

An investigation of rapid earthquake characterization using single station waveforms and a convolutional neural network

Anthony Lomax

ALomax Scientific, Mouans-Sartoux, France. e-mail: anthony@alomax.net

Alberto Michelini and Dario Jozinović

Istituto Nazionale di Geofisica e Vulcanologia (INGV), Roma, Italy. e-mail: alberto.michelini@ingv.it, djozinovi@gmail.com

Abstract

Effective early-warning, response and information dissemination for earthquakes and tsunamis require rapid characterization of an earthquake's location, size and other parameters. This characterization is mainly provided by real-time seismogram analysis using established, rule-based, seismological procedures. With the advent of powerful machine learning (ML) tools to make predictions from large data sets with little or no rule-based knowledge, a basic question for earthquake monitoring is how ML performs to detect and characterize general earthquakes using short, single-station seismogram waveforms?

To investigate this question, we adapt a ML, deep convolutional neural network (CNN) for local earthquake detection and epicentral classification using single station waveforms (Perol et al., 2018) to form ConvNetQuake_INGV, a CNN to characterize earthquakes at any distance (local to far teleseismic). ConvNetQuake_INGV operates on 50sec, 3-component, broadband, single-station waveforms to detect seismic events and obtain binned, probabilistic estimates of the distance, azimuth, depth and magnitude of the event. The best performance of ConvNetQuake_INGV is obtained using a last convolutional layer with fewer nodes than the number of output classifications, a form of "information bottleneck".

ConvNetQuake_INGV can detect and characterize earthquakes over a broad range of distances and magnitudes, but errors can be large and there are indications of overfitting of the CNN training data. We find weak evidence that the CNN is performing more than high-dimensional regression and pattern recognition, and is generalizing information or "learning", to provide useful predictions for new events represented by little or no training data.

ConvNetQuake_INGV, though not yet a practical monitoring tool, allows investigation of the performance of CNN and ML in general for rapid, automated, earthquake detection and characterization using short, single-station waveforms. We expect that real-time ML procedures like ConvNetQuake_INGV, perhaps incorporating rule-based knowledge, will ultimately prove valuable for rapid source characterization of earthquakes for earthquake response and tsunami early warning.

1 Introduction

There is increasing demand for earthquake and tsunami early warning and for rapid earthquake characterization to allow immediate mitigation actions, to aid in emergency response, and to provide public information (Allen and Kanamori, 2003; Weber, E., et al., 2007; Tsushima et al., 2011; Newman et al., 2011; Bernardi et al., 2015). Currently, this characterization is mainly provided by real-time analysis of seismogram waveforms using empirical and physics rule-based procedures based on over a century of seismological study (Melgar et al., 2016; Lomax and Michelini, 2012; Allen et al., 2009). The increasing availability of machine learning (ML) tools to make predictions from large and complex data sets with little or no rule-based knowledge (e.g Mehta et al., 2018) has lead to application of ML to a wide variety of waveform based earthquake monitoring problems (e.g., Wang and Teng, 1995; Gentili and Michelini, 2006; Russell et al., 2010; Ruano et al., 2014; Michelini and Lomax, 2017; Rouet-Leduc et al., 2017; Perol et al., 2018; Ross et al., 2018). One of the most general ML approaches is the deep learning, convolutional neural network (CNN; LeCun et al., 2015; Goodfellow et al., 2016; Mehta et al., 2018; Perol et al., 2018) which represents relations between input, intermediate and output quantities as a large network of simple multiply and add operations. A basic question for earthquake monitoring is how might a CNN performs to detect and characterize earthquake location and magnitude when operating directly on short time windows of single-station seismogram waveforms that have only minimal, basic pre-processing?

Advantages of the CNN for seismogram waveform analysis are 1) the CNN can operate directly on the waveforms, with little pre-processing and without feature extraction such as energy detection, time-series transformation, or frequency-domain analysis, 2) the CNN architecture is shift invariant and so not sensitive to or dependent on the time position of features such as P and S wave arrivals in the waveform, and 3) unlike standard regression, the CNN is not limited by assumed and simplified mathematical relations between quantities. These characteristics of the CNN help simplify and generalize the CNN algorithm, reduce the number of algorithm parameters and, in principle, increase the range of event distances and sizes for which the algorithm is valid. However, the CNN does not make explicit use of existing, physics-based or empirical seismological knowledge of relations between earthquake sources and seismic waveforms, such as attenuation, S-P time difference, seismic travel-times, wave polarization angles, or synthetic waveform modeling. This raises questions on the efficiency and accuracy of the CNN approach, and also if the CNN can “learn”, in some sense, any of this knowledge or remains effectively a high-dimensional regression and pattern recognition procedure.

Perol et al. (2018) introduce ConvNetQuake, a CNN for local, microseismic earthquake detection and location using single station waveforms. They show that a CNN operating on 10sec, 3 component, 100Hz, normalized broadband waveform segments from individual stations can detect and classify the epicentral location of nearby earthquakes, outperforming other methods for waveform-based event detection, including waveform matching techniques (Gibbons and Ringdal, 2006; references in Perol et al., 2018). They also show that ConvNetQuake performs well in detecting events whose waveforms are dissimilar to any in the CNN training data

set; this “generalization” of detection ability suggests a possible “learning” of general features of seismogram waveforms for events in the study area.

Here we adapt the procedures and code of Perol et al. (2018) to form ConvNetQuake_INGV, a CNN to detect and determine the location and magnitude of global earthquakes at any distance (local to far teleseismic) over a large range of magnitudes using single station waveforms. ConvNetQuake_INGV is trained with 50sec, 3 component, 20Hz, broadband waveforms from all stations within the regional, MedNet network (Mazza et al., 2008) to: 1) detect seismic events, and 2) obtain binned, probabilistic estimates of the distance, azimuth, depth and magnitude of an event. We examine 1) how well the trained ConvNetQuake_INGV can detect and characterize earthquakes over a large range of distances and magnitudes, and 2) if the CNN is performing more than a high-dimensional regression and is generalizing information or “learning”, so that useful predictions are obtained for events in regions with little or no training data.

This implementation ConvNetQuake_INGV is not proposed as a practical tool for earthquake monitoring, but instead to enable an initial investigation of the performance of CNN and ML in general for rapid, automated, earthquake detection and characterization.

2 The ConvNetQuake_INGV algorithm

ConvNetQuake_INGV, derived from ConvNetQuake (Perol et al., 2018), implements a CNN to detect earthquake events and determine their location and magnitude. ConvNetQuake_INGV is trained through supervised learning using waveform windows containing a diverse set of known event and noise waveforms.

For ConvNetQuake_INGV, we modify the procedures and codes of Perol et al. (2018) and develop new tools to retrieve events, and noise and event waveforms from FDSN web-services [www.fdsn.org/webservices]. For details see (E) sections S1, S2 and S3 available in the electronic supplement to this article.

Event and waveform data

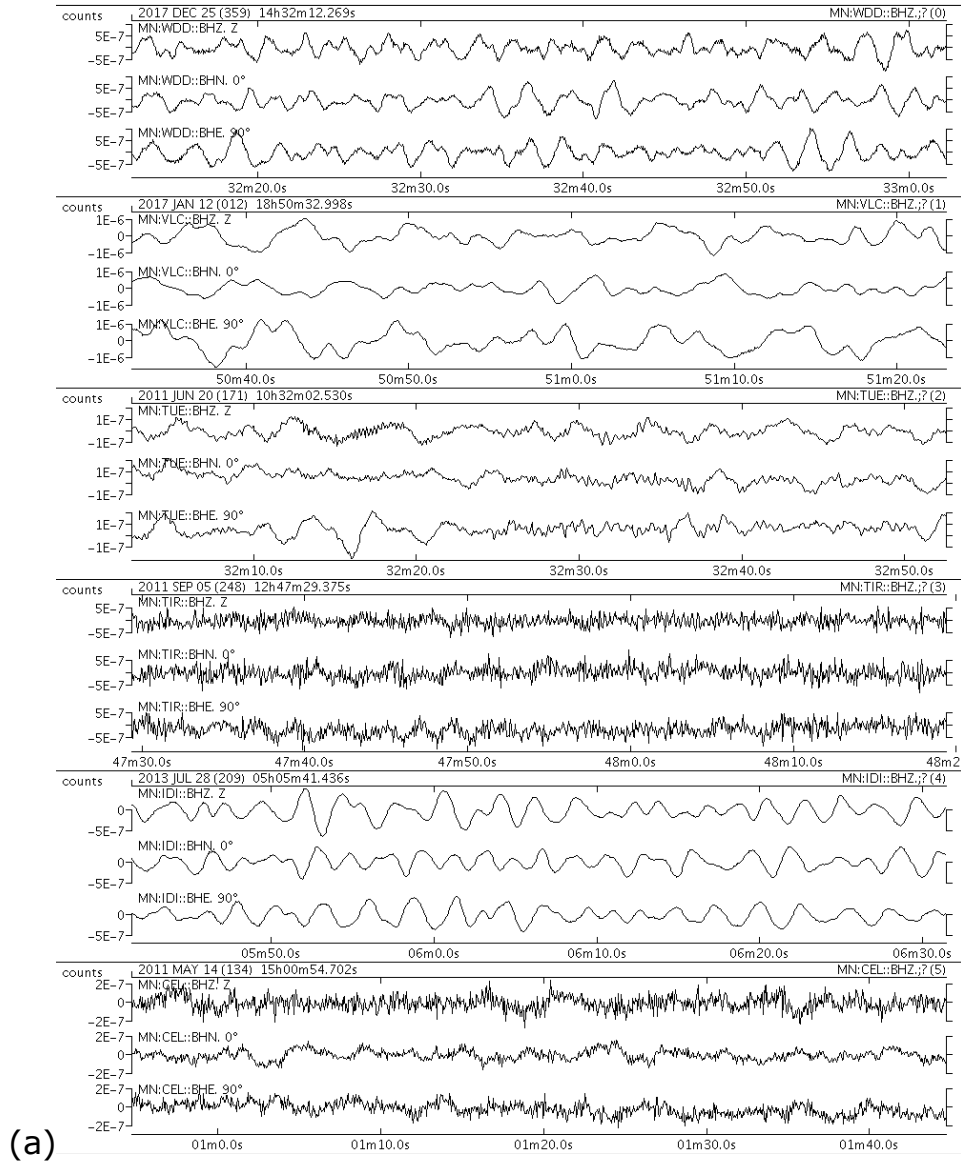
3-component (BHZ/N/E) waveforms which span the distance, azimuth, depth and magnitude range of target events for application of the trained ConvNetQuake_INGV are obtained using random events and stations for training and validate data sets, and for all available stations and random events for a test data set. Noise waveforms are extracted in time windows between random, consecutive events. Each waveform is labelled with binned classifications for: event or noise, station-event distance (0-180°), station-event azimuth (0-360°, 10° step), event magnitude (0-10, 0.5mu step) and event depth (0-700km). The bin steps for distance and depth increase geometrically to give higher weight and precision to nearer and shallower events. (E) Table S1 listing the binned, target labels is available in the electronic supplement to this article.

Waveform quality control and pre-processing includes:

1. checking that event waveforms have signal-to-noise ratio (SNR) greater than a specified threshold, and that noise waveforms have SNR less than a specified threshold,

2. trimming event waveforms to start 5 sec before the predicted P arrival time (using the ak135 model) and event and noise waveforms to a 50 sec total window length,
3. normalizing to the global maximum of all 3 traces, store the normalization value, *stream_max*, to aid in CNN magnitude estimation.

Figures 1 shows examples of noise and event waveforms.



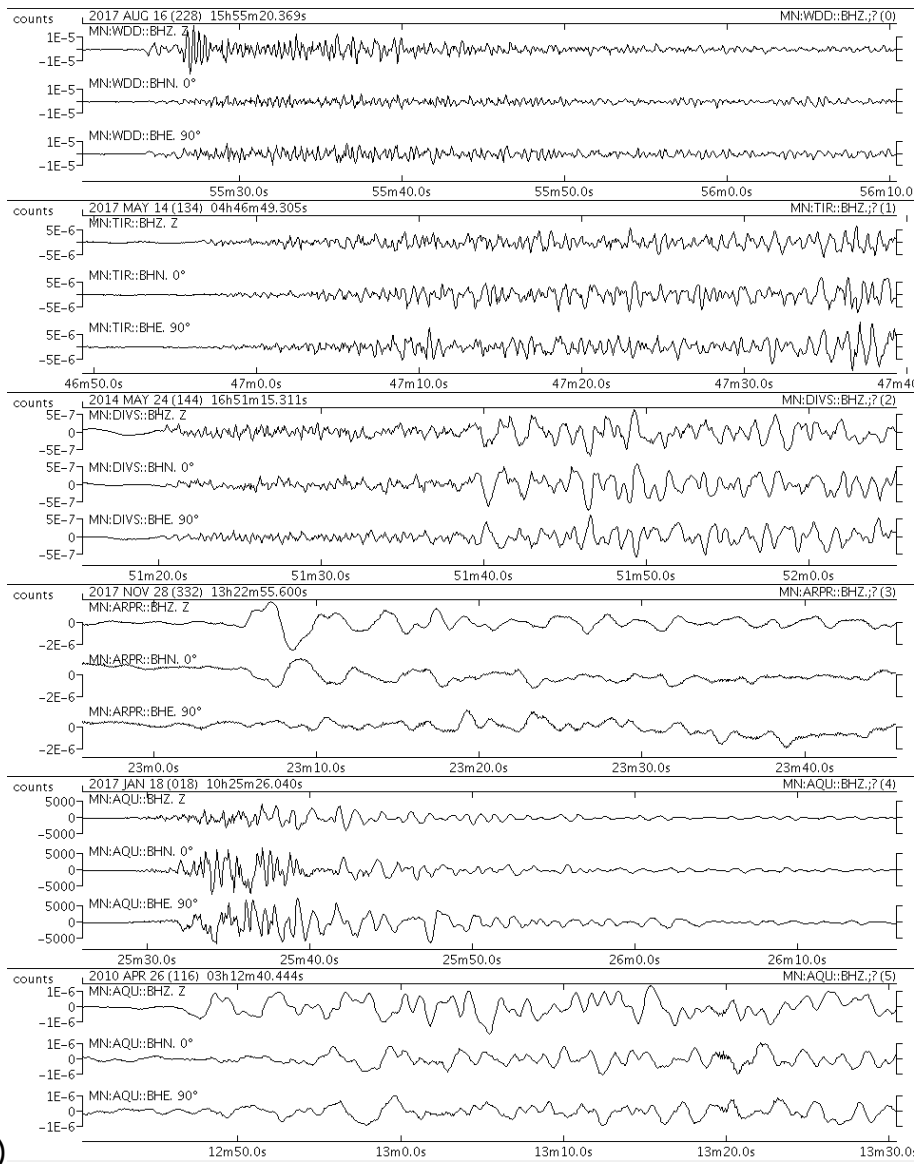


Figure 1. Examples of (a) noise waveforms and (b) event waveforms before normalization

The use in ConvNetQuake_INGV of a window length of 45 sec after the predicted P arrival entails that only waveforms for stations within about 3.5° (about 400km) of an event will contain S waves. The vast majority of regional waveforms will contain only the P wave and following P coda phases, while the teleseismic waveform signal will typically be dominated by the direct arriving P pulses and possibly following phases such as pP, sP, PcP and PP. Presumably, using longer records of many minutes which contain the S wave would greatly improve the ConvNetQuake_INGV distance estimation, but there is little practical use for single-station earthquake characterization after more than a few minutes.

Neural Network Architecture

The CNN architecture for ConvNetQuake_INGV (Figure 2) is similar to that of ConvNetQuake (Perol et al., 2018): an input layer with 3 components of normalized noise or event waveforms of 1000 samples (features), a feed-forward stack of 9

convolutional layers each with 32 CNN channels and half the number of features as the preceding layer giving a last set of 64 nodes (32 channels * 2 features), and 1 fully connected layer derived from a flattened copy of the last convolutional layer. Additionally, in ConvNetQuake_INGV, the normalization factor for the traces, *stream_max*, is appended to the fully connected layer to provide amplitude information for magnitude classification. The final layer contains 1 “no event” node, 50 distance nodes, 20 magnitude nodes, 20 depth nodes, and 36 azimuth nodes, for a total of 127 nodes.

Note that the last convolutional layer has fewer nodes than the final classification layer, which produced smoother results and much less overfitting than did networks with fewer convolutional layers and 128 or more nodes in the last convolutional layer ((E) section S2).

Each channel of the 9 convolutional layers is obtained by convolving the channels of the previous layer with a bank of linear 1D filters, summing, adding a bias term, and applying a point-wise nonlinearity through a rectified linear (ReLU) activation function (Mehta et al., 2018). The softmax function (Mehta et al., 2018) is applied to the output fully connected layer class scores to obtain a properly normalized probability distribution over classification bins. See Perol et al. (2018) for more details and mathematical definitions of these operations.

According to Perol et al. (2018), the use of small filters (with kernel size 3) between convolutional layers means that the first layers respond to localized, higher frequency features in the waveforms, while the deeper layers gain an exponentially increasing receptive field over the input signal and respond to broader, longer period features. The fully connected layer allow the network to combine multiple parts of the signal such as P and S waves to generate a class score and allows detection of events independently of their position within the waveform window. Thus the convolutional layers effectively perform a frequency analysis of the waveform, while the fully-connected layer applies a temporal analysis.

CovNetQuake_INGV Architecture

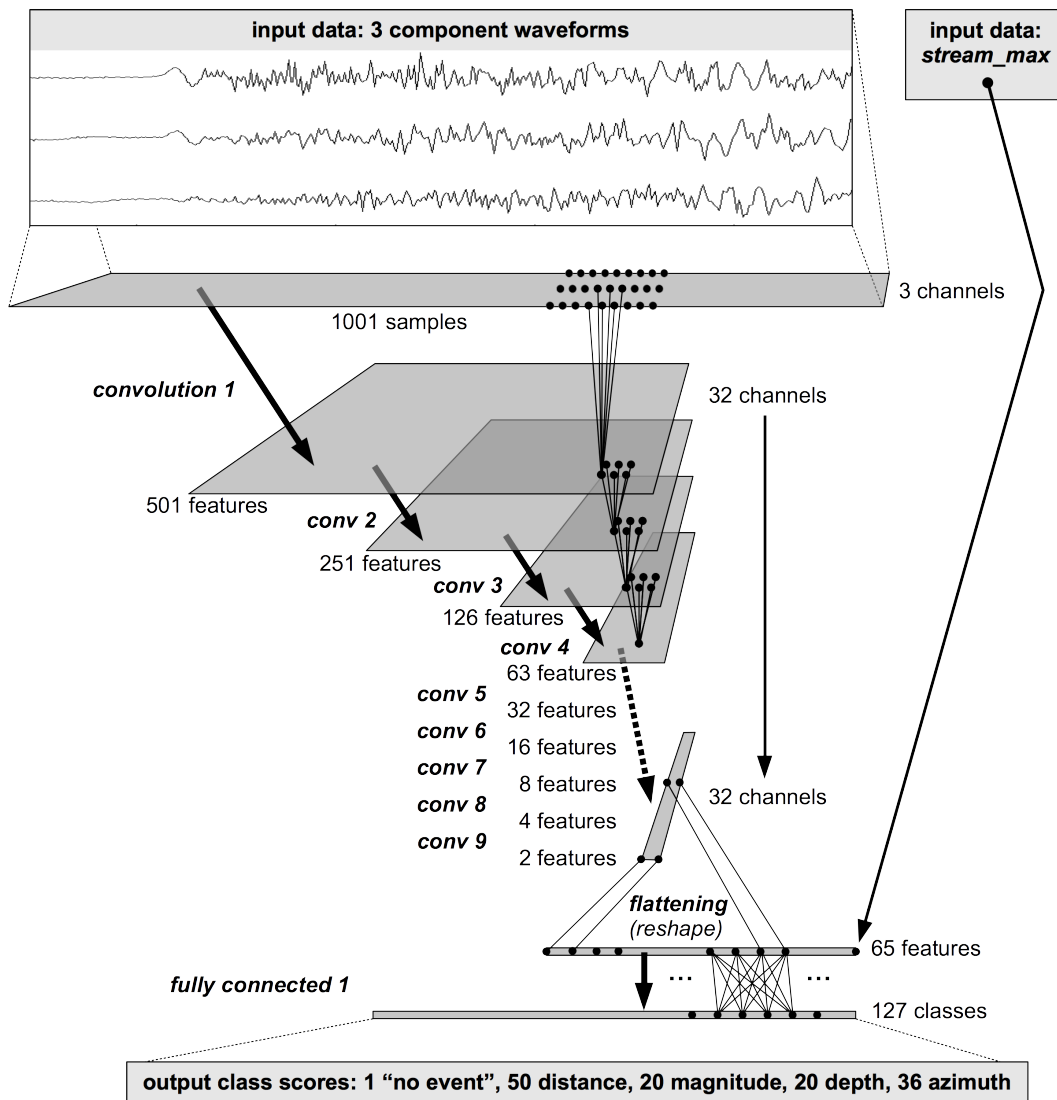


Figure 2. Network Architecture. The final architecture for ConvNetQuake_INGV differs from ConvNetQuake in that it has a flattened layer with 65 features (including stream-max value) and a final fully connected layer containing 127 "no event", distance, magnitude, depth and azimuth classifications.

For training, ConvNetQuake_INGV uses an L2-regularized cross-entropy loss (misfit) function and the Adam Optimizer algorithm, as in ConvNetQuake (Perol et al., 2018; Mehta et al., 2018; Kingma and Ba, 2017).

Overfitting is a major problem for regression and learning algorithms, such as CNN, with a large number of coefficients, since any data set is necessarily finite and usually small relative to ideal data sets spanning the full, true data space (Mehta et al., 2018). During training, an efficient ML algorithm may first under-fit (estimate similar class values for many waveform samples) with relatively large error, then, as training proceeds, provide better and better fit with smaller error. Ultimately, this process can lead to overfitting (fitting of details and noise in the training samples), and not to fitting general features of the data that would give good performance for new data. Overfitting can be identified by applying a trained CNN

to a test data set that has not been used for training or for validation (iteratively assessing the training results and modifying the CNN parameters). In the case of ConvNetQuake_INGV, overfitting may be indicated by high-probability, outlier classification of distance, azimuth, depth or magnitude bins far from the true test event values and often towards classifications corresponding to high training event density.

3 Application to local, regional and global events

Training

We train ConvNetQuake_INGV using 15200 event (Figure 3) and 10724 noise waveforms from MedNet stations (BHZ/N/E components) for events from 2010-2018 from 0-180° with the following lower magnitude limits as a function of event epicentral distance:

$M \geq 3.0$	$\Delta \leq 2^\circ$
$M \geq 4.0$	$2^\circ < \Delta \leq 20^\circ$
$M \geq 5.0$	$20^\circ < \Delta \leq 70^\circ$
$M \geq 6.0$	$70^\circ < \Delta \leq 180^\circ$

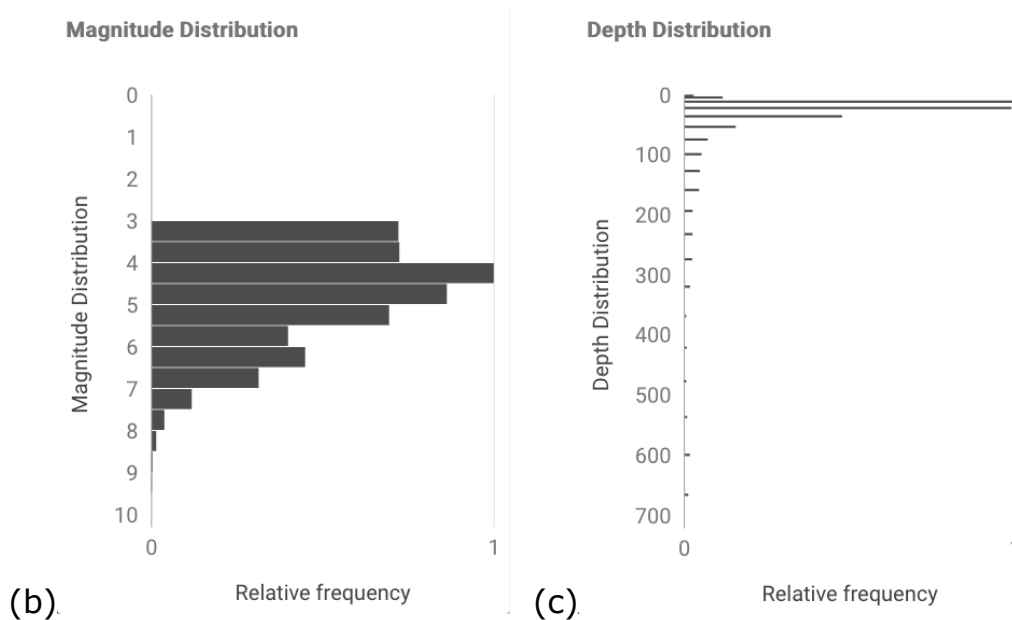
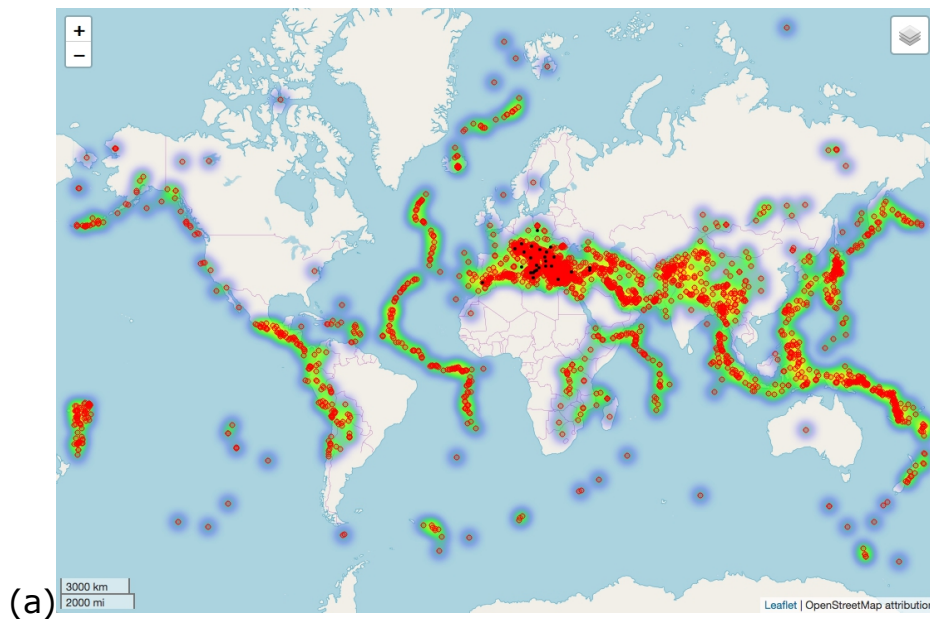


Figure 3. Training event data set (2010-2018): (a) Event epicenter map showing events (red dots), event heat (density) map and MedNet stations (black dots). (b) Magnitude distribution. (c) Depth distribution.

The training ran for over 1M steps, and required around 1.5 days using a CPU (4-core, 3.5 GHz Intel Core i7).

Figure 4 shows the evolution of detection, distance, azimuth, depth and magnitude accuracy and overall loss during training. The detection accuracy (correct detections / total number of events and noise waveforms; Goodfellow et al., 2016) rises quickly in training and reaches close to 1.0 after around 100-200K training steps. The accuracy for distance, azimuth, depth and magnitude, which only quantify if predicted classifications are exactly the same as the true class, rise at lower rates and towards lower asymptotes than detection accuracy.

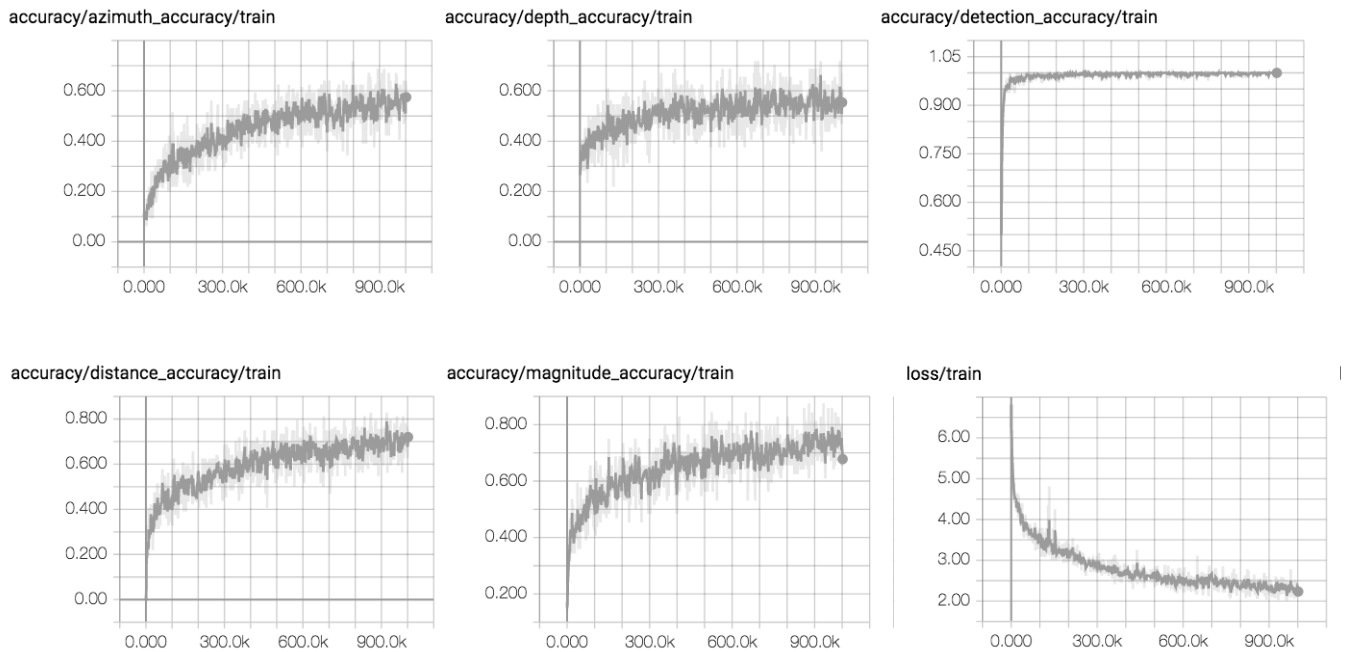


Figure 4. Evolution of accuracy and L2-regularized, cross-entropy loss during training plotted as a function of training step up to around 1M steps. Accuracy is the ratio of number of correct classifications to the total number of input data (all waveforms for detection, event waveforms otherwise).

The training loss (Figure 4) is the sum of L2-regularized cross-entropy losses for distance (including correct or incorrect detections), and for azimuth, depth and magnitude (ignoring noise events). The training loss decreases asymptotically during training.

The validation data set consists of 1773 event waveforms and 1198 noise waveforms, selected as for training but not used in training. Figure 5 shows scatter diagrams for the validation data set of predicted versus true classification values for distance, depth, magnitude and azimuth. Overall the predictions match well the true values, though there are erroneous results for distance at greater than about 5° , and for many values of azimuth and depth. Similar diagrams for the training data set ((E) Figure S1) shows very good agreement between predictions and true classification values, indicating that the ConvNetQuake_INGV training is able to fit the training data nearly perfectly but also may be indicating overfitting.

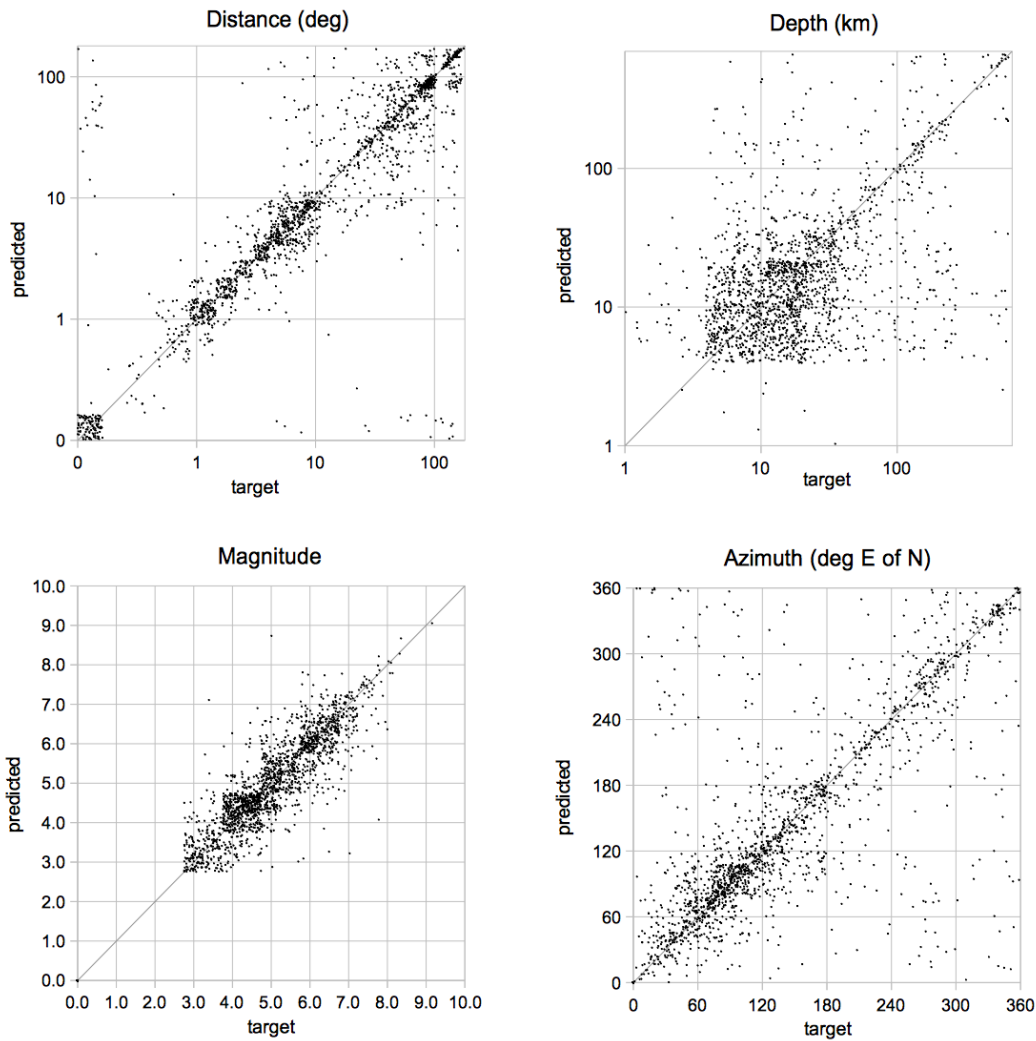


Figure 5. Predicted vs true validation data set (2010-2018) classification results for distance, depth, magnitude and azimuth, after training for 500k steps. Predicted and true (target) values are perturbed randomly within their respective bins to avoid overlapping data points and aid in visualization of spatial density of points. For distance, few scattered points along the predicted (false positive) and target (missed events) axes show the high accuracy of detection (scatter points near the origin).

Results on independent test event data

The following presents results for test events from 2009, which were not used for training or validation. The test data set consists of 1003 event waveforms and 621 noise waveforms from MedNet stations (BHZ/N/E components) from 2009 at 0-180° distance with the following lower magnitude limits as a function of event epicentral distance:

$M \geq 4.0$	$\Delta \leq 2^\circ$
$M \geq 5.0$	$2^\circ < \Delta \leq 20^\circ$
$M \geq 6.0$	$20^\circ < \Delta \leq 50^\circ$
$M \geq 6.5$	$50^\circ < \Delta \leq 180^\circ$

Figure 6 shows the geographic and magnitude distributions of the 2009 test event set. For each event, all available MedNet station traces which satisfy the waveform quality control were used as test data.

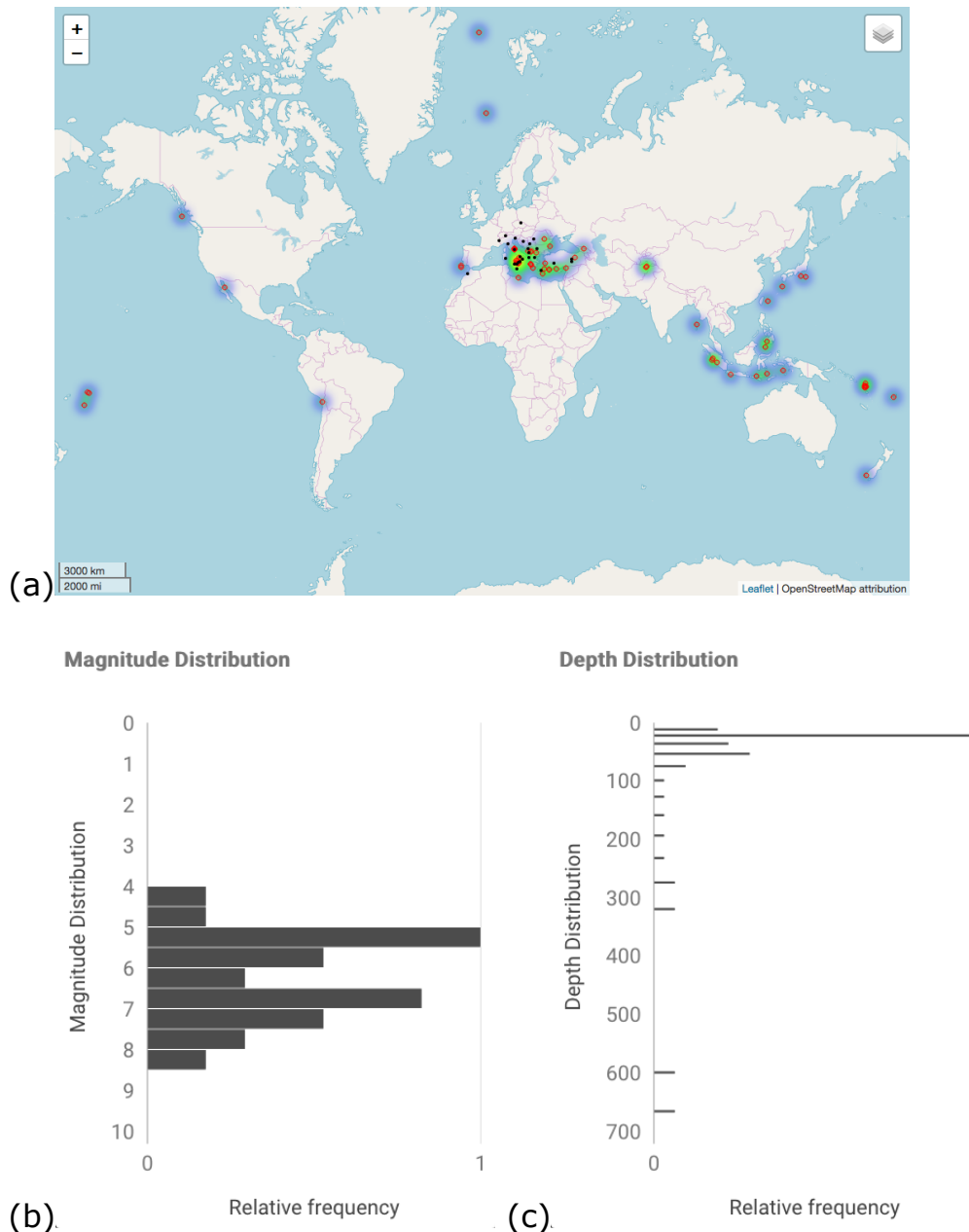


Figure 6. Test event data set (2009): (a) Event epicenter map showing events (red dots), event heat (density) map and MedNet stations (black dots). (b) Magnitude distribution. (c) Depth distribution.

ConvNetQuake_INGV trained with 500k steps is applied to the test events from 2009. Figure 7 shows the predicted vs. true classification results for the 2009 test data set. These results are similar to, but notably more scattered than the validation results (Figure 5). It is likely that the test events, drawn from a time window not in the training data, contain few waveforms similar to those in the training data, while the validation events, drawn from the same time window as the training data, include many aftershock, swarm and other events with waveforms similar to those in the training data. Thus ConvNetQuake_INGV may not generalize

well to event waveforms that are not well represented in the training data set, and is not likely to perform better on future event waveforms than on the test data set.

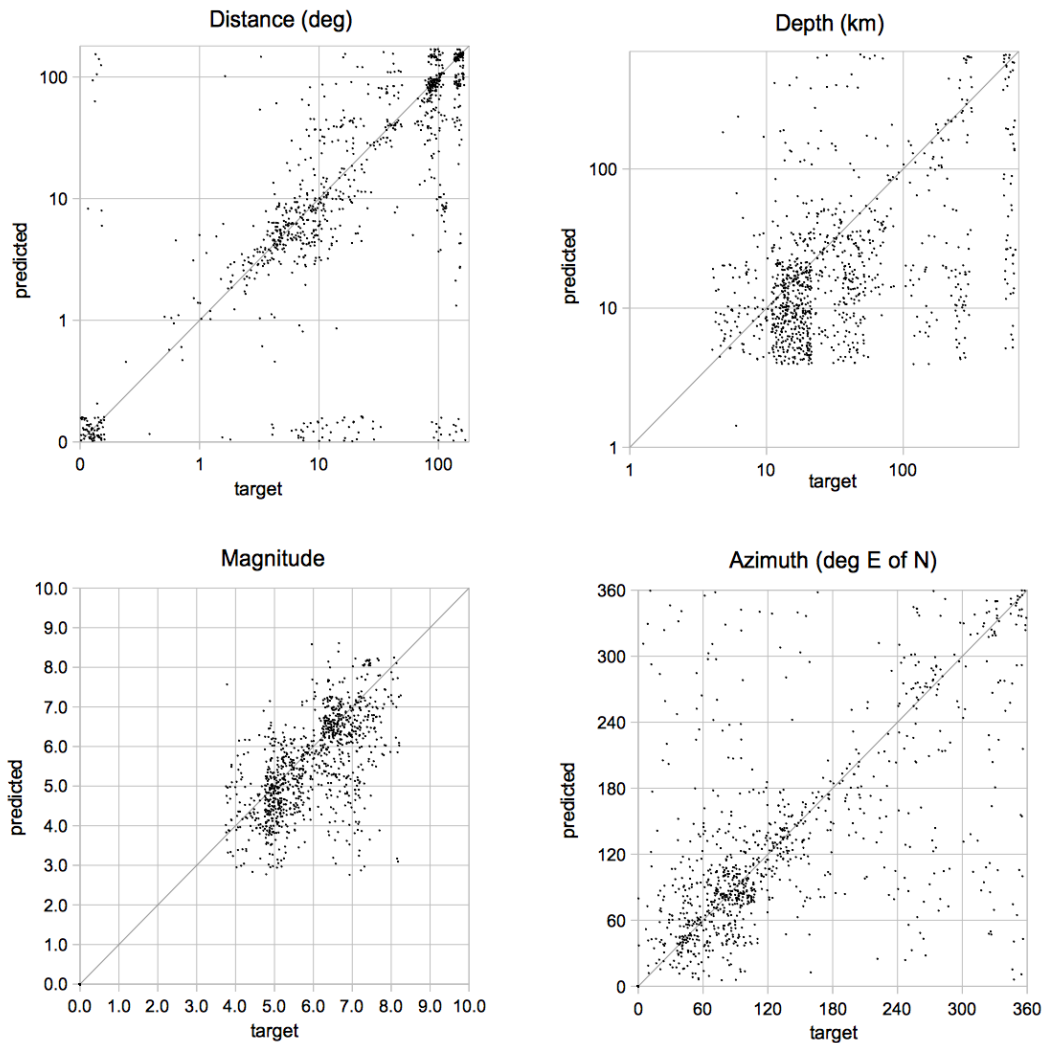


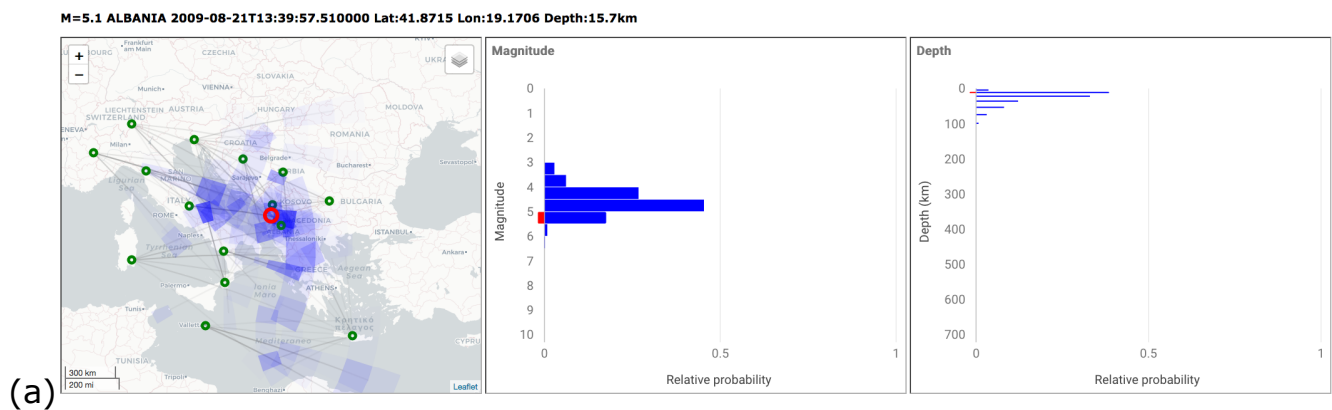
Figure 7. Predicted vs true classification results for the test data set (year 2009; 999 event waveforms and 623 noise waveforms) for distance, depth, magnitude and azimuth, after training for 500k steps. Predicted and true (target) values are perturbed randomly within their respective bins to avoid overlapping data points and aid in visualization of spatial density of points. For distance, few scattered points along the predicted (false positive) and target (missed events) axes show the high accuracy of detection (scatter points near the origin).

We consider performance measures (Goodfellow et al., 2016) for the 2009 test data set after 500k training steps. The detection accuracy (correct detections / total number of events and noise waveforms) is 0.87, the detection precision (number of correct event detections / total number of predicted event detections) is 0.97, the recall (number of correct event detections / total number of events) is 0.81, and the F1 statistic ($2 * \text{precision} * \text{recall} / (\text{precision} * \text{recall})$) is 0.88. These statistics suggest a good to very good event detection performance for ConvNetQuake_INGV. We do not analyze results for the test data set after 1M training steps because they showed slightly worse accuracy and precision, and more high-probability outlier classifications (evidence of overfitting) than after 500k steps.

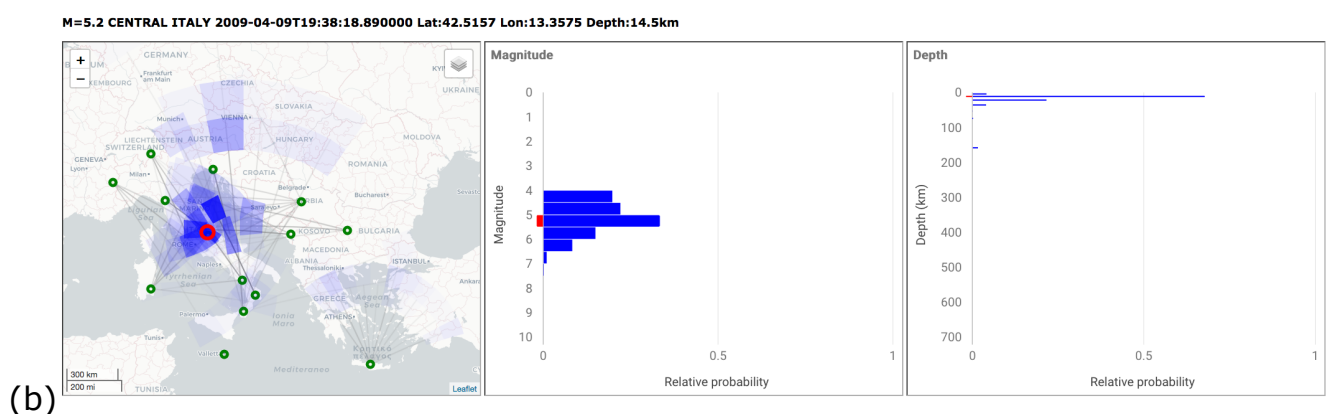
Figures 8 show plots of distance-azimuth bin (epicenter) softmax probabilities, and magnitude and depth histograms with comments for selected test events. The complete set of results for the 2009 test event results are available (E) as plot files and as interactive html source files in the electronic supplement to this article.

[NOTE: for publication, all Figure 8 captions will be grouped together separate from the figure panels.]

Figure 8. Epicenter, magnitude and depth results after 500k training steps with qualitative assessment of event constraint. Shown for each event: Epicenter probability map (left) with true epicenter (red circle), stations with available waveforms (green circles), predicted epicenter probability for station distance-azimuth bins (blue patches; patch intensity shows relative probability; grey line connects to corresponding station). Histograms over all stations of predicted, relative probabilities for magnitude (middle) and depth (right) from all stations; red bar on the left of each histogram shows true magnitude (middle) and depth (right).

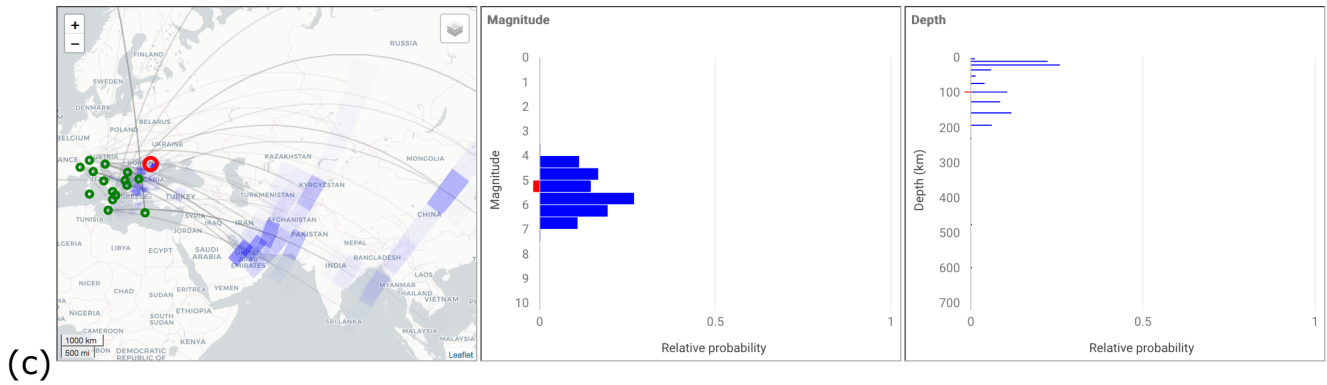


(a) Moderately well constrained local/regional event: epicenter, magnitude and depth are moderately well constrained around the true values.



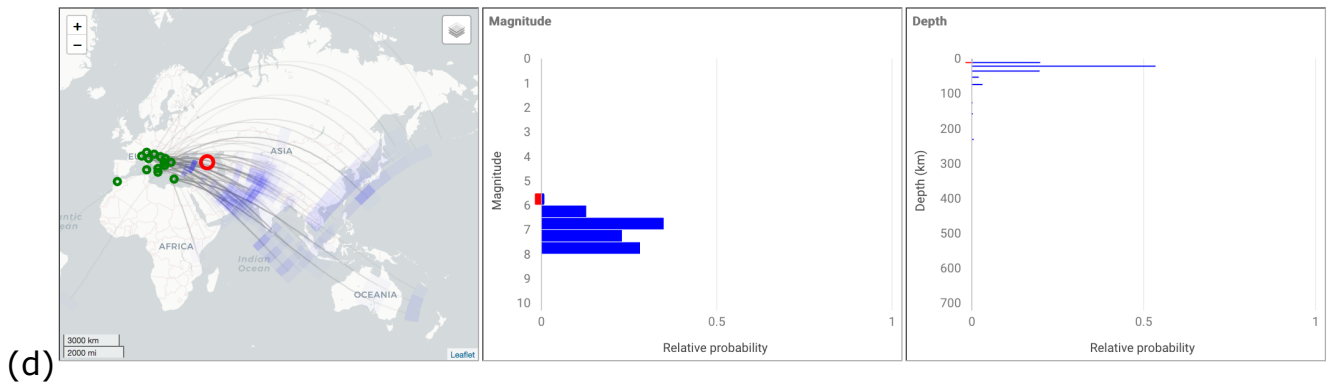
(b) Well constrained local/regional event: epicenter, and depth are well constrained around the true values for most stations, with outlier results for a few stations. Magnitude is moderately well constrained.

M=5.2 ROMANIA 2009-04-25T17:18:48.870000 Lat:45.7039 Lon:26.5365 Depth:102.4km



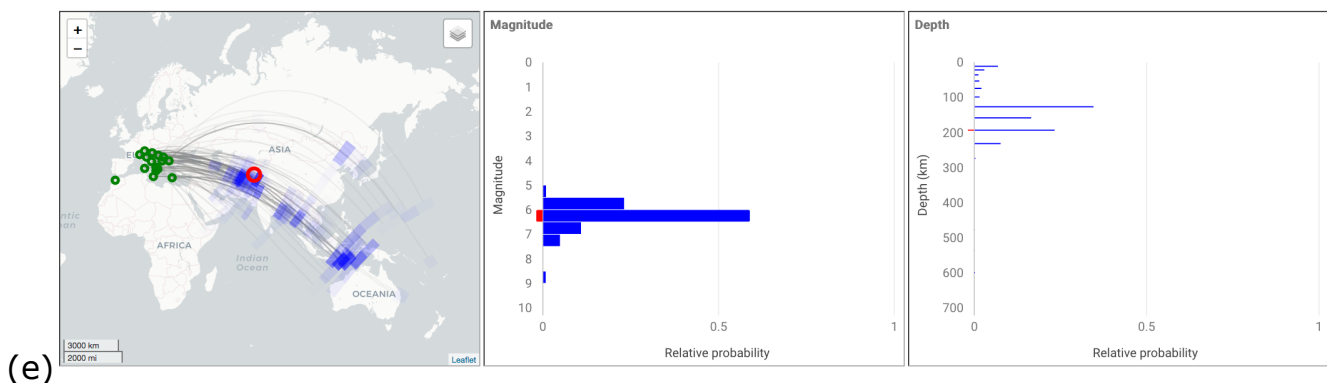
(c) Poorly constrained, local/regional event: epicentral distance, azimuth and depth are poorly constrained around the true values, while magnitude is moderately constrained. High relative probability distance-azimuth outlier bins far from the epicenter and with large azimuth error, the high relative probability for epicenters around Iran/Afghanistan (likely a result of the high training event density in this area, see Figure 3a), and high probabilities for shallow depth while the true depth is intermediate at 100km, are indications of overfitting.

M=6.0 NORTHWESTERN CAUCASUS 2009-09-07T22:41:37.160000 Lat:42.5829 Lon:43.4523 Depth:14.6km



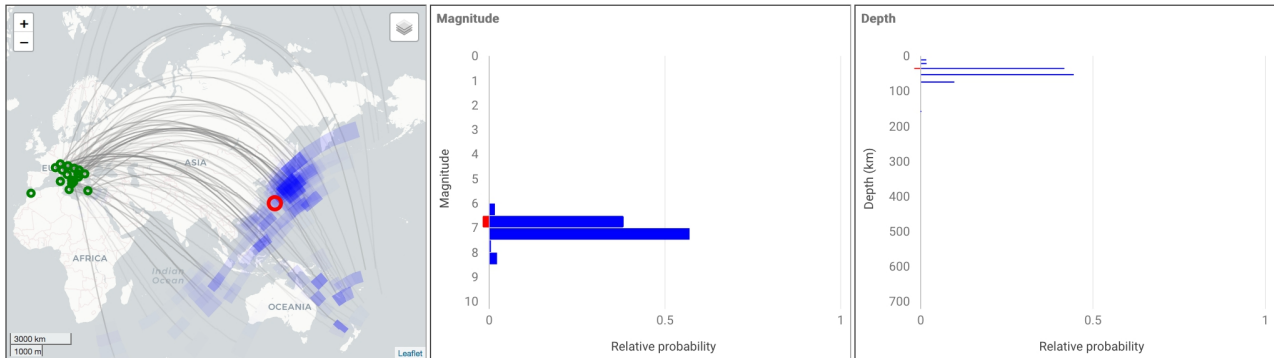
(d) Poorly constrained, far regional event: epicentral distance and magnitude are poorly constrained around the true values, while azimuth is moderately well constrained, and event depth is well constrained. Moderate relative probability distance-azimuth outlier bins far from the epicenter in areas of high training event density (see Figure 3a) are indications of overfitting.

M=6.1 AFGHANISTAN-TAJIKISTAN BORD REG. 2009-10-22T19:51:27.600000 Lat:36.5187 Lon:71.0121 Depth:188.6km



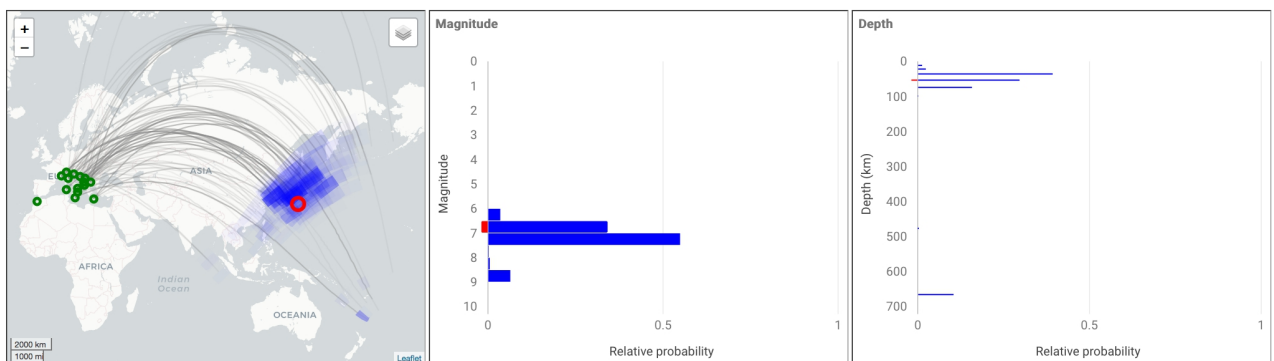
(e) Moderately constrained, intermediate depth, far-regional event: azimuth is well constrained, depth and magnitude are moderately well constrained around the true values, distance has poor constraint) The high relative probability for epicenters towards and around Indonesia, partly due to the high training event density in this area (Figure 4a), is an indication of overfitting.

M=6.8 RYUKYU ISLANDS, JAPAN 2009-10-30T07:03:39.120000 Lat:29.1677 Lon:129.9847 Depth:33.9km



(f) Moderately constrained teleseismic event: distance, magnitude and depth are well constrained near the true values, azimuth is poorly constrained. The high relative probability for epicenters biased towards Japan, likely a result of the high training event density around Japan (Figure 4a) and of weak azimuth constraint, is an indication of overfitting. Higher probability epicenter bins with too-large distance, but correct azimuth in the southwest Pacific may be due to a combination of overfitting and a real similarity in P waveform for this event and events at far teleseismic distances.

M=6.6 SOUTHEAST OF HONSHU, JAPAN 2009-08-12T22:48:52.100000 Lat:32.8133 Lon:140.4279 Depth:61.8km



(f) Well constrained teleseismic event: distance, azimuth, magnitude and depth are well constrained around the true values but there are some outliers.

Analysis of epicentral errors

An analysis of the error in epicenter estimates for the test event data set provides a more quantitative assessment of ConvQuakeNet_INGV and helps understand if it is performing more than a high-dimensional regression and may be extrapolating information and "learning" from the training data set. We use an extended set of 4074 test events from 2007-2009, selected otherwise with the same criteria as for the 2009 test data.

Figure 9 shows distance and azimuth mis-location errors for individual station distance-azimuth estimate for the 2007-2009 test events after 500k training steps. A plot of distance error vs. event distance (Figure 9a) shows an overall increase in distance error with distance up to around 30° true epicentral distance, and similarly distributed distance errors at greater distances. The variation in error at all distances is large, over an order of magnitude, but the majority of distance errors are the same or smaller than the true event distance except at very small distances and around $3.5\text{--}40^\circ$. A sharp and strong increase in maximum error beyond about 3.5° indicates reduced distance accuracy when there is no S arrival in the waveform window (beyond about 3.5° or 400km, where the S onset is about 40 sec after the P arrival). The larger errors around $3.5\text{--}40^\circ$ may also be related to the complexity of the first 45 sec of regional waveforms consisting of P and P coda waves strongly affected by the Mohorovicic discontinuity and upper mantle phase triplications (Bai and Kennett, 2001).

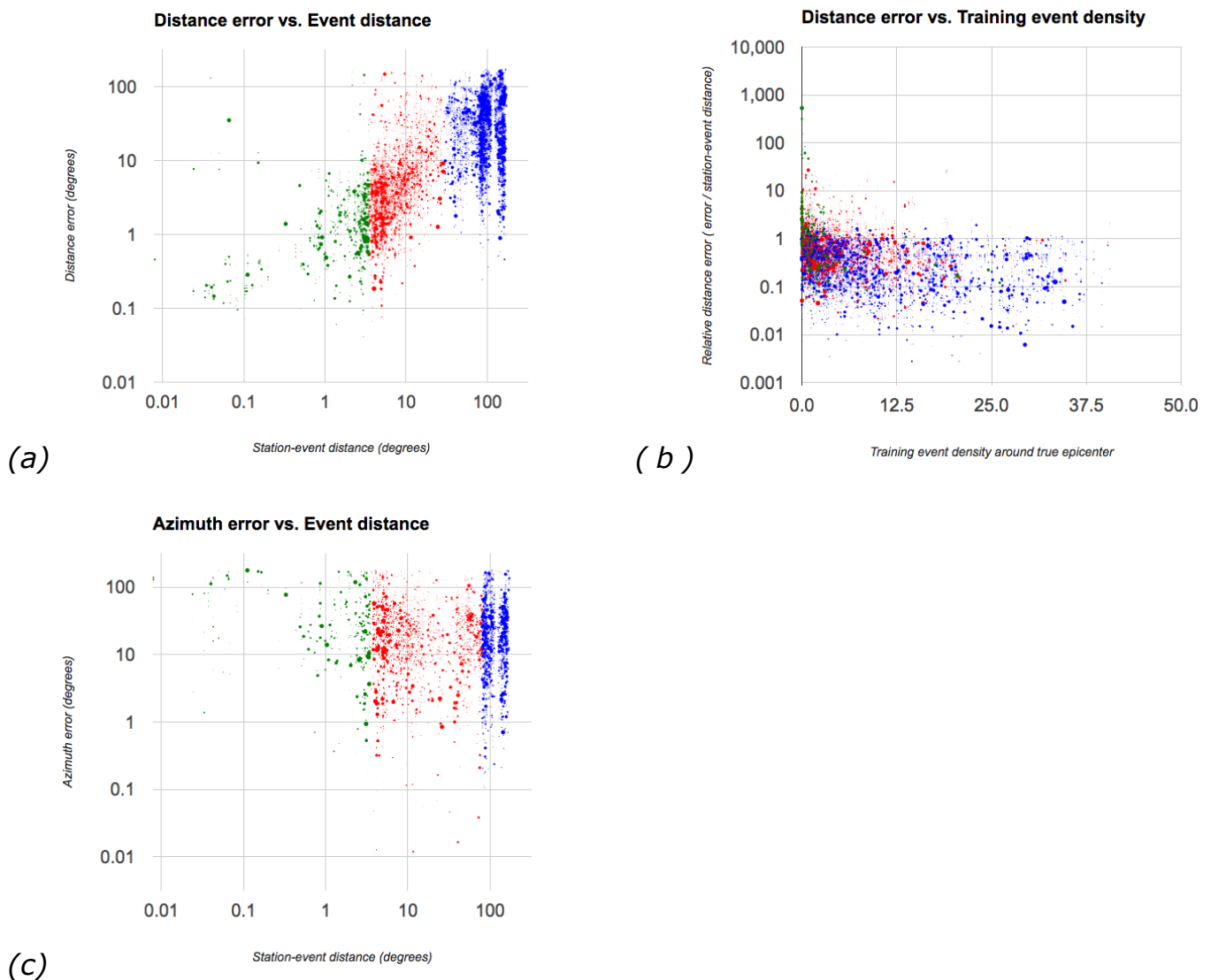


Figure 9. Errors analysis for 2007-2009 test events for individual station waveforms. Point size shows relative probability for the corresponding station distance-azimuth bin. (a) (b) Green, red and blue points show stations at less than

3.5°, 3.5-30°, or greater than 30°, respectively from the true epicenter. (a) Distance error (difference between predicted and true station epicentral distance) for individual stations as a function of true epicentral distance. (b) Relative distance error (distance between predicted and true epicenters for individual stations / true epicentral distance) as a function of station training event density around true epicenter (Equation 1). (c) Station to event azimuth error (difference between predicted and true azimuth) as a function of epicentral distance; green, red and blue points show stations at less than 3.5°, 3.5-80°, or greater than 80°, respectively from the true epicenter.

Another way to examine the epicentral errors is with regards to the density, ρ , of training events near each test event, here defined as a Gaussian-blur, distance-weighted sum over all training events, i ,

$$\rho = \sum_i e^{-\Delta_i^2/2\sigma^2} , \quad (1)$$

where Δ is the distance between the training event, i , epicenter and the test event epicenter, and σ is 1/20 the mean over the epicentral distances for all stations providing waveforms for the event.

A plot of relative distance error as a function of training event density ρ (Figure 9b) shows a mild, overall reduction of distance error with increasing ρ , showing that more training data provides more constraint in fitting nearby test events. However, there is an indication of increased relative distance error for epicentral distances $< 30^\circ$ at low training event density ($\rho < \sim 6$), possibly related to the complexity and variability of local and regional waveforms.

A plot of azimuth error vs. event distance (Figure 9c) shows reduced error for true epicentral distances between about 5-80°. At the lower end of this range, the lack of S arrivals in the waveform window may aid in azimuth determination on the first P pulse, while beyond 10-20° the initial teleseismic P waveform can have a relatively simple form (Bai and Kennett, 2001) with well defined P polarization azimuth. There is decreasing constraint on azimuth with increasing distance since the P wave incidence angle at the station steepens and signal amplitude on the horizontal components decreases. This decreasing constraint would explain the apparent increasing azimuth error from about 20-80deg, while the lack of direct P after 90-100° and very steep incidence for PKP beyond about 140° would explain the jump to large azimuth errors beyond $\sim 80^\circ$.

Large azimuth error at the smallest distances ($< \sim 2^\circ$) may be the result of local event waveform variability, where the signal is dominated by S waves whose polarization is mainly dependent on source mechanism and crustal structure, and not station to epicenter azimuth.

4 Discussion

We have introduced and analyzed a CNN (ConvNetQuake_INGV) in order to explore the potential of ML to use single station seismograms for event detection and for predicting event distance, azimuth, depth and magnitude for earthquakes at local, regional and teleseismic distances.

After training, ConvNetQuake_INGV performs very well for the training data set, as is to be expected, and performs well for the validation data set, but this data spans the same time range as the training data and may include similar waveforms. The performance with the more independent test data set is most important, as it indicates the likely performance of ConvNetQuake_INGV on waveforms for future events.

For the 2009 test data set ConvNetQuake_INGV performs very well for event detection, and shows moderately good overall performance for predicting event distance, azimuth, depth and magnitude for the test data set. But single station results often show large error in predictions, so the current configuration and training of ConvNetQuake_INGV is not yet a practical tool for application to future event waveforms.

Evidence of possible overfitting the training data is shown by:

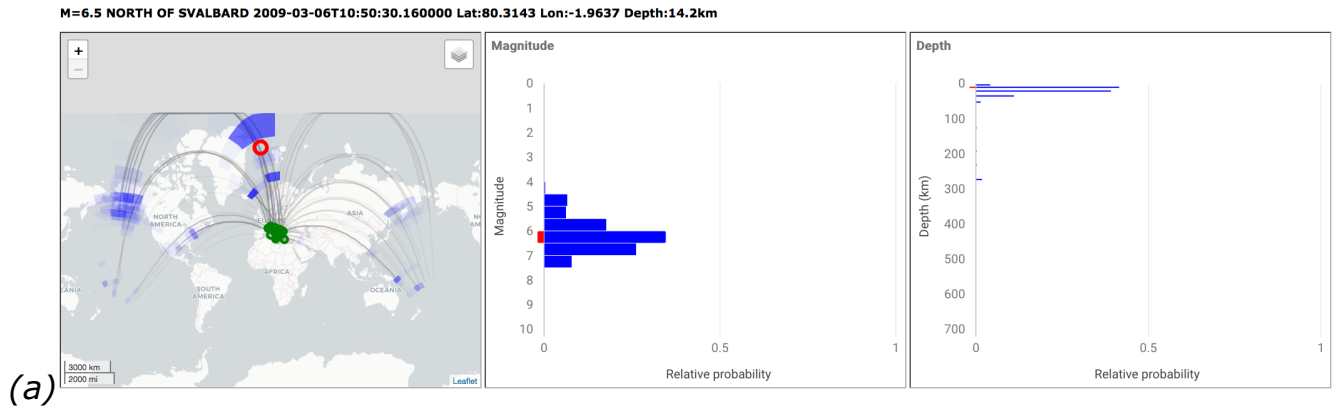
1. Larger number of incorrect predicted vs true classification values for the validation dataset relative to the training dataset, and for the test dataset relative to the validation dataset.
2. Results for individual events for the test datasets often show relatively high probability for outlier classification bins far from the true values; outliers often occur in areas with a high density of training events.
3. The test data results after 1M training steps show more high-probability outlier classifications than after 500k steps.

If ConvNetQuake_INGV is generalizing beyond the information in the training data set through interpolation and “learning” of waveform characteristics that determine event distance, azimuth, magnitude and depth, then application of ConvNetQuake_INGV to waveforms from new and future events can provide useful characterisation of the events.

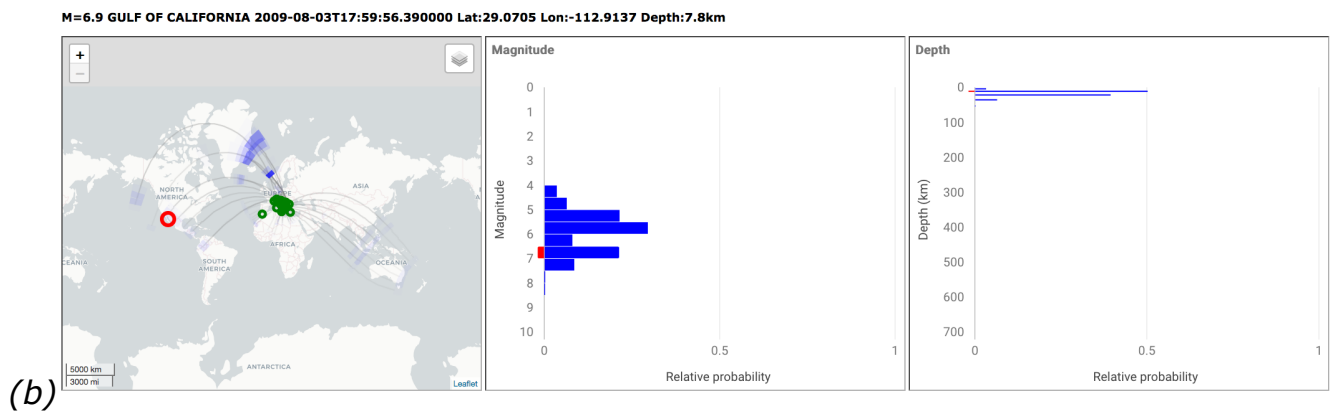
We look for evidence of generalization by examining test events (Figure 7a) for which the training event density (Figure 4a) is very low. For such test events it is unlikely that a good determination of distance, azimuth, magnitude or depth by ConvNetQuake_INGV is due to effective direct matching of the test waveforms to training event waveforms. The evidence (Figure 10) is mixed, as it shows good constraint on depth for shallow events (but most training events are shallow), moderate to poor constraint on magnitude, and mixed constraint on distance and azimuth, with clear outliers related to high training event density and CNN overfitting. There is clearly some generalization of the determination of event parameters, but overall the results for test events in regions with few training events are not as well constrained as many other test events.

[NOTE: for publication, all Figure 10 captions will be grouped together separate from the figure panels.]

Figure 10. Epicenter, magnitude and depth results for test events in areas of low training event density. Results after 500k training steps; see Figure 8a for caption details.

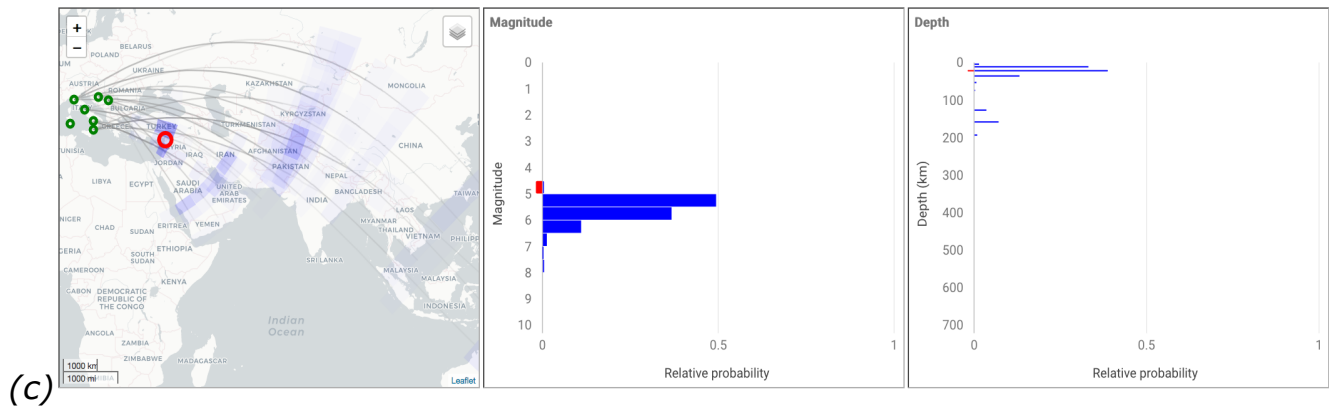


(a) Results for this event show poor distance constraint around the true value, moderate definition of the correct event azimuth, and outlier epicenter bins with high relative probability are in areas of high training event density (Figure 4a) such as Alaska, Kamchatka and SW Pacific. Magnitude and depth are well constrained around the true values.



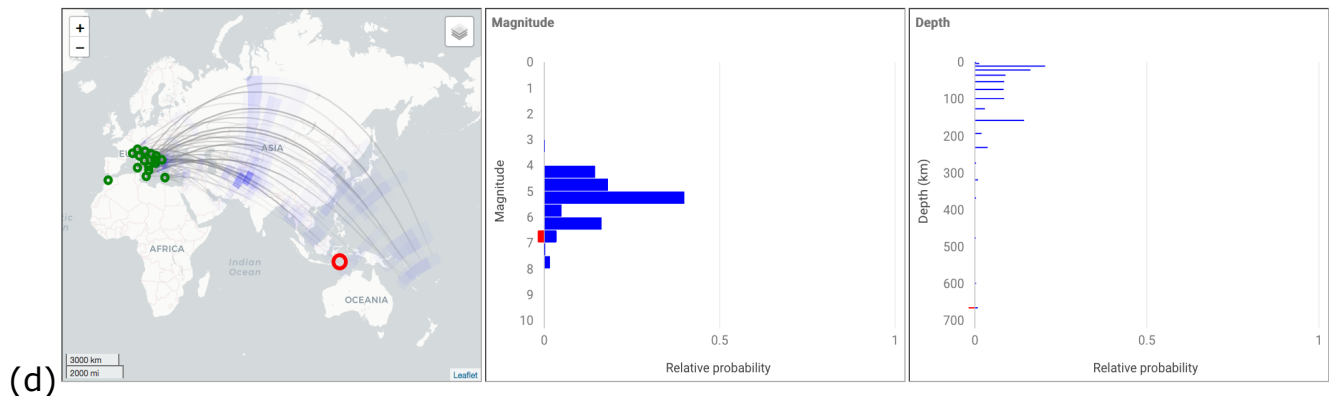
(b) Distance is poorly constrained with low probability around the true values and high-probability outliers, azimuth and magnitude are poorly constrained and depth is well constrained.

M=5.0 JORDAN - SYRIA REGION 2009-06-17T04:29:14.930000 Lat:36.1321 Lon:36.0173 Depth:16.8km



(c) Azimuth is moderately constrained around the true value, distance and magnitude are poorly constrained, though several moderate probability distance-azimuth bins group around the true epicenter.

M=6.9 BANDA SEA 2009-08-28T01:51:19.860000 Lat:-7.2021 Lon:123.4606 Depth:640.1km



(d) Very deep event. Azimuth is moderately constrained around the true values, distance, magnitude and depth are very poorly constrained.

The issues of overfitting, generalization and learning are important. The incorrect classification results and weak evidence of generalization may be due to the well known instability of CNN's with respect to small variations in input data (Szegedy et al. 2013; Rosenfeld et al. 2018). This instability may be a limitation or challenge in the application of CNN to highly variable and noisy data such as short waveforms for general earthquakes.

ConvNetQuake_INGV has a last convolutional layer with fewer nodes (64) than the final classification layer (127); this architecture produced smoother results and much less overfitting than did networks with fewer convolutional layers. This result corresponds to the concepts of "information bottleneck" (Tishby and Zaslavsky, 2015) and "minimal representation" of data (Iten et al. 2018) in neural-networks, where the optimal architecture includes layers with as simple as possible a structure to carry compactly all relevant information in the input data needed to constrain the output quantities.

Further development and study of ConvNetQuake_INGV or other related ML methods might involve specially selected training and test data sets. For example,

training data sets excluding events with specific parameters intervals, such as a distance range, would allow evaluation if the ML method can correctly recover the true parameter values for test events with parameters within these intervals. Detailed examination of local event data sets where the S arrival is always present in the waveform window (stations within about 3.5° or 400km for maximum S onset within 40 sec after the P arrival) would show if ML methods can accurately recover the relation between S-P time and epicentral distance, as is suggested by the analysis of epicentral errors in Figure 9.

5 Conclusions

Though ConvNetQuake_INGV is not yet a practical monitoring tool, it allows investigation of the performance of CNN and ML in general for rapid, automated, earthquake detection and characterization using short, single-station waveforms.

ConvNetQuake_INGV can detect and characterize earthquakes over a broad range of distances and magnitudes, but characterization errors can be large and there are indications of overfitting of the CNN training data. There is weak evidence that the CNN is performing more than high-dimensional regression and pattern recognition, and is generalizing information or “learning”, so that useful predictions are obtained for new events in regions with little or no training data.

The performance of ConvNetQuake_INGV was improved by using a last convolutional layer with fewer nodes than the number of output classifications. This architecture may efficiently compresses and transmits relevant information in the input data that best constrains the output quantities.

We expect that real-time ML procedures like ConvNetQuake_INGV, perhaps in a hybrid form incorporating empirical and physics rule-based knowledge, will ultimately prove valuable for rapid source characterization of earthquakes for earthquake response and tsunami early warning.

6 Data and resources

The ConvNetQuake_INGV software is open-source and available at https://github.com/alomax/ConvNetQuake_INGV.

Seismograms were obtained using the Federation of Digital Seismograph Networks (FDSN) Web Services of INGV at http://cnt.rm.ingv.it/webservices_and_software (last accessed August 2018)

Event meta-data used in this study were obtained using the FDSN Web Services of the IRIS Data Management Center at www.iris.edu (last accessed August 2018)

The ConvNetQuake software was obtained from <https://github.com/tperol/ConvNetQuake> (last accessed August 2018).

ConvNetQuake_INGV uses ObsPy (Beyreuther et al., 2010; Lion et al., 2015; <http://obspy.org>) for waveform retrieval and processing, and TensorFlow (Abadi et al. 2015; <https://www.tensorflow.org> last accessed September 2018) for CNN implementation, running and analysis.

7 Acknowledgments

This research was supported by the EC ARISTOTLE project ECHO/SER/2015/722144. We used Folium (<http://python-visualization.github.io/folium>), Leaflet (<https://leafletjs.com>), Google Charts (<https://developers.google.com/chart>) and TensorFlow (Abadi et al. 2015) for charts and figures, SeisGram2K (<http://www.alomax.net/software>) for seismogram analysis and figures, and OpenOffice.org for spreadsheet calculations, charts and figures.

8 References

- Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng (2015). TensorFlow: Large-scale machine learning on heterogeneous systems <https://arxiv.org/abs/1603.04467>.
- Allen, R. M., and H. Kanamori (2003). The Potential for Earthquake Early Warning in Southern California, *Science*, 300-5620, pp. 786-789, doi:10.1126/science.1080912
- Allen, R. M., P. Gasparini, O. Kamigaichi, and M. Böse (2009). The status of earthquake early warning around the world: An introductory overview. *Seismol. Res. Lett.* 80, 682–693.
- Bai, C.Y., and B. Kennett (2001). Phase identification and attribute analysis of broadband seismograms at far-regional distances, *J. Seismology*, 5, 217–231, <https://doi.org/10.1023/A:1011436421196>
- Bernardi, F., A. Lomax, A. Michelini, V. Lauciani, A. Piatanesi, and S. Lorito (2015). Appraising the Early-est earthquake monitoring system for tsunami alerting at the Italian candidate Tsunami Service Provider, *Nat. Hazards Earth Syst. Sci.*, 15, 2019-2036, doi:10.5194/nhess-15-2019-2015 (<http://www.nat-hazards-earth-syst-sci.net/15/2019/2015/>)
- Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.* 81, no. 3, 530–533.
- Gentili, S., and A. Michelini (2006). Automatic picking of P and S phases using a neural tree, *J. Seismol.* 10, 39–63.
- Gibbons, S.J., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.* (2006) 165, 149–166 doi: 10.1111/j.1365-246X.2006.02865.x
- Goodfellow, I., Y. Bengio and A. Courville (2016). *Deep Learning*, MIT Press, 775pp., <http://www.deeplearningbook.org>

- IRIS (2012). SEED Reference Manual. Standard for the Exchange of Earthquake Data. SEED Format Version 2.4, 224pp., https://www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf
- Iten, R., T. Metger, H. Wilming, L. del Rio, R. Renner (2018). Discovering physical concepts with neural networks, <https://arxiv.org/abs/1807.10300>
- Kingma, D., and J. Ba (2017). Adam: A Method for Stochastic Optimization, Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015, <https://arxiv.org/abs/1412.6980>
- LeCun, Y. , Y. Bengio, and G. Hinton (2015). Deep learning. *Nature* 521, 436–444.
- Lion, K., M. Tobias, B. Robert, B. Moritz, L. Thomas, C. Corentin, and W. Joachim (2015). ObsPy: A bridge for seismology into the scientific Python ecosystem, *Comput. Sci. Discov.* 8, no. 1, 014003.
- Lomax, A. and A. Michelini (2012). Tsunami early warning within 5 minutes, *Pure and Applied Geophysics*, 170, 1385-1395, doi: 10.1007/s00024-012-0512-6.
- Mazza, S., M. Olivieri, A. Mandiello, and P. Casale (2008). The Mediterranean Broad Band Seismographic Network Anno 2005/06, *Earthquake Monitoring and Seismic Hazard Mitigation in Balkan Countries*, 133–149.
- Mehta, P., M. Bukov, C-H. Wang, A.G.R. Day, C. Richardson, C.K. Fisher and D. J. Schwab (2018, n.d.). A high-bias, low-variance introduction to Machine Learning for physicists, <https://arxiv.org/abs/1803.08823> [physics.comp-ph]
- Melgar, D., R. M. Allen, S. Riquelme, J. Geng, F. Bravo, J. C. Baez, H. Parra, S. Barrientos, P. Fang, Y. Bock, M. Bevis, D. J. Caccamise II, C. Vigny, M. Moreno, and R. Smalley Jr. (2016). Local tsunami warnings: Perspectives from recent large events, *Geophys. Res. Lett.*, 43, doi:10.1002/2015GL067100.
- Michelini, A., and A. Lomax (2017). Rapid assessment of tsunami impact from real-time seismology and geographic, historical other datasets using machine learning, EGU, April 2017, Vienna, EGU2017-11456. (Abstract: <http://meetingorganizer.copernicus.org/EGU2017/EGU2017-11456-1.pdf>)
- Newman, A.V., Hayes, G., Wei, Y., and Convers, J. (2011). The 25 October 2010 Mentawai tsunami earthquake, from real-time discriminants, finite-fault rupture, and tsunami excitation, *Geophys. Res. Lett.*, 38, L05302, doi:10.1029/2010GL046498.
- Perol, T., Gharbi, M., and Denolle, M. (2018). Convolutional neural network for earthquake detection and location, *Science Advances*, 4:2, doi:10.1126/sciadv.1700578 (<http://advances.sciencemag.org/content/4/2/e1700578.full>)
- Rosenfeld, A., R. Zemel, J. K. Tsotsos (2018). The Elephant in the Room, <https://arxiv.org/abs/1808.03305>
- Ross, Z. E., Meier, M.-A., and Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *J. Geophys. Res.-Solid Earth*, 123, 5120–5129. <https://doi.org/10.1029/2017JB015251>

- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., and Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.*, 44, 9276–9282. <https://doi.org/10.1002/2017GL0746>
- Ruano, A.E., Madureira, G., Barros, O., Khosravani, H.R., Ruano, M.G., Ferreira, P.M. (2013). Seismic detection using support vector machines, *Neurocomputing*, 135, 273-283, <http://dx.doi.org/10.1016/j.neucom.2013.12.020>
- Russell, S., Vaidya, S., and Le Bras, R. (2010). Machine learning for Comprehensive Nuclear-Test-Ban Treaty monitoring, *CTBTO Spectrum* 14.
- Szegedy, C., W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, R. Fergus (2013). Intriguing properties of neural networks, <https://arxiv.org/abs/1312.6199>
- Tishby, N., and N. Zaslavsky. (2015). Deep learning and the information bottleneck principle. 2015 IEEE Information Theory Workshop (ITW). doi:10.1109/itw.2015.7133169
- Tsushima, H., Hirata, K., Hayashi, Y., Tanioka, Y., Kimura, K., Sakai, S., Shinohara, M., Kanazawa, T., Hino, R., and Maeda, K. (2011). Near-field tsunami forecasting using offshore tsunami data from the 2011 off the Pacific coast of Tohoku Earthquake, *Earth Planets Space*, 63, 821–826.
- Wang, J. and Teng, T.-L. (1995). Artificial neural network-based seismic detector, *Bull. Seism. Soc. Am.*, 85, 308-319.
- Weber, E., Convertito, V., Iannaccone, G., Zollo, A., Bobbio, A., Cantore, L., Corciulo, M., Di Crosta, M., Elia, L., Martino, C., Romeo, A., Satriano, C. (2007). An advanced seismic network in the southern Apennines (Italy) for seismicity investigations and experimentation with earthquake early warning, *Seismol. Res. Lett.*, 78, 622–634. doi:10.1785/gssrl.78.6.622

Anthony Lomax
ALomax Scientific
320 Chemin des Indes
06370 Mouans-Sartoux
France

Alberto Michelini and Dario Jozinović
Istituto Nazionale di Geofisica e Vulcanologia (INGV)
Via di Vigna Murata, 605
00143 Roma, Italy