiv preprint arXiv:1712.09913.
- NCEDC, 2014. *Northern California Earthquake Data Center*. UC Berkeley Seismological Laboratory. Dataset.
- Noh, H., Hong, S. & Han, B., 2015. Learning deconvolution network for semantic segmentation in *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, Nice, pp. 1520–1528.
- Powers, D.M.W., 2011. Evaluation: from precision, recall and F-measure to Roc, informedness, markedness & correlation, *J. Mach. Learn. Technol.*, **2**(1), 37–63.
- Ronneberger, O., Fischer, P. & Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation, *Miccai*, **9351**, 234–241.
- Ross, Z.E. & Ben-Zion, Y., 2014. Automatic picking of direct *P*, *S* seismic phases and fault zone head waves, *Geophys. J. Int.*, **199**(1), 368–381.
- Saragiotis, C.D., Hadjileontiadis, L.J. & Panas, S.M., 2002. PAI-S/K: a robust automatic seismic *P* phase arrival identification scheme, *IEEE*, **40**(6), 1395–1404.
- Sleeman, R. & Van Eck, T., 1999. Robust automatic *P*-phase picking: an online implementation in the analysis of broadband seismogram recordings, *Phys. Earth planet. Inter.*, **113**, 265–275.