

LA-UR-18-29660 (Accepted Manuscript)

## Pairwise Association of Seismic Arrivals with Convolutional Neural Networks

McBrearty, Ian William  
Delorey, Andrew A.  
Johnson, Paul Allan

Provided by the author(s) and the Los Alamos National Laboratory (2019-05-09).

**To be published in:** Seismological Research Letters

**DOI to publisher's version:** 10.1785/0220180326

**Permalink to record:** <http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-18-29660>

**Disclaimer:**

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA00001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

# Pairwise Association of Seismic Arrivals with Convolutional Neural Networks

Ian W. McBrearty<sup>1</sup>, Andrew A. Delorey<sup>1</sup>, Paul A. Johnson<sup>1</sup>

January 15, 2019

**Corresponding Author:** Ian W. McBrearty<sup>1</sup> (imcbrearty@lanl.gov).

**Author Affiliations:** <sup>1</sup> Geophysics Group, Los Alamos National Laboratory – Los Alamos, NM, USA.

**Mailing Address:** D446 Los Alamos Natl. Lab P.O. Box 1663, Los Alamos, NM 87545

**Key Words:** Earthquake Detection, Earthquake Arrival Association, Convolutional Neural Networks

## Abstract

Correctly determining the association of seismic phases across a network is crucial for developing accurate earthquake catalogs. Nearly all established methods use travel time information as the only criterion for determining associations, and in problems where earthquake rates are high and many false arrivals are present, many standard techniques may fail to resolve the problem accurately. As an alternative approach, in this work we apply convolutional neural networks (CNN's) to the problem of associations; we train CNN's to read earthquake waveform arrival pairs between two stations and predict the binary classification of whether the two waveforms are from a common source, or different sources. Applying the method to a large training dataset of previously cataloged earthquakes in Chile, we obtain  $> 80\%$  true positive prediction rates for high frequency data ( $> 2$  Hz) and stations separated in excess of 100 km. As a secondary benefit, the output of the neural network can also be used to infer predicted phase types of arrivals. The method is ideally applied in conjunction with standard travel time based association routines, and can be adapted for arbitrary network geometries and applications, so long as sufficient training data is available.

## INTRODUCTION

Seismic data are routinely processed to develop earthquake catalogs and generate records of seismicity for source regions all around the world. Earthquake catalogs are inherently incomplete, and the magnitude of completion reflects an inability to accurately catalog small magnitude events. Small events occur more frequently than larger ones, and generally have lower signal-to-noise ratio arrivals, which can easily be mistaken with false arrivals. A primary complication of identifying small earthquakes, in addition to phase picking, is that high rates of small magnitude events produce arrivals which can occur within short intervals of each other, and may be overlapping in time across a seismic network. When many arrivals are present, it can be difficult to determine which sets of arrivals are from a common source. It can also be difficult to determine which arrivals are

26 false picks, and how many distinct earthquakes have occurred. Developing accurate catalogs of  
27 small earthquakes requires determining the associations of phases across a network, along with the  
28 auxiliary questions of identifying false arrivals, and determining the number of earthquakes.

29 Existing association methods largely rely on travel time information and an assumed sparsity  
30 of sources in time and space. Observed relative arrival times of phases (between stations) are com-  
31 pared with theoretical move-out times to group sets of arrivals that are consistent with a realistic  
32 source. Some methods use grid-search (Johnson et al., 1997) or iterative techniques (Sippl et al.,  
33 2013) to optimize the number of arrivals paired with earthquakes, while simultaneously trying to  
34 use as few events as possible to fit the data. Other notable techniques include a Bayesian associa-  
35 tion method (Arora et al., 2013), a waveform-based pairwise association technique built upon the  
36 Fingerprint and Similarity Thresholding (FAST) method (Bergen and Beroza, 2018), and the use  
37 of recurrent neural networks applied to arrival time data (Ross et al., 2018c). With the exception  
38 of template matching and some beamforming and waveform coherency methods, which combine  
39 association along with event detection (Ringdal and Kväerna, 1989; Gibbons and Ringdal, 2006;  
40 Grigoli et al., 2013; Bergen and Beroza, 2018), the waveforms of individual arrivals are rarely used  
41 to assist in determining associations. Raw waveforms are difficult to use for associations because  
42 path effects from arbitrary source locations are typically not well known and mask information in  
43 the waveforms that would indicate a common source (Aki and Richards, 2002).

44 Convolutional neural networks (CNN’s) have recently been applied to arrival time picking of  
45 earthquakes with significant promise (Perol et al., 2018; Ross et al., 2018b; Wu et al., 2018; Zhu  
46 and Beroza, 2018), and to a number of other problems relevant to geophysics and seismology  
47 (Lu et al., 2018; Ross et al., 2018a; Zhu et al., 2018). CNN’s use a cascading series of convo-  
48 lution operations to map input images or signals directly to arbitrary output targets, with min-  
49 imal preprocessing, and without relying on manually chosen feature inputs. Rather than using  
50 pre-specified features, CNN’s excel at extracting and using hidden information contained in large  
51 preexisting training datasets to learn (or calibrate) the filter (or kernel) weights inside the net-

52 work, such that it can correctly predict the input-output relationships of the training dataset to  
53 high accuracy, and ideally generalize to new, unseen data (LeCun et al., 2015; Schmidhuber, 2015;  
54 Goodfellow et al., 2016). In addition to automatically determining complex functional mappings  
55 from existing datasets, a strength of CNN’s is that arbitrary dimensional input-output relationships  
56 can be learned, which allows seismogram records (of arbitrary pre-specified length) to be directly  
57 mapped to answer specific, targeted questions.

58 In this paper, we present a deep neural network approach applied to the problem of phase  
59 associations. We implement a CNN to take pairs of earthquake arrival waveforms on two different  
60 stations, and predict whether the two waveforms are from the same source (label = 1), or different  
61 sources (label = 0). By using large, preexisting earthquake catalogs, databases of paired waveforms  
62 (between stations) to common sources are readily available. An advantage of this approach is that  
63 the raw seismic data can directly be used to predict associations, leveraging diagnostic information  
64 contained in the waveforms that is often unused for this purpose.

## 65 **DATA**

66 In this study we use broadband seismic data collected on the Integrated Plate Boundary Ob-  
67 servatory Chile (IPOC). We take a catalog from northern Chile which includes 866,842 earth-  
68 quakes between years 2013 - 2017, spanning the region between Lats.  $[18^{\circ}S, 24^{\circ}S]$  and Lons.  
69  $[67^{\circ}W, 72^{\circ}W]$  (Fig. 1). From each earthquake we extract waveform pairs between stations that  
70 have an arrival pick associated with the event. For each arrival, we take 6 seconds of data from 1.5  
71 s preceding to 4.5 s following the inferred arrival time, for all three components. For each station,  
72 we also extract waveforms of all ‘false picks’ that were detected by an energy-based single station  
73 triggering algorithm, but ultimately not associated to any earthquake in the catalog. All extracted  
74 waveforms are normalized to have unit particle-motion energy while preserving the relative energy  
75 between each component. All data is band-passed between 2 - 22 Hz and sampled at 50 Hz.

76 For demonstration we use one station, PB02, as the ‘master station’, and extract a separate  
77 database for several other stations of the network with paired arrivals to this one. Each waveform  
78 pair is marked as either P-P or S-S, indicating which phase types comprise the pair. Each station  
79 pair has a variable number of associated picks, but on average  $\sim 550,000$  waveform pairs are  
80 available between each set of stations, and for each phase type. The precise number of sample  
81 points (of P-P, S-S, and false phase types) are given in Table S1. A schematic of the waveform  
82 extraction and database building steps is given in Fig. 2.

## 83 METHODS

84 Convolutional neural networks are composed by stacking several (or many) convolutional lay-  
85 ers, followed by a number of fully connected layers which map the flattened feature vector of  
86 the last convolutional layer to the final output. Between each layer, downsampling and nonlinear  
87 activations are used, which enable the extraction of information from the input and its nonlinear  
88 mapping to the targeted output (Goodfellow et al., 2016). The fundamental computational unit of  
89 a CNN is the convolutional filter (or kernel) acting on a tensor, which may be defined by

$$x_i^{(n+1)}(t) = \sigma\left(\sum_j (x_j^{(n)}(t) * w_{ij}^{(n+1)}(t)) + b_i^{(n+1)}\right), \quad (1)$$

90 for 1D time-series, which maps the tensor  $x^{(n)}$  at layer  $n$  to the  $i^{th}$  channel of layer  $n + 1$ , using  
91 the  $(n + 1)^{th}$  layers  $i^{th}$  set of filters  $w_{ij}^{(n+1)}(t)$  and bias term  $b_i^{(n+1)}$ . The  $\sigma(\cdot)$  operator represents  
92 any nonlinear activation, and  $*$  represents the convolution operation. The full capacity of a CNN  
93 (i.e., ability to produce complex mappings) is achieved through using many filters simultaneously,  
94 and iteratively mapping each layers current tensor state to the subsequent layers, following the  
95 calculations of equation (1). The fully connected layers at the end of the CNN resemble a standard,  
96 non-convolutional artificial neural network (ANN), which map the features extracted from the

97 convolutional steps to the final output, using a non-convolutional functional analog of equation  
98 (1).

99 Through some trial and error and monitoring of the validation test metrics, we chose a four  
100 layer CNN with two fully connected layers. Between each convolutional layer, max pooling is  
101 used. The rectified linear unit (ReLU) activation is used for all nodes inside the network, and the  
102 sigmoid activation is used as the activation for the final node. The ReLU and sigmoid activations  
103 are defined by

$$ReLU(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}, \text{ and,} \quad (2a)$$

$$Sigmoid(x) = \frac{1}{(1 + \exp(-x))}. \quad (2b)$$

104 The output of the CNN is the activation at the last node, which is bounded between 0 and 1 by the  
105 sigmoid activation. As input to the CNN, we concatenate the three component slices of data from  
106 each station across channels, such that the input size is six 1D channels of fixed length (300 data  
107 points per channel). Details of the number of filters and the kernel sizes used are given in Fig. 3.

## 108 **Training**

109 We split events in the catalog by absolute time into training (years: 2013 - 2015), validation  
110 (year: 2016), and testing (year: 2017) bins (Fig. 1). We train a distinct CNN for each station  
111 and phase pair (P-P or S-S). For each of the CNN's, positive labels are all correctly associated  
112 waveforms of the phase types chosen, and false label pairs can be of several varieties: waveforms  
113 of opposite phases (i.e., P-S, S-P), waveforms from two different earthquakes, and one or more  
114 of the waveforms being a false pick from the database of all false arrivals (Fig. 2). We note that

115 with this labelling scheme, for  $n$  earthquakes there are  $\mathcal{O}(n)$  and  $\mathcal{O}(n^2)$  possible true, and false  
116 combinations, respectively. Of the IPOC network, we only include station pairs that have at least  
117 200,000 paired waveforms available in the training dataset (Table S1).

118 We train each CNN with stochastic gradient descent (SGD), using the ADAM optimizer (Kingma  
119 and Ba) and the automatic differentiation capabilities of PyTorch (Paszke et al., 2017). The loss  
120 function we minimize is the binary cross-entropy loss, and is computed over each mini-batch of  $N$   
121 samples, as defined by

$$H(p, q) = -\frac{1}{N} \sum_{i=1}^N (q \cdot \log(p) + (1 - q) \cdot \log(1 - p)) , \quad (3)$$

122 where  $q$  is the target label (0 or 1), and  $p$  is the predicted label. Using automatic differentiation,  
123 the loss (equation (3)) is differentiable with respect to all of the free parameters of the CNN. For  
124 each mini-batch, we randomly choose 750 correctly paired waveforms from the database, and 750  
125 randomly chosen false associations. SGD is terminated after the validation error stabilizes, which  
126 generally occurs after  $\sim 10,000$  mini-batch updates (Figs. S1,2).

## 127 **EXPERIMENTS**

128 Trained models are evaluated over the test dataset by computing the precision, recall, and F1  
129 quality metrics (Powers, 2011). Precision ( $P$ ) is the ratio of true positives to true positives plus  
130 false positives, recall ( $R$ ) is the ratio of true positives to true positives plus false negatives, and F1  
131 is a summary statistic given by the quantity  $2 \cdot (P \cdot R)/(P + R)$ . We round outputs of the CNN  
132 to their nearest class to infer predicted labels (i.e., all outputs  $> 0.5$  denote a positive predicted  
133 label, and those  $\leq 0.5$  denote a negative predicted label). Across all experiments we find relatively  
134 high levels of precision and recall, with median values of F1 of 0.869 and 0.879, for P-, and S-  
135 wave models, respectively. Median precision and recall quantities are within +/- 0.03 units of one

136 another, indicating there is no significant imbalance between false positives and false negatives  
137 (Table 1). Most station pairs have similar performance, though performance gradually weakens as  
138 interstation distances increase (Table 1, Fig. S3).

139 As a more complete summary of model performance we also compute the receiver operating  
140 characteristic (ROC) curves of the CNN's predictions in testing. ROC curves compute the rate of  
141 false positives to false negatives as the discrimination threshold of a classifier varies between 0 and  
142 1; they are frequently used to evaluate a binary classifiers overall performance (Powers, 2011). The  
143 area under the ROC curve (AUC) for a perfect model is 1, while a model that performs equivalently  
144 to a random guess is 0.5. In our experiments we measure an AUC of 0.944, and 0.943 for the P-,  
145 and S-wave models, respectively (Fig. S4).

146 We probe the trained neural networks with a few additional tests. First, we bin all earthquakes  
147 in the testing dataset by their magnitude at several increasing quantile values of the catalogs magni-  
148 tude distribution. From each bin, we compute the true positive rate over 30,000 randomly selected  
149 positive samples, with randomly selected station pair and phase type combinations. We find a small  
150 but notable improvement of 4 - 6% in the true positive prediction rate in going from the smallest  
151 to the largest magnitudes (Fig. S5).

152 Secondly, we evaluate which sets of falsely associated waveform pairs are most likely to pro-  
153 duce false positives. For this, we iterate over the following procedure: (1) choose a random earth-  
154 quake waveform recorded on PB02, (2) sample another randomly chosen set of 1,000 earthquakes  
155 that are falsely associated with it (from arbitrary station and phase pair combinations), and (3)  
156 compute the CNN predictions between these arrival pairs. For all events with predictions  $> 0.85$   
157 (i.e., falsely associated waveform pairs that produce highly confident, false positive predictions),  
158 we record the differential measurement of the spatial locations of these events with the original  
159 earthquake used. We iterate this process 1,000 times and record the full database of differential  
160 spatial locations. The histograms of these spatial differences have well defined maxima at nearly

161 zero offset (Figs. S6,7). This indicates that falsely associated waveform pairs from similar source  
162 coordinates (and offset in time) have the highest tendency of producing false positives. We note  
163 that to make these results robust, we sample the 1,000 trial earthquakes at each iteration with prob-  
164 abilities inversely proportional to the spatial density estimate of earthquake locations in the catalog  
165 (Epanechnikov, 1969), so that dense earthquake regions do not bias the residuals towards zero  
166 mean.

## 167 **EXAMPLE APPLICATION**

168 As a demonstration of applying the CNN's to all stations of the network simultaneously, we  
169 apply the neural networks to twenty days of continuous data between January 1 - 20<sup>th</sup>, 2017. We  
170 compute the CNN predictions for each arrival of station PB02 with arrivals from other stations of  
171 the network that occur within +/- 50 seconds. We treat all arrivals as having an unknown phase,  
172 and compute the predictions of both the P- and S-phase trained models. All arrival connections  
173 that have predictions  $> 0.5$  are recorded, and for any two arrivals compared only the maximum  
174 prediction between the P- and S-phase models is retained. Each arrival on station PB02 is further  
175 restricted to have at most one connection to each other station; the pair that's chosen is the sample  
176 with the maximum predicted association value between itself and all other arrivals of a fixed station  
177 (if there are multiple). These steps create a database of predicted, phase-dependent associations  
178 between arrivals on station PB02 and arrivals across the rest of the network over continuous time.  
179 Fig. 4b shows an example set of predicted associations following this procedure, which is given  
180 alongside the ground truth result recorded in the catalog (Fig. 4a). In this example we note that  
181 many of the earthquakes have a large number of correctly associated arrivals, and in addition many  
182 correct phase types (i.e., P or S) have also been determined.

183 In its current form pairwise association likelihood does not by itself resolve the full association  
184 problem. However we can assess how well the CNN predictions agree with the ground truth

185 catalog with a simple analysis. We count the number of earthquakes in the catalog that have  
186  $\geq n_1$  P-P and  $\geq n_2$  S-S phase associations, and then calculate the proportion of these events  
187 that are recovered by the same criteria with the CNN’s predictions. We vary  $n_1$  and  $n_2$  over a  
188 range of values to obtain a spectrum of success and failure rates. While ad hoc, this gives a view  
189 into how random or structured the CNN predictions are with respect to applications across the  
190 entire network. Evaluating the reconstruction rates over January, 2017, and letting  $n_1$  and  $n_2$  vary  
191 between one and four (five is the maximum with only six stations available), we find that for the  
192 lowest values of  $n_1$  and  $n_2$ , rates of recovered events are high (e.g.,  $> 80\%$ ), and they fall off to  
193  $\sim 70\%$  recovery rate for the most restrictive thresholds (Fig. S8). Using these same criteria, we  
194 also compute the number of events with  $\geq n_1$  P-P, or  $\geq n_2$  S-S phase connections obtained from  
195 this process that were not originally in the ground truth catalog. Reported as a proportion of the  
196 number of total events in the ground truth catalog, we find that rates of previously uncatalogued  
197 detections vary between  $\sim 0.4$  to  $\sim 0.05$  as  $n_1$  and  $n_2$  vary between one and four (Fig. S9).

## 198 **DISCUSSION**

199 In this paper we have presented the possibility of directly ‘memorizing’ pairwise associations  
200 between stations based on the waveforms themselves, using a deep neural network approach which  
201 can be trained on large preexisting datasets of pairwise associated waveforms. We posed the prob-  
202 lem as a binary classification task, and trained two separate models for P-P and S-S phase asso-  
203 ciations. We found no significant differences in performance between the two models (Fig. S3).  
204 The biggest influence on performance metrics were the interstation distances, and the number of  
205 training data points. We found that station pairs with too few training samples (e.g.,  $< 200,000$ )  
206 did not learn robust models for the architecture we chose (Fig. 3). As interstation distances in-  
207 creased, performance quality decreased, however there is a possibility that increased numbers of  
208 falsely associated waveform pairs are contained in the catalog between more distant stations.

209 The method we have implemented could be expanded in several ways. Rather than use a  
210 different CNN for both phase types (P-P, S-S), it seems appropriate to use multi-class classification,  
211 and use a single CNN to predict any combination of phase types, possibly including mixed phase  
212 combinations (e.g., P-S, S-P). Using a single fixed station as the master station was only done  
213 for demonstration; logically a CNN could be trained for all station pairs. If false positive and  
214 false negative predictions are uncorrelated between distinct station pair CNN mappings, using the  
215 redundancy of all  $n^2$  station pairs for  $n$  stations could dramatically improve the reliability of the  
216 outputs. Further, by using all interstation pairs, many station pairs would have small interstation  
217 distances, which could lead to improved prediction results.

218 A suitable use of this method would be to augment other association techniques, rather than  
219 function as a standalone algorithm. Relative arrival time information is a hard constraint on which  
220 sets of arrivals may be associated across a seismic network, and it would be ideal to combine  
221 the insights from travel time based metrics along with information contained in the waveforms  
222 themselves, perhaps in a Bayesian framework (Arora et al., 2013). A valuable way to interpret  
223 the output of this method would be as a mathematical graph, where each arrival pair has an edge-  
224 weight that varies as a function of predicted association likelihood (and phase type). Such graphs  
225 could be optimized to obtain sparse, representative structures (Nemhauser and Wolsey, 1988), and  
226 other graph metrics such as node centrality and connectivity could be used to identify false arrivals  
227 and potentially the number of distinct earthquakes occurring in a time interval. Determining how  
228 pairwise association likelihoods and travel time based association metrics could be combined into  
229 a standalone association algorithm is a potentially promising direction of future research.

230 This method shares some similarities to waveform similarity based detection algorithms. Tem-  
231 plate matching and autocorrelation, among other waveform similarity based methods (Gibbons and  
232 Ringdal, 2006; Brown et al., 2008; Yoon et al., 2015), use identifiable repeating signals to detect the  
233 occurrence of new events. We found that (distinct) events occurring in similar locations were the  
234 most likely to produce false positives for our neural networks (Figs. S5,6). This result suggests that

235 the CNN's may be learning an understanding of the underlying path effects, or Greens functions  
236 connecting arbitrary source-receiver pairs, which is the same mechanism through which classical  
237 waveform similarity detection algorithms succeed. Key differences however are that we focus on  
238 association in this work, rather than detection, and that a single CNN is learning some understand-  
239 ing of all, or at least many path effects, whereas in standard template matching a different template  
240 is used for all known source-receiver pairs. By processing all waveforms through a single neural  
241 network, we expect there may be gains in the CNN's ability to generalize to variants of waveforms  
242 that it has never been trained on (e.g., differing moment tensors, source coordinates, source time  
243 functions), and for which traditional waveform similarity metrics (e.g., cross-correlation, L2-norm  
244 distance) may not be able to adapt to.

## 245 **CONCLUSIONS**

246 Determining earthquake associations is a non-trivial task, given that in general, several auxil-  
247 iary problems must be considered and resolved simultaneously. Determining the number of events,  
248 the phase types of arrivals (including identifying false arrivals), and the associations between ar-  
249 rivals to common sources are all interrelated problems. Methods that enable us to use new forms  
250 of data and information may prove invaluable in assisting with these tasks, and we believe pre-  
251 dicting pairwise association likelihoods directly from the raw waveforms is a valuable step in this  
252 direction.

## 253 **DATA AND RESOURCES**

254 This work included data from the CX seismic network, obtained from the GEOFON data cen-  
255 tre. The catalog used in the analysis is a currently unpublished catalog of northern Chile, available  
256 upon request. All data processing is done in MATLAB and Python, and the neural networks are

257 trained using PyTorch.

## 258 **ACKNOWLEDGMENTS**

259 This work was funded by Institutional Support (LDRD) at the Los Alamos National Laboratory.  
260 We thank Joan Gomberg for helpful discussions.

## 261 **References**

- 262 Aki, K. and Richards, P. G. (2002). *Quantitative seismology*.
- 263 Arora, N. S., Russell, S., and Sudderth, E. (2013). Net-visa: Network processing vertically inte-  
264 grated seismic analysis. *Bulletin of the Seismological Society of America*, 103(2A):709–729.
- 265 Bergen, K. J. and Beroza, G. C. (2018). Earthquake fingerprints: Extracting waveform features for  
266 similarity-based earthquake detection. *Pure and Applied Geophysics*, pages 1–23.
- 267 Brown, J. R., Beroza, G. C., and Shelly, D. R. (2008). An autocorrelation method to detect low  
268 frequency earthquakes within tremor. *Geophysical Research Letters*, 35(16).
- 269 Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *The-*  
270 *ory of Probability & Its Applications*, 14(1):153–158.
- 271 Gibbons, S. J. and Ringdal, F. (2006). The detection of low magnitude seismic events using array-  
272 based waveform correlation. *Geophysical Journal International*, 165(1):149–166.
- 273 Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT  
274 press Cambridge.

275 Grigoli, F., Cesca, S., Amoroso, O., Emolo, A., Zollo, A., and Dahm, T. (2013). Automated  
276 seismic event location by waveform coherence analysis. *Geophysical Journal International*,  
277 196(3):1742–1753.

278 Johnson, C. E., Lindh, A. G., and Hirshorn, B. F. (1997). Robust regional phase association.  
279 Technical report, US Geological Survey,.

280 Kingma, D. and Ba, J. Dp kingma and j. ba, adam: A method for stochastic optimization, arxiv:  
281 1412.6980. *Adam: A Method for Stochastic Optimization*.

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

283 Lu, P., Morris, M., Brazell, S., Comiskey, C., and Xiao, Y. (2018). Using generative adversarial  
284 networks to improve deep-learning fault interpretation networks. *The Leading Edge*, 37(8):578–  
285 583.

286 Nemhauser, G. L. and Wolsey, L. A. (1988). Integer programming and combinatorial optimiza-  
287 tion. *Wiley, Chichester. GL Nemhauser, MWP Savelsbergh, GS Sigismondi (1992). Constraint*  
288 *Classification for Mixed Integer Programming Formulations. COAL Bulletin*, 20:8–12.

289 Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A.,  
290 Antiga, L., and Lerer, A. (2017). Automatic differentiation in pytorch. In *NIPS-W*.

291 Perol, T., Gharbi, M., and Denolle, M. (2018). Convolutional neural network for earthquake de-  
292 tection and location. *Science Advances*, 4(2):e1700578.

293 Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness,  
294 markedness and correlation.

295 Ringdal, F. and Kväerna, T. (1989). A multi-channel processing approach to real time network  
296 detection, phase association, and threshold monitoring. *Bulletin of the Seismological Society of*  
297 *America*, 79(6):1927–1940.

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

300 Ross, Z. E., Meier, M.-A., Hauksson, E., and Heaton, T. H. (2018b). Generalized seismic phase  
301 detection with deep learning. *arXiv preprint arXiv:1805.01075*.

302 Ross, Z. E., Yue, Y., Meier, M.-A., Hauksson, E., and Heaton, T. H. (2018c). Phaselink: A deep  
303 learning approach to seismic phase association. *arXiv preprint arXiv:1809.02880*.

304 Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61:85–  
305 117.

306 Sippl, C., Schurr, B., Yuan, X., Mechie, J., Schneider, F., Gadoev, M., Orunbaev, S., Oimah-  
307 madov, I., Haberland, C., Abdybachaev, U., et al. (2013). Geometry of the pamir-hindu kush  
308 intermediate-depth earthquake zone from local seismic data. *Journal of Geophysical Research:*  
309 *Solid Earth*, 118(4):1438–1457.

310 Wu, Y., Lin, Y., Zhou, Z., Bolton, D. C., Liu, J., and Johnson, P. (2018). Deepdetect: A cascaded  
311 region-based densely connected network for seismic event detection. *IEEE Transactions on*  
312 *Geoscience and Remote Sensing*, (99):1–14.

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

315 Zhu, W. and Beroza, G. C. (2018). Phasenet: A deep-neural-network-based seismic arrival time  
316 picking method. *arXiv preprint arXiv:1803.03211*.

317 Zhu, W., Mousavi, S. M., and Beroza, G. C. (2018). Seismic signal denoising and decomposition  
318 using deep neural networks. *arXiv preprint arXiv:1811.02695*.

**Author Mailing Addresses:**

320 **Ian W. McBrearty:** D446 Los Alamos Natl. Lab P.O. Box 1663, Los Alamos, NM 87545

321 **Andrew A. Delorey:** D446 Los Alamos Natl. Lab P.O. Box 1663, Los Alamos, NM 87545

322 **Paul A. Johnson:** D446 Los Alamos Natl. Lab P.O. Box 1663, Los Alamos, NM 87545

Table 1: Testing evaluation metrics. Precision, recall, and F1 are given separately for each station, and for both phase types. Stations are sorted by epicentral distance from the master station, PB02. Metrics are the average result over 250,000 random samples from the testing data set, which consist of half positive and half negative labelled samples. Median metrics over all station pairs are given on the bottom row.

Station	Distance (km)	P wave			S wave		
		P	R	F1	P	R	F1
PB07	45	0.930	0.864	0.896	0.903	0.896	0.900
PB01	52	0.910	0.877	0.893	0.897	0.897	0.897
PB03	82	0.879	0.860	0.869	0.874	0.885	0.879
PB08	152	0.887	0.820	0.853	0.866	0.846	0.856
PB11	174	0.824	0.830	0.827	0.844	0.837	0.841
Median:		0.887	0.860	0.869	0.874	0.885	0.879

## 323 List of Figures

324	1	Locations of 15,000 earthquakes randomly chosen from the catalog between training (blue circles), validation (black circles), and testing (red circles) datasets. Seismic stations used in this study are marked by triangles, station names are horizontally aligned with their respective stations. Squares represent other stations of the IPOC network used in the development of the catalog. Box in inset of upper right corner marks the study region in Chile. . . . .	19
325			
326			
327			
328			
329			
330	2	An example of extracting pairwise waveforms between stations, and the labelling scheme of true and false association pairs. (top) Seismic records of two stations, with P- and S-arrivals marked, and all 16 possible pairwise connections drawn between arrivals of either station. Filled in lines denote true associations for the four cases where either P-P, or S-S associations of the same earthquake are connected. Dotted lines denote false associations, and include connections between earthquake pairs widely spaced in time or space. (bottom) Insets of the extracted, normalized waveform pairs, along with their assigned label, for four example connections marked above. . . . .	20
331			
332			
333			
334			
335			
336			
337			
338			
339	3	Schematic of the CNN architecture. Light blue box, red boxes, and green boxes denote input data, convolutional layers, and max pooling operators, respectively. Purple boxes denote dense (or fully connected) layers. Kernel sizes and number of convolutional filters in each convolutional layer given above red boxes. ReLU and sigmoid activations are marked below each layer in which they are used. Tensor sizes at each layer of the network are denoted by tuples of (rows, columns, channels). . . . .	21
340			
341			
342			
343			
344			
345			
346	4	An example of the predicted associations across the network, compared with the ground truth. (a) Ground truth associations in catalog. Only arrivals associated to an earthquake are displayed. (b) Predicted associations from CNN's. Each arrival on station PB02 is compared with arrivals on the other stations within +/- 50 seconds, and only the maximum association prediction > 0.5 (of either phase type) is retained. In this panel, all arrivals in the dataset, including false arrivals, are displayed. (c) Example earthquake waveforms and associations from panel (b). In (a,b,c) blue lines denote P-P phase connections, red lines denote S-S phase connections. Green circles mark arrivals on the master station, PB02, blue circles mark arrivals on all other stations. Arrivals and waveforms are plotted on the y-axis at the latitude coordinate of the station they're detected on. . . . .	22
347			
348			
349			
350			
351			
352			
353			
354			
355			
356			

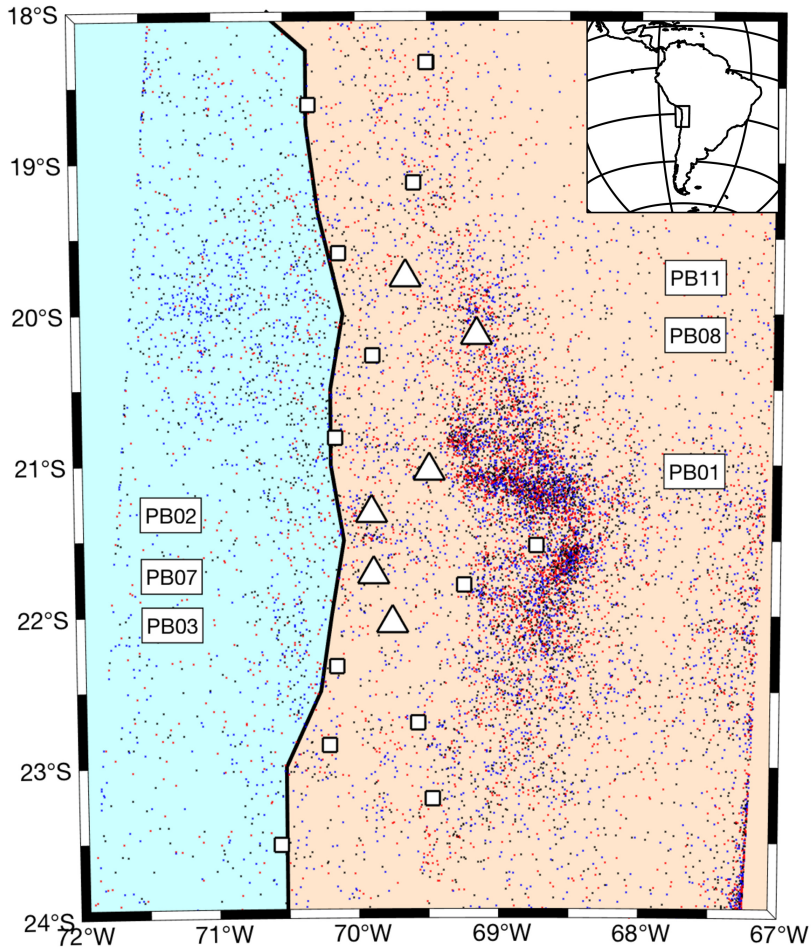


Figure 1: Locations of 15,000 earthquakes randomly chosen from the catalog between training (blue circles), validation (black circles), and testing (red circles) datasets. Seismic stations used in this study are marked by triangles, station names are horizontally aligned with their respective stations. Squares represent other stations of the IPOC network used in the development of the catalog. Box in inset of upper right corner marks the study region in Chile.

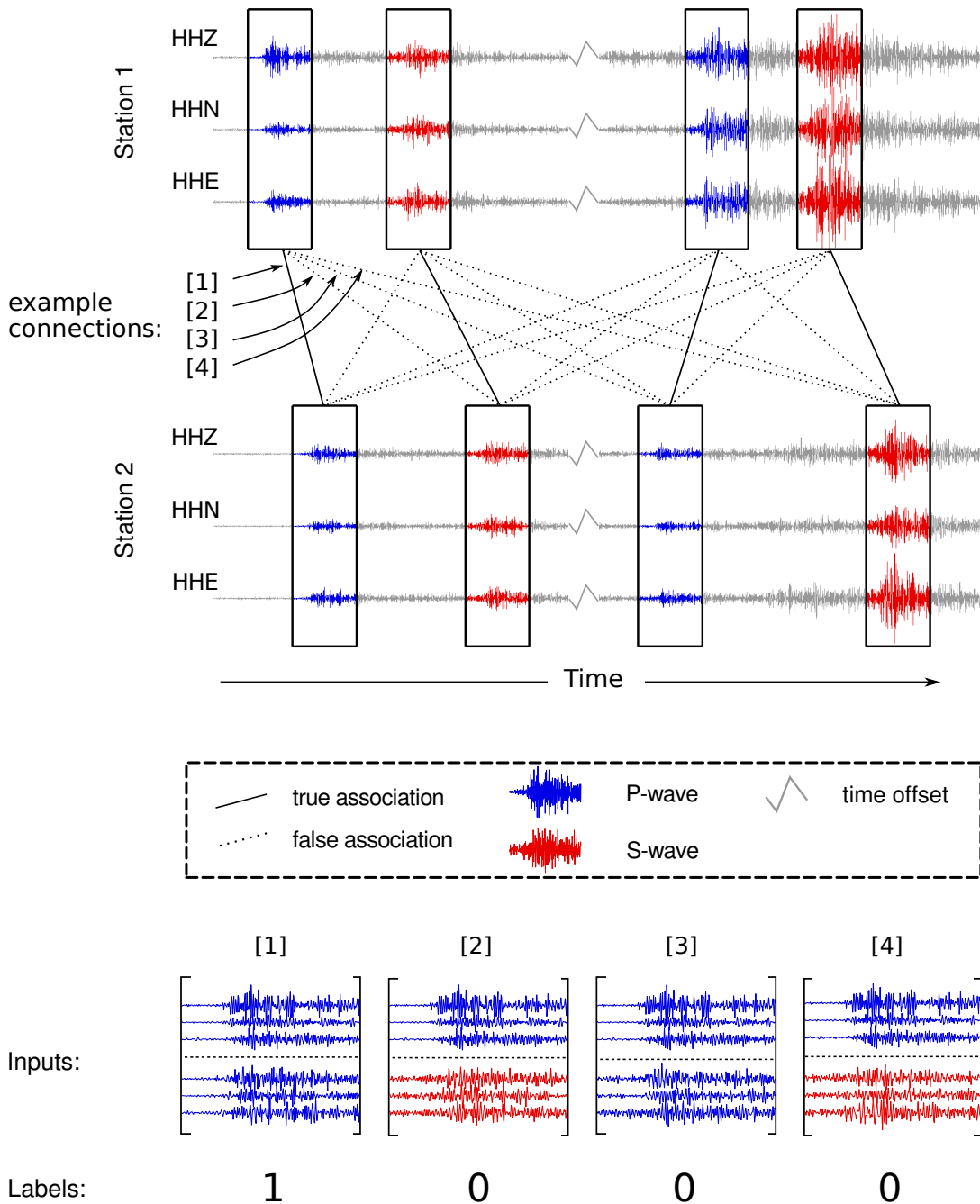


Figure 2: An example of extracting pairwise waveforms between stations, and the labelling scheme of true and false association pairs. (top) Seismic records of two stations, with P- and S-arrivals marked, and all 16 possible pairwise connections drawn between arrivals of either station. Filled in lines denote true associations for the four cases where either P-P, or S-S associations of the same earthquake are connected. Dotted lines denote false associations, and include connections between earthquake pairs widely spaced in time or space. (bottom) Insets of the extracted, normalized waveform pairs, along with their assigned label, for four example connections marked above.

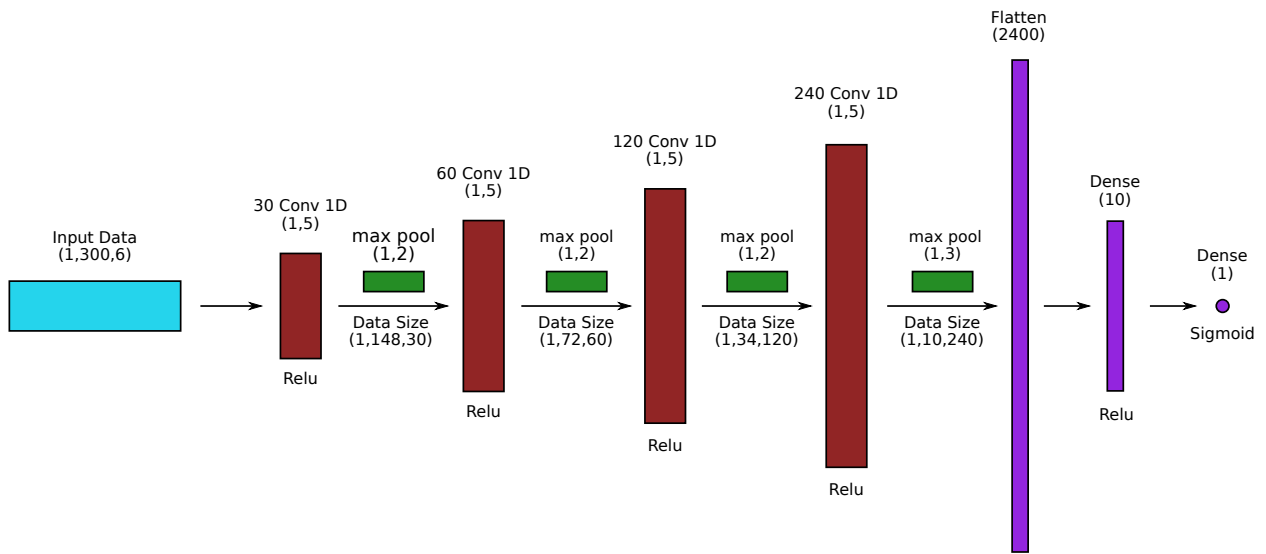


Figure 3: Schematic of the CNN architecture. Light blue box, red boxes, and green boxes denote input data, convolutional layers, and max pooling operators, respectively. Purple boxes denote dense (or fully connected) layers. Kernel sizes and number of convolutional filters in each convolutional layer given above red boxes. ReLU and sigmoid activations are marked below each layer in which they are used. Tensor sizes at each layer of the network are denoted by tuples of (rows, columns, channels).

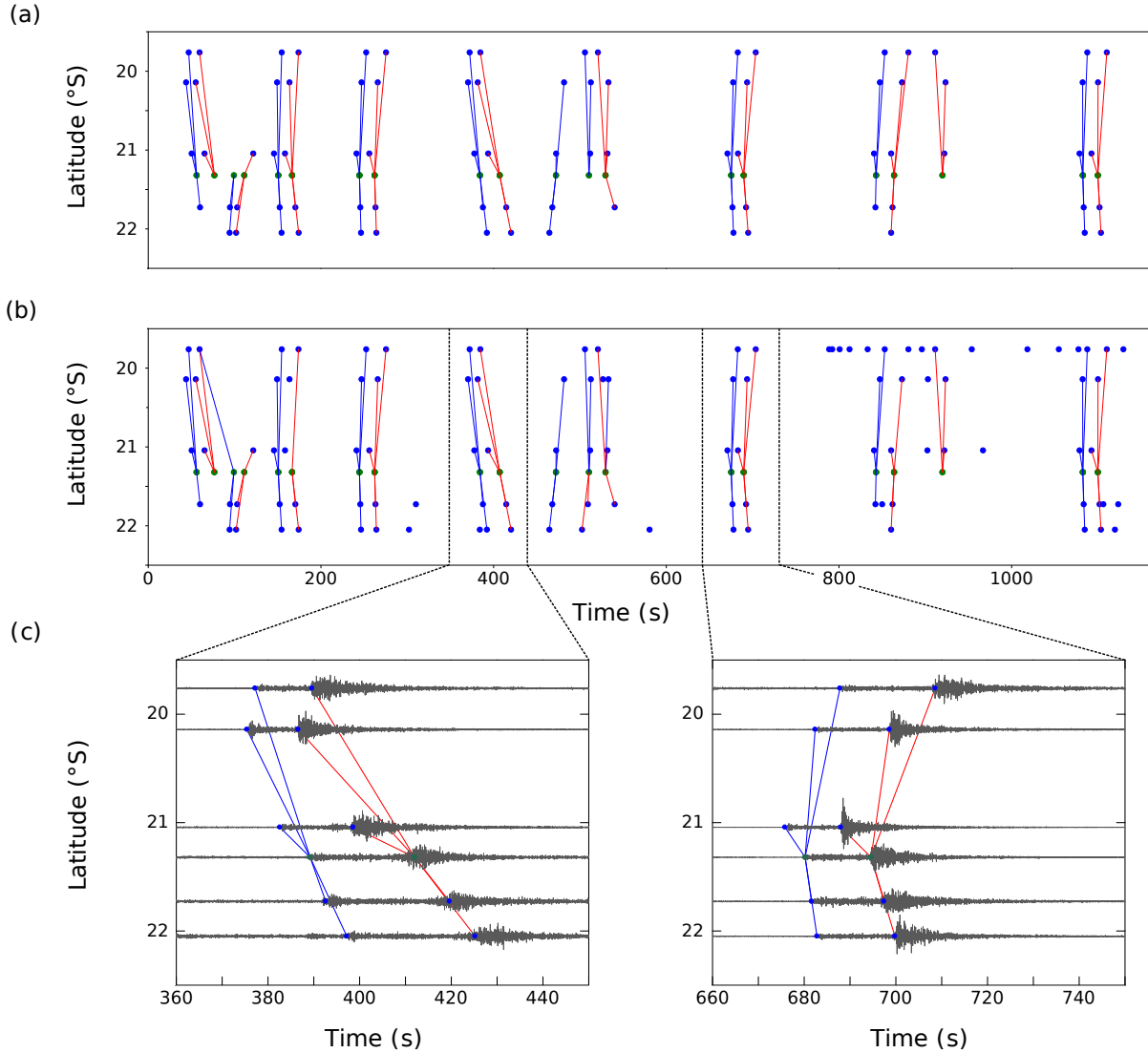


Figure 4: An example of the predicted associations across the network, compared with the ground truth. (a) Ground truth associations in catalog. Only arrivals associated to an earthquake are displayed. (b) Predicted associations from CNN's. Each arrival on station PB02 is compared with arrivals on the other stations within  $\pm 50$  seconds, and only the maximum association prediction  $> 0.5$  (of either phase type) is retained. In this panel, all arrivals in the dataset, including false arrivals, are displayed. (c) Example earthquake waveforms and associations from panel (b). In (a,b,c) blue lines denote P-P phase connections, red lines denote S-S phase connections. Green circles mark arrivals on the master station, PB02, blue circles mark arrivals on all other stations. Arrivals and waveforms are plotted on the y-axis at the latitude coordinate of the station they're detected on.