

Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network

Dario Jozinović^{1,2}, Anthony Lomax,³ Ivan Štajduhar⁴ and Alberto Michelini¹

¹Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605, 00143 Rome, Italy. E-mail: djozinovi@gmail.com

²Department of Science, Università degli Studi Roma Tre, Via Ostiense, 159, 00154 Rome, Italy

³ALomax Scientific, 320 Chemin des Indes, 06370 Mouans-Sartoux, France

⁴Department of Computer Engineering, Faculty of Engineering, University of Rijeka, 51000 Rijeka, Croatia

Accepted 2020 May 8. Received 2020 April 23; in original form 2020 February 11

SUMMARY

This study describes a deep convolutional neural network (CNN) based technique to predict intensity measurements (IMs) of earthquake ground shaking. The input data to the CNN model consists of multistation, 3C acceleration waveforms recorded during the 2016 Central Italy earthquake sequence for $M \geq 3.0$ events. Using a 10 s window starting at the earthquake origin time, we find that the CNN is capable of accurately predicting IMs at stations far from the epicentre which have not yet recorded the maximum ground shaking. The CNN IM predictions do not require previous knowledge of the earthquake source (location and magnitude). Comparison between the CNN model predictions and those obtained with the Bindi *et al.* GMPE (which requires location and magnitude) shows that the CNN model features similar error variance but smaller bias. Although the technique is not strictly designed for earthquake early warning, we find that it can provide useful estimates of ground motions within 15–20 s after earthquake origin time depending on various setup elements (e.g. times for data transmission, computation, latencies). The technique has been tested on raw data without any initial data pre-selection in order to closely replicate real-time data streaming. When noise examples were included with the earthquake data the CNN was found to be stable, accurately predicting the ground shaking intensity corresponding to the noise amplitude.

Key words: Europe; Neural networks fuzzy logic; Time-series analysis; Waveform inversion; Earthquake early warning; Earthquake ground motions.

1 INTRODUCTION

Rapid assessment of earthquake generated ground motions is a fundamental task of earthquake monitoring to provide information crucial to disaster response and public information. In recent years, rapid analysis has become of primary importance since advances in communication technology and the advent of social media have enabled public dissemination of (near) real-time information on events and their associated impact. To meet this challenge and to mitigate the impact of earthquakes, the seismological community has developed a number of earthquake early warning (EEW) systems (e.g. see Allen *et al.* 2009; Satriano *et al.* 2011 for reviews; Minson *et al.* 2018). These EEW systems are designed to detect earthquakes very rapidly (in a few seconds) and to provide early warning on the impending ground motion at selected target points. EEW systems are developed as regional EEW systems (e.g. Kohler *et al.* 2017) or on-site EEW systems (e.g. Spallarossa *et al.* 2019).

On a longer timeline than conventional EEW, the ShakeMap software (Wald *et al.* 1999) was developed with the primary purpose

of providing accurate maps of strong ground motion. These maps, typically available within 5–10 min after an earthquake, allow disaster risk managers (DRMs) to make preliminary assessments of the shaking impact. ShakeMap generates maps of ground shaking using earthquake source parameters (location, magnitude and the finite fault if available), intensity measurements (PGA, PGV, etc.), ground motion models (GMMs) and Vs30 maps as a proxy to account for site amplifications. Ground motion interpolation is performed at points of the map which do not have ground motions recorded (e.g. Worden *et al.* 2018). The ShakeMap software has not been designed for EEW—typically shakemaps become available only when the first location and magnitude estimation are available. However, the ShakeAlert system for EEW in the western U.S. has been recently upgraded to include the determination of ground motion using the same region-specific ground-motion prediction equations that are used by ShakeMap implementations in California, Oregon, and Washington (Given *et al.* 2018). In general, the time required to produce the first shakemaps depends on several factors such as data availability, and transmission and processing

latencies, and the map accuracy depends on the density of the stations and on the quality of the data available (e.g. for the 2016 Central Italy $M = 6$ August 24 main shock, the first ShakeMaps were provided 6 min after origin time; Faenza *et al.* 2016).

In this study, in an effort to quantify as quickly as possible the level of ground shaking in a given area, we aim at predicting very rapidly the ground shaking intensity at a predefined set of seismic stations within a given seismic area by using a machine learning (ML) approach. The idea is to use waveforms up until only a few seconds after origin time when only a few nearby stations have recorded the first P waves to predict the ground shaking intensity at more distant stations. The proposed technique does not require knowledge of the source parameters and it utilizes only a training set of earthquake waveforms recorded at a pre-configured network of recording stations. Implementation of the technique could be valuable when seeking a few seconds of warning to issue alerts for critical infrastructure points for potential failure such as highway or high velocity railways bridges, gas-pipelines, industrial plants handling high risk and polluting chemicals, hospitals or schools.

We use the information contained in the recorded waveforms by adopting a convolutional neural network (CNN) model to predict intensity measurements (IMs) at distant stations using only the initial N seconds after the origin time of the recordings at nearby stations. This implies that, depending on the value of N , the source receiver-geometry, and the duration of strong motion, waveforms for stations near the epicenter may contain the signal corresponding to their maximum (labelled) IM, whereas the waveforms for more distant stations may end before the seismogram peak value is reached. The assumption is that the CNN model will be able to learn from the patterns of signal and noise across the input vector (i.e. relative amplitudes of the waveform signal in the window), and be able to make predictions on the maximum level of ground motion at the farther stations that have not yet recorded the peak amplitude values. The characteristics of the seismic wave arrivals across the input vector (i.e. the network waveform pattern) are informing the CNN model about the earthquake characteristics (e.g. location, magnitude, etc.), which the model then can use to learn a kind of locally calibrated GMPE, directly from the observed data. Though this approach may not compete with the rapidity of an EEW system, it is expected to provide predictions of the ground motions similar or better to those obtained using EEW techniques and possibly the very first shakemaps, but at an earlier time.

While ML techniques have been used in seismology for almost three decades (e.g. Chiaruttini *et al.* 1989; Dysart & Pulli 1990), they have become more intensively used only in recent years (for more details see the reviews by Kong *et al.* 2018 and Bergen *et al.* 2019). ML has been applied to numerous problems: fast magnitude determination (e.g. Ochoa *et al.* 2018), ground motion prediction from source parameters (e.g. Alavi & Gandomi 2011; Derras *et al.* 2014), earthquake detection (e.g. Reynen & Audet 2017; Mousavi *et al.* 2019, Rojas *et al.* 2019), insight into physics of labquakes (Hulbert *et al.* 2019) and many other seismological applications. A specific type of ML modelling technique, the CNN, has been used widely for waveform analysis. A CNN is a type of artificial neural network which uses convolutions as the fundamental building block for learning proper feature extraction, coupled with the potential for modelling highly non-linear classification/regression interactions between the variables. Perol *et al.* (2018) used a CNN to detect and locate earthquakes, by working directly on the seismic waveforms, without feature extraction. Similarly, Lomax *et al.* (2019) used a CNN for single-station, earthquake location, magnitude and depth determination from local to teleseismic scale lengths. Zhu & Beroza

(2018) used a CNN to determine P - and S -wave arrival picks, also by analysing single-station seismic waveforms. Mousavi & Beroza (2020) designed a network that consists of convolutional and recurrent layers for magnitude estimation. Kriegerowski *et al.* (2018) created an earthquake location algorithm, by applying a CNN for analysing multistation (10 stations) waveforms of clustered earthquakes. The successful applications of CNNs cited here, and the problem analysed in Kriegerowski *et al.* (2018), which is similar to the problem we address (analysis of multistation waveforms), shows that CNN modelling is a viable approach for analysing multistation waveforms towards rapid and accurate IM predictions.

2 DATA

The *input* data used in this study consist of three-component earthquake waveforms data from the 2016 Central Italy sequence (Chiaraluce *et al.* 2017) recorded by 39 stations in the epicentral area and its surroundings. The data set has been selected because of the dense network of stations in the area, the large number of earthquakes and the importance of the sequence from the seismological perspective. We use earthquakes in the study area bounded by latitude [42°, 43.75°] and longitude [12.3°, 14°] which occurred from 1 January 2016 to 29 November 2016. All the events occur within crustal depths in the range $1.6 \text{ km} < D < 28.9 \text{ km}$. Using these criteria, 915 earthquakes with magnitude $M \geq 3$ have been used (Figs 1 and 2). In the same study area there are 86 recording stations from the networks IV and XO available. In this study, we have selected a subset of 39 stations, which had at least 700 earthquakes recorded on the station (Fig. 3). The selected stations were all belonging to the IV network. Together with the event data, we have also selected 1037 examples of noise data, recorded from 30th August to 30th September 2016 recorded at all the selected stations. The criterion adopted for selecting the noise data consisted of windowing the continuous waveforms with a start time at least 180 s after and 20 s before the origin time of any earthquake in the area, available in INGV earthquake catalogue. The waveform data were downloaded using the INGV FDSN web services for HN* (acceleration) and HH* and EH* (velocity) channels, where * ∈ [E,N,Z]. The data were processed to remove the instrument response. Velocity data were differentiated to acceleration. When necessary, the data have been resampled to 100 Hz. For $M < 4$ earthquakes, the HH and EH channels were used after differentiation and for earthquakes with $M \geq 4.0$ the HN channels were used. This criterion avoided possible signal saturation and the use of true acceleration recordings for the larger magnitude earthquakes. If the velocimeter data were not available at one station with a colocated accelerometer, the latter recording was used. For certain stations and for some earthquakes, the waveform data were completely missing and it was chosen to fill them with zeros. To this regard, there are different ways of filling missing data in ML (Garca-Laencina *et al.* 2010) and we chose to adopt zeroes as a natural way of applying the station dropout technique with the expectation of improving the generalization properties of the model (similar to Kriegerowski *et al.* 2018). The earthquake recordings are missing randomly only for the earthquakes which occurred after the Amatrice August 24th $M 6.0$ main shock. The 125 earthquakes that occurred before the Amatrice earthquake were not recorded by the temporary stations which were installed immediately after this event, so for those stations the data are not missing randomly.

All data have been extracted using a 50 s window starting at origin time. We use origin time as a convenience for aligning the waveforms. In practice, however, any reference time before the

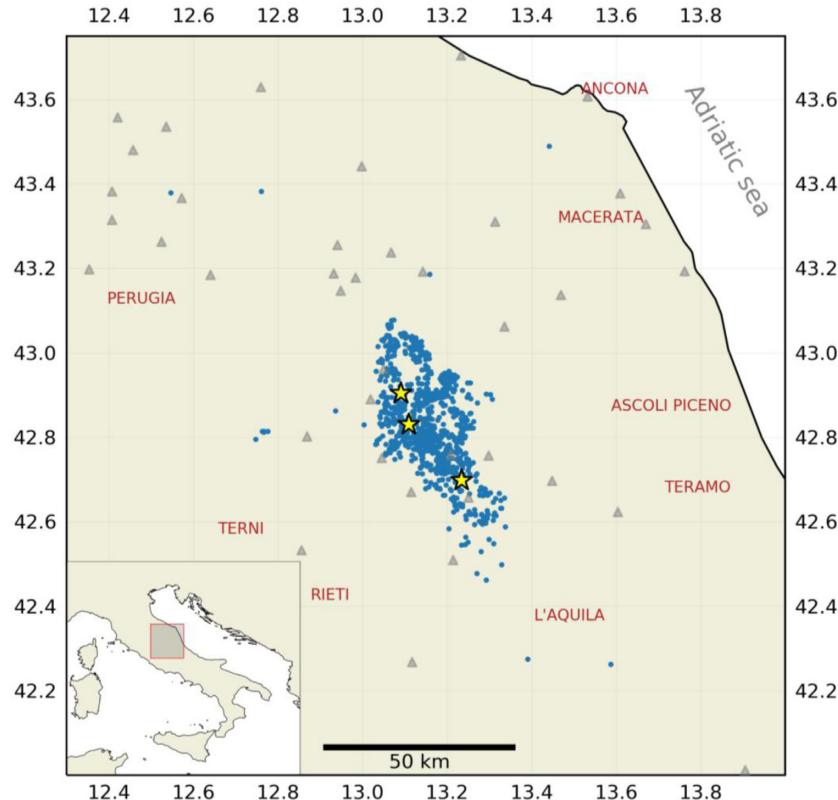


Figure 1. Blue points represent earthquake epicenter locations, stars represent the three largest events of the 2016 Central Italy sequence and the light grey triangles represent stations.

first recorded P wave at the first recording station could be used instead. Data that started more than 0.1 s (178 cases) after origin time have been manually inspected, and the faulty ones (instrument malfunctioning; 24 cases) were removed. Those that were retained (154 cases) were prepended with zeros between origin time and the start of data.

The target data consisted of the IMs associated with each recording: peak ground acceleration (PGA), peak ground velocity (PGV), spectral acceleration (SA) at 0.3, 1 and 3 s periods. The outputs of our model always need to have the dimensions (39, 5), regardless of the number of stations for which we have the data. To fill the missing target data (i.e. missing true IMs values) in the training process we use predictions calculated using USGS ShakeMap software, with the latest configuration for Italy (Michelini *et al.* 2019), since having an estimate of ground motions at the stations with missing data is more representative of real rapid response or EEW application, as opposed to predicting zeros or some randomly sampled value. ShakeMap predictions are also used in the test data set to assess model performance for the stations which had no true IM values recorded. ShakeMaps have been calculated using the IMs from all available stations. This approach, however, can introduce some error since the ShakeMap prediction may be incorrect given the uncertainties in the estimation of the ground motion. We note, however, that Michelini *et al.* (2019) have shown that nearly no systematic bias in the prediction of the maximum IMs was observed with the new ShakeMap configuration giving us good confidence that the values inserted as target values are statistically relevant. The target data for the noise waveforms were the maximum amplitudes recorded at the station inside the window used as input. In the case of missing inputs in the noise waveforms, the target has been set to zero.

3 METHOD AND TRAINING

The ML model adopts a CNN developed using Keras (see Data and Resources). Input to the model is a combination of all the waveform data (all 39 stations) for a given earthquake (Fig. 4), for an input array size (39, N , 3), where 39 is the number of stations, N is the number of samples (with a sampling rate of 100 samples per second) and 3 is the number of components. The ordering of the stations is always preserved. All the waveform data for each earthquake start at the event origin time. The data are normalized by the input maximum (i.e. the largest amplitude observed across all stations within the time window), and this maximum is saved as the normalization value which is later inserted into the network. It is recommended to normalize the input data for improved CNN performance (LeCun *et al.* 2012). However, we note that it is important to retain the maximum amplitude of the input waveforms since they are crucial to accurately predict the target peak IMs. To overcome these limitations we normalize the input waveform by its maximum value, and insert the normalization value into the last fully connected layer (following Lomax *et al.* 2019). In order to verify the effectiveness of our normalization schema, we compared the training performed using the (i) non-normalized inputs, (ii) normalized inputs without retaining maximum amplitude information and (iii) normalized inputs with the insertion of the maximum amplitude value into the last fully connected layer. The latter approach was found to produce the best CNN model results in terms of the averaged fivefold mean squared error of the CNN model. The outputs are arrays of size (39, 5), where 39 is the number of stations, and 5 is the number of predicted IMs per station. The base-10 logarithm has been applied to all the IMs (i.e. $\log_{10} \text{IM}$).

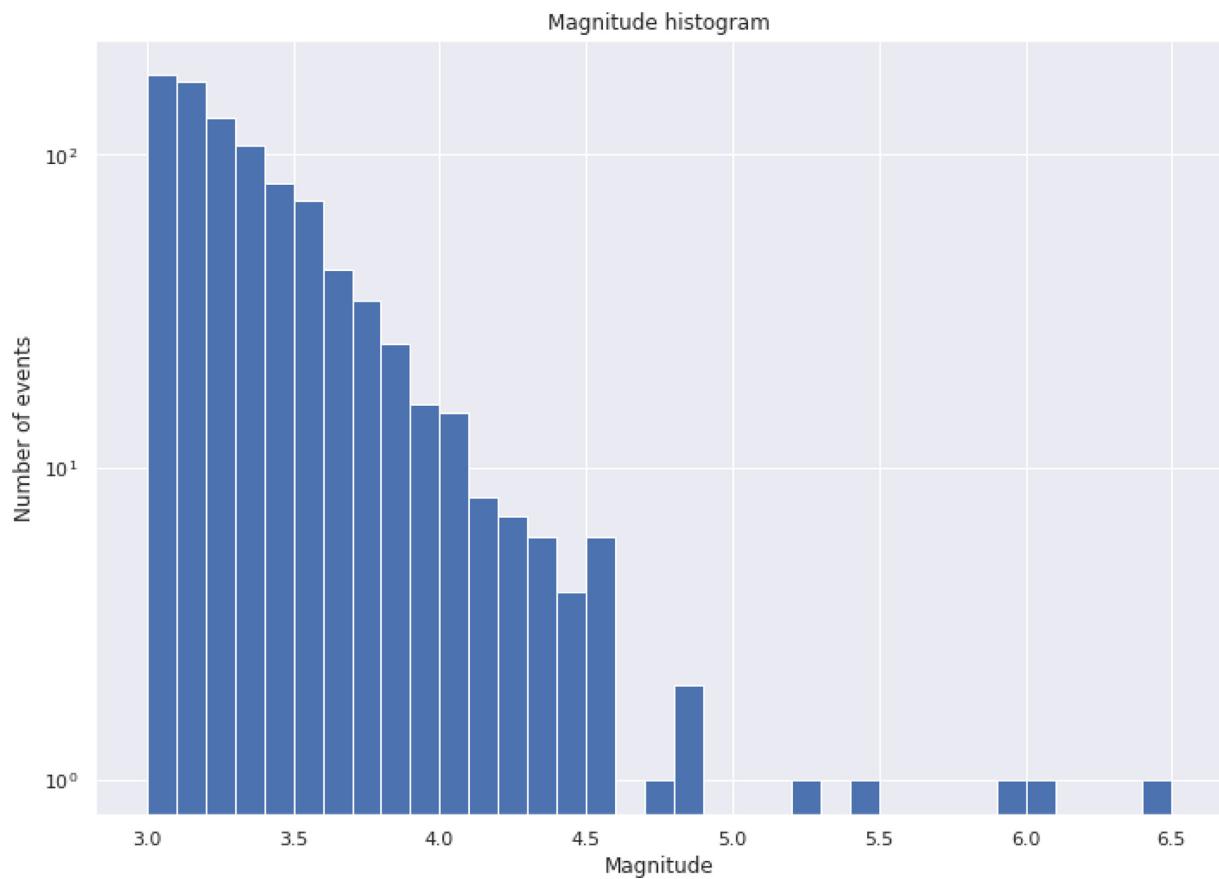


Figure 2. Earthquake-magnitude distribution of the selected earthquakes.

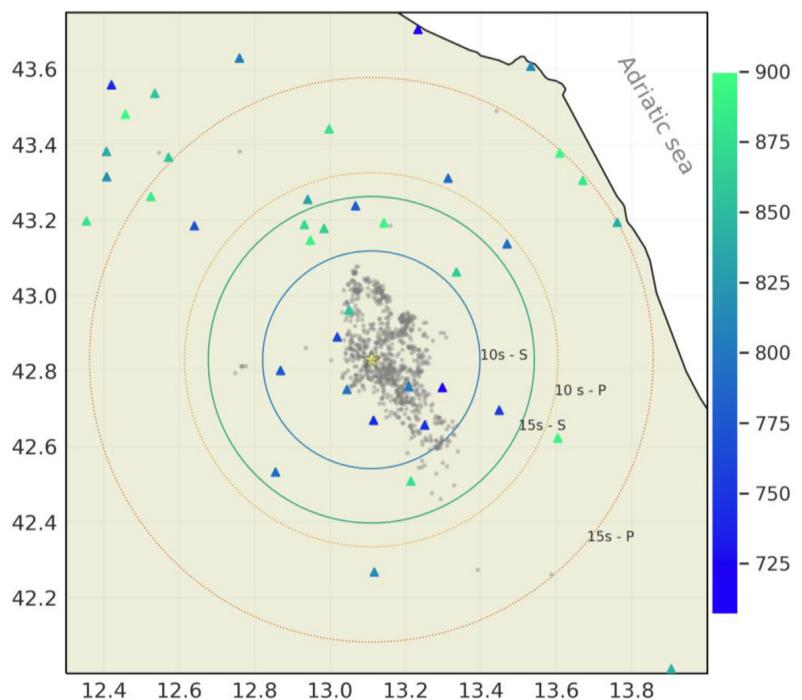


Figure 3. Stations used, with colours showing the number of earthquakes recorded per station. The circles represent the approximate travel times of direct P arrivals using $V_P = 5.5$ and $V_S = 3.2$ of the biggest event in sequence ($M 6.5$ 2016/10/30; shown with a star). Earthquakes as light grey points.

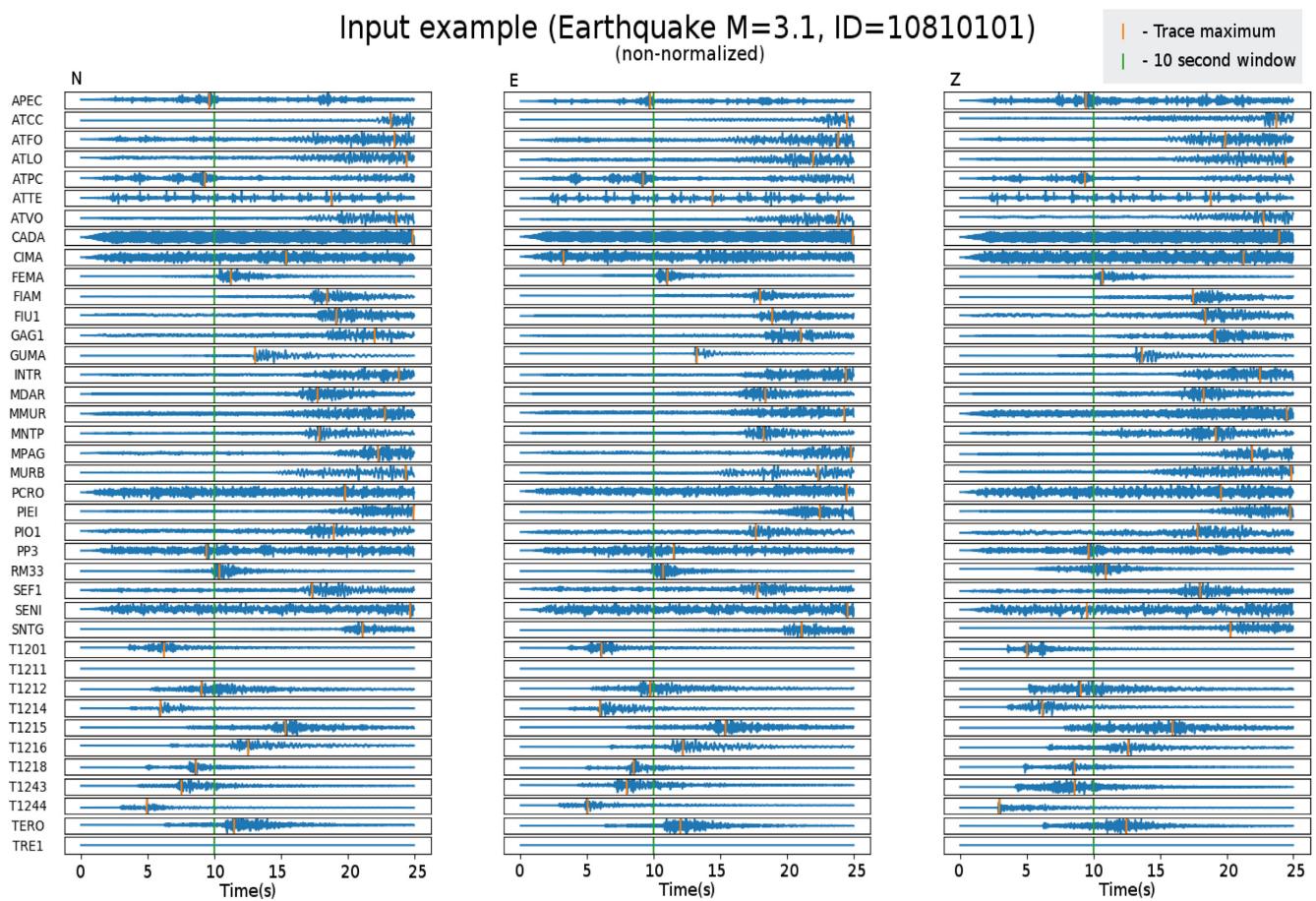


Figure 4. Non-normalized input example. Orange vertical lines show the trace peak value. Green vertical lines show the end of the 10-s window.

The CNN model architecture is based on Kriegerowski *et al.* (2018) with minor modifications (Fig. 5). The first two convolutional layers, having 32 and 64 filters, respectively, have filter kernels with height of 1 and width of 125 and 250, respectively, whose purpose is to learn the temporal patterns station by station. The third convolutional layer, having 64 channels and kernels with height 39 and width 5, gathers the cross station information. The first two convolutional layers have the stride (1,2), and the third layer has the stride (39,5). The last convolutional layer is then flattened and concatenated with the normalization value of the input, and then fed to a fully connected layer. Finally, the last fully connected layer produces an array of size (39,5) with continuous values inside. The ReLU activation function, $a = \max(0, x)$, has been used in all layers, except the last layer in which linear activation has been used. L2 regularization has been applied to the convolutional layers with a regularization rate of $\lambda = 10^{-4}$. Furthermore, a 40 per cent dropout rate has been applied before the last fully connected layer (Goodfellow *et al.* 2016). The data had been split randomly into training (80 per cent) and test (20 per cent) subsets. For evaluation of the performance of the CNN model fivefold cross-validation has been used, which splits the randomly permuted data set into 5 equally sized disjunct subsets, and uses each of them as the test set to the model trained on the remaining 4 subsets. The batch size used for optimization was 5, and the mean squared error (MSE) function was used as the loss function in the model. Hyperparameter values for model optimization were based on Kriegerowski *et al.* (2018) with minor modifications. The model was trained for 12 epochs (the training history plot is shown in Fig. S5 of the electronic

supplement). The training took approximately 3 min on an Nvidia GTX 1060 6GB for a 10 s input window.

4 RESULTS

4.1 Window length

In the first part of this study, we explored the window length after the origin time needed to make reliable predictions of the maximum values of the IMs. We used window lengths of 7, 10 and 15 s. The results are shown only for PGA (Fig. 6), as other IMs follow similar trends. For the 7 s window, the MSE of the residuals between the base-10 logarithms of observed and predicted values [i.e. $\log_{10}(\text{IM}_{\text{true}}/\text{IM}_{\text{predicted}})$] was 0.228. For the 10 s window, the MSE is reduced to 0.176, and for the 15 s window, the misfit was further reduced to 0.165. A significant drop in performance is thus occurring for shorter windows.

The vertical stride of the observed values at approximately -0.9 derived from the occasional bad input waveforms (examples in the Figs S1 and S2 in the electronic supplement). In what follows, we adopt a 10 s window as a good compromise between accuracy of the predictions and the timeliness.

4.2 Model performance

The performance of the CNN model has been evaluated using five-fold cross-validation. The results from all five test sets were then

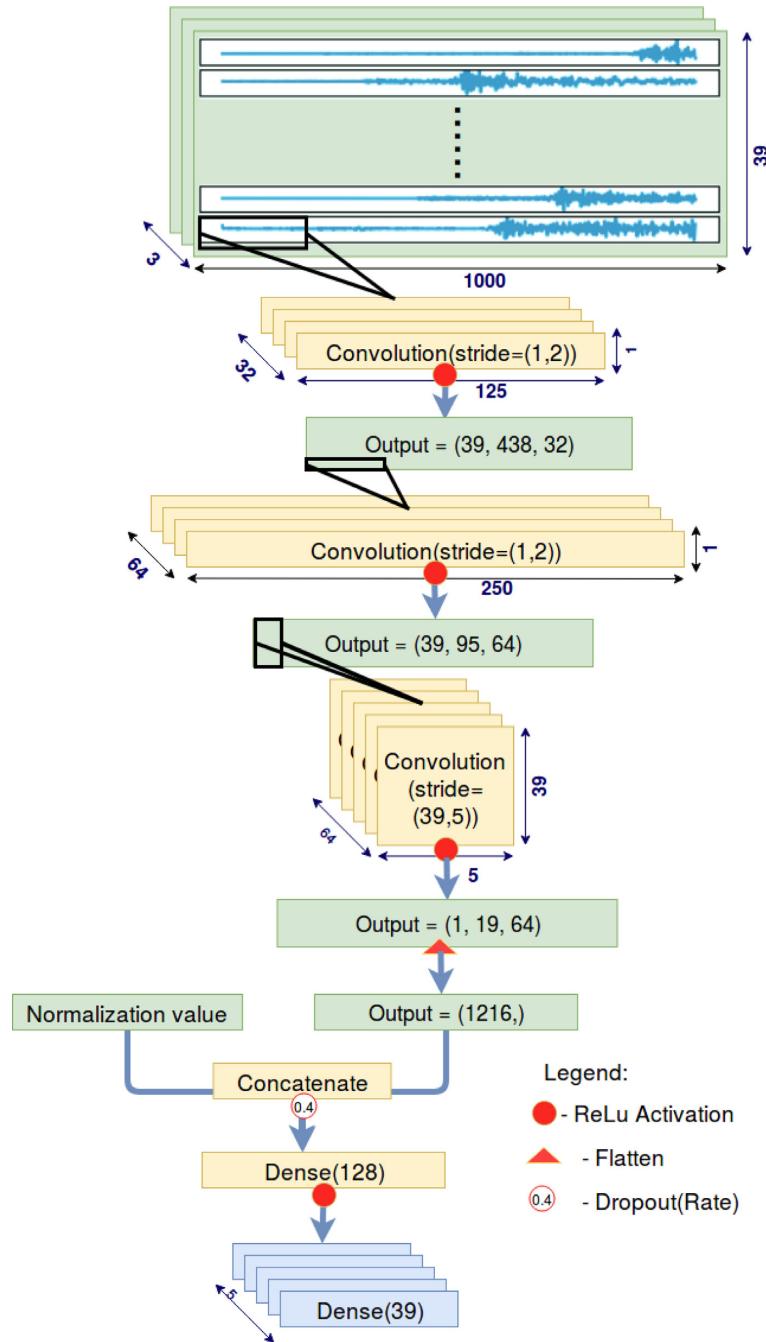


Figure 5. The architecture of the CNN model used. Boxes shaded yellow represent filter banks and operators, whereas boxes shaded green represent activations. Parenthesized vectors denote the dimensions of the outputs in question (height, width and depth).

averaged, and are presented next. The target values are the recorded IMs (93 per cent), and the ShakeMap predictions of IMs that were used when no recordings were available (7 per cent). The residuals [$\log_{10}(\text{IM}_{\text{obs}}/\text{IM}_{\text{predicted}})$] have been calculated together with their mean, median and standard deviation. Large residual values [$\log_{10}(\text{IM}_{\text{obs}}/\text{IM}_{\text{predicted}} > |1|)$] were removed resulting in 87.49 per cent of the data kept for the ShakeMap predictions and 92.00 per cent for the observed IMs Table 1 and Fig. 7 show the performance of the CNN model on the observed values, along with the ShakeMap predictions.

As noted above, the ShakeMap predictions are used for the training set when no observed data are available, and the input is set to a

zero value time series. In this regard, we find it notable the capacity of the adopted CNN algorithm to learn how to treat those stations that occasionally may be missing data and the target value consists of the ShakeMap predictions. In general, a slight positive bias in the CNN predictions of IM values is observed, indicating that the CNN model is slightly underpredicting the observed values with the exception of the ShakeMap predicted SA(3.0).

To test the CNN model performance against a baseline case, we compare our results to the predictions obtained using the GMPE by Bindi *et al.* (2011) calibrated for Italy. For the IM predictions using Bindi *et al.* (2011), we used the INGV catalog magnitude and location (see data and resources) and the IM predictions are corrected

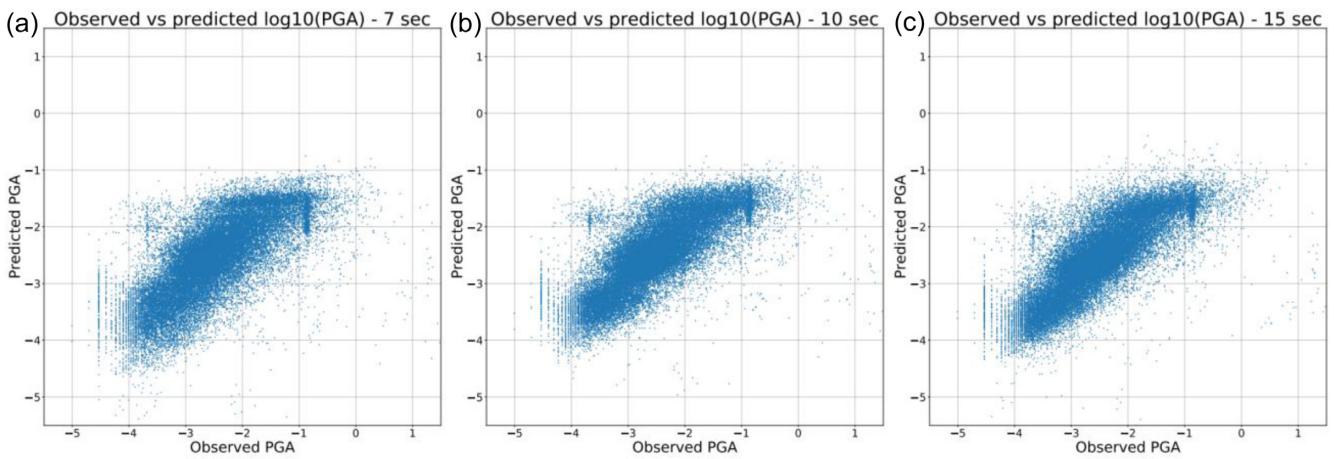


Figure 6. $\log_{10}(\text{PGA}_{\text{observed}})$ versus $\log_{10}(\text{PGA}_{\text{predicted}})$ for different time window lengths: (a) 7 s length, (b) 10 s length and (c) 15 s length. The units are $\log(\text{m s}^{-2})$.

Table 1. IM residual statistics $\log_{10}(\text{IM}_{\text{observed}}/\text{IM}_{\text{predicted}})$ for the CNN predictions for the observed IMs (for the stations having recorded data) and ShakeMap predictions (for the stations that had no recorded data). The last three columns show the values obtained using the Bindi *et al.* (2011) GMPE to estimate the IMs

IM	Observed median	Observed mean	Observed STD	ShakeMap median	ShakeMap mean	ShakeMap STD	GMPE median	GMPE mean	GMPE STD
PGA	0.038	0.035	0.346	0.059	0.038	0.372	0.013	0.017	0.352
PGV	0.036	0.034	0.338	0.043	0.041	0.380	-0.174	-0.151	0.33
SA(0.3)	0.031	0.031	0.34	0.056	0.046	0.37	-0.284	-0.252	0.359
SA(1.0)	0.029	0.034	0.338	0.001	0.017	0.374	-0.207	-0.198	0.303
SA(3.0)	0.019	0.027	0.374	-0.037	-0.012	0.404	0.026	0.083	0.368

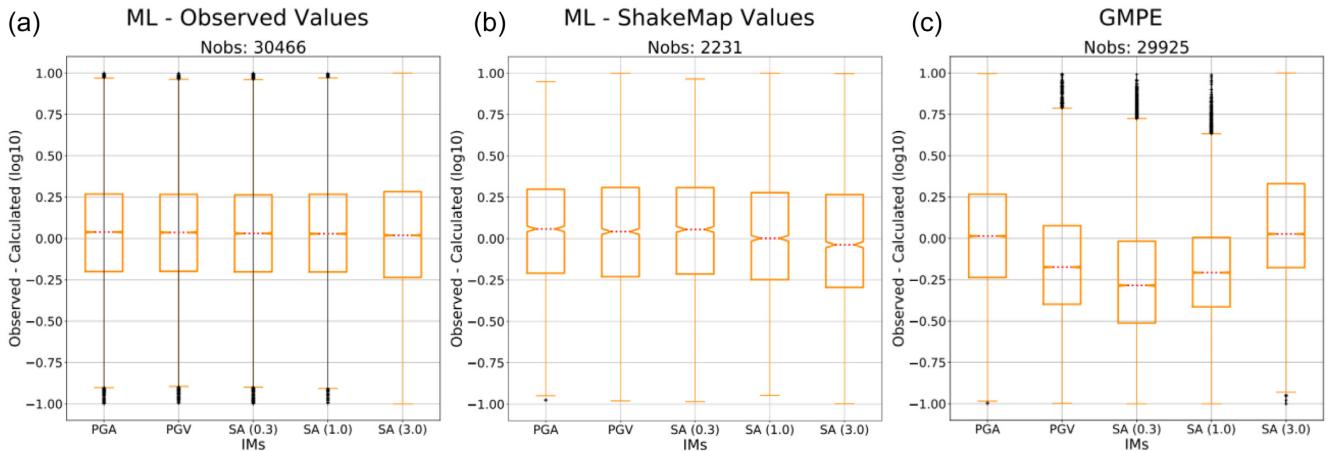


Figure 7. Boxplots for the residuals $\log_{10}(\text{IM}_{\text{observed}}/\text{IM}_{\text{predicted}})$: (a) CNN model results for the observed IMs, (b) CNN model results for the ShakeMap predicted IMs and (c) GMPE results.

appropriately using the station EC8 site classes. No between-event correction (Al Atik *et al.* 2010), however, has been applied when predicting the ground motion using the GMPE by Bindi *et al.* (2011). The outliers have been discarded using the same criteria used earlier which left us with 90.3 per cent of the data. For the CNN model performance, only the results obtained for the observed IM values are significant. The residuals $\log_{10}(\text{IM}_{\text{observed}}/\text{IM}_{\text{predicted}})$ have been calculated for the results obtained with the GMPE model, in the same way as for the CNN model.

The mean, median and standard deviation of the differences between the logarithms of observed and predicted values,

$\log_{10}(\text{IM}_{\text{observed}}/\text{IM}_{\text{predicted}})$, for the CNN and the GMPE models are reported in Table 1 and they are shown in Fig. 7. We find that the median values of $\log_{10}(\text{IM}_{\text{observed}}/\text{IM}_{\text{predicted}})$, which are an expression of the model bias, are significantly reduced in the CNN model, especially for PGV, PSA(0.3) and PSA(1.0), yet they are slightly higher for PGA. Concerning the standard deviation, we find that the values are comparable for the two models, and are in agreement with those reported by Bindi *et al.* (2011) who reported a standard deviation between 0.34 and 0.38 $\log_{10}(\text{IM}_{\text{observed}}/\text{IM}_{\text{predicted}})$.

The variation of the residuals with magnitudes (shown in Fig. S6 of the electronic supplement) shows that the bias of the residuals

$[\log_{10}(\text{IM}_{\text{obs}}/\text{IM}_{\text{predicted}})]$ increases with magnitude, that is that the CNN model is underpredicting the values for the larger magnitudes. This is to be expected, given the large imbalance in the magnitudes of the training examples (Fig. 2). In contrast, the variations of the residuals with epicentral distance (shown in Fig. S7 of the electronic supplement) show no clear trend.

When applied to the three largest events in the sequence (using fivefold cross-validation), with magnitudes of $M = [5.9, 6.0, 6.5]$, the CNN model performance deteriorates (Fig. 8), compared to the model performance on events with smaller magnitude. These results are due to the unbalanced dataset with few training examples at large magnitude (Fig. 2), which makes learning at larger magnitudes difficult. More specifically, the poor performance for the $M 6.0$ event can be also attributed to the use of only 24 station waveforms in the input data, as no temporary stations had yet been installed in the area before the 24 August 2016, $M 6.0$ main event. This same reasoning, however, does not hold for the performance differences between the $M 6.5$ and $M 5.9$ events (33 and 35 stations, respectively) with the former one performing much more poorly than the latter. One possible and partial explanation for this performance disparity could be ascribed to the difference in the pattern of the input waveforms (Figs S3 and S4). Since the $M 5.9$ event is located to the north where there is a larger number of stations, this event features more stations with ‘significant’ earthquake P wave (or even S wave for the nearest ones) signal information within the 10 s window when compared to the $M 6.5$ earthquake. However, given that these are only two examples, it’s hard to make a quantitative analysis that would confirm all the points discussed here. More in general terms, the rather poor predictive performance observed for the $M 6.0$ and $M 6.5$ earthquakes derives primarily from the lack of earthquake waveform data at the larger magnitudes. We have also trained the CNN model only on the events before the Amatrice $M 6.0$ earthquake (388 events), and then tested it on the Amatrice $M 6.0$ earthquake and the events that happened after (528 events). We found that the CNN model performs worse than the model trained on the complete dataset, especially for larger values of IMs. This mainly can be explained by the low number of events available for training, and partly also by the 125 events for which the 9 temporary stations were missing. In Figs S8(a) and (b) of the electronic supplement a comparison between the number of station waveforms in the inputs and the performance of the CNN model predictions has been shown. When the comparison is done on the residuals of the CNN model predictions and observed PGA targets, increasing the number of station waveforms in the inputs reduces bias and has no clear effect on the standard deviation. When the same comparison is done on the residuals calculated with ShakeMap predicted targets, there is also a bias reduction, and no clear trend in standard deviation.

4.3. Including noise examples in the data

It is important to evaluate the CNN model performance in the presence of noise-only data, to understand the model performance if used on real-time streamed data, without an earthquake detector paired to the CNN model. For noise-only data the predicted IM’s should reflect the maximum noise level of the input waveforms at each station. We added 1037 examples of the noise data to both the training and the test subset, and again we used the maximum PGA of the input waveforms as the training and test target variables. The results (Fig. 9) show that the CNN model is also able to predict the target IMs of the noise data reasonably well. The target data for

the non-existing noise waveforms (for which inputs are filled with zeros) are zero, and they are not shown in the plot.

5 DISCUSSION

In this study, we have shown that a CNN model can be used to accurately predict earthquake IMs at recording stations using only raw, multistation waveforms with a 10 s time window starting at the earthquake origin time. That is, from the very first recordings at stations near the epicenter, it is possible to predict accurately the IMs at farther stations that have not yet recorded the earthquake signal or its maximum values. To this end, we exploit the 3C station data (i.e. waveforms) and the pattern of the waveforms of the recording network without any knowledge concerning the earthquake location or magnitude. The output targets, the 5 different IMs, are estimated by the CNN model for all the 39 stations as a regression problem. This is similar to a GMPE model where the input parameters are the direct ground motion observations from the network of stations (i.e. the pattern of waveform ground motion), instead of earthquake source and station site parameters.

Input data consist of the recorded acceleration waveforms, while the target variables are PGA, PGV and SA at 0.3, 1 and 3 s periods. The CNN model generally performs equally well regardless of the IM type. As expected, model performance improves when longer time windows (e.g. 15 s time windows) are used as input.

Because of the Gutenberg-Richter distribution of earthquake magnitudes, our data set is severely unbalanced with a much larger frequency of smaller events than larger ones. As shown in Fig. 8, this makes it difficult to predict the IMs of larger events. A possible solution to this problem could follow from the use of data augmentation (e.g. Chollet 2018) by introducing more training data for the larger magnitudes after applying random transformations to the existing larger events. Another alternative can consist of calculating synthetic seismograms at the same recording stations for additional (hypothetical) larger events in the same area. Another approach towards mitigating problems due to small data sets is the use of transfer learning methods (Pan and Yang 2010.). These methods allow the CNN model to subspecialize by training on a smaller data set, using a pre-trained model which was previously trained on a larger data set. Following Chollet (2018), we could even use a training network developed elsewhere and for different purposes, but with a very large data set, and use it to our goal of IM prediction. Similarly, the network developed in this study could be adopted as a pre-trained model for other areas where IM predictions are required.

To test the methodology, we have used raw waveforms without any preliminary data cleaning, in order to simulate real-time analysis where missing or erroneous data due to equipment malfunctioning or data transmission problems is common. To compensate for these deficiencies, we found that replacing missing data with zero values and the adoption of ShakeMap predicted target data values is suitable to ensure that the target data exist for all cases during learning. We have shown that by adopting this strategy, the CNN model is still capable of predicting the IMs fairly accurately. In practice, this is somewhat similar to applying the station dropout technique (Kriegerowski *et al.* 2018) to the training procedure. We have found that the CNN model was able to accurately predict the ground motions at the stations with missing data, albeit with a slightly lower accuracy when compared to the IMs of the stations which have no missing data. This all suggests that the CNN model is able to learn the seismic wave propagation characteristics and can compensate

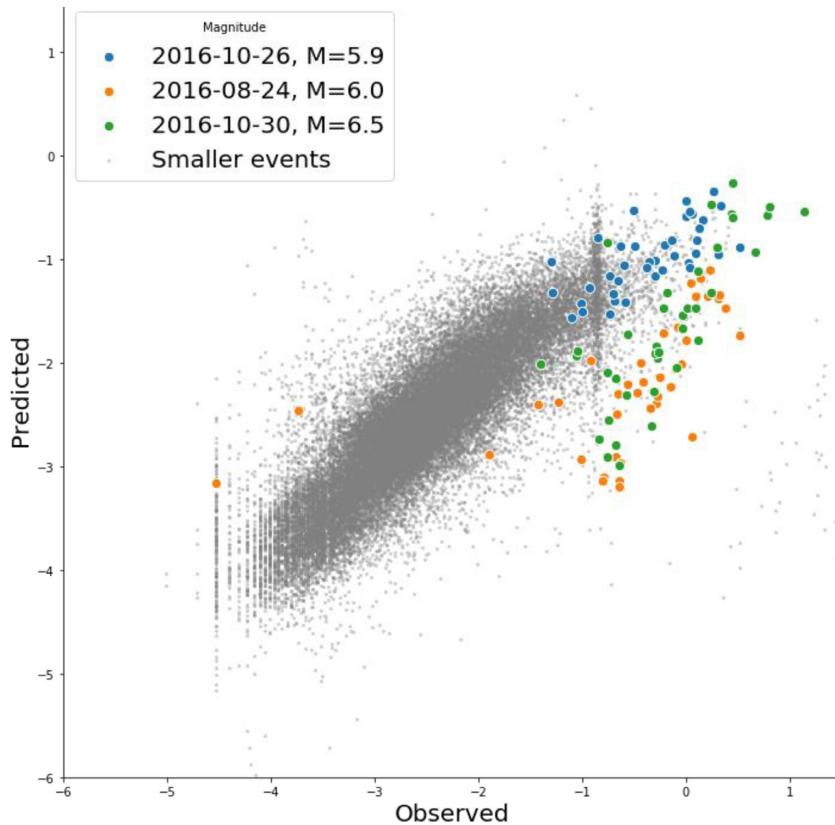


Figure 8. $\log_{10}(\text{PGA}_{\text{observed}})$ versus $\log_{10}(\text{PGA}_{\text{predicted}})$ for the three largest events in the sequence (using 10 s waveforms). The vertical stripe visible at approximately $-0.9 \log_{10}(\text{PGA}_{\text{observed}})$ results from data logger equipment malfunctioning

for the presence of missing data when using imperfect target data, like the ShakeMap predictions, for learning.

To verify the predictive performance of our CNN model, and compare it against the results of a more conventional way of predicting the IMs, we have used the GMPEs proposed by Bindi *et al.* (2011) which are adopted in the configuration of the ShakeMap for Italy (Michelini *et al.* 2019) to predict the IMs at our selected stations. The results shown in Fig. 7 indicate the CNN model to be superior in terms of IM residuals. This follows from (i) the CNN model being independent of earthquake magnitude uncertainties (i.e. no need for between-event correction terms), (ii) our source-receiver geometry featuring similar wave paths within a rather small area, and all earthquakes occurring approximately in the same source region and (iii) the ability of the CNN model to capture local site anomalies and compensate for them. We note that these points are all taken into account by CNN modelling procedure automatically by learning directly from the 3C waveforms.

As noted above, the earthquakes used in these experiments are concentrated spatially which significantly reduces the variability at a given station resulting from different paths, and this may be one of the main reasons for the good quality of the results obtained. We are confident, however, that by using a bigger data set, with the different spatial distribution of epicenters, the CNN model will likely be able to learn the path and site characteristics of each station and for each zone where the earthquakes occur. This aspect will be addressed in future studies encompassing larger areas and data sets.

Our results with earthquake and noise only data (Fig. 9), have shown that the methodology is capable to predict the ground motion adequately. This fact combined with the translational invariance characteristics of convolutional neural networks suggest that the

technique could be used in a real-time setup with data streamed continuously to predict the IMs seamlessly at farther stations as the earthquake nearest stations start recording the first earthquake generated signals.

6 CONCLUSIONS

A CNN model has been used to predict IMs (PGA; PGV; PSA 0.3 s, 1 s, 3 s) with satisfying accuracy adopting multistation, 3C 10 s window waveforms starting at origin time. The waveforms come from a data set recorded by 39 stations from a set of 915 earthquakes with $M \geq 3.0$ and 1037 only noise examples, for a total of 33 855 earthquake waveforms, and 40 443 noise waveforms.

The IM predictions do not require previous knowledge of earthquake location, distance and magnitude. When using longer time windows the performance of the CNN improved but the selected 10 s window of our setup appears to be a good compromise between accuracy and timeliness. No feature extraction or pre-processing was to be applied on the data because CNN models do it automatically. The performance of the CNN model does not depend on the IM type.

We have found that the CNN model was able to predict with satisfying accuracy the IMs for stations which had no input data when training. Comparison between the CNN results and those obtained with the Bindi *et al.* (2011) ground motion model, calibrated with earthquakes in Italy, has shown that the CNN model does not suffer from prediction bias while exhibiting similar variance of the residuals.

The performance of the CNN model degrades for large magnitudes ($M > 5$) because of the small number of earthquakes (5)

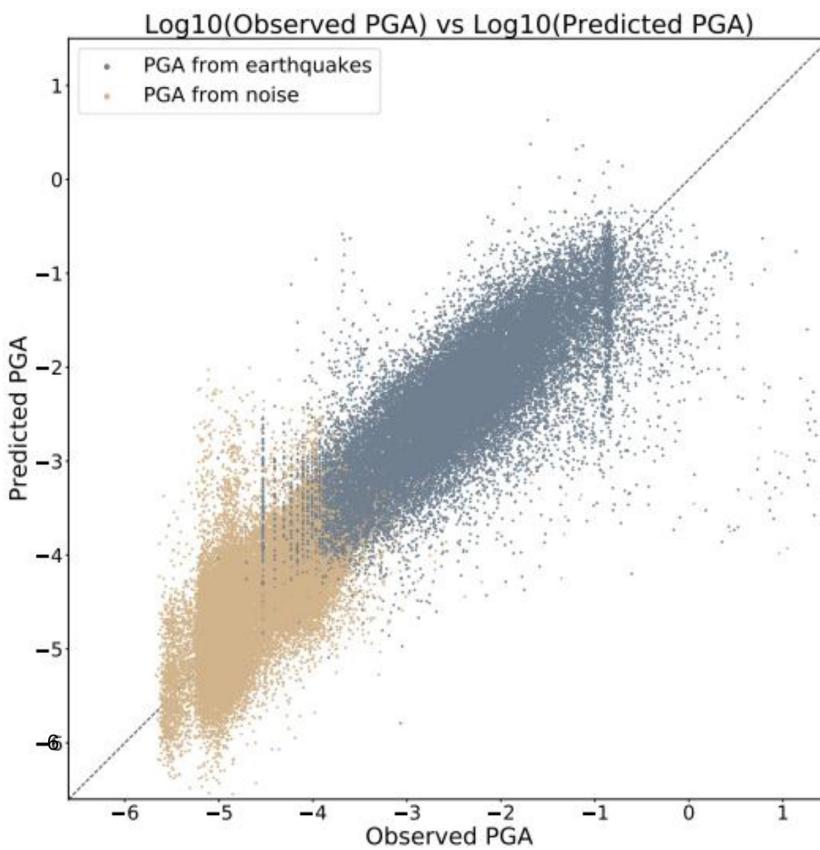


Figure 9. $\log_{10}(\text{PGA}_{\text{observed}})$ versus $\log_{10}(\text{PGA}_{\text{predicted}})$ for training with noise and event data. The vertical stripe visible at approximately $-0.9 \log_{10}(\text{PGA}_{\text{observed}})$ results from data logger equipment malfunctioning.

available for training. We have also observed that the CNN model performs well when noise-only data were included in the data set. These results indicate that the proposed analysis could be implemented in a real-time analysis configuration. Although the technique is not strictly designed for earthquake early warning, we found that it could provide useful estimates of ground motions within 15–20 s after the earthquake origin time depending on the data transmission infrastructure, latencies and processing time requirements.

The study uses a specific set of stations and earthquakes concentrated spatially and it could be applied to larger areas with more widespread seismicity as long as enough data are available for training. The main purpose of this study was to show that the technique can provide quite satisfactory results in terms of predicted ground motion only by learning patterns of multistation 3C waveforms.

7 DATA AND RESOURCES

Earthquake catalogue and waveform data have been downloaded through the INGV FDSN web services (http://terremoti.ingv.it/en/webservices_and_software; INGV Seismological Data Centre 2006; Emersito Working Group 2018). Waveforms have been downloaded and processed using python library Obspy (Beyreuther *et al.* 2010). IMs for the stations with no data have been calculated using USGS ShakeMap 4 (http://usgs.github.io/shakemap/sm4_index.html). The CNN model has been developed using the Keras Python Deep Learning library (Chollet *et al.* 2015). The code and the data of this paper will be available on <https://github.com/djozinovi/CNNpredIM>.

ACKNOWLEDGEMENTS

The research has been funded by SERA EU project (Seismology and Earthquake Engineering Research Infrastructure Alliance for Europe; contract n. 730900). It has also greatly benefited from a travel grant for a short-term scientific mission provided by G2NET, COST action CA17137. The authors would like to thank the operators of RSN (INGV) and the USGS ShakeMap team of developers. The authors thank the reviewers Dr S. Mostafa Mousavi and Dr Dino Bindi for their reading of the manuscript and their many comments and suggestions.

Author contribution statement: The topic of this study was conceived by Alberto Michelini, with contributions from Anthony Lomax and Dario Jozinović. Dario Jozinović prepared, downloaded and processed the earthquake data; Alberto Michelini prepared and processed the ShakeMap data. The CNN model has been set up and trained by Dario Jozinović with contribution from Ivan Štajduhar. The results have been analysed by Dario Jozinović with contribution by Alberto Michelini. The final manuscript has been written by Dario Jozinović and Alberto Michelini, with contributions by Anthony Lomax and Ivan Štajduhar.

REFERENCES

- Alavi, A.H. & Gandomi, A.H., 2011. Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing, *Comput. Struct.*, **89**(23–24), 2176–2194.
- Allen, R.M., Gasparini, P., Kamiguchi, O. & Bose, M., 2009. The status of earthquake early warning around the world: an introductory overview, *Seismol. Res. Lett.*, **80**(5), 682–693.

- Atik, L.A., Abrahamson, N., Bommer, J.J., Scherbaum, F., Cotton, F. & Kuehn, N., 2010. The variability of ground-motion prediction models and its components, *Seismol. Res. Lett.*, **81**(5), 794–801.
- Bergen, K.J., Johnson, P.A., Maarten, V. & Beroza, G.C., 2019. Machine learning for data-driven discovery in solid Earth geoscience, *Science*, **363**(6433), doi:10.1126/science.aau0323.
- 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.
- Bindi, D., Pacor, F., Luzi, L., Puglia, R., Massa, M., Ameri, G. & Paolucci, R., 2011. "Ground motion prediction equations derived from the Italian strong motion database, *Bull. Earthq. Eng.*, **9**(6), 1899–1920.
- Chiaraluce, L. *et al.*, 2017. The 2016 central Italy seismic sequence: a first look at the mainshocks, aftershocks, and source models, *Seismol. Res. Lett.*, **88**(3), 757–771.
- Chollet, F., 2018. *Deep Learning with Python*. Manning Publications.
- Chollet, F. *et al.*, 2015. Keras. <https://keras.io>.
- Chiaruttini, C., Roberto, V. & Saitta, F., 1989. Artificial intelligence techniques in seismic signal interpretation, *Geophys. J. Int.*, **98**(2), 223–232.
- Derras, B., Bard, P.Y. & Cotton, F., 2014. Towards fully data driven ground-motion prediction models for Europe, *Bull. Earthq. Eng.*, **12**(1), 495–516.
- Dysart, P.S. & Pulli, J.J., 1990. Regional seismic event classification at the NORESS array: seismological measurements and the use of trained neural networks, *Bull. Seism. Soc. Am.*, **80**(6B), 1910–1933.
- Faenza, L., Lauciani, V. & Michelini, A., 2016. The ShakeMaps of the Amatrice, M6, earthquake, *Ann. Geophys.*, **59**.
- Garca-Laencina, P.J., Sancho-Gómez, J.-L. & Figueiras-Vidal, A.R., 2010. Pattern classification with missing data: a review, *Neural Comp. Appl.*, **19**(2), 263–282.
- Given, D.D. *et al.*, 2018. Revised technical implementation plan for the ShakeAlert system—an earthquake early warning system for the West Coast of the United States: U.S. Geological Survey Open-File Report 2018-1155, 42 p. <https://doi.org/10.3133/ofr20181155>.
- Goodfellow, I., Bengio, Y. & Courville, A., 2016. *Deep Learning*. MIT Press.
- Hulbert, C., Rouet-Leduc, B., Johnson, P.A., Ren, C.X., Rivière, J., Bolton, D.C. & Marone, C., 2019. Similarity of fast and slow earthquakes illuminated by machine learning, *Nat. Geosci.*, **12**(1), 69.
- INGV Seismological Data Centre. 2006. Rete Sismica Nazionale (RSN), Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy, .
- EMERSITO Working Group. 2018. Rete sismica del gruppo EMERSITO, sequenza sismica del 2016 in Italia Centrale. Istituto Nazionale di Geofisica e Vulcanologia (INGV), .
- Kohler, M.D. *et al.*, 2017. Earthquake early warning ShakeAlert system: west coast wide production prototype, *Seismol. Res. Lett.*, **89**(1), 99–107.
- Kong, Q., Trugman, D.T., Ross, Z.E., Bianco, M.J., Meade, B.J. & Gerstoft, P., 2018. Machine learning in seismology: turning data into insights, *Seismol. Res. Lett.*, **90**(1), 3–14.
- Kriegerowski, M., Petersen, G.M., Vasyura-Bathke, H. & Ohrnberger, M., 2018. A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms, *Seismol. Res. Lett.*, **90**(2A), 510–516.
- LeCun, Y.A., Bottou, L., Orr, G.B. & Müller, K.R., 2012. Efficient Back-Prop, in *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*, Vol. **7700**, eds Montavon, G., Orr, G.B., & Müller, K.R., Springer.
- Lomax, A., Michelini, A. & Jozinović, D., 2019. An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network, *Seismol. Res. Lett.*, **90**(2A), 517–529.
- Michelini, A., Faenza, L., Lanzano, G., Lauciani, V., Jozinović, D., Puglia, R. & Luzi, L., 2019. The new ShakeMap in Italy: progress and advances in the last 10 Yr, *Seismol. Res. Lett.*, **91**(1), 317–333.
- Minson, S.E., Meier, M.A., Baltay, A.S., Hanks, T.C. & Cochran, E.S., 2018. The limits of earthquake early warning: timeliness of ground motion estimates, *Sci. Adv.*, **4**(3), eaao5054.
- Mousavi, S.M. & Beroza, G.C., 2020. A machine-learning approach for earthquake magnitude estimation, *Geophys. Res. Lett.*, **47**(1), doi: 10.1029/2019GL085976.
- Mousavi, S.M., Zhu, W., Sheng, Y. & Beroza, G.C., 2019. CRED: a deep residual network of convolutional and recurrent units for earthquake signal detection, *Sci. Rep.*, **9**(1), 1–14.
- Ochoa, L.H., Niño, L.F. & Vargas, C.A., 2018. Fast magnitude determination using a single seismological station record implementing machine learning techniques, *Geod. Geodyn.*, **9**(1), 34–41.
- Pan, S.J. & Yang, Q., 2010. A Survey on Transfer Learning, *IEEE Trans. Knowl. Data Eng.*, **22**, 1345–1359.
- Perol, T., Gharbi, M. & Denolle, M., 2018. Convolutional neural network for earthquake detection and location, *Sci. Adv.*, **4**(2), e1700578.
- Reynen, A. & Audet, P., 2017. Supervised machine learning on a network scale: Application to seismic event classification and detection, *Geophys. J. Int.*, **210**(3), 1394–1409.
- Rojas, O., Otero, B., Alvarado, L., Mus, S. & Tous, R.T., 2019. Artificial neural networks as emerging tools for earthquake detection, *Computación y Sistemas*, **23**(2), 335–350.
- Satriano, C., Wu, Y.-M., Zollo, A. & Kanamori, H., 2011. Earthquake early warning: concepts, methods and physical grounds, *Soil Dyn. Earthq. Eng.*, **31**(2), 106–118.
- Spallarossa, D., Kotha, S.R., Picozzi, M., Barani, S. & Bindi, D., 2019. On-site earthquake early warning: a partially non-ergodic perspective from the site effects point of view, *Geophys. J. Int.*, **216**(2), 919–934.
- Wald, D.J., Quitoriano, V., Heaton, T.H., Kanamori, H., Scrivner, C.W. & Worden, C.B., 1999. TriNet "ShakeMaps": rapid generation of peak ground motion and intensity maps for earthquakes in southern California, *Earthq. Spectra*, **15**(3), 537–555.
- Worden, C.B., Thompson, E.M., Baker, J.W., Bradley, B.A., Luco, N. & Wald, D.J., 2018. Spatial and spectral interpolation of ground-motion intensity measure observations, *Bull. Seism. Soc. Am.*, **108**(2), 866–875.
- Zhu, W. & Beroza, G.C., 2018. PhaseNet: a deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.*, **216**(1), 261–273.

SUPPORTING INFORMATION

Supplementary data are available at [GJI](https://doi.org/10.1093/gji/gjaa222) online.

Figure S1. Bad earthquake data.

Figure S2. Bad earthquake data.

Figure S3. Waveforms for $M = 5.9$ earthquake. Green is the 10-s marker and orange the PGA marker.

Figure S4. Waveforms for $M = 6.5$ earthquake. Green is the 10-s marker and orange the PGA marker.

Figure S5. Training and test loss during the training epochs.

Figure S6. Variation of residuals with magnitude.

Figure S7. Variation of residuals with epicentral distance.

Figure S8. Mean and standard deviation of the PGA residuals, calculated for each number of stations that have the data in the input waveforms. In the figure (a) the results are calculated for the stations with observed targets (i.e. the stations which have the input data). In the figure (b) the results are calculated for the stations with ShakeMap predicted targets (i.e. the stations for which the input data are missing).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.