Geophysical Journal International

Geophys. J. Int. (2022) **231,** 2186–2204
Advance Access publication 2022 August 18
GJI Seismology

https://doi.org/10.1093/gji/ggac325

Learning source, path and site effects: CNN-based on-site intensity prediction for earthquake early warning

Hongcai Zhang,[1,2] Diego Melgar,[3] Valerie Sahakian,[3] Jake Searcy[4] and Jiun-Ting Lin[3]

[1] *Seismic Monitoring Center, Fujian Earthquake Agency, CEA, Fuzhou 350003, China. E-mail: zhanghc@fjea.gov.cn*
[2] *Marine Seismic Observation Group, Institute of Xiamen Marine Seismology, CEA, Xiamen* 361021*, China*
[3] *Department of Earth Sciences, University of Oregon, Eugene, OR* 97403-1272*, USA*
[4] *Research Advanced Computer Services, University of Oregon, Eugene, OR* 97403-1272, USA

Accepted 2022 August 11. Received 2022 July 23; in original form 2022 March 17

SUMMARY

To provide timely and accurate seismic alerts for potential users during the earthquake early warning (EEW) process, several algorithms have been proposed and implemented. Some of the most common rely on the characterization of the earthquake magnitude and location, and then use a ground motion model to forecast shaking intensity at a user's location. It has been noted that with this approach the scatter in the forecasted intensities can be significant and may affect the reliability and usefulness of the warnings. To ameliorate this, we propose a single station machine learning (ML) algorithm. We build a four-layer convolutional neural network (CNN), named it CONIP (Convolutional neural network ONsite Intensity Prediction), and test it using two data sets to study the feasibility of seismic intensity forecasting from only the first few seconds of a waveform. With only limited waveforms, mainly *P* waves, our CONIP model will forecast the on-site seismic intensity. We find that compared with existing methods, the forecasted seismic intensities are much more accurate. To understand the nature of this improvement we carry out a residual decomposition and quantify to what degree the ML model learns site, regional path, and source information during the training. We find that source and site effects are easily learned by the algorithm. Path effects, on the other hand, can be learned but will depend largely on the number, location, and coverage of stations. Overall, the ML model performance is a substantial improvement over traditional approaches. Our results are currently only applicable for small and moderate intensities but, we argue, could in future work be supplemented by simulations to supplement the training data sets at higher intensities. We believe that ML algorithms will play a dominant role in the next generation of EEW systems.

Key words: Machine learning; Convolutional neural network; Residual decomposition.

1 INTRODUCTION

The main goal of an earthquake early warning (EEW) system is to issue alerts or notifications to people and vulnerable systems likely to experience potentially damaging ground shaking (Allen & Melgar 2019). Many of the presently operating EEW systems accomplish this through an indirect approach. They estimate source parameters (hypocentre and magnitude) rapidly from the first few seconds of a waveform or set of waveforms. After that, ground motion for target areas is forecasted by applying a suitable ground-motion model (Goltz & Flores, 1997; Alcik *et al.* 2009; Allen *et al.* 2009; Hsiao *et al.* 2009; Kamigaichi *et al.* 2009; Wu *et al.* 2013; Cuéllar *et al.* 2014; Hoshiba & Ozaki 2015; Zhang *et al.* 2016; Given *et al.* 2018; Kohler *et al*. 2016). Recently, it has become commonplace to evaluate the performance of EEW algorithms based on their ability to forecast shaking and not on the quality of the obtained source parameters (Kuyuk & Motosaka 2009; Kodera *et al*. 2016; Meier 2017; Kilb *et al.* 2021; Saunders *et al.* 2022). This is reasonable because it is ultimately shaking that leads to damage and loss of life. In these evaluations, it has been found that although only a few parameters are required for the point source approach, estimation of ground motion can have significant scatter (Meier 2017). This is mostly because any bias in the source parameters, combined with the uncertainties in the GMMs, leads to potentially large ground motion prediction errors. Uncertainties in GMMs arise in part because most are developed with the assumption that ground motion is ergodic (Anderson & Brune 1999), which means that the distribution of ground motions over time at a given site is the same as their spatial distribution over all sites for the same magnitude, distance and site condition (Baltay *et al.* 2017). As a result, propagation through heterogeneous media with varying attenuation properties, complex site effects, directivity and other non-ergodic processes is not captured well. Efforts to move away from the ergodic assumption are still an area of active research (Walling 2009; Dawood & Adrian 2013; Villani & Abrahamson 2015; Landwehr *et al.* 2016; Kotha *et al.* 2017; Sahakian *et al.* 2019; Kuehn & Abrahamson 2020) and would improve EEW approaches that rely on a GMM (Spallarossa *et al.* 2019).

Another important outstanding issue is that most GMM-based EEW ground-motion prediction methods assume a point source model which can lead to unsatisfactory performance for major earthquakes. In addition to the significant underestimation in large magnitude events, the ground motion of multiple simultaneous earthquakes is often overpredicted. Some proportion of events during the intense aftershock sequences following such large events are not captured. Once more, an example of this can be found in the Tohoku earthquake sequence. To address these many issues, several methods have been proposed. In PreSEIS (Böse *et al.* 2008), a two-layer feed-forward neural network is used to estimate the expansion of the evolving rupture. Yamada (2014) proposed a near/far-source classification approach to identify the fault rupture geometry. Similarly, Böse *et al.* (2012b) use the image recognition technique to detect fault ruptures and developed the FinDer algorithm.

Recognizing this fundamental difficulty as well as the fact that it is the ground motion that is of fundamental interest to EEW, Hoshiba (2013) proposed a real-time ground motion prediction method based on the Kirchhoff–Fresnel boundary integral equation. By real-time wavefield monitoring and assumptions on the direction of the wave propagation, ground motion at various locations can be predicted without knowing the hypocentre and magnitude of the event. On the basis of this method, Hoshiba & Aoki (2015) proposed a numerical shake prediction approach by applying a data assimilation technique. Tamaribuchi *et al.* (2014) use the integrated particle filter (IPF) method to address overpredictions caused by hypocentre mislocation that results from multiple simultaneous earthquake events. Kodera *et al.* (2016, 2018) proposed the propagation of local undamped motion (PLUM) algorithm to minimize underpredictions for very large earthquakes. These methods have been tested offline in the United States, Cochran *et al.* (2019) assessed the PLUM algorithm in southern California and found PLUM can be configured to successfully issue alerts using lower I_{MMI} trigger thresholds. Minson *et al.* (2020) present a retrospective analysis of PLUM performance and explore three potential regional alerting strategies based on the July 2019 M_{w} 6.4 and 7.1 Ridgecrest earthquakes, that study shows PLUM initial shaking forecasts can be faster than the point source algorithm of ShakeAlert and is able to accurately forecast shaking and proved to be largely immune to high-data latencies that degraded ShakeAlert's performance. These algorithms all offer improvements over the classical source parameter/GMM combination approach, but often require a dense observation network and can have difficulty with sparse station coverage.

Meanwhile, the amount of available seismic data has increased significantly with the rapid growth of networks worldwide. As a typical data-driven discipline (Kong *et al.* 2019) this has led to new methodological advances, and there are potentials to find patterns and new significant features in waveforms from these massive data sets. Leveraging this, multiple machine learning (ML) algorithms have been proposed in seismology. These use either hand-selected physical features or deep learning approaches to automatically extract features to detect events (Wiszniowski *et al.* 2014; Zhang *et al.* 2014; Yoon *et al.* 2015; Rabin *et al.* 2016; Trugman & Shearer 2017; Bergen & Beroza 2018; Perol *et al.* 2018; Ross *et al.* 2019; Chen 2020), pick phases (Li *et al.* 2018; Ross *et al.* 2018a, b, c; Yin *et al.* 2018; Zhu & Beroza 2019), estimate ground motion (Böse *et al.* 2012a; Derras *et al.* 2012, 2014; Alimoradi & Beck 2015; Asim *et al.* 2017; Khoshnevis & Taborda 2018; Ochoa *et al.* 2018; Klimasewski *et al.* 2021), provide early warnings (Böse *et al.* 2014; Kong *et al.* 2016a,b; Fauvel *et al.* 2020; Münchmeyer *et al.* 2021a, b), or even, and controversially, earthquake prediction and forecasting (Asencio-Cortés *et al.* 2017, 2018 ; Rafiei & Adeli 2017; Rouet-Leduc *et al.* 2017; DeVries *et al.* 2018).

In this study, instead of estimating the earthquake magnitude (Lomax *et al.* 2019; Mousavi & Beroza 2020a, b) and locating the event (Kriegerowski *et al*. 2018; Mousavi *et al*. 2019) to then forecast ground motion, we propose and demonstrate an on-site algorithm that uses information from a single station to forecast intensity at that site. We accomplish this by using a convolutional neural network (CNN) approach. In our algorithm, which we call CONIP (Convolutional neural network On-site Intensity Prediction), only the first few seconds of a waveform (mainly *P* waves) are used to train the ML model. This is consistent with the existing on-site early warning methods. To achieve the needs of real-time processing inherent in an EEW system, our algorithm starts the seismic intensity forecasting from the first *P*-wave detection and then updates the prediction second by second. Thus, it provides a progressive seismic intensity prediction approach which we believe will be very suitable for the EEW process. This is still a typical on-site approach, but unlike the τ_{c} and P_{d} algorithm (Wu *et al.* 2003, 2006; 2007; Wu & Kanamori 2005, 2008; Zollo *et al.* 2006, 2010; Peng *et al.* 2014, 2017), we did not extract feature parameters. Rather, we estimate seismic intensity directly from available waveforms. Additionally, we did not need to detect events (that can be done with another standard ML or STA/LTA (Allen 1978) algorithms), which we believe may provide higher accuracy earthquake alerts to end-users. In fact, several attempts had been done during the past few years. Käufl (2015) developed a neural network-based methodology to determine source parameters and predict ground motion directly. Jozinović *et al.* (2020) show the capability of a CNN model to predict the intensity measurements (IMs), they use a fixed 10 s window starting at the earthquake origin time to predict IMs at stations far from the epicentre. Based on this, they adopt the 'transfer learning' (TL) methodologies to counter the lack of data in central-western Italy (Jozinović *et al.* 2022). Münchmeyer *et al.* 2021a,b) proposed a deep-learning-based transformer earthquake alerting model (TEAM), and evaluate its performance in Japan and Italy. Their study shows TEAM performed significantly better than the conventional algorithm such as for earthquake location and magnitude estimation, thus it is applicable to real-time prediction and provides realistic uncertainty estimates based on probabilistic inference. Similarly, we compare the performance of our approach to an idealized on-site EEW system where the magnitude and locations of the events are very well defined from a high-quality catalogue and assume that they could be known to this high degree of accuracy in real-time. We then estimate intensities using this ideal knowledge of the event (which is not possible currently in real-time) combined with a high-quality regional GMM. We will show that, when judged by the errors in the intensity forecast, the ML approach far outperforms the ideal on-site algorithm. To understand why this superior performance is obtained, we use a ground-motion residual decomposition approach (Atik *et al.* 2010; Baltay *et al.* 2017;

Sahakian *et al.* 2018) and show that much of this improvement can be attributed to the algorithm learning source, path and site effects specific to each region where it is applied. Trying to provide more accurate MMI by choosing ML algorithms has been attempted with varying degrees of success before (Kuyuk & Motosaka 2009; Böse *et al.* 2012a, 2014), but few studies explained why ML models potentially perform better. What do they really learn during training, and which parts of the physical process are the most difficult to learn by using this approach? In this paper, beyond proposing a new algorithm we provide some answers to these questions by relying on residual decomposition.

Finally, we note that our approach is currently limited to moderate magnitudes and shaking intensities because that is what is most widely available in the database of events for training. However, we consider this to be one more step toward fully harnessing the power of ML algorithms for EEW to learn source, site and path effects, and to generalize this to arbitrarily large earthquakes and ground motions.

2 DATA SETS

Voluminous and high-quality data sets are the core part of the success of any ML approach. To facilitate the study, several seismic ML data sets had been published and used widely, such as STEAD (Mousavi *et al.* 2019), INSTANCE (Michelini *et al.* 2021) and DiTing (Ming *et al*. 2022). As most of the existing data sets are large-scale (e.g. STEAD), it will be hard to learn useful information for a specific small area even though massive records are contained in the data set. Some of the data sets (e.g. INSTANCE and DiTing) were created much later than our work and records provided in some data sets are raw data (e.g. DiTing), and the instrument response was not removed, which make it almost impossible to calculate seismic intensity from the records directly. For these reasons, here we build two data sets by using historical events that occurred from January 2000 to December 2019 in California. We download waveforms from the Incorporated Research Institutions for Seismology (IRIS) for two adjacent areas in California with high seismic activity. The first is smaller and is focused around Parkfield (PK) and the second is a large swath of Southern California centred around Los Angeles (LA, Fig. 1). We train and test the performance of the algorithm in these two regions as they provide examples of what is possible with very dense (LA) and with a relatively sparse (PK) network. Events with magnitude greater than 2.0 are included in our study, this yields 10 827 and 12 715 historical events available in these two regions during this period. After data download, we remove the instrument response and detrend all traces. Phase arrivals provided by IRIS are used to trim data from the *P*-wave arrival. For training, we choose a signal-to-noise ratio (SNR) criteria to keep records with acceptable quality. The SNR is calculated based on the ratio of root mean square of the waveform before and after *P*-wave arrival in a 3 s time window. Records with SNR $\geq$ 3 were used during the data set construction, model training and testing.

Initially, for each earthquake, we download any available accelerogram and broad-band seismogram in the study region. For the big events, broad-band sensors are likely to clip at a short distance, which will no longer be suitable at which point strong ground motion records are needed. But most of the events downloaded are small magnitude, thus the SNR for a large portion of the accelerograms is significantly lower than seismograms at the same hypocentral distance. As a result, for this study we focus only on broad-band seismograms leaving for each of these two study regions, 13 115 (PK) and 19 144 (LA) three-channel records to include, and the range of epicentral distance in the data sets is 0.2–190.9 km (PK) and 0.9–227.6 km (LA), respectively. These come from 17 (PK) and 204 (LA) stations and for 4024 (PK) and 925 (LA) events separately (Fig. 2). The magnitude range in each data set is 2.0–5.96 (PK) and 2.0–5.44 (LA), respectively. It is worth mentioning that, as most of the events we used here usually only provide M_L scale, in order to facilitate subsequent analysis (see Section 2), we convert them to moment magnitude (M) scale by following (Ross *et al.* 2018) study. For each record, we extract peak ground velocity (PGV) and peak ground acceleration (PGA, after differentiation) from the two horizontal components (computed using the geometric mean of both components) and calculate MMI by using the statistical relationship provided by (Worden *et al.* 2012), therefore, the range of MMI in the data sets are –2.3 to 6.0 (PK) and –1.4 to 7.2 (LA), respectively. Like PGA and PGV, MMI is a more widely used physical quantity that reflects the of ground motion intension, as it's directly related to the potential damages during an earthquake, we set it as the training target of our CONIP model. Finally, we save all data into an HDF5 file and use this to train CONIP. In Fig. 2, we show the magnitude and MMI distribution of records selected in our two data sets. We have more larger magnitude (and/or higher MMI) records in the LA data set, this also allows us to compare model performance for these two different event distributions.

3 METHODS

3.1 CNN model

A CNN is a class of deep neural networks, most commonly applied to image and video recognition, image classification, medical image analysis, natural language processing and financial time-series (Krizhevsky *et al.* 2012). CNNs extract features automatically and are able to learn them effectively. This is very different from the traditional hand-engineered feature extraction mode. A CNN consists of multiple layers, usually an input and an output layer, as well as hidden layers. The core mathematical operation, convolution, is usually done in the hidden layers, other operations, such as activation, pooling and fully connectedness are all carried out in hidden layers. As a result, we only need to design the network according to our actual needs without particular attention as to which features the network should learn.

3.2 Model structure

The neural network that is used by CONIP contains four convolutional layers, two maxpooling layers, one flattening layer and two fully connected layers (Fig. 3). The input to the network is the 100 Hz three channel waveform data for a particular station/event pair and the output is the predicted MMI at that same site. We train our network to produce predictions that are as accurate as possible across the data sets. We use the open-source neural-network library Keras to construct the network. For training, we use a mini-batch size of 64 and set the learning rate to 0.001. We use the mean square error (MSE) loss function to supervise the accuracy of the prediction results and the Adam optimizer. Other technical details of each convolution layer (e.g. filters used in each convolution layer) can be found in Table 1. For every second that has elapsed after *P*-wave arrival, we use the same CNN model architecture to predict the on-site MMI. The only difference is the shape of input data for each time window whose length (TW) will increase accordingly. Thus,

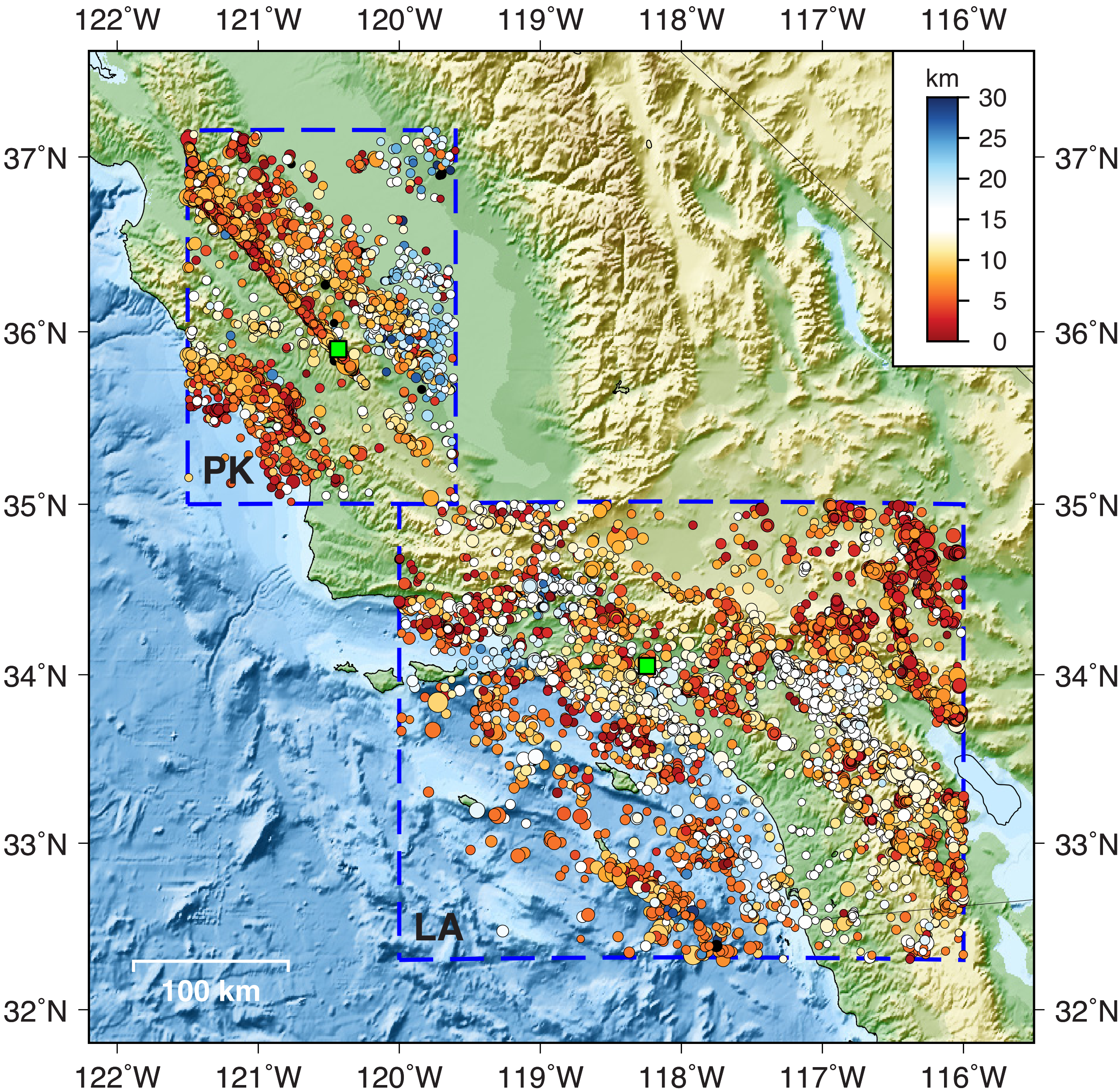

Figure 1. Historical events distribution of two adjacent regions [Parkfield (PK) and Los Angeles (LA)] studied in our paper. Circle size show the magnitude of each event and colours represent the depth of events (only in the electronic version), triangles represent broad-band stations used in our study.

we have a 1 s CNN model, that makes predictions based on the first second of *P*-wave data, a 3 s model, a 5 s model, and so on. As most of the peak ground motion values will be available within 10 s in our data sets, we only trained models out to the first 10 s. Of course, for larger events, this training window can be expanded while keeping the model architecture the same.

3.3 Model training

For training, we first normalize all waveforms by their peak amplitudes and then randomly split the data into independent training (75 per cent) and testing data sets (25 per cent). Meanwhile, for the training data set, 80 per cent of the records are used to train the model and 20 per cent of records were selected to validate the results. The model was trained with only an 8-core i7 CPU laptop. Since we do not use graphical processing units (GPU) during the model training, the entire training process of our model is somewhat time-consuming. On average, completing one training epoch takes several seconds of CPU wall-time. For our modestly sized data set, the time spent training is tolerable.

To prevent overfitting, we apply the 'early-stop' strategy. We halt training as soon as the validation loss no longer improves for five consecutive epochs. Then the epoch with the lowest loss value on the validation set was selected as the best model. For our study, we typically need 50–60 epochs to obtain the 'lowest loss' CNN model. As mentioned earlier in our paper, we start on-site MMI prediction at the *P*-wave detection time and will update the prediction every second. As the time window (TW) length increases, the shape of input data will increase accordingly, so we train separate models

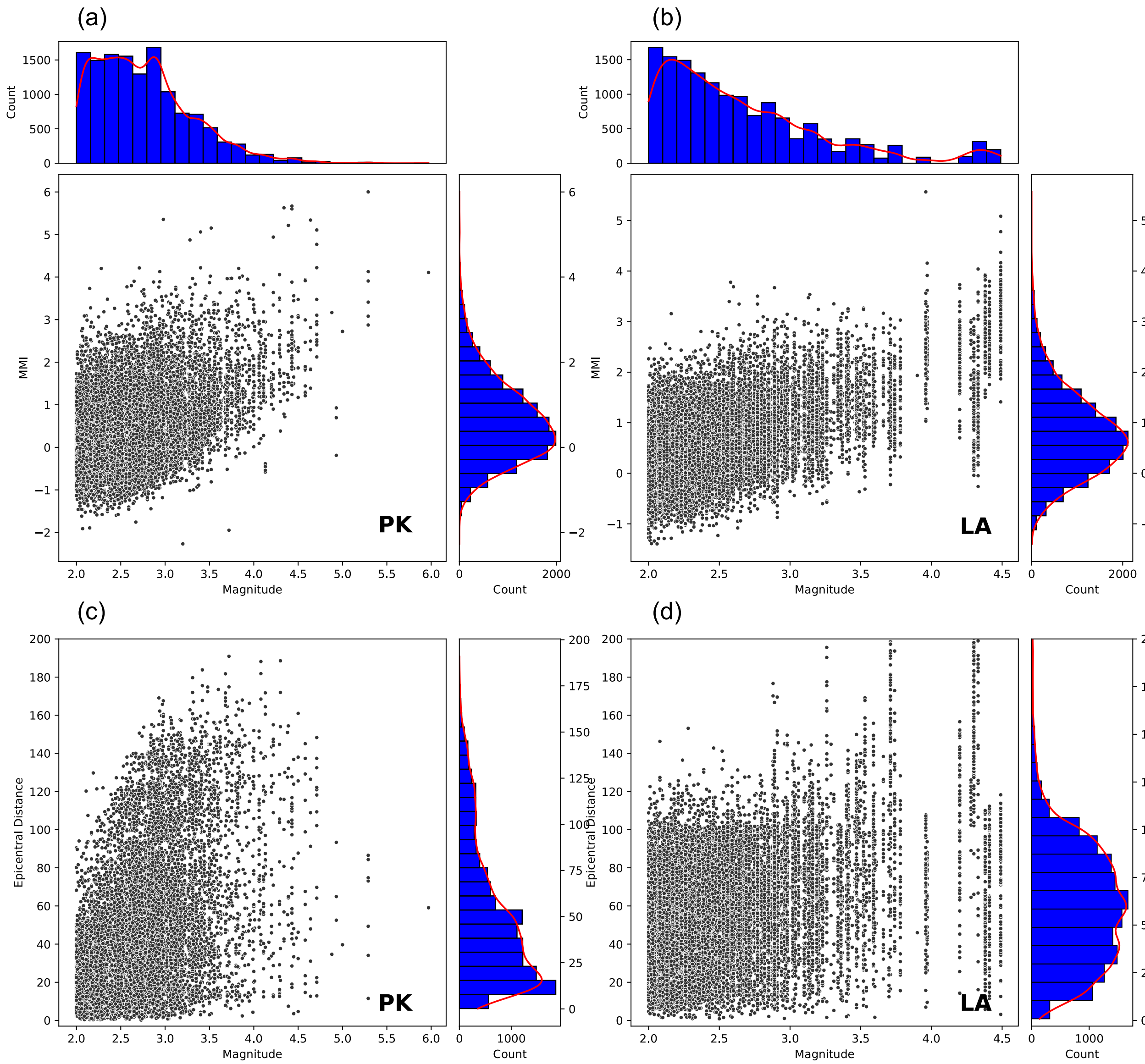

Figure 2. Statistical of selected seismic records distribution in two data sets (PK and LA) separately. These two data sets include 13 115 (a) and 19 144 (b) records, respectively. The bottom left-hand panel in (a) and (b) shows the distribution of magnitude versus MMI in each data set, and the up and right panels in (a) and (b) are the magnitude and MMI histograms, respectively. Similarly, the bottom left-hand panel in (c) and (d) shows the distribution of magnitude versus epicentral distance in each data set and the right-hand panels in (c) and (d) are the epicentral distance histograms.

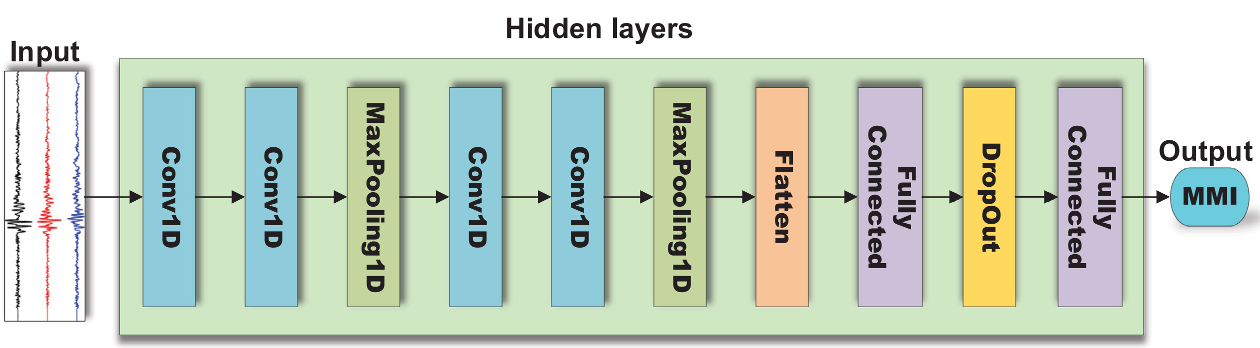

Figure 3. Cartoon depicting the CNN workflow for on-site MMI prediction, where Conv1D means 1-D convolution layer, MaxPooling1D means max pooling operation for 1-D temporal data. In addition to these layers, fully connected layer, dropout layer and flatten layer are essential components of a CNN model.

Table 1. Detail of the layers that used in our study.

Name	L1 Conv1D	L2 Conv1D	L3 MaxPooling1D	L4 Conv1D	L5 Conv1D	L6 MaxPooling1D	L7 Dense	L8 Dropout	L9 Dense
Size/rate	9	9	3	9	9	3	128	0.5	1
Filters	128	128	-	64	64	-	-	-	-
Activation	ReLU	ReLU	-	ReLU	ReLU	-	ReLU	-	Linear

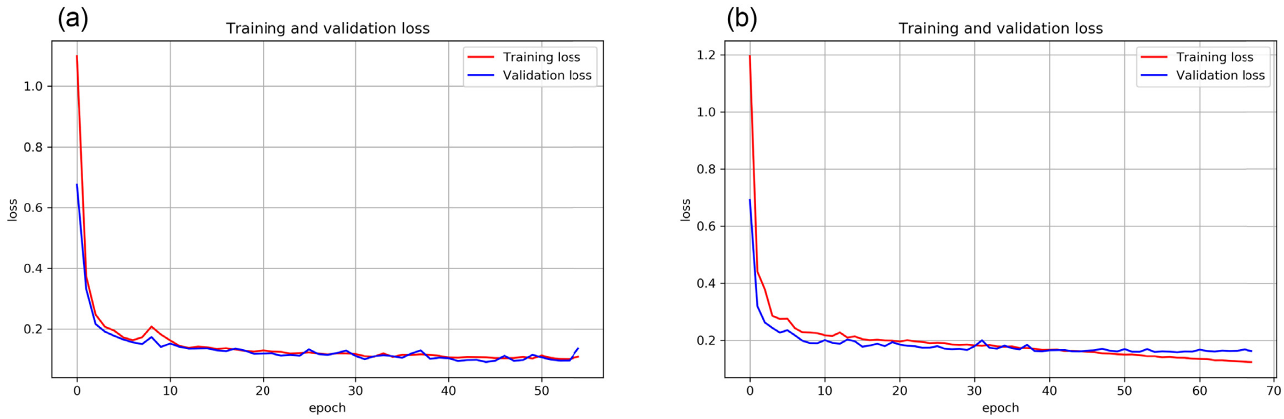

Figure 4. Model training and validation loss changes with epochs for (a) PK and (b) LA separately. The training and validation losses decline rapidly with epochs increase and gradually stabilized after 40–60 epochs for both PK and LA region.

for each TW based on the same model architecture. *A priori*, one expects that, when a longer TW is chosen, MMI predictions should be more accurate than with a shorter TW. As a result, based on the same CNN structure, we should be able to provide a progressive approach to forecasting MMI. Existing on-site EEW algorithms source parameters and ground motion forecasts are usually generated from a 3-s TW. Because of this, in this paper, as a benchmark for comparison with existing approaches, we focus mostly on the CNN model with a 3 s TW. We note that only records for which the peak values were not reached during a specific TW length were used to train each model.

To evaluate performance, we use a loss history curve. In Fig. 4, we show the results for the training and validation data sets as a function of the number of training epochs for TW = 3 s in both PK and LA data sets. As shown in this figure, both the training and validation loss history have highly consistent trends. At the early stage of the model training, the model loss is always high but decreases rapidly after a few epochs. This pattern is seen in both PK and LA data sets, which reflects that the CNN model used in our study is suitable. After 50–60 epochs, validation loss gradually stabilizes, and we can then obtain the best model for a particular TW.

3.4 Idealized on-site algorithm

In order to compare the performance of CONIP against traditional EEW approaches, we also build an idealized on-site algorithm. For each event, we assume the magnitude and location from the COMCAT catalogue and use the BSSA14 GMM (Boore *et al.* 2014) to forecast shaking intensity at each individual site. The BSSA14 GMM is part of the NGA-West2 family of ground models and was designed specifically for use in California. For each site, we also consider the Vs30 values from the hillslope proxy map provided by the USGS (Allen & Wald 2009).

3.5 Residual decomposition

To understand specifics about the performance of both the idealized on-site EEW approach and CONIP, we quantify how much of the misfit between observed and predicted intensities is due to errors in characterizing the source, site, and path effects. For this, we use a residual decomposition approach. The residual, δ_{ij}, for a specific event, i, and station, j, pair is defined as the difference between an observed ($I_{\mathrm{obs},ij}$) and a predicted ($I_{\mathrm{pred},ij}$) intensity such that

$$\delta_{ij} = \ln(I_{\mathrm{obs},ij}) - \ln(I_{\mathrm{pred},ij}). \quad (1)$$

The observations are determined from the measured waveforms and the predictions are either from the idealized EEW algorithm where magnitude and location are perfectly known or from the CNN approach.

We thus have two sets of residuals between which we can compare: GMM from the traditional EEW approach, and CONIP from our algorithm. Residuals reflect a misfit between what is observed and what is expected, and it is possible to decompose them to separate which part of the discrepancy can be ascribed to the source, to the path and to the site. We use a mixed-effects maximum likelihood approach (Pinheiro & Bates 2000), similar to Abrahamson & Youngs (1992) and Stafford (2014)), which solves for *fixed effects*—terms that are unaffected by data set selection, and *random effects*—biases that arise due to the data set selection. Typically, with an intensity measure as the predictive parameter, the remaining terms in the GMM functional form (such as magnitude and distance) are inverted for fixed effects (Brillinger & Preisler 1984, 1985; Abrahamson & Youngs 1992; Sahakian *et al.* 2018), and individual event (and occasionally site) corrections as random effects. The benefit of using a maximum likelihood mixed-effects approach as opposed to a pooled, ordinary least-squares inversion, is that the resulting values of each random effect consider the uncertainties of other random effects (Abrahamson & Youngs 1992; Sahakian *et al.* 2018).

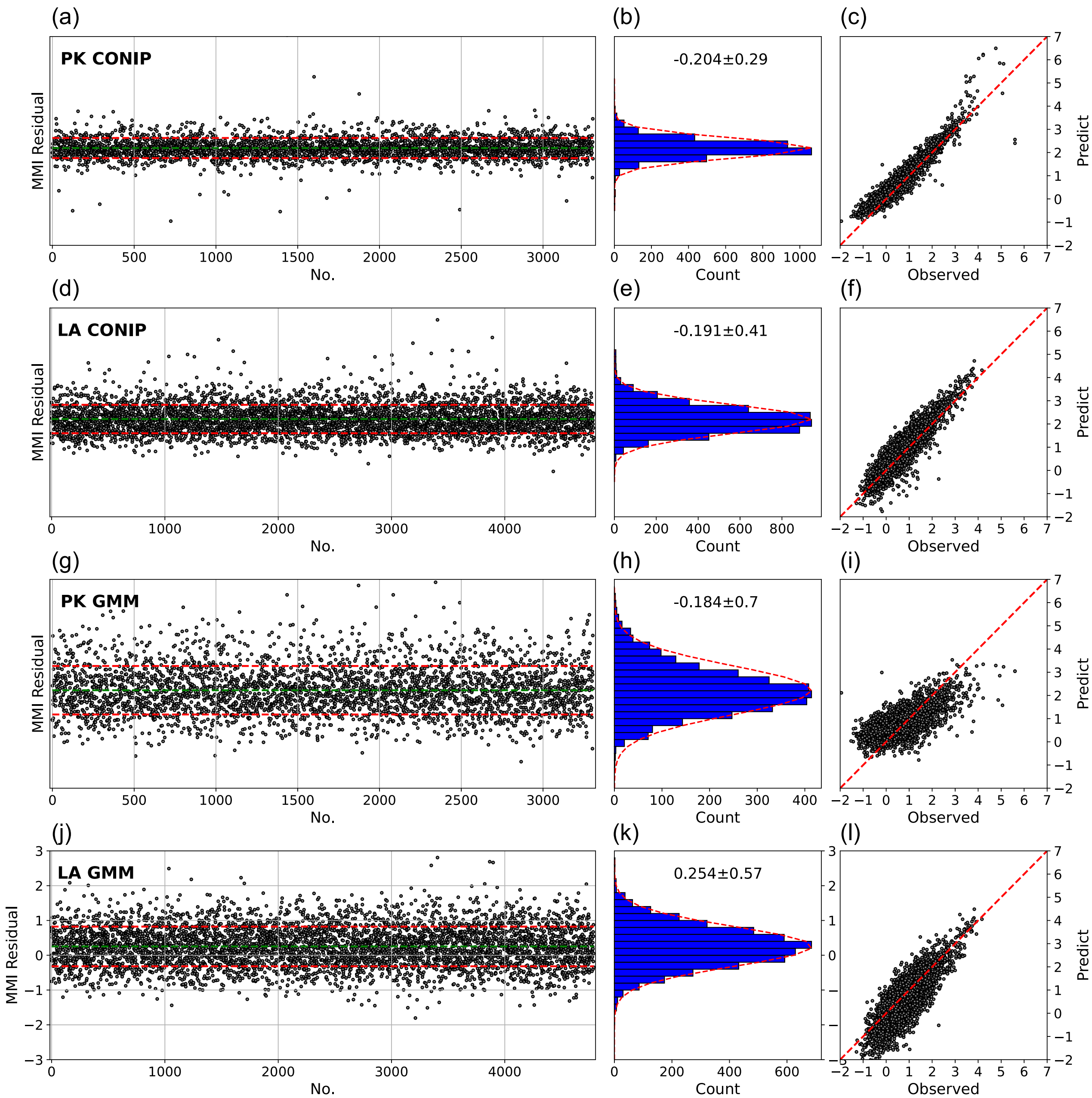

Figure 5. MMI prediction performance. Panels (a) and (c) are prediction residual (observe minus predict) of validation data sets for PK and LA model separately, (e) and (g) are prediction residual by using ground motion prediction equation (GMM) for PK and LA on validation data sets separately. The right-hand panel shows the histogram of each scatter.

Because we compare two different types of models and do not regress the observations with a functional form (i.e. we are not trying to build a ground motion model), we use the total residual (δ_{ij}) as data. The residual for any event-station pair (δ_{ij}) can be represented as a combination of a bias term a for the entire set of residuals, a correction or term for each event i (δE_i), a correction term for each site j (δS_j) and a remaining term (δWS_{ij}) that represents the remaining uncertainty.

$$\delta_{ij} = a + \delta E_i + \delta S_j + \delta WS_{ij}. \quad (2)$$

Here, the bias term a is the only fixed effect. We solve for 'event' (δE_i) and 'site' (δS_j) terms as random effects. The remaining residual, δWS_{ij}, is generally considered a combination of the path correction for the ith earthquake recorded at the jth station (δP_{ij}), as well as remaining aleatory uncertainty (δW^0_{ij}):

$$\delta WS_{ij} = \delta P_{ij} + \delta W^0_{ij}. \quad (3)$$

We do not attempt to separate these latter two terms, but rather assume that most of δWS_{ij}, represents effects along the path, and take this to be our 'path term'.

It is worth noting that in the case of the traditional EEW approach, the GMM residual population here is compared to the BSSA14 model as described in the above section. This model was developed on events $M > 3$, but we extrapolate do $M2$ for our data set. While not ideal, we require this magnitude range to develop our CNN, and desire to perform a direct comparison between our residuals, and those of the more traditional approach.

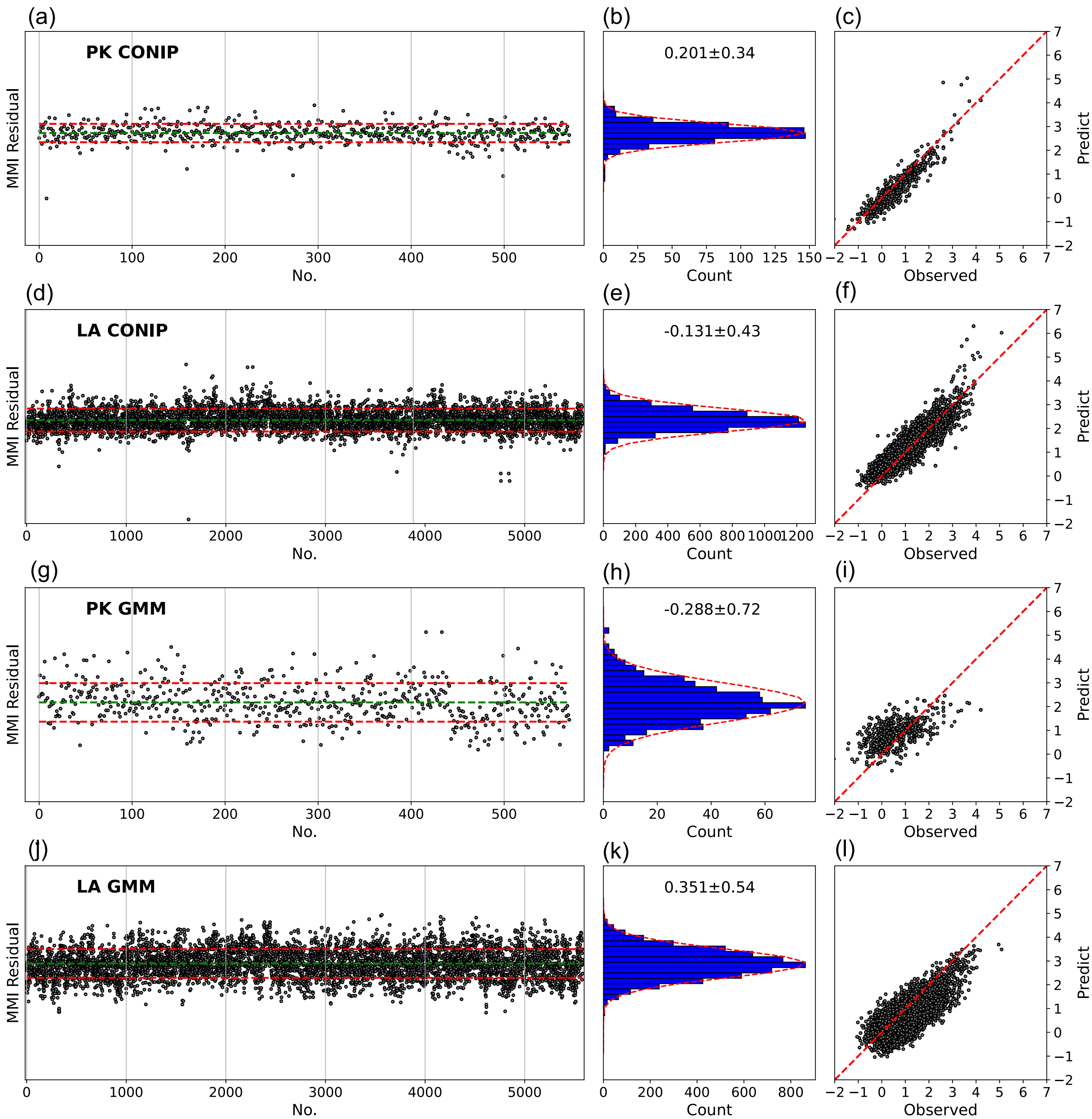

Figure 6. MMI prediction residual for unseen data (January to June 2020) in two regions. Panels (a) and (c) are prediction residuals (observe minus predict) when the trained model is applied to unseen data for PK and LA separately, and (e) and (g) are prediction residuals by using the GMM-based method. The right-hand panel shows a histogram of each population.

4 RESULTS

4.1 Model prediction performance on testing data sets

After training is complete, we verify the MMI prediction accuracy of the model by using the testing data sets. In Figs 5(a)–(d), we plot the MMI prediction residual (observed minus predicted) for both data sets. To make comparisons with traditional methods (e.g. τ_c and P_d methods), only TW = 3 s was analysed in our paper. Overall, MMI prediction residuals for CONIP for both LA and PK are relatively small, although they retain a systematic but very slight overestimation for both data sets. The mean and standard deviation of the total prediction residual for each test data set is –0.204 ± 0.29 and –0.191 ± 0.41 separately. Figs 5(e)–(h) shows the same analysis but for the idealized on-site GMM-based MMI prediction approach. Recall that here we assume that the event location and magnitude are the same as in the earthquake catalogue. The results show much higher residuals with means and standard deviations of –0.184 ± 0.70 and 0.254 ± 0.57 for the PK and LA data sets. In addition, we observed a few outliers in Figs 5(a) and (c). After verifying our data set, we find all of them are caused by instrumental artefacts in the waveforms. Since these account for a very small part of the data set, we did not deliberately remove them. In the supplemental material, we provide a full list of the MMI

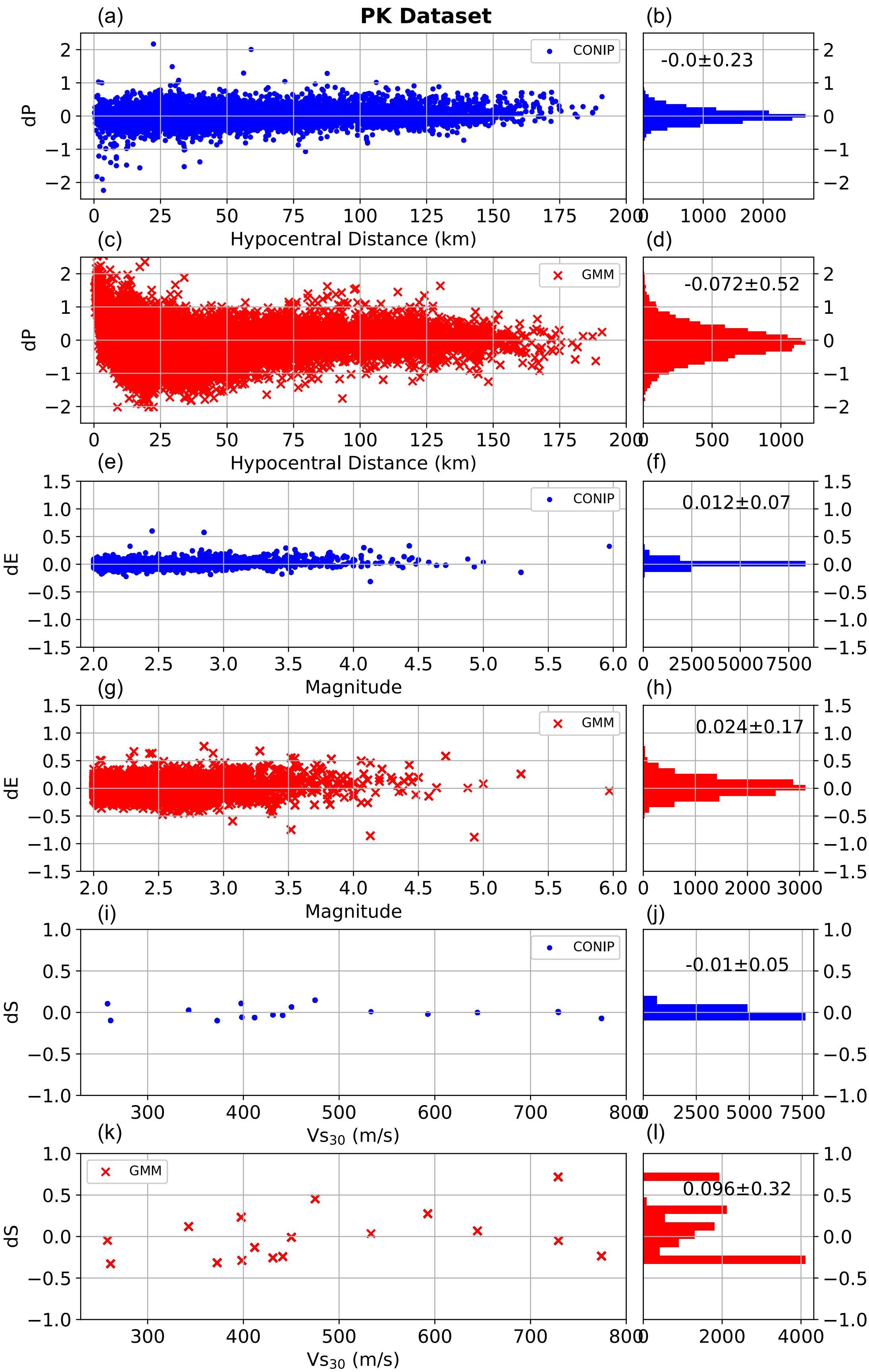

Figure 7. Path (δWS_{ij}), source (δE_i) and site (δS_j) component in MMI prediction residuals for PK data set. Panels (a), (e) and (i) scatter plots for our CONIP method and (c), (g) and (k) are scatter plots for the GMM method, the right-hand panel shows the histogram of each scatter.

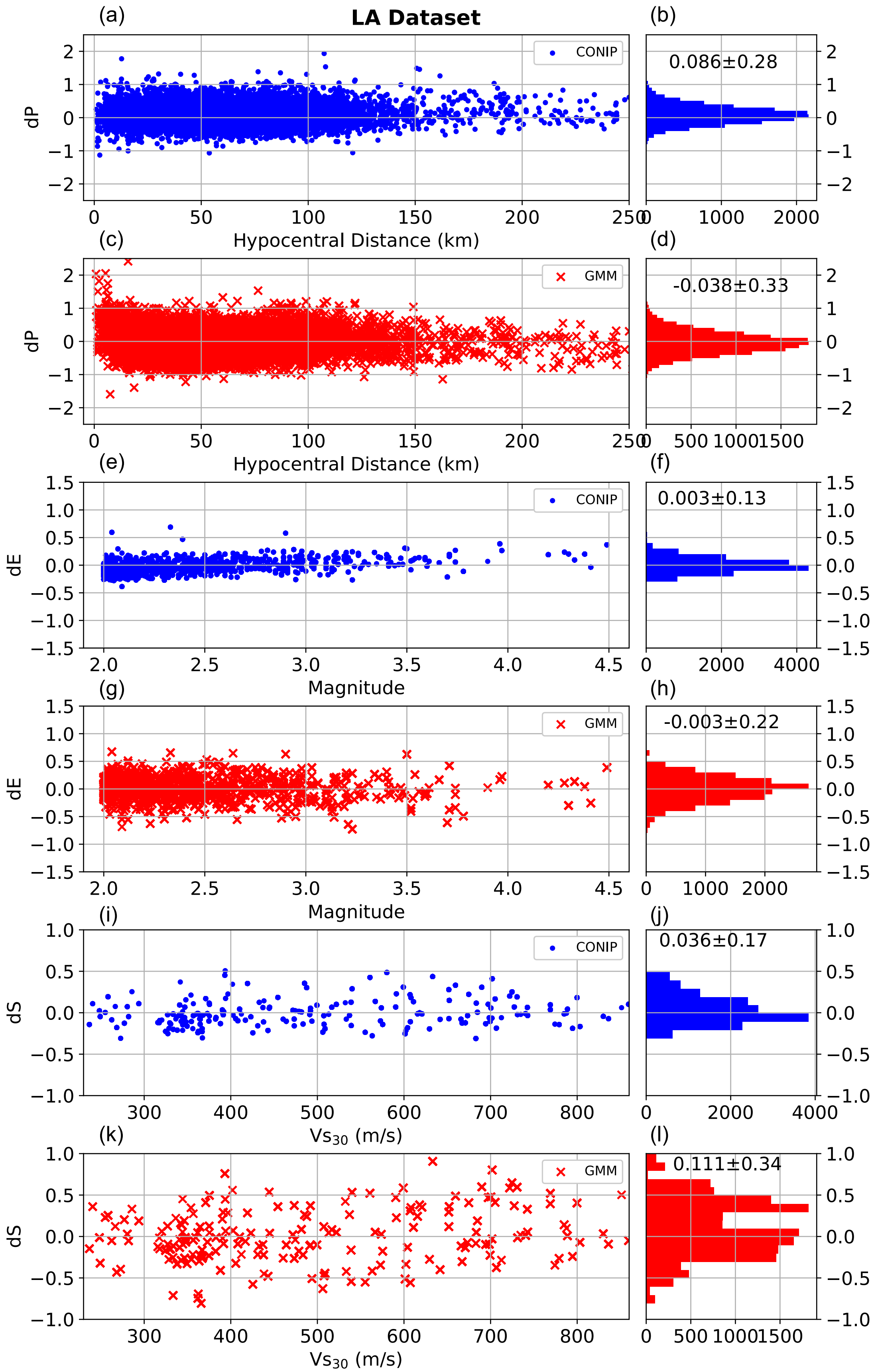

Figure 8. Similar to Fig. 7, but for the LA data set.

prediction for using GMM and CONIP in both PK and LA regions. We upload the MMI prediction result for each station–event pair to the supplemental material.

4.2 Model prediction performance on unseen records

To further demonstrate and verify the practicality and reliability of the trained model, we download more recent events from IRIS (January to June 2020), and we use the same records selection criteria ($M > 2.0$, SNR > 3, and broad-band seismograph only) and same data pre-processing process to build two new testing data sets. After that, we test the CONIP prediction performance. As shown in Figs 6(a)–(d), with these completely new records that were never seen during the model training, the model is still capable of quite accurately estimating MMI by using the first 3 s waveform, and MMI estimation residuals are 0.106 ± 0.36 and –0.131 ± 0.43, respectively. But for the GMM-based seismic intensity prediction method (Figs 6e–h), the MMI estimation residuals are –0.288 ± 0.72 and 0.351 ± 0.54, respectively. The MMI prediction of each unseen record is also uploaded to the supplemental material.

4.3 Residual decomposition

Using the mixed-effects approach outlined above we decomposed the MMI prediction residuals into three categories, path effects, d*P*, source effects, d*E* and site effects, d*S*, for both regions and for both CONIP and the idealized on-site approach (Figs 7 and 8). Compared with the GMM-based method [subfigure (c), (g) and (k) in Figs 7 and 8], CONIP clearly shows much better performance. We observe significantly less scatter in the residuals whether in the path, the source, or the site aspects. We also observe an unbiased residual distribution in both data sets, which is a very important feature of the CNN approach as this will also result in a more robust ground motion prediction performance when applied to actual processing systems.

Fig. 7 shows that in PK data set the scatter for CONIP, compared to the idealized approach, is two times smaller for d*P*, for d*E* the scatter is three times smaller, and for d*S* the scatter is eight times smaller. The mean of the MMI prediction residuals for each contribution are 0.0 ± 0.23 (d*P*), 0.012 ± 0.07 (d*E*) and –0.01 ± 0.05 (d*S*) for our CONIP, and 0.0 ± 0.46, 0.056 ± 0.22 and 0.097 ± 0.41 for the GMM-based method correspondingly. And in the LA data set (Fig. 8), the scatter in d*P* is 1.2 times smaller, for d*E* it is two times smaller and for d*S* is three times smaller for CONIP. The mean MMI prediction residual for each aspect is 0.0 ± 0.27 (d*P*), 0.018 ± 0.11 (d*E*) and 0.008 ± 0.11 (d*S*) for our CNN method, and 0.0 ± 0.33, –0.009 ± 0.22 and 0.091 ± 0.35 for the GMM-based method correspondingly.

5 DISCUSSION

5.1 Improvements over traditional on-site algorithms

In comparison to typical on-site early warning methods, the CNN-based approach does not need to know the hypocentre or the magnitude of the earthquake to forecast shaking intensity (Böse *et al.* 2012a). This leads to significant improvements in the method's ability to generate more accurate seismic intensity estimation (Fig. 5) with the same limited waveform. From an applications perspective, the main difference is we estimate MMI directly, rather than obtaining seismic intensity estimation based on source parameters and a GMM, thus error propagation (from the source parameters to the GMM) will be significantly reduced.

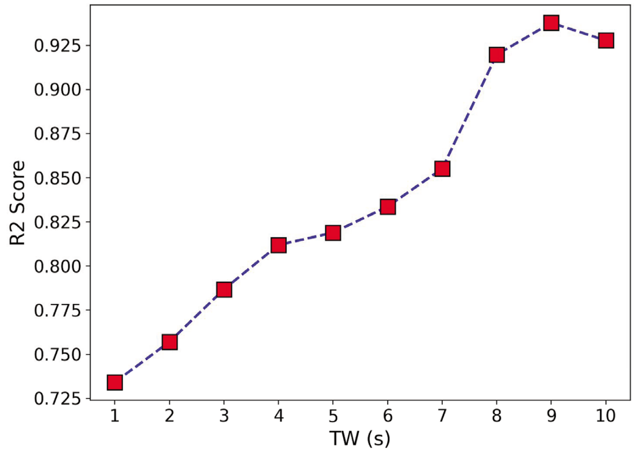

Figure 9. R^2 score changes with time windows (TW) grow. R^2 score gradually increased with TW expansion, which represents the model fit better when more waveform were available.

In this paper, we graded the performance of the algorithms by studying the MMI estimation residuals (Fig. 5). For the ideal on-site algorithm case, we used the BSSA14 GMM relationship (Fig. 5) proposed by Boore *et al.* (2014), This is one of the NGA-West2 GMMs (Bozorgnia *et al.* 2014) and is specifically designed for use in California. When using the GMM we assume the earthquake magnitude and hypocentre from the COMCAT catalogue and as such that they are already very well constrained. We also obtained Vs30 for each site by using the terrain-slope proxy approach and a high-quality digital elevation model (DEM), specific Vs30 values for each station can be found in the supplemental material. As shown in this figure (Fig. 5), even in this ideal situation, we observed a significantly better MMI prediction performance from CONIP over the GMM-based method. The mean residuals for these two approaches are –0.184 ± 0.70 and 0.254 ± 0.57, respectively. In actual real-time performance, a traditional on-site algorithm would indubitably perform worse. There would be magnitude and location errors resulting in a larger scatter in the MMI estimation.

This improvement is to be expected. It is well known that GMMs are usually generated by using data at multiple stations during different events and in various source regions, thus it is a regional or even global average. In contrast, as the CNN model, we present here is trained based on local event data, it learns more local features leading to better MMI estimation. It is possible that a GMM that was estimated specifically from the same restricted region produces a smaller prediction residual. However, as we discuss in the next section because the vast majority of GMMs are still obtained assuming ergodicity, it is still unlikely that a regional model would perform better than our CNN-based method.

CONIP can also provide progressive MMI prediction when new data become available. This can better meet the actual demands of a system during the EEW process. The on-site MMI prediction starts as soon as the *P* wave is detected, we then update the prediction second by second until the peak value is reached. Theoretically, with the TW increase, MMI estimation accuracy from the CNN model should also increase. To exemplify this accuracy, we calculate the R^2 score, which represents the proportion of the variance for a dependent variable that is explained by an independent variable, or variables, in a regression model. The R^2 score is between 0 and 1, where 0 means the model was very poorly fit and 1 means a perfect

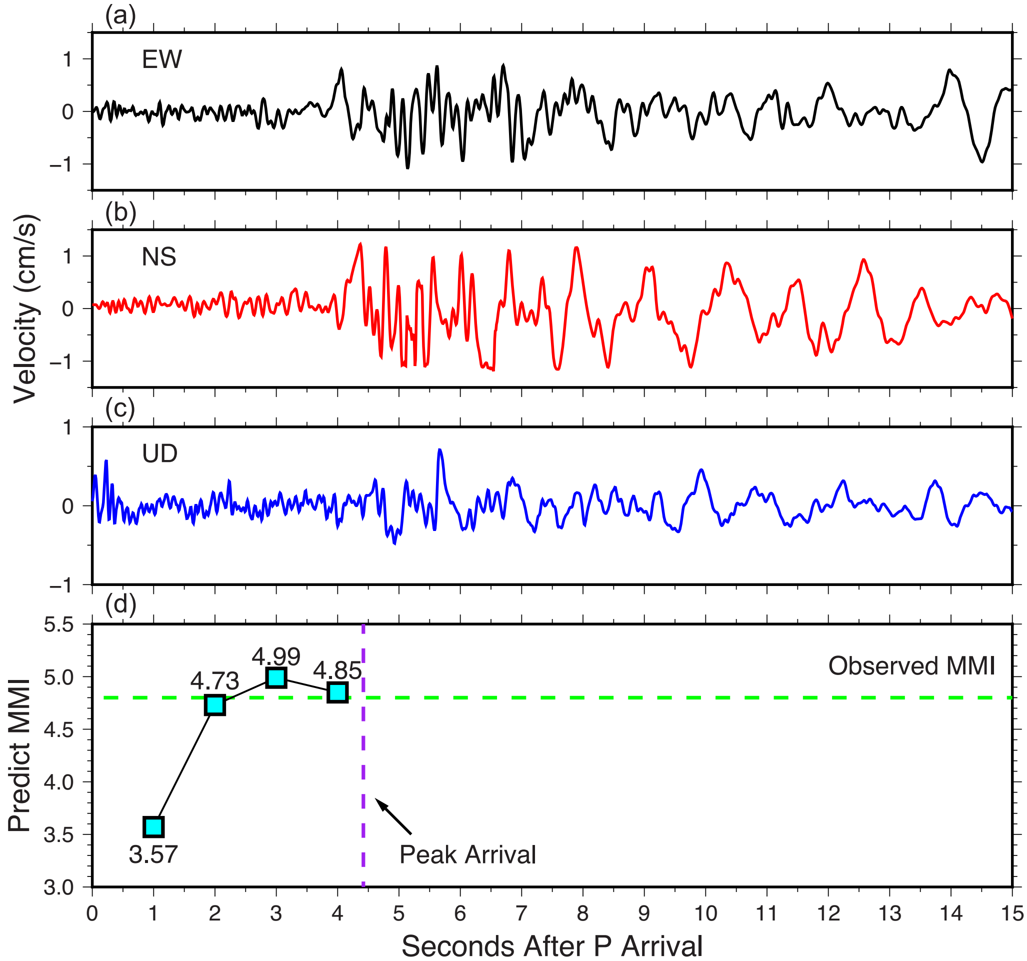

Figure 10. On-site MMI real-time prediction for EEW. Station code: NC.BBGB, event ID: 73 292 360, catalogued magnitude: 4.71, focal depth: 9.7 km and hypocentre distance: 22.4 km. Panels (a), (b) and (c) are the three-channel velocity history and (d) shows the predicted MMI changes with TW.

fit. As shown in Fig. 9, even with only the first second waveform the R^2 score is around 0.73, and gradually increased with TW, which is also consistent with our expectations. In Fig. 10, we randomly select a record to show the potential real-time MMI prediction performance of CONIP. As can be seen from this figure, with just the first few seconds' waveforms, we are able to predict the MMI value with very good accuracy. For the first 4 s, the progressive MMI prediction is 3.57, 4.73, 4.99 and 4.85, respectively while the observed MMI is 4.82. Note that we just provide MMI prediction for the first 4 s and PGV of these record arrivals at 4.48 s. Although this leading time is quite small, as the MMI prediction is very accurate, CONIP will be still useful for part users. For example, factories can automatically stop the power supply to avoid more losses caused by the earthquake, high speed trains can automatically activate emergency braking procedures and slow down their speed to prevent derailment.

5.2 What has the model learned?

As noted, the CNN model far outperforms the traditional on-site algorithms even in the ideal situation when magnitude and hypocentre are very well known. It is evident that CONIP extracted and learned more useful features that lead to more accurate MMI estimation under the same situation.

To understand what the model learned during the training we applied the MMI prediction residual decomposition (Figs 7 and 8; Sahakian *et al.* 2018; Baltay *et al.* 2017). Fig. 7 shows the comparison between path, source, and site residuals for the PK model. The improvement in all three is dramatic (between a factor of 3 and a factor of 8). Figs 11, 12 and 13 show the geographic distribution of each of these terms. In the case of the source residuals, in the GMM approach, there are systematic clusters of events that have either high or low residuals, indicative of some particularities of the source processes for those cluster (such as elevated or lowered stress drops). Though some of these same patterns are still somewhat recognizable in the CONIP residuals, many of them have disappeared. This indicates that CONIP has learned to recognize these effects from the waveform data. Likewise, the site effects are greatly reduced, and sites that show significant site effects (despite the GMM including Vs30) have now much smaller residuals. Once more CONIP has learned that some sites systematically amplify or dampen shaking intensity.

More impressive is the factor of 2 reductions in the path effect residuals for PK. In Fig. 13, we have binned the model domain in 5 km cells and connected each site and path via a straight line. We then obtain the mean value of dP, the path residual, inside of each cell from all the rays that cross it. This is a first-order approximation to try to understand which paths systematically amplify or dampen intensity. We only plot cells with more than 50 rays crossing it. The

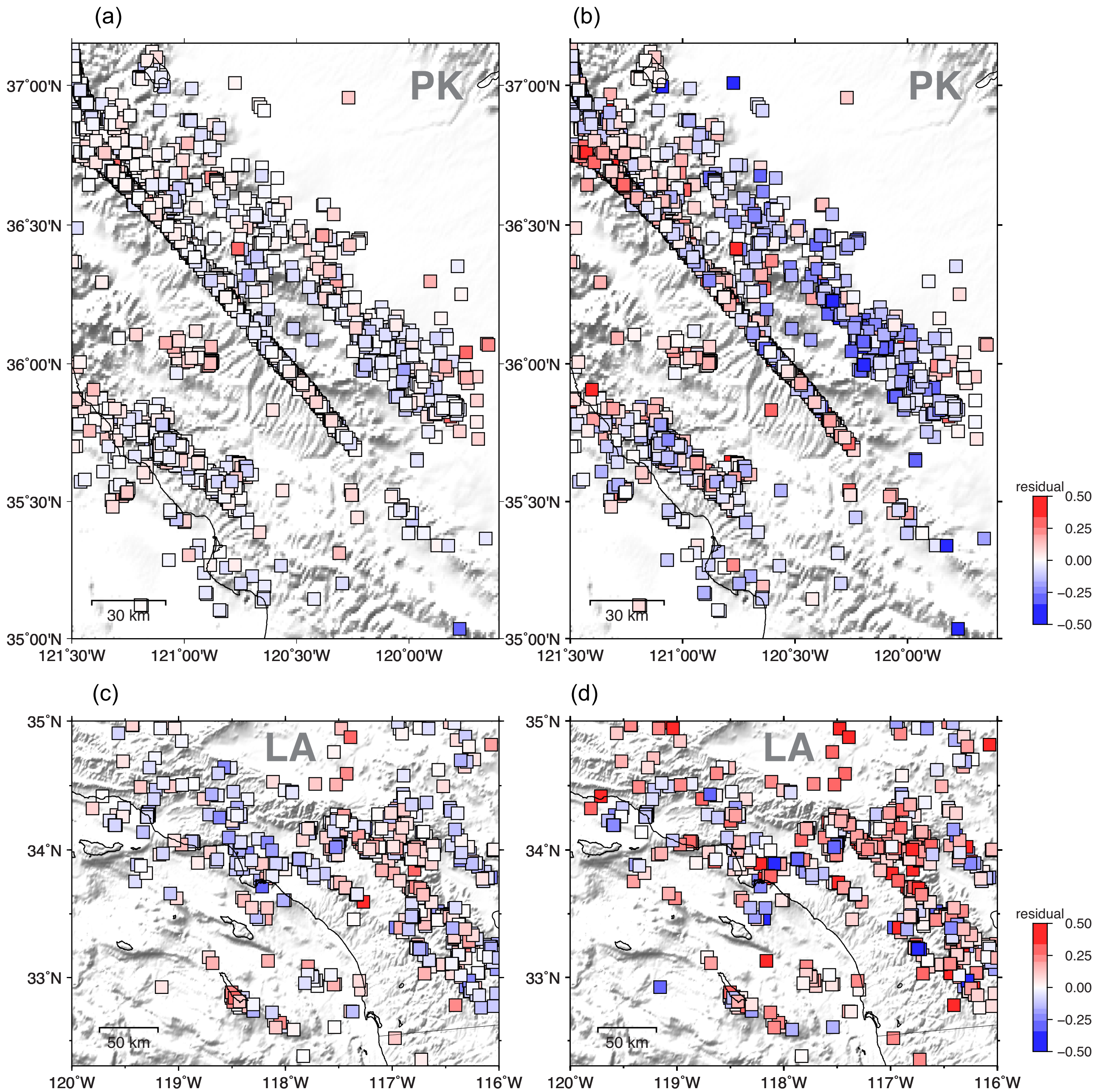

Figure 11. Comparison of source component d*E* in the MMI prediction residuals for PK [(a) and (b)] and LA [(c) and (d)] separately. Panels (a) and (c) are for our CONIP method, (b) and (d) for GMM-based method.

figure reveals that, for example, paths that cross the San Andreas fault have systematically lower residuals; the ground motions are lower than expected. This likely shows that the fault damage zone dampens energy in the frequency range of the intensity measures studied here, perhaps due to lower *Q* due to a reduced shear modulus (Li *et al.* 2006; Cochran *et al.* 2009), or from scattering attenuation induced by heterogeneous velocity structure (Sahakian *et al.* 2019). Likewise, in the mountains either side of the fault, lower attenuation leads to paths with systematically higher residuals. In CONIP, while these patterns are still faintly seen, they have much smaller values. In particular, the effect of the SAF dampening the intensities disappears almost completely.

The results for the LA model are similar (Fig. 8) with significant reductions in the source effects (Fig. 11) and site effects (Fig. 12), in particular compared to the site residuals of Parker *et al.* (2020). However, the reduction in the path effect residual is more modest. The scatter is reduced from 0.33 in the GMM approach to only 0.27 in CONIP. This is also apparent in the geographic distribution of the path residuals (Fig. 13). For example, the elevated residuals for paths that cross the Transverse Ranges and the San Bernardino Mountains are lower, yes, but remain elevated in the CONIP results. In the LA model most of the improvement comes from the source and site effects being much smaller. This is likely due to the dimensions of the geographic domain and the number of rays that cross each of

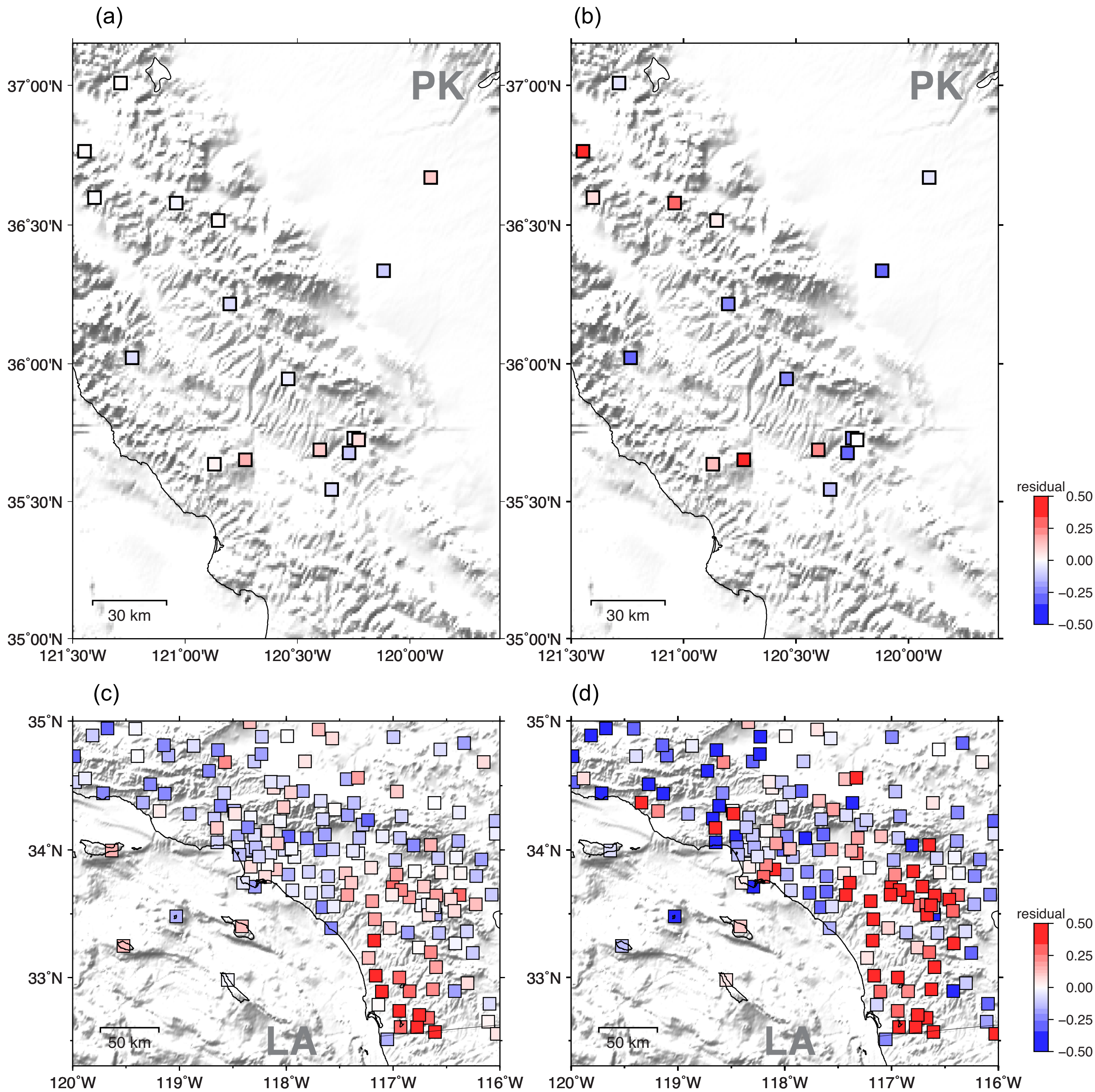

Figure 12. Comparison of site component d*S* in the MMI prediction residuals for PK [(a) and (b)] and LA [(c) and (d)] separately. panels (a) and (c) are for our CONIP method, (b) and (d) for GMM-based method.

the cells. Figs 13(c) and (f) also shows that in the smaller PK model we have on average 265 rays crossing any given cell. Meanwhile, in LA this number drops to 131. It is likely that there isn't enough data in the bigger LA Model region to make a more substantial improvement in the path effects.

Despite this, CONIP still far outperforms the idealized GMM approach. The reduction in the site and source effects is still meaningful enough that it leads to an almost factor of 3 improvement (Fig. 8) in the overall performance in the LA model. Similarly, the results also show that even when the network is relatively sparse, as in the PK model, as long as there are enough events for training the performance can also be significantly improved. These results agree with the findings of a side-by-side comparison of ML ground-motion estimation and traditional empirical GMMs (Klimasewski *et al.* 2021). Though less sophisticated, this work found that artificial neural nets (ANN) seem to learn regional site and path behaviour, providing more site-specific estimates of ground motion.

5.3 Present limitations of the approach

As shown in our paper, the CNN-based intensity prediction method proposed is very reliable, and our study shows that advanced ML algorithms have good prospects for application in the field of EEW. However, there remain numerous improvements to be made. As

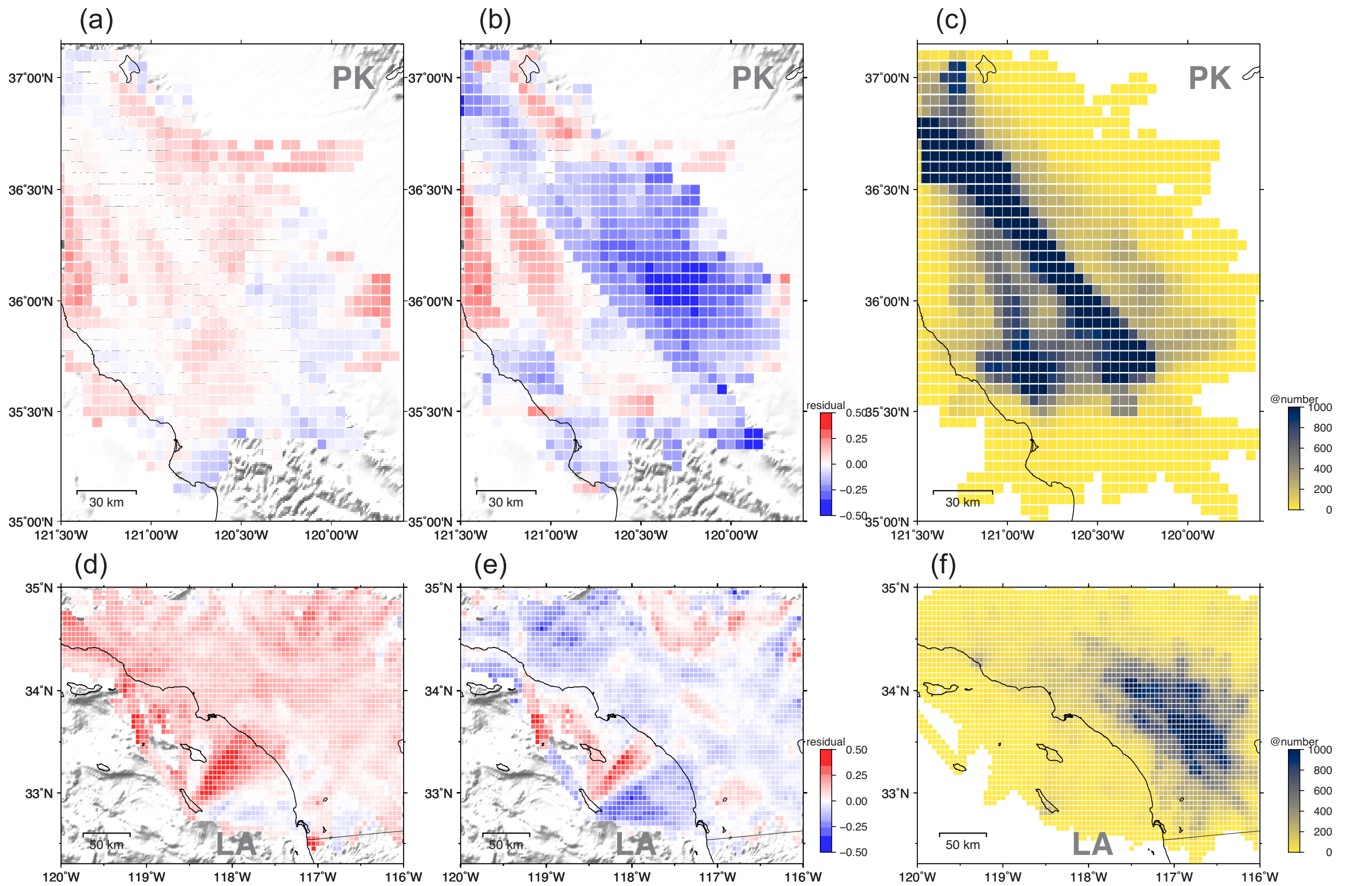

Figure 13. Comparison of path component dP in the MMI prediction residuals and for PK [(a) and (b)] and LA [(d) and (e)] separately. panels (a) and (d) are for CNN-based method, (b) and (e) for our CONIP method. Panels (c) and (f) are contour maps of the number of ray traces for path effect decomposite analysis for PK (c) and LA (e) data set separately.

with most on-site early warning algorithms, for sequences of events especially during aftershocks that follow a bigger main shock, our method may not be able to perform well. The premise behind the CONIP approach is that the input (limited) waveform is the initial stage of the earthquake. From this, CONIP will predict the seismic intensity. When multiple events occur simultaneously in a short period of time, it will be difficult to distinguish the coda of one event from the initial *P* waves of another. Training with more complex data sets can potentially ameliorate this.

Readers may have noted that we set the borders of the two study regions somewhat arbitrarily. For a truly operational setting considering major geological boundaries when training regional algorithms would likely be warranted. One of the research goals of the study is to assess whether the ML models trained from these two adjacent regions share a certain consistency, and what factors lead to the differences. We set the dimensions of the LA region almost two times larger than PK, as we also want to find if the ML model is still robust in different spatial ranges. Our findings are that after choosing a different scope, the trained models are very different so for operations, this is certainly an issue that will warrant careful thought. Additionally, larger domains will have more varied path effects and so will require larger training data sets (as is the case with the LA domain).

Perhaps the largest current limitation of the approach we show here lays in the magnitude (and intensity) of the training data. We are limited by actual available historical events in the catalogues and data sets. In this work, most of the events that we used during the CNN training are smaller magnitude events and moderate intensities, with most of the stations at local to regional distances from the epicentre. Therefore, as currently formulated whether the algorithm will generalize well to high-intensity records is unknown and perhaps unlikely. Additionally, it is by nature a region-dependent algorithm that learns the specifics of source, site, and paths in said region. Its performance outside of the training region will also likely be limited. To expand the algorithms' reliability to higher intensities and larger magnitude events it is likely that simulation will play a big role. Fully deterministic modelling of ground motion for large ($\sim M7$) continental earthquakes is still a computationally very intensive task but it is becoming possible to model out to frequencies as high as 10 Hz (Rodgers *et al.* 2020). Future methodological improvements and GPU computing promise to make these approaches more accessible. Semi-stochastic methods for large earthquakes (Frankel *et al.* 2018) also continue to improve and can now include realistic *P* waves in addition to the strong-shaking component (Goldberg & Melgar 2020), these could also be potentially used to generate synthetic waveforms for training. Whatever the solution, the results we detail here are an initial step, we firmly believe that ML approaches will generate more accurate MMI estimations for other regions.

As we mentioned repeatedly in our paper, the CNN-based method is just a single station on-site approach, which is easier to implement. As currently coded, it cannot benefit from valuable information provided by other sites in the network also measuring the ground motion wavefield. Future improvements will include a network-based CNN.

6 CONCLUSION

Based on advances in ML algorithms we constructed a four layers CNN model to predict on-site seismic intensity in real-time. To test the approach, we built two separate data sets and demonstrated good performance in both. Specifically, when compared with an ideal on-site algorithm that uses the best catalogue information available for each earthquake, our CNN-based method produces forecasts of seismic intensity that are much more accurate and from only limited portions of the *P* wave part of the waveform. We decomposed the MMI prediction residuals for both the CNN approach and the ideal on-site algorithm. The results show that the CNN model easily learns site and source effects. It also learns path effects, but this is a harder task that depends largely on having enough sources to station pairs that sample many potential paths in a given region. Overall, this makes up the shortcoming of the GMM-based seismic intensity estimation methods which assume ergodicity of ground motions. Our method is currently only applicable to small and moderate intensities since that is what is most commonly available in the catalogue of events, but we argue it could in future work be supplemented with simulations. In short, while this is a proof of concept, we believe the research shows that ML-based forecasting of shaking intensities has good prospects for success in future EEW systems.

ACKNOWLEDGMENTS

We thank two anonymous reviews and the editor, Andrew Valentine, for many helpful suggestions for improving and clarifying this work. The support provided by the China Scholarship Council (CSC) during a visit of Zhang Hongcai (201904190007) to the University of Oregon is acknowledged. This research was supported by the Science for Earthquake Resilience project (XH20028Y).

DATA AVAILABILITY

All the seismographs can be downloaded online (http://ds.iris.edu, last accessed August 2021), Figures in this article were prepared using the Generic Mapping Tools (GMT, http://gmt.soest.hawaii.edu) and Matplotlib (https://matplotlib.org). We use the Vs30 model provided by USGS (https://earthquake.usgs.gov/data/vs30, last accessed August 2021). Our CNN model and python code are available at https://github.com/ZhangHCFJEA/GroundMotionCNN. git, and our training data sets are available at https://zenodo.org/record/6626709#.YqG-LHZByMo.

CONFLICT OF INTEREST

The authors declare no conflicts of interest relevant to this study.

REFERENCES

Abrahamson, N.A. & Youngs, R.R., 1992. A stable algorithm for regression analyses using the random effects model, *Bull. seism. Soc. Am.,* **82,** 505–510.

Alcik, H., Ozel, O., Apaydin, N. & Erdik, M., 2009. A study on warning algorithms for Istanbul earthquake early warning system, *Geophys. Res. Lett.,* **36,** doi:10.1029/2 008GL036659.

Alimoradi, A. & Beck, J.L., 2015. Machine-learning methods for earthquake ground motion analysis and simulation, *J. Eng. Mech.,* **141**(4), doi:10.1061/(ASCE)EM.1943-7889.0000869.

Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces, *Bull. seism. Soc. Am.,* **68,** 1521–1532

Allen, R.M. & Melgar, D., 2019. Earthquake early warning: advances, scientific challenges, and societal needs, *Annu. Rev. Earth planet. Sci.,* **47,** 361–388.

Allen, R.M., Gasparini, P., Kamigaichi, O. & Böse, M., 2009. The status of earthquake early warning around the world: an introductory overview, *Seismol. Res. Lett.,* **80,** 682–693.

Allen, T.I. & Wald, D.J., 2009. On the use of high-resolution topographic data as a proxy for seismic site conditions (V_S30), *Bull. seism. Soc. Am.,* **99,** 935–943

Anderson, J.G. & Brune, J.N., 1999. Probabilistic seismic hazard analysis without the ergodic assumption, *Seismol. Res. Lett.,* **70,** 19–28.

Asencio-Cortés, G., Martínez-Álvarez, F., Troncoso, A. & Morales-Esteban, A., 2017. Medium–large earthquake magnitude prediction in Tokyo with artificial neural networks, *Neural Comput. Applicat.,* **28,** 1043–1055.

Asencio-Cortés, G., Morales-Esteban, A., Shang, X. & Martínez–Álvarez, F., 2018. Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure, *Comput. Geosci.,* **115,** 198–210

Asim, K.M., Martínez-Álvarez, F., Basit, A. & Iqbal, T., 2017. Earthquake magnitude prediction in Hindukush region using machine learning techniques, *Nat. Hazards,* **85,** 471–486.

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,** 794–801

Baltay, A.S., Hanks, T.C. & Abrahamson, N.A., 2017. Uncertainty, variability, and earthquake physics in ground-motion prediction equations, *Bull. seism. Soc. Am.,* **107,** 1754–1772.

Bergen, K.J. & Beroza, G.C., 2018. Detecting earthquakes over a seismic network using single-station similarity measures, *Geophys. J. Int.,* **213,** 1984–1998.

Boore, D.M., Stewart, J.P., Seyhan, E. & Atkinson, G.M., 2014. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes, *Earthq. Spectra,* **30,** 1057–1085

Bozorgnia, Y. *et al.*, 2014. NGA-West2 research project. *Earthq. Spectra,* **30**(3), 973–987.

Böse, M., Wenzel, F. & Erdik, M., 2008. PreSEIS: a neural network-based approach to earthquake early warning for finite faults, *Bull. seism. Soc. Am.,* **98,** 366–382.

Böse, M., Heaton, T.H. & Hauksson, E., 2012a. Rapid estimation of earthquake source and ground-motion parameters for earthquake early warning using data from a single three-component broadband or strong-motion sensor, *Bull. seism. Soc. Am.,* **102,** 738–750.

Böse, M., Heaton, T.H. & Hauksson, E., 2012b. Real-time finite fault rupture detector (FinDer) for large earthquakes, *Geophys. J. Int.,* **191,** 803–812.

Böse, M., Graves, R.W., Gill, D., Callaghan, S. & Maechling, P.J., 2014. CyberShake-derived ground-motion prediction models for the Los Angeles region with application to earthquake early warning, *Geophys. J. Int.,* **198,** 1438–1457.

Brillinger, D.R. & Preisler, H.K., 1984. An exploratory analysis of the Joyner-Boore attenuation data, *Bull. seism. Soc. Am.,* **74,** 1441–1450.

Brillinger, D.R. & Preisler, H.K., 1985. Further analysis of the Joyner-Boore attenuation data, *Bull. seism. Soc. Am.,* **75,** 611–614.

Chen, Y., 2020. Automatic microseismic event picking via unsupervised machine learning, *Geophys. J. Int.,* **222,** 1750–1764 .

Cochran, E.S., Bunn, J., Minson, S.E., Baltay, A.S., Kilb, D.L., Kodera, Y. & Hoshiba, M., 2019. Event detection performance of the plum earthquake early warning algorithm in southern California, *Bull. seism. Soc. Am.,* **109,** 1524–1541.

Cochran, E.S., Li, Y.G., Shearer, P.M., Barbot, S., Fialko, Y. & Vidale, J.E., 2009. Seismic and geodetic evidence for extensive, long-lived fault damage zones, *Geology,* **37,** 315–318

Cuéllar, A. *et al.*, 2014. The Mexican Seismic Alert System (SASMEX): its alert signals, broadcast results and performance during the M7.4 Punta Maldonado earthquake of March 20th, 2012, in *Early Warning for Geological Disasters, Advanced Technologies in Earth Sciences,* Chapter 4, pp. 71–87, eds Wenzel, F. & Zschau, J., Springer-Verlag.

Jozinović, D., Lomax, A., Štajduhar, I. & Michelini, A., 2020. Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network, *Geophys. J. Int.,* **222**(2), 1379–1389.

Jozinović, D., Lomax, A., Štajduhar, I. & Michelini, A., 2022. Transfer learning: improving neural network based prediction of earthquake ground shaking for an area with insufficient training data, *Geophys. J. Int.,* **229**(1), 704–718.

Dawood, H.M. & Rodriguez-Marek, A., 2013. A method for including path effects in ground-motion prediction equations: an example using the Mw 9.0 Tohoku earthquake aftershocks method for including path effects in GMPEs using Mw 9.0 Tohoku earthquake aftershocks, *Bull. seism. Soc. Am.,* **103,** 1360–1372.

Derras, B., Bard, P.Y., Cotton, F. & Bekkouche, A., 2012. Adapting the neural network approach to PGA prediction: an example based on the KiK-net data, *Bull. seism. Soc. Am.,* **102,** 1446–1461.

Derras, B., Bard, P.Y. & Cotton, F., 2014. Towards fully data driven ground-motion prediction models for Europe, *Bull. Earthq. Eng.,* **12,** 495–516.

DeVries, P.M.R., Víegas, F., Wattenberg, M. & Meade, B.J., 2018. Deep learning of aftershock patterns following large earthquakes, *Nature,* **560,** 632–634.

Fauvel, K. *et al.* 2020. A distributed multi-sensor machine learning approach to earthquake early warning, *Proc. AAAI Conf. Artificial Intellig.,* **34,** 403–411.

Frankel, A., Wirth, E., Marafi, N., Vidale, J. & Stephenson, W., 2018. Broadband synthetic seismograms for magnitude 9 earthquakes on the Cascadia megathrust based on 3D simulations and stochastic synthetics. Part 1: methodology and overall results, *Bull. seism. Soc. Am.,* **108,** 2347–2369

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, Open-File Report 2018-1155, US Geological Survey.

Goldberg, D.E. & Melgar, D., 2020. Generation and validation of broadband synthetic P waves in semistochastic models of large earthquakes, *Bull. seism. Soc. Am.,* **110,** 1982–1995.

Goltz, J.D. & Flores, P.J., 1997. Real-time earthquake early warning and public policy: a report on Mexico City's Sistema de Alerta Sismica, *Seismol. Res. Lett.,* **68,** 727–733.

Hoshiba, M., 2013. Real-time prediction of ground motion by Kirchhoff-Fresnel boundary integral equation method: extended front detection method for earthquake early warning, *J. geophys. Res.,* **118,** 1038–1050.

Hoshiba, M. & Aoki, S., 2015. Numerical shake prediction for earthquake early warning: data assimilation, real-time shake mapping, and simulation of wave propagation, *Bull. seism. Soc. Am.,* **105,** 1324–1338.

Hoshiba, M. & Ozaki, T., 2015. Earthquake early warning and Tsunami Warning of the Japan Meteorological Agency, and their performance in the 2011 off the Pacific Coast of Tohoku Earthquake (M_w9:0), in *Early Warning for Geological Disasters, Advanced Technologies in Earth Sciences,* pp. 1–28, eds Wenzel, F. & Zschau, J., Springer-Verlag.

Hsiao, N.C., Wu, Y.M., Shin, T.C., Li, Z. & Teng, T.L., 2009. Development of earthquake early warning system in Taiwan, *Geophys. Res. Lett.,* **36,** doi:10.1029/2008GL036596.

Kamigaichi, O. *et al.*, 2009. Earthquake early warning in Japan: warning the general public and future prospects, *Seismol. Res. Lett.,* **80,** 717–726.

Käufl, P.J., 2015 Rapid probabilistic source inversion using pattern recognition, *Doctoral dissertation,* University Utrecht.

Khoshnevis, N. & Taborda, R., 2018. Prioritizing ground-motion validation metrics using semisupervised and supervised learning prioritizing ground-motion validation metrics using semisupervised and supervised learning, *Bull. seism. Soc. Am.,* **108,** 2248–2264.

Kilb, D. *et al.*, 2021. The PLUM earthquake early warning algorithm: a retrospective case study of West Coast, USA, data. *J. geophys. Res.,* **126**(7), e2020JB021053, doi:10.1029/2020JB021053.

Klimasewski, A., Sahakian, V.J. & Thomas, A.M., 2021. Comparing performance of artificial neural networks with traditional ground-motion models for small-magnitude earthquakes in Southern California, *Bull. seism. Soc. Am.,* **111,** 1577–1589.

Kodera, Y., Saitou, J., Hayashimoto, N., Adachi, S., Morimoto, M., Nishimae, Y. & Hoshiba, M., 2016. Earthquake early warning for the 2016 Kumamoto earthquake: performance evaluation of the current system and the next-generation methods of the Japan Meteorological Agency, *Earth, Planets Space,* **68,** 68–202.

Kodera, Y., Yamada, Y., Hirano, K., Tamaribuchi, K., Adachi, S., Hayashimoto, N., Morimoto, NM.M. & Hoshiba, M., 2018. The propagation of local undamped motion (PLUM) method: a simple and robust seismic wavefield estimation approach for earthquake early warning, *Bull. seism. Soc. Am.,* **108,** 983–1003.

Kohler, M.D. *et al.*, 2016. Earthquake early warning ShakeAlert system: west coast wide production prototype, *Seismol. Res. Lett.,* **89,** 99–107.

Kong, Q., Allen, R.M. & Schreier, L., 2016. MyShake: initial observations from a global smartphone seismic network, *Geophys. Res. Lett.,* **43,** 9588–9594.

Kong, Q., Allen, R.M., Schreier, L. & Kwon, Y.W., 2016. MyShake: a smartphone seismic network for earthquake early warning and beyond, *Sci. Adv.,* **2**(2), e1501055, doi:10.1126/sciadv.1501055.

Kong, Q., Inbal, A., Allen, R.M., Lv, Q. & Puder, A., 2019. Machine learning aspects of the MyShake global smartphone seismic network, *Seismol. Res. Lett.,* **90,** 546–552.

Kotha, S.R., Bindi, D. & Cotton, F., 2017. From ergodic to region-and site-specific probabilistic seismic hazard assessment: method development and application at European and Middle Eastern sites, *Earthq. Spectra,* **33,** 1433–1453.

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.

Krizhevsky, A., Sutskever, I. & Hinton, G., 2012. ImageNet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems 25 (NIPS 2012),* Vol. **102,** eds Pereira, F., Burges, C.J., Bottou, L. & Weinberger, K.Q., NIPS.

Kuehn, N.M. & Abrahamson, N.A., 2020. Spatial correlations of ground motion for non-ergodic seismic hazard analysis. *Earthq. Eng. Struct. Dyn.,* **49**(1), 4–23.

Kuyuk, H.S. & Motosaka, M., 2009. Real-time ground motion forecasting using front-site waveform data based on artificial neural network, *J. Disaster Res.,* **4,** 260–266.

Landwehr, N., Kuehn, N.M., Scheffer, T. & Abrahamson, N., 2016. A non-ergodic ground-motion model for California with spatially varying coefficients, *Bull. seism. Soc. Am.,* **106,** 2574–2583.

Li, Y.G., Chen, P., Cochran, E.S., Vidale, J.E. & Burdette, T., 2006. Seismic evidence for rock damage and healing on the San Andreas fault associated with the 2004M6.0 Parkfield earthquake, *Bull. seism. Soc. Am.,* **96,** S349–S363.

Li, Z., Meier, M., Hauksson, E., Zhan, Z. & Andrews, J., 2018. Machine learning seismic wave discrimination: application to earthquake early warning, *Geophys. Res. Lett.,* **45,** 4773–4779.

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.

Meier, M., 2017. How "good" are real-time ground motion predictions from earthquake early warning systems? *J. geophys. Res.,* **122,** 5561–5577.

Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović1, D. & Lauciani, V., 2021. INSTANCE–the Italian seismic dataset for machine learning. *Earth Syst. Sci. Data,* **13**(12), 5509–5544.

Ming, Z., Zhuowei, X., Shi, C. & Lihua, F., 2022. DiTing: a large-scale Chinese seismic benchmark dataset for artificial intelligence in seismology. *Earthq. Sci.,* **35,** 1–11.

Minson, S.E., Saunders, J.K., Bunn, J.J., Cochran, E.S., Baltay, A.S., Kilb, D.L., Hoshiba, M. & Kodera, Y., 2020. Real-time performance of the PLUM earthquake early warning method during the 2019 M6.4 and 7.1 Ridgecrest, California, Earthquakes, *Bull. seism. Soc. Am.,* **110,** 1887–1903.

Mousavi, S.M. & Beroza, G.C., 2020a. Bayesian-deep-learning estimation of earthquake location from single-station observations. *IEEE Trans. Geosci. Remote Sens.,* **58**(11), 8211–8224.

Mousavi, S.M. & Beroza, G.C., 2020b. A machine-learning approach for earthquake magnitude estimation, *Geophys. Res. Lett.,* **47**(1), e2019GL085976, doi:10.1029/2019GL085976.

Mousavi, S.M., Sheng, Y., Zhu, W. & Gregory, C.B., 2019. STanford EArthquake Dataset (STEAD): a global data set of seismic signals for AI. *IEEE Access,* **7,** 179464–179476.

Münchmeyer, J., Bindi, D., Leser, U. & Tilmann, F., 2021a. The transformer earthquake alerting model: a new versatile approach to earthquake early warning, *Geophys. J. Int.,* **225**(1), 646–656.

Münchmeyer, J., Bindi, D., Leser, U. & Tilmann, F., 2021b. Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network, *Geophys. J. Int.,* **226**(2), 1086–1104.

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,** 34–41.

Parker, G.A., Baltay, A.S., Rekoske, J. & Thompson, E.M., 2020. Repeatable source, path, and site effects from the 2019 M 7.1 Ridgecrest earthquake sequence. *Bull. seism. Soc. Am.,* **110**(4), 1530–1548.

Peng, C.Y., Yang, J.S., Xue, B., Zhu, X.Y. & Chen, Y., 2014. Exploring the feasibility of earthquake early warning using records of the 2008 Wenchuan earthquake and its aftershocks, *Soil Dyn. Earthq. Eng.,* **57,** 86–93

Peng, C.Y., Yang, J.S., Zheng, Y., Zhu, X.Y., Xu, Z.Q. & Chen, Y., 2017. New τ_c regression relationship derived from all P wave time windows for rapid magnitude estimation, *Geophys. Res. Lett.,* **44,** 1724–1731.

Perol, T., Gharbi, M. & Denolle, M., 2018. Convolutional neural network for earthquake detection and location, *Sci. Adv.,* **4**(2), e1700578, doi:10.1126/sciadv.1700578.

Pinheiro, J.C. & Bates, D.M., 2000. *Statistics and Computing. Mixed-Effects Models in S and S-PLUS,* Springer-Verlag.

Rabin, N., Bregman, Y., Lindenbaum, O., Ben-Horin, Y. & Averbuch, A., 2016. Earthquake-explosion discrimination using diffusion maps, *Geophys. J. Int.,* **207,** 1484–1492.

Rafiei, M.H. & Adeli, H., 2017. NEEWS: a novel earthquake early warning model using neural dynamic classification and neural dynamic optimization, *Soil Dyn. Earthq. Eng.,* **100,** 417–427.

Rodgers, A.J., Pitarka, A., Pankajakshan, R., Sjögreen, B. & Petersson, N.A., 2020. Regional-scale 3D ground-motion simulations of Mw7 earthquakes on the Hayward Fault, Northern California resolving frequencies 0–10 Hz and including site-response corrections, *Bull. seism. Soc. Am.,* **110,** 2862–2881

Ross, Z.E., Trugman, D.T., Hauksson, E. & Shearer, P.M., 2019. Searching for hidden earthquakes in Southern California, *Science,* **364,** 767–771.

Ross, Z.E., Meier, M. & Hauksson, E., 2018a. P wave arrival picking and first-motion polarity determination with deep learning, *J. geophys. Res.,* **123,** 5120–5129.

Ross, Z.E., Meier, M., Hauksson, E. & Heaton, T.H., 2018b. Generalized seismic phase detection with deep learning, *Bull. seism. Soc. Am.,* **108,** 2894–2901.

Ross, Z.E., Ben-Zion, Y., White, M.C. & Vernon, F.L., 2018c. Analysis of earthquake body wave spectra for potency and magnitude values: implications for magnitude scaling relations, *Geophys. J. Int.,* **207,** 1158–1164.

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C.J. & Johnson, P.A., 2017. Machine learning predicts laboratory earthquakes, *Geophys. Res. Lett.,* **44,** 9276–9282.

Sahakian, V.J., Baltay, A., Hanks, T., Buehler, J., Vernon, F., Kilb, D. & Abrahamson, N., 2018. Decomposing leftovers: event, path, and site residuals for a small-magnitude Anza region GMPE decomposing leftovers: event, path, and site residuals for a small-magnitude Anza Region GMPE, *Bull. seism. Soc. Am.,* **108,** 2478–2492

Sahakian, V.J., Baltay, A., Hanks, T.C., Buehler, J., Vernon, F.L., Kilb, D. & Abrahamson, N.A., 2019. Ground motion residuals, path effects, and crustal properties: a pilot study in southern California, *J. geophys. Res.,* **124,** 5738–5753.

Saunders, J.K., Minson, S.E. & Baltay, A.S. 2022. How low should we alert? Quantifying intensity threshold alerting strategies for earthquake early warning in the United States. *Earth's Future,* **10**(3), e2021EF002515, doi:10.1029/2021EF002515.

Spallarossa, D., Kotha, S.R., Picozzi, M., Barani., S. & Bindi., D., 2019. Onsite earthquake early warning: a partially non-ergodic perspective from the site effects point of view, *Geophys. J. Int.,* **216,** 919–934.

Stafford, P.J., 2014. Crossed and nested mixed-effects approaches for enhanced model development and removal of the ergodic assumption in empirical ground-motion models, *Bull. seism. Soc. Am.,* **104,** 702–719

Tamaribuchi, K., Yamada, M. & Wu, S., 2014. A new approach to identify multiple concurrent events for improvement of earthquake early warning, *Zisin,* **67,** 1324–1338.

Trugman, D.T. & Shearer, P.M., 2017. GrowClust: a hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish Springs and Sheldon, Nevada, earthquake sequences, *Seismol. Res. Lett.,* **88,** 379–391.

Villani, M. & Abrahamson, N.A., 2015. Repeatable site and path effects on the ground-motion sigma based on empirical data from southern California and simulated waveforms from the CyberShake platform, *Bull. seism. Soc. Am.,* **105,** 2681–2695.

Walling, M.A., 2009. *Non-Ergodic Probabilistic Seismic Hazard Analysis and Spatial Simulation of Variation in Ground Motion,* University of California.

Wiszniowski, J., Plesiewicz, B.M. & Trojanowski, J., 2014. Application of real time recurrent neural network for detection of small natural earthquakes in Poland, *Acta Geophys.,* **62,** 469–485.

Worden, C.B., Gerstenberger, M.C., Rhoades, D.A. & Wald, D.J., 2012. Probabilistic relationships between ground-motion parameters and modified Mercalli intensity in California, *Bull. seism. Soc. Am.,* **102,** 204–221

Wu, Y.M., Chen, D.Y., Lin, T.L., Hsieh, C.Y., Chin, T.L., Chang, W.Y., Li, W.S. & Ker, S.H., 2013. A high-density seismic network for earthquake early warning in Taiwan based on low-cost sensors, *Seismol. Res. Lett.,* **84,** 1048–1054.

Wu, Y.M. & Kanamori, H., 2005. Experiment on an onsite early warning method for the Taiwan early warning system, *Bull. seism. Soc. Am.,* **95,** 347–353

Wu, Y.M. & Kanamori, H., 2008. Exploring the feasibility of onsite earthquake early warning using close-in records of the 2007 Noto Hanto earthquake, *Earth, Planets Space,* **60,** 155–160

Wu, Y.M., Yen, H.Y., Zhao, L., Huang, B.S. & Liang, W.T., 2006. Magnitude determination using initial P waves: a single station approach, *Geophys. Res. Lett.,* **33**(5), doi:10.1029/2005GL025395.

Wu, Y.M., Kanamori, H., Allen, R.M. & Hauksson, E., 2007. Determination of earthquake early warning parameters, τ_c and P_d, for southern California, *Geophys. J. Int.,* **170,** 711–717

Wu, Y.M., Teng, T.L., Shin, T.C. & Hsiao, N.C., 2003. Relationship between peak ground acceleration, peak ground velocity, and intensity in Taiwan, *Bull. seism. Soc. Am.,* **93,** 386–396

Yamada, M., 2014. Estimation of fault rupture extent using near-source records for earthquake early warning, in *Early Warning for Geological Disasters, Advanced Technologies in Earth Sciences,* pp. 29–47, eds Wenzel, F. & Zschau, J., Springer-Verlag.

Yin, L., Andrews, J. & Heaton, T., 2018. Rapid earthquake discrimination for earthquake early warning: a bayesian probabilistic approach using three-component single-station waveforms and seismicity forecast rapid earthquake discrimination for earthquake early warning, *Bull. seism. Soc. Am.,* **108,** 2054–2067.

Yoon, C.E., O'Reilly, O., Bergen, K.J. & Beroza, G.C., 2015. Earthquake detection through computationally efficient similarity search, *Sci. Adv.,* **1,** e1501057.

Zhang, H.C., Jin, X., Wei, Y.X., Li, J., Kang, L.C., Wang, S.C., Huang, L.Z. & Yu, P.Q., 2016. An earthquake early warning system in Fujian, China, *Bull. seism. Soc. Am.,* **106,** 755–765.

Zhang, J., Zhang, H.J., Chen, E.H., Zheng, Y., Kuang, W.H. & Zhang, X., 2014. Real-time earthquake monitoring using a search engine method, *Nat. Commun.,* **5,** 56–64.

Zhu, W.Q. & Beroza, G.C., 2019. PhaseNet: a deep-neural network-based seismic arrival-time picking method, *Geophys. J. Int.,* **216,** 261–273.

Zollo, A., Lancieri, M. & Nielsen, S., 2006. Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records, *Geophys. Res. Lett.,* **33**(263), doi:10.1029/2006GL027795.

Zollo, A., Amoroso, O., Lancieri, M., Wu, Y.M. & Kanamori, H., 2010. A threshold-based earthquake early warning using dense accelerometer networks, *Geophys. J. Int.,* **183,** 963–974