JGR Solid Earth

RESEARCH ARTICLE
10.1029/2021JB022703

Special Section:
Machine learning for Solid Earth observation, modeling and understanding

Early Warning for Great Earthquakes From Characterization of Crustal Deformation Patterns With Deep Learning

J.-T. Lin[1], **D. Melgar**[1], **A. M. Thomas**[1], **and J. Searcy**[2]

[1]Department of Earth Sciences, University of Oregon, Eugene, OR, USA, [2]Research Advanced Computing Services, University of Oregon, Eugene, OR, USA

Key Points:

- We use synthetic data to train a deep learning model to predict magnitude from crustal deformation patterns in simulated real-time
- The model, M-LARGE, has an accuracy of 99% and accurately estimates the magnitude of five real large events, outperforming other methods
- M-LARGE's rapid and accurate magnitude prediction suggesting that significant warning times are possible during real large earthquakes

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J.-T. Lin,
jiunting@uoregon.edu

Citation:

Lin, J.-T., Melgar, D., Thomas, A. M., & Searcy, J. (2021). Early warning for great earthquakes from characterization of crustal deformation patterns with deep learning. *Journal of Geophysical Research: Solid Earth*, *126*, e2021JB022703. https://doi.org/10.1029/2021JB022703

Received 28 JUN 2021
Accepted 25 SEP 2021

Abstract

Although infrequent, large (Mw7.5+) earthquakes can be extremely damaging and occur on subduction and intraplate faults worldwide. Earthquake early warning (EEW) systems aim to provide advanced warning before strong shaking and tsunami onsets. These systems estimate earthquake magnitude using the early metrics of waveforms, relying on empirical scaling relationships of abundant past events. However, both the rarity and complexity of great events make it challenging to characterize them, and EEW algorithms often underpredict magnitude and the resulting hazards. Here, we propose a model, M-LARGE, that leverages deep learning to characterize crustal deformation patterns of large earthquakes for a specific region in real-time. We demonstrate the algorithm in the Chilean Subduction Zone by training it with more than six million different simulated rupture scenarios recorded on the Chilean GNSS network. M-LARGE performs reliable magnitude estimation on the testing data set with an accuracy of 99%. Furthermore, the model successfully predicts the magnitude of five real Chilean earthquakes that occurred in the last 11 years. These events were damaging, large enough to be recorded by the modern high rate global navigation satellite system instrument, and provide valuable ground truth. M-LARGE tracks the evolution of the source process and can make faster and more accurate magnitude estimation, significantly outperforming other similar EEW algorithms. This is the first demonstration of our approach. Future work toward generalization is outstanding and will include the addition of more training and testing data, interfacing with existing EEW methods, and applying the method to different tectonic settings to explore performance in these regions.

Plain Language Summary

Great earthquakes are infrequent but devastating natural disasters. To mitigate their effects, earthquake early warning (EEW) systems aim to provide advance warning of strong shaking and tsunami. However, many of the most sophisticated EEW algorithms operating globally have a difficult time characterizing large earthquakes quickly and accurately enough to issue a meaningful warning—this is most evident from the failure of EEW during the 2011 M9 Tohoku Oki, Japan earthquake. Here, we propose a model, M-LARGE, that learns earthquake's surface deformation patterns from millions of simulations for a specific region, and then applies it to unseen events in the same region. Our model shows a high accuracy of 99% performing on the testing data set and accurately estimates the magnitude of five real large historical events in Chile. The M-LARGE outperforms currently operating similar EEW algorithms. This is the first demonstration of our approach, we note that for actual EEW operation, future work including the addition of more training and testing data, interacting with existing EEW methods, and testing the method in alternative tectonic settings are necessary prior to real-time operations.

1. Introduction

Following earthquake initiation, most earthquake early warning (EEW) algorithms provide initial hazard predictions based on the character of the first arriving P-waves, which is the earliest information available. However, it is well known that this approach will routinely struggle during large magnitude earthquakes owing to magnitude saturation, or underestimation, a current limitation of such EEW systems. As an example of this, the Japanese EEW system misidentified the 2011 Mw9.0 Tohoku-oki earthquake as only an Mw8.1 for the first hour after rupture (Hoshiba et al., 2011). Saturation occurs for a couple of reasons. First, inertial-based instruments (seismometers) that record earthquakes in the near-field tend to distort large, low-frequency, signals radiated from large earthquakes, making the data unreliable when real-time automatic processing is considered (Bock & Melgar, 2016; Boore & Bommer, 2005; Larson, 2009). Second, methods that directly calculate earthquake magnitude from body waves or surface wave amplitudes focus on particular frequency bands of the source spectrum, which saturate during large (Mw7.5+) events (Geller, 1976). Third, large earthquakes have durations of several minutes and early onset signals (i.e., the first few seconds of waveforms), utilized by EEW systems, might not contain enough information to forecast the final magnitude of large events (Goldberg et al., 2019; Ide, 2019; Meier et al., 2016, 2017; Melgar & Hayes, 2017; Rydelek & Horiuchi, 2006). Magnitude saturation has consequences for downstream applications that rely on rapid magnitude determination, specifically in the 2011 Tohoku-oki case both forecasts of the expected shaking and the tsunami amplitudes were drastically underpredicted (Colombelli et al., 2013; Hoshiba & Ozaki, 2014).

In recent years, a number of EEW algorithms have attempted to perform more accurate and faster magnitude calculations. For example, it is possible to match shaking patterns in real-time to the expected geometric extension of the causative fault (Böse et al., 2012; Hutchison et al., 2020). Another approach is to forego complete characterization of the earthquake, and simply take the observed shaking wavefield at a particular instant in time, and forecast its time evolution into the future (Cochran et al., 2019; Kodera et al., 2018). Furthermore, the advent of widespread high rate global navigation satellite system (HR-GNSS) networks have enabled a new class of EEW algorithms based on measurements of crustal deformation and are particularly well suited to identifying large magnitude earthquakes (Crowell et al., 2013; Grapenthin et al., 2014; Kawamoto et al., 2016; Minson et al., 2014). Noteworthy among these methods is the Geodetic First Approximation of Size and Time (GFAST) algorithm, which is primarily based on the scaling of peak ground displacement (PGD) and is currently operating in the U.S. EEW system for large earthquakes (Crowell et al., 2013, 2016).

More recently, advances in computing power and technologies have enabled the use of machine learning and deep learning algorithms (LeCun et al., 2015). These have also been demonstrated to provide significant improvements in other data-rich seismological applications such as earthquake detection, phase picking, and association (Kong et al., 2019; Mousavi et al., 2020; Perol et al., 2018; Ross et al., 2018; Zhu & Beroza, 2019). Earthquake magnitude estimation is a popular application of deep learning, and multiple studies have demonstrated their fast and accurate magnitude prediction (Lomax et al., 2019; Mousavi & Beroza, 2020; van den Ende & Ampuero, 2020).

Despite the sophistication of these existing algorithms, many of which are employed in some of the most advanced EEW systems worldwide (such as the U.S. and Japan) (Kodera et al., 2020; Murray et al., 2018), each of them has limitations. For example, the seismic wavefield-based approaches overcome saturation at the expense of short warning times, typically of the order of ∼10–20 s (Kodera et al., 2018). Meanwhile, PGD-based approaches avoid saturation but can struggle when earthquakes have very long or unilateral ruptures (Williamson et al., 2020) and can grossly overpredict the magnitudes of these kinds of events. Furthermore, the deep learning algorithms previously proposed that show good performance on magnitude estimation, do not focus on large (Mw7.5+) or even very large (Mw9.0+) magnitude earthquakes. Such events are the most important targets of an EEW system. At the root of these difficulties is that every large earthquake is different from the next. Each can, and likely will, have a different starting location, rupture velocity, slip distribution, and radiated seismic energy that evolves in a complex way as the rupture unfolds. All of these properties fundamentally affect EEW system performance and are difficult if not impossible to predict prior to earthquake occurrence. As such, developing algorithms that can reliably characterize this complexity from surface observations in real-time has proven to be challenging.

In spite of this diversity of earthquake characteristics, advances in seismic and geodetic instrumentation over the last 30 years have allowed observation and synthesis of the basic kinematic behaviors of large ruptures (Hayes, 2017; Vallée & Douet, 2016; Ye et al., 2016). Additionally, the location and geometry of the faults on which many large earthquakes are expected to occur are well known (Hayes et al., 2018). By combining these observations, it is now possible to efficiently simulate the rupture process of many potential earthquakes in a realistic way and to predict their expected seismic and geodetic signatures (Frankel et al., 2018; Goldberg & Melgar, 2020; Melgar et al., 2016; Pitarka et al., 2020). Another important improvement, specifically in the case of HR-GNSS, is that noise models for real-time data have been proposed (Geng

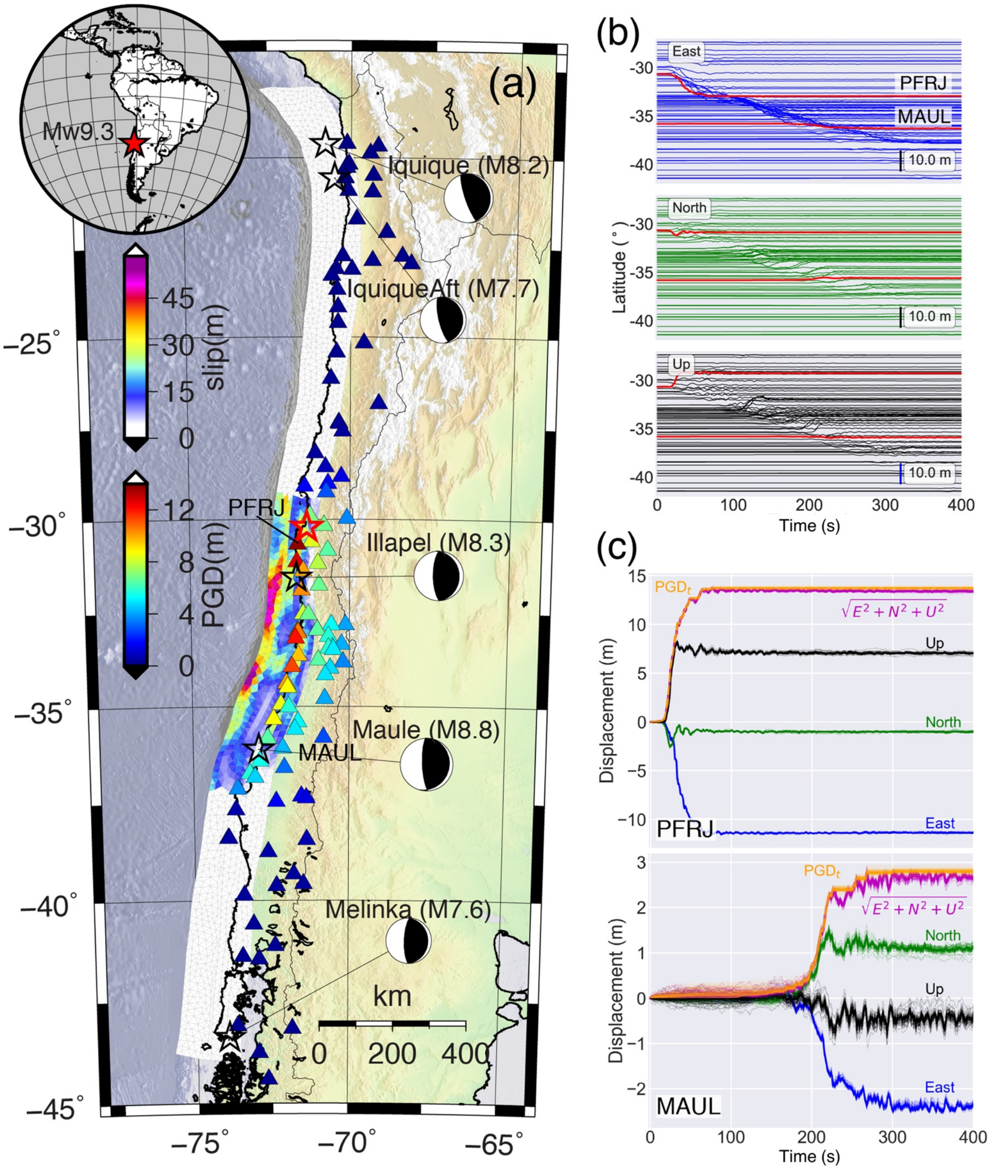

Figure 1. Map of the Chilean subduction zone, example rupture scenario, and resulting HR-GNSS waveforms. (a) Slip distribution of a synthetic Mw9.3 earthquake. GNSS stations (triangles) are color coded by their peak ground displacement. Focal mechanisms of 5 large events that have occurred since 2010. Red and black stars with focal mechanisms represent the hypocenter of the Mw9.3 rupture scenario and of the historical earthquakes, respectively. (b) Simulated three-component GNSS time series sorted by latitude. Bold red lines denote the records at stations PFRJ and MAUL. (c) Time series at stations PFRJ and MAUL. Thin lines denote the GNSS noise introduced in the Data and Methods section (see Section 2.1).

et al., 2018; Melgar et al., 2020). HR-GNSS displacements are a derived product and there can be significant differences between real-time and post-processed solutions. This improvement enables adding noise to any simulated waveform, making the waveforms more realistic.

Here, we will show how to leverage deep learning, the aforementioned earthquake simulations and their associated HR-GNSS waveforms to characterize earthquake magnitude in real-time. As a demonstration, we apply this approach to the Chilean Subduction Zone, which has a dense real-time GNSS network and assess its performance on five recent large-magnitude earthquakes that have occurred there (Figure 1).

2. Data and Methods

2.1. Rupture Scenarios and Synthetic Waveforms

The Chilean Subduction Zone on the west coast of South America is nearly 3,000 km long and accommodates 78–85 mm/yr of convergence between the Nazca and South American plates (DeMets et al., 2010). It regularly hosts large magnitude earthquakes, including five Mw7.5+ events in the last 10 years (Riquelme et al., 2018). Chile has a real-time HR-GNSS network with more than 120 stations currently in operation (Báez et al., 2018) and provides an excellent testbed for our proposed approach.

For generating the kinematic ruptures, we use the Slab2.0 3D slab geometry of Hayes et al. (2018). We utilize the Chilean slab model from its southern terminus to ∼100 km north of the Chile/Peru border. We limit the seismogenic depth to 55 km consistent with the down-dip extent of recently observed large earthquakes (Ruiz & Madariaga, 2018). The resulting geometry spans a nearly 3,000 km long, 200 km wide fault. The entire fault is then gridded into a total of 3,075 triangular subfaults using a finite element mesher, the average length and width of the subfault vertices is ∼12 km.

We generate the 36,800 ruptures on this geometry spanning the magnitude range Mw7.2 to Mw9.4 using the stochastic approach first described by Graves and Pitarka (2010) with modifications proposed by LeVeque et al. (2016) to avoid the use of Fourier transformations (Figure S1). The magnitudes of the scenarios are uniformly distributed with allowing a small perturbation so that the exact magnitude spans a wider range (Figure S2). The goal here is not to obey the Guttenberg-Richter frequency magnitude distribution but rather to generate a meaningful large and varied number of ruptures to expose M-LARGE to a sufficient variety of sources. The process of generating one particular rupture and its associated waveforms is described in detail in Melgar et al. (2016) and is summarized here: once the target magnitude is selected, we define the length and width of fault for that particular rupture. We make a random draw from a probabilistic length, L and width, W, scaling law (Blaser et al., 2010). L and W are obtained from a random draw from the log-normal distributions

$$log\left(L\right) \sim N\left(-2.37 + 0.57M_w, \sigma_L\right), \tag{1}$$

$$log\left(W\right) \sim N\left(-1.86 + 0.46M_w, \sigma_W\right), \tag{2}$$

with standard deviations defined in the original work of Blaser et al. (2010). The objective is to obtain a length and width that is consistent with the behavior seen in earthquakes worldwide while retaining the observed variability as well. The probabilistic scaling law thus ensures that for a given magnitude, we do not always employ the same fault dimensions. Detailed statistics on the resulting fault dimensions for all simulated ruptures can be seen in Figure S1. After the fault dimensions are defined, we randomly select a hypocentral location on the megathrust from a uniform spatial distribution. We do not take into account the variability in along-strike plate convergence rates or any information pertaining to which parts of the megathrust are considered more or less likely to experience a rupture.

Having selected the hypocentral location, we next generate the slip pattern and GNSS waveforms. For this we use the Karhunen-Loeve (KL) expansion method (LeVeque et al., 2016; Melgar et al., 2016). The process is separated into the following steps: (a) generate the stochastic slip pattern, (b) define rupture kinematics, and (c) forward model the resulting GNSS waveforms using a Green's function approach. Additional details are provided in Text S1 in Supporting Information S1. Finally, to make the synthetic data more realistic, we introduce noise into the displacement waveform characteristics using a known real-time GNSS noise model (Melgar et al., 2020), which was computed from the analysis of one-year-long real HR-GNSS observations spanning a large region. The reference noise model provides expected spectra of noise that vary from the 1st percentile or "low" noise model, continuously through the 50th percentile "median" noise model and up to the 90th percentile "high" noise model. For each waveform, we randomly select the percentile noise model and add it to the displacement data. It is worthwhile to note that we only assume the amplitude spectrum of noise, we keep the phase spectrum random. The addition of noise guarantees that the resulting time-domain waveform varies with each realization. In this way, we guarantee a large variability of noise and quality in the stations as is routinely seen in true real-time operations.

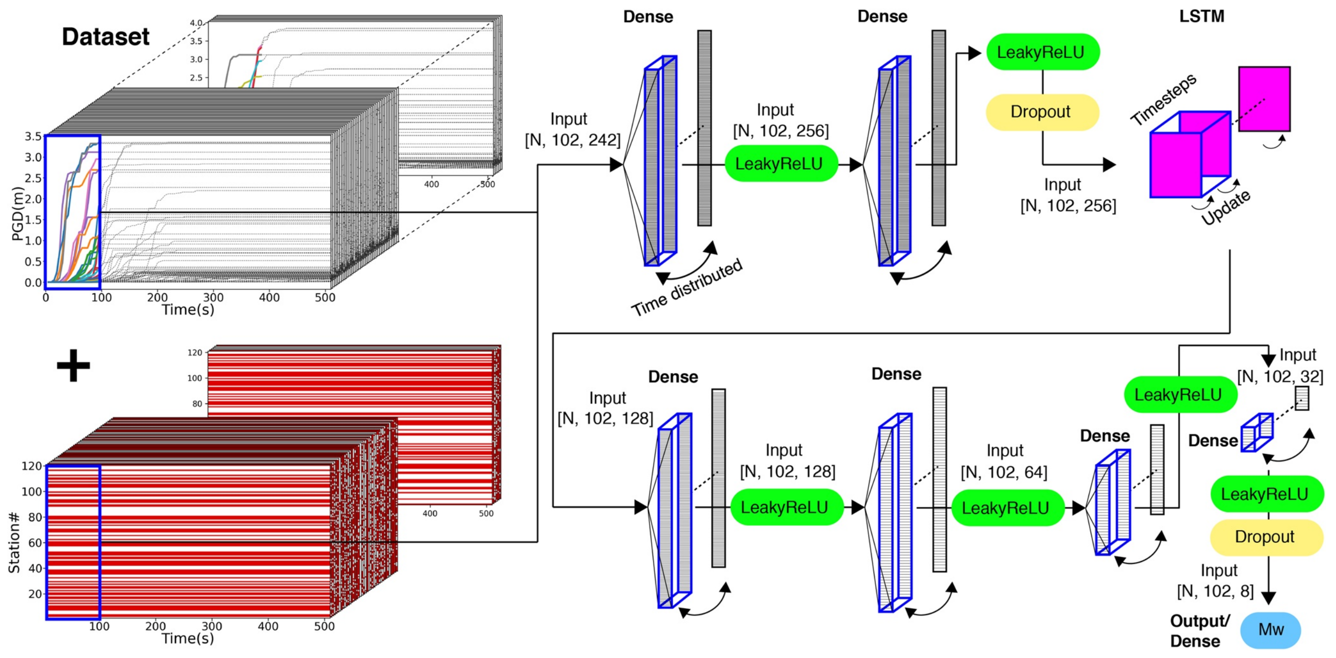

Figure 2. M-LARGE model architecture showing the input as the time-dependent PGD values from the GNSS stations plus the station on or off (existence) codes. Detailed parameter values are listed in Table S1. Blue rectangles mark the input PGD time series (i.e., 100 s) from all the available stations with their existence codes and the participating layers.

To ensure that the synthetic waveforms are realistic, we validate the data by comparing the simulated peak ground displacement against what is expected from PGD-Mw scaling (Melgar et al., 2015; Ruhl et al., 2019). This is shown in Figure S3, we find that the synthetic PGD pattern matches the scaling based on real observations at hypocentral distance ∼100 km and Mw from Mw7.7 to Mw8.7. We note that misfit between modeled and expected values of PGD increases at Mw greater than Mw9.0 or hypocentral distance smaller than 10 km. This is due to the fact that the PGD regressions are constructed from databases of real events; large earthquakes (i.e., Mw9.0+) and very close observations are comparatively rare in those databases (Melgar et al., 2016; Ruhl et al., 2019). The large misfit is also due to the limitation of point source assumption in PGD-Mw scaling laws where finiteness of large events needs to be considered. All the resulting synthetic rupture scenarios and GNSS waveforms are publicly available on Zenodo (https://doi.org/10.5281/zenodo.5015610) (Lin et al., 2020).

2.2. M-LARGE: Model Architecture

For time-dependent earthquake magnitude prediction, we employ a deep learning model, called Machine Learning Assessed Rapid Geodetic Earthquake model (M-LARGE) by linking the input time series recorded at each GNSS station to the time-dependent Mw for each rupture derived from the integration of the source time function (STF). We attempted to use simple architectures such as artificial neural networks (Figure S4) but ultimately found they did not have enough flexibility to capture the crustal deformation behavior. Our final model is composed of seven fully connected layers and a unidirectional long-short term memory (LSTM) recurrent layer (Hochreiter & Schmidhuber, 1997), which iteratively predicts Mw using the current and previous HR-GNSS observations across the network (Figure 2; Table S1). We adopted this model architecture because LSTMs are ideal for processing sequential data, which allows M-LARGE to update magnitude predictions as the rupture progresses. Additionally, it does not require a priori source information (such as the hypocenter) typically required by other rapid modeling methods (e.g., Crowell et al., 2018). Note that the dense layers only connect the feature values at the same time channel rather than all the features, which would include future times as well. Dropouts are applied to prevent overfitting during the training process (Srivastava et al., 2014). We use a Leaky ReLU function with a slope of 0.1 at negative values (Maas et al., 2013), which is an adaptation of the regular ReLU (i.e., slope of 0 at negative values) (Glorot et al., 2011) for the activation of dense layers. Finally, the last layer is connected to an ReLU function to output a current magnitude prediction, and the goal is to minimize the mean square error (MSE) contributed from the magnitude misfits at every epoch (Figure 2).

2.3. M-LARGE: Input Features, Output Labeling, and Model Training

To train the model, we use the 36,800 synthetic ruptures described in Section 2.1 and split them into training (70% or 25,760 ruptures), validation (20% or 7,360 ruptures), and testing data (10% or 3,680 ruptures) (Figure S2). The separation of validation and testing data ensures that the final model, which is selected based on the minimum validation loss, will not be biased during the model testing stage. Note that these are the number of rupture scenarios, the actual input data consider different station recording combinations and GNSS noise. In our case, we generate more than 6 million training data from the original training data set and 8,192 testing data from the testing data set (i.e., each rupture has ∼2 different station recording combinations and GNSS noise).

We use the PGD time series from all the stations as the model input

$$\mathrm{PGD}(t) = \max\left(\sqrt{E(t)^2 + N(t)^2 + Z(t)^2}\right), \tag{3}$$

where E(t), N(t), and Z(t) represents the East, North, and vertical component of the GNSS displacement time series starting from the earthquake origin (i.e., $t = 0$), respectively. We use PGD because it has a clear relationship to earthquake magnitude (Crowell et al., 2016; Melgar et al., 2015; Ruhl et al., 2019). We introduce feature scaling, which is commonly applied in machine learning, to avoid large feature values dominating smaller ones, making the model convergence difficult. The PGD time series is first clipped at a minimum of 0.01 m and scaled logarithmically. This is done so that during this rescaling process, the zero-valued data do not diverge to negative infinity. For the model output, we use the time integration from the real STF, convert it to the moment magnitude scale, and rescale this by multiplying the value by 0.1 for computational efficiency. Both the input and output time series are decimated to 5 s sampling so that we obtain Mw updates in 5 s increments.

To increase the variability of the data, we apply data augmentation by introducing realistic HR-GNSS noise as described in Section 2.1. We also randomly select some GNSS records to discard to simulate station outages and network variability yielding more than 6 million earthquake and station scenarios used for 50,000 training steps (Figure S5). The station incompleteness is necessary to simulate station outages, which commonly occur in real-time. For a given network, not all stations are operational all of the time, the algorithm should be resilient to this. To distinguish between existing and nonoperational stations, we add an additional "station existence" feature channel for every site. We set the value to zero to simulate a station outage and set it to 0.5 if the station is working normally. This last value is arbitrary, but it is a good practice to set it to the same scale of other features (i.e., in our feature, 0.5 is 3.16 m in the original scale, so that their values are comparable).

A total of 121 stations (Figure 1) and 102 time steps (i.e., 5 s sampling for 510 s of signal duration) of data are employed. Data incompleteness is included by randomly removing stations up to a maximum of 115 stations (i.e., a minimum of only 6 stations remaining). We also set a minimum training threshold of Mw7.5 and require at least 4 stations to locate within 3 degrees of the hypocenter. This is to make sure that the training data still carry some clear near-field information, otherwise the algorithm could introduce a bias because of the similar far-field values but different labeled magnitudes. Note that the hypocenter is the only necessary information for the data augmentation step. During the training or actual running process, no hypocenter information is needed. Here, we also note that M-LARGE does not detect the onset of an event because our intention here is not to build a detection algorithm, but to determine the correct magnitude when an earthquake is obviously detected. GNSS data are usually noisy enough that event detection from the real-time data can lead to many false positives (Kawamoto et al., 2016). Rather M-LARGE requires triggering, ostensibly by a seismic system as is common in other GNSS algorithms (e.g., Crowell et al., 2018). The noise in GNSS data is greater than that in seismic data and many algorithms have been demonstrated for the detection of the onset of events using inertial recordings (Perol et al., 2018; Ross et al., 2018; Zhu & Beroza, 2019) so a system that relies on seismometers for triggering is still the most robust.

One assumption we make is that there is no travel time delay due to the propagation of seismic waves from source-to-station in the feature and label pairs. Given the proximity of the Chilean subduction zone to the Chilean GNSS network (Figure 1), we assume any rupture would be recorded soon after the origin. In fact, in most events of our simulations the first arrival occurs by 20 s, and only 6% of rupture scenarios have arrivals later than 20 s. The latency is only a small fraction of the 510 s time series duration. Finally, we save the training weights every 5 epochs and use the model, which has the minimum validation loss as the final model (Figure S5). The code base is publicly available and can be obtained at https://doi.org/10.5281/zenodo.4527253 (Lin, 2021).

2.4. Comparison to Other Geodetic EEW Algorithms

Our main point of comparison for assessing whether M-LARGE is an improvement will be GFAST (Crowell et al., 2016), which predicts magnitude from GNSS observations, and it is also one of the most stable GNSS EEW methods currently operating in the U.S. EEW system (i.e., ShakeAlert). It uses the PGD observations from HR-GNSS time series. When a hypocenter is confirmed by a seismic method, the magnitude is calculated based on the PGD-Mw scaling relationship (Crowell et al., 2016; Melgar et al., 2015; Ruhl et al., 2019). To ensure the data contain PGD information and not noise, a 3 km/s travel-time filter is added into the algorithm, and the model only predicts Mw when at least 4 stations have valid information.

GFAST is not the only GNSS modeling approach; there are other proposed algorithms that utilize near-field GNSS data to rapidly estimate earthquake magnitude. To further compare with M-LARGE, we also run the Global Positioning System-based centroid moment tensor (GPSCMT) method, which utilizes the near-field static offset term from the GNSS records to calculate magnitude, moment tensor, and centroid location (Lin et al., 2019; Melgar et al., 2012). Unlike the GFAST approach, GPSCMT does not require hypocenter information, instead it grid-searches every predefined centroid location and solves for the moment tensor. We take the same subfault meshes used by M-LARGE as the potential centroid locations for the GPSCMT algorithm. Both the performance of GFAST and GPSCMT are shown in the next section.

3. Results

3.1. M-LARGE Performance on Testing Data Set

The performance of M-LARGE on the testing data set is shown in Figure 3. To quantify how well the model performs on testing data, we define a correct prediction as one within ±0.3 units of the target magnitude (i.e., time-dependent magnitude) and calculate the model accuracy (Figure 3b). Within these bounds, the model performs well with a high accuracy of 95% after 60 s, which increases to 99% by 120 s. The standard deviation of the magnitude misfits are 0.15, 0.1, and 0.09 at 60, 120, and 360 s, respectively. The accuracy change from 120 to 360 s is not significant; however, this additional time improves some underestimations of long source duration events with magnitude greater than Mw8.5 (Figure 3a).

We compare this statistic to the GFAST algorithm by using the same testing data set as M-LARGE. Note that for GFAST, we remove those predictions with $\mathrm{Mw} = 0$ due to the four-station minimum and only show the data that have nonzero values (Figure 3b). In this example, over 30% of the testing events have $\mathrm{Mw} = 0$ prediction at 60 s, these predictions require additional time to converge compared to our model. Despite this removal, we find that GFAST has a lower accuracy of 62% at 60 s, which slowly increases to 78% by 120 s and to 80% by 360 s. In comparison to M-LARGE's maximum accuracy of 99%, GFAST's accuracy saturates at 86% by 215 s. The standard deviation of the magnitude predictions of GFAST are also larger, which are 0.22, 0.21, and 0.22 at 60, 120, and 360 s, respectively, about 2 times more scatter than the M-LARGE performance. To summarize, M-LARGE reaches 80% accuracy 5 times faster than GFAST and has half the scatter on average.

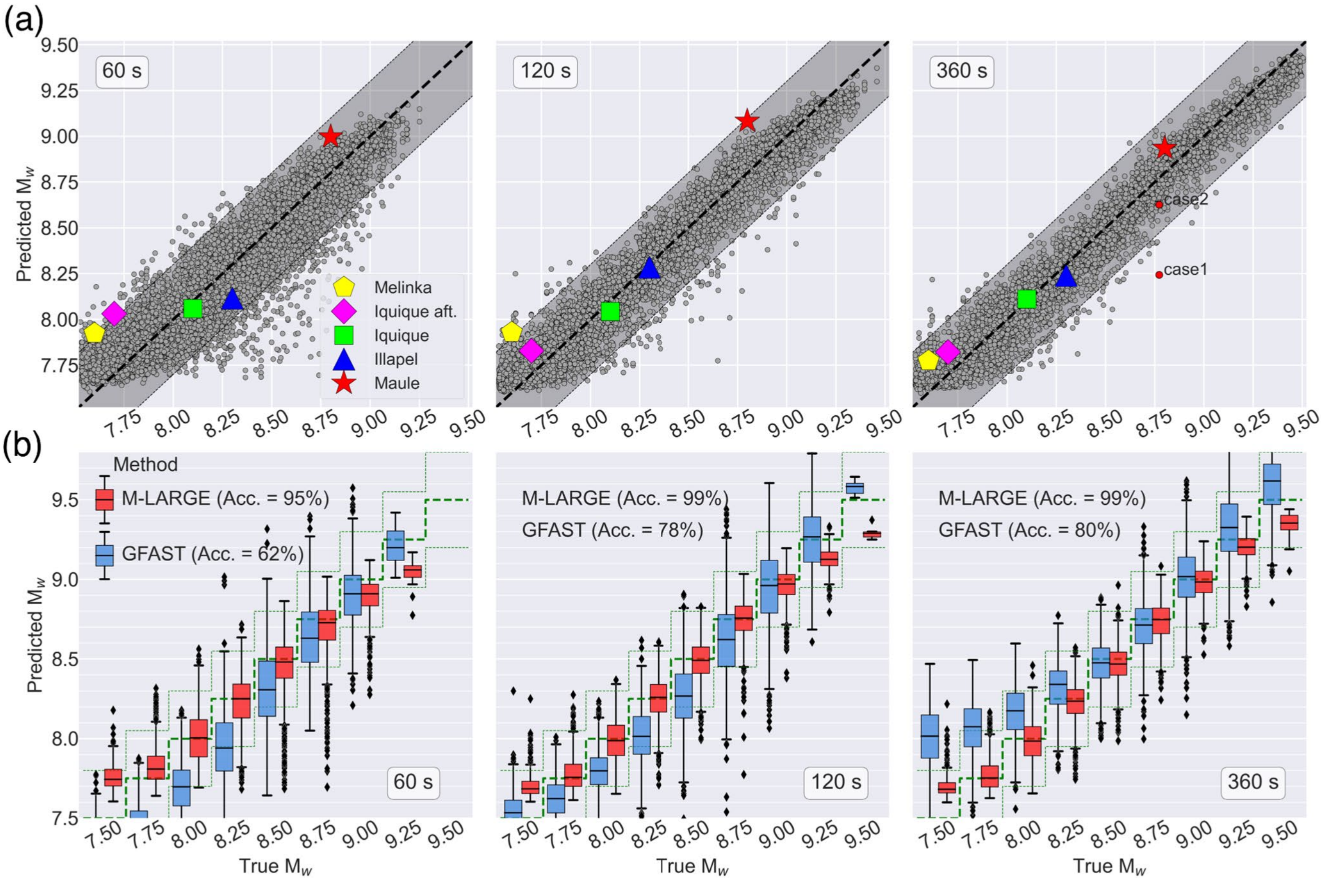

Figure 3. Model performance on testing data set and on real events. (a) (from left to right) snapshots of M-LARGEs performance at 60, 120, and 360 s, respectively. Gray dots show the Mw predictions compared to the time-dependent magnitude. Black dashed line represents the 1:1 line; shaded area represents the ±0.3 magnitude range. Colored markers denote the M-LARGE-predicted Mw and their final Mw for 5 real events is shown in Figure 1. The red dots with labels are the scenarios, which will be discussed in Section 3.3. (b) Comparison of the GFAST (blue) and M-LARGE (red) predicted magnitudes at 60,120, and 360 s for different magnitude bins. Model accuracies at 60, 120, and 360 s are shown in text. The thick and thin green dashed lines show the 1:1 and ±0.3 reference for each magnitude bin.

Furthermore, we compare the performance between M-LARGE and the GPSCMT (Figure S6). Again, M-LARGE significantly outperforms the GPSCMT, where the accuracies are only 41%, 28%, and 28% at 60, 120, and 360 s, respectively. Noting that the GPSCMT performs with overall much larger scatter, lower accuracy, and systematic overestimations. This has been noted before, that a point source has limited ability on recovering the deformation of large offshore events (e.g., Melgar et al., 2013). Thus, without additional constraints, the model accuracy of GPSCMT method is about 40% according to our testing data set.

3.2. M-LARGE Performance on Real Large Earthquakes

To further assess the performance of M-LARGE, we apply the model to five large historical events in the Chilean Subduction Zone with HR-GNSS records, which were not used for training (Figure 4). For each of these earthquakes, different numbers of GNSS sites were available. For the 2010 Mw8.8 Maule earthquake, the model only takes 40 s to reach the ±0.3 magnitude unit criteria. M-LARGE also successfully predicts the final magnitude of the 2014 Mw8.1 Iquique and the 2015 Mw8.3 Illapel earthquakes at 20 and 60 s, respectively. For the 2014 Mw7.7 Iquique aftershock and the 2016 Mw7.6 Melinka earthquakes, both the M-LARGE predictions slightly overshoot the true magnitude at 30 s, but soon correct downward to their actual magnitude ranges. Compared to the performance of GFAST, which underestimates the Mw7.7 Iquique aftershock by 0.3 magnitude units and overestimates the Melinka earthquake by 0.6 magnitude units, our model shows more robust results on those smaller magnitude events where large GNSS noise dominated

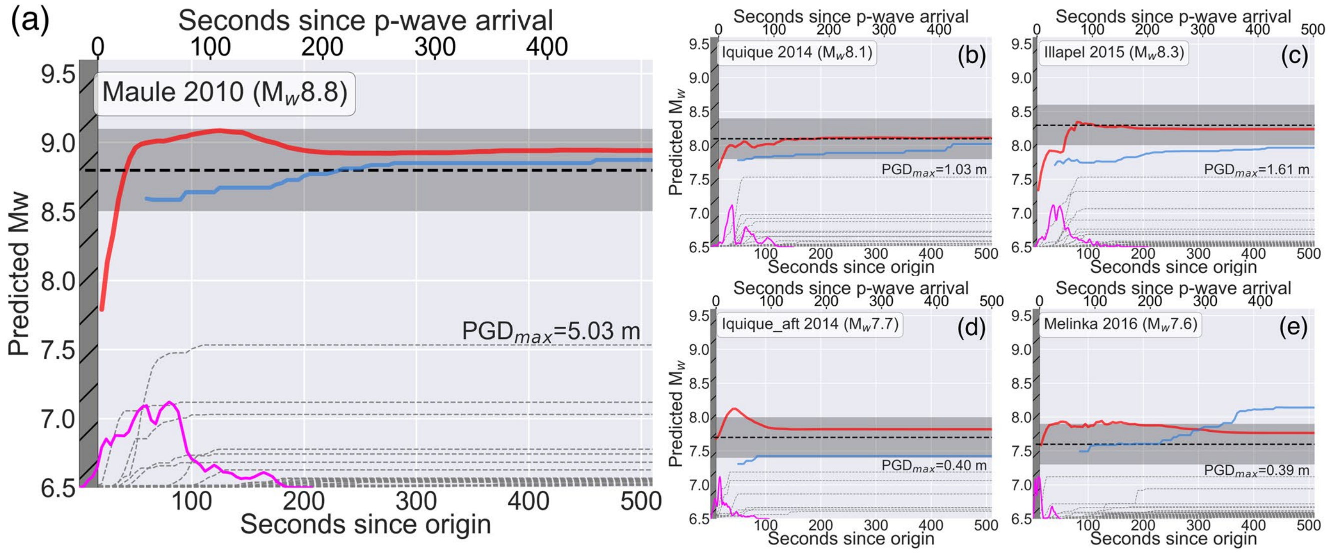

Figure 4. M-LARGE performance on real Chilean earthquakes compared to GFAST. (a) The 2010 Maule Mw8.8 earthquake. Red and blue lines show the M-LARGE and GFAST prediction, respectively. Black dashed line and gray shaded areas represent the true Mw and the ±0.3 magnitude unit range. Thin dashed lines show the peak ground displacement waveforms from the 2010 Maule earthquake (Figure 1). Magenta line represents the event source time function from the USGS finite fault product. Hatched dark gray area is the time period prior to the arrival of the P-wave at the closest site where no information on the rupture is available. (b–e) Same as (a) but for the 2014 Mw8.1 Iquique earthquake, the 2015 Mw8.3 Illapel earthquake, the 2014 Mw7.7 Iquique aftershock, and the 2016 Mw7.6 Melinka earthquake, respectively.

the data across the whole network spanning a 3,000 km long subduction zone (Figure 1). We also note that the performance is bounded by the delay times prior to the P-wave arrival at the closest stations. For example the Maule earthquake, where most of the presently operating closest stations did not exist in 2010, the first arrival time was 17 s. Considering these delay times, useful predictions are made as soon as the signals are recorded but the lowest uncertainties are anticipated after ∼30 s. The timing of the successful prediction correlates with the source duration, data sparsity or coverage, and the signal-to-noise ratio of the event, which will be further discussed in Sections 3.3 and 3.4, and in the discussion section.

3.3. M-LARGE Testing: Imperfect or Sparse Data

To understand how M-LARGE performs on imperfect data, we test M-LARGE on two different recording scenarios of the same rupture from the testing data set. One scenario has poor station coverage whereas the other has excellent station coverage (i.e., red dots in Figure 3a) (Figure 5). In the poor station coverage example (i.e., Case 1 in Figure 5), almost all the near-field data are missing and M-LARGE fails to estimate the true magnitude. It is not until far-field stations begin recording data that M-LARGE upgrades its moment estimate closer to the lower bound of the actual magnitude (Figures 5b and 5c). In contrast, for the good station coverage example, abundant near-field data are used to accurately characterize the rupture process and M-LARGE predicts the actual magnitude successfully (Figures 5b and 5d). This suggests that data sparsity in the near-field plays an important role for the accuracy and timeliness of the predictions. The clear implication is that having more stations closer to the source improves M-LARGE's performance.

3.4. M-LARGE Testing: STF Complexity

To understand M-LARGEs performance as a function of source complexity, we choose four different characteristic source time function shapes (i.e., symmetric, bimodal, early, and late skewed) and analyze the results. Figure 6 shows examples of each of these characteristic STFs. The models successfully converge to their actual magnitude. We use the time-to-peak STF (hereafter, peak time τ_p) and centroid time (τ_{cent}) as proxies to measure the model performance on different STF shapes. We define the time to corrected prediction (τ_c) as the time when the model successfully predicts the final magnitude with a misfit smaller than

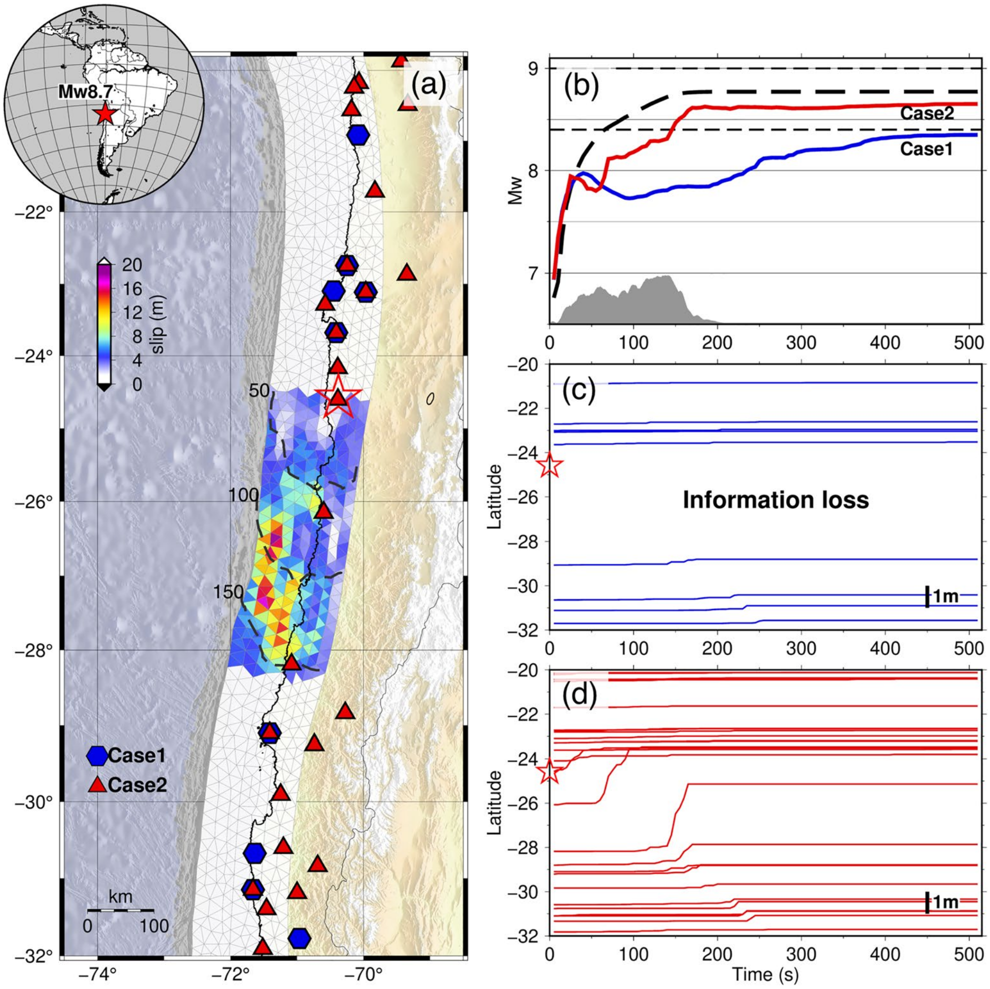

Figure 5. M-LARGE prediction tests with two different station distributions. (a) Rupture scenario of an Mw8.7 earthquake with the station distribution of Case1 (blue hexagon) and Case2 (red triangle). Red star denotes the hypocenter. Black dashed lines show the 50 s rupture time contours. (b) M-LARGE predictions for Case 1 (blue line), Case 2 (red line), and the actual Mw (dashed line) calculated from the STF (gray area). (c) Peak ground displacement data of Case 1 sorted by latitude, red star denotes the hypocenter latitude. Panel (d) similar to (c), but data of Case 2.

0.3 (Figure 6). We calculate the correlation coefficient (CC) between these metrics (Figure S7) and find that the $\tau_c - \tau_p$ has the weakest correlation with CC = 0.66 compared to $\tau_c - \tau_{\mathrm{cent}}$ of CC = 0.9 and $\tau_c - \tau_{\mathrm{duration}}$ of CC = 0.85. This suggests that the τ_p, a proxy of STF's shape, does not significantly affect the timing and accuracy of the M-LARGE estimations. On the other hand, the performance relies on the actual moment release because it is trained to map the STF directly. We note that the CC of $\tau_c - \tau_{\mathrm{cent}}$ is slightly higher than the $\tau_c - \tau_{\mathrm{duration}}$, this is because of the effect of the tolerated magnitude, which we will discuss in Section 4.2.

4. Discussion

4.1. Rupture Scenario as a Representation of the Real World

One of the challenges for developing algorithms to characterize large earthquakes is the rarity of real events. However, while large earthquakes are infrequent, their rupture kinematics are not unexpected. They are the same as what we imagine in more modestly sized events. If a comprehensive and realistic rupture scenario data set can be generated, then for training a machine learning algorithm, large earthquakes are no longer “rare.” We can rely on the synthetic data as a sufficient representation of the real world. To test this, we

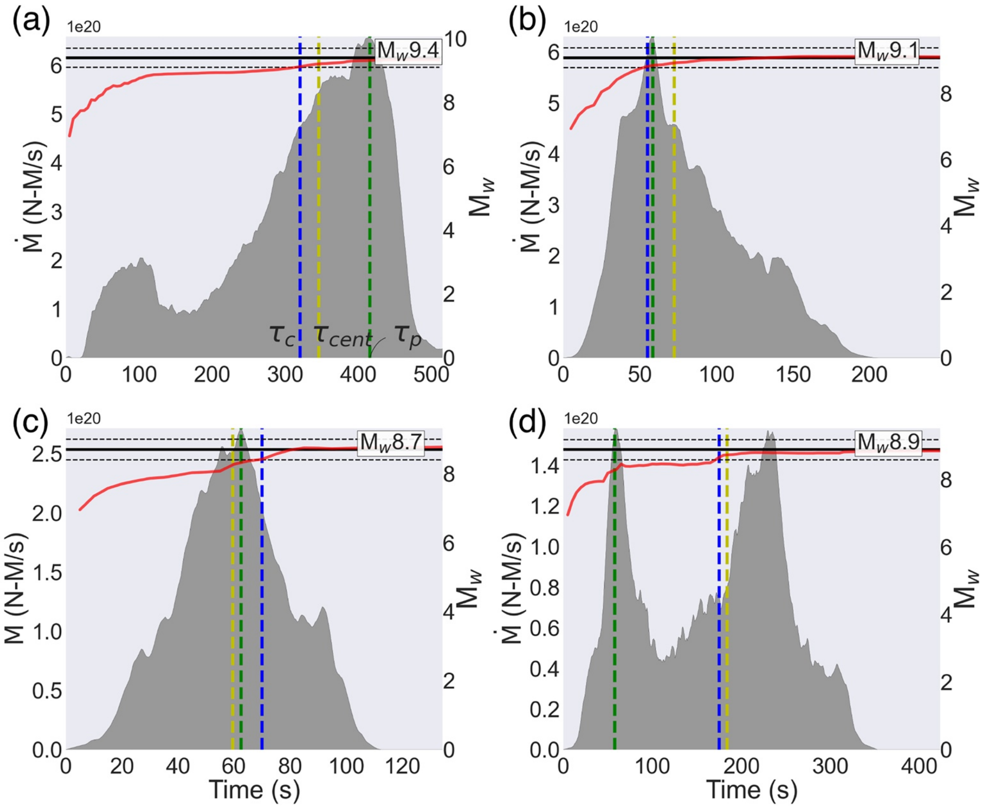

Figure 6. Example plot for the time to corrected prediction (τ_c), centroid time (τ_{cent}), and peak time (τ_p) for four different cases. Red and black lines show the M-LARGE predictions and final magnitudes, respectively. Black dashed lines denote the ±0.3 magnitude unit range. Panel (a) shows the case of late rupturing, where the source focuses at the end of the rupture, (b) shows the case with early rupturing, where the source focuses at the beginning of the rupture, (c) nearly symmetric (triangular) source time function, and (d) shows the case of two rupture asperities.

compare the records of the 2010 Mw8.8 Maule earthquake to the 25,760 scenario events in the training data set. These rupture scenarios have no a priori knowledge of the source behavior of the Maule earthquake, but simply follow the assumptions in Section 2.1. In Figure S8, we use a grid-search to select 5 best-fitting events from the data set to the Maule records, and they all have similar moment magnitudes (i.e., events with similar PGD patterns have similar fault slips). This suggests that large earthquakes kinematics are not unmodeled, and given sufficient realistic simulations, rarity is no longer an issue for training.

4.2. Timing of Final Magnitude Estimation

Although the timeliness of the final magnitude assessment is intimately tied to the evolution of the STFs (i.e., whether the event grows faster or slower), we find that frequently the time to correct predictions do not follow the exact STF behavior. The reason for this time variation is mainly due to the effect of the ± 0.3 tolerance. A successful prediction can occur earlier than the actual source duration and at the lower bound of the magnitude tolerance (e.g., Figure 6) resulting in an earlier prediction. While the effect of the magnitude tolerance depends on the shape of the individual STF, which is nontrivial to our stochastic simulations; however, by simply assuming the STF as a triangular function (i.e., rise and fall-off rates are the same), we can estimate this being 71% of the original duration time based on the scaling of Duputel et al. (2013) (a detailed derivation is provided in the Text S2 in Supporting Information S1). For example, on average, an Mw9.0 event takes ~170 s to rupture, while it only takes 120 s to rise to the acceptable Mw threshold of Mw8.7. Our model result shows an even shorter magnitude determining time, which is about 25%–50% of the source duration (Figure 7a). This advance in time is when we consider sources that follow a nonsym-

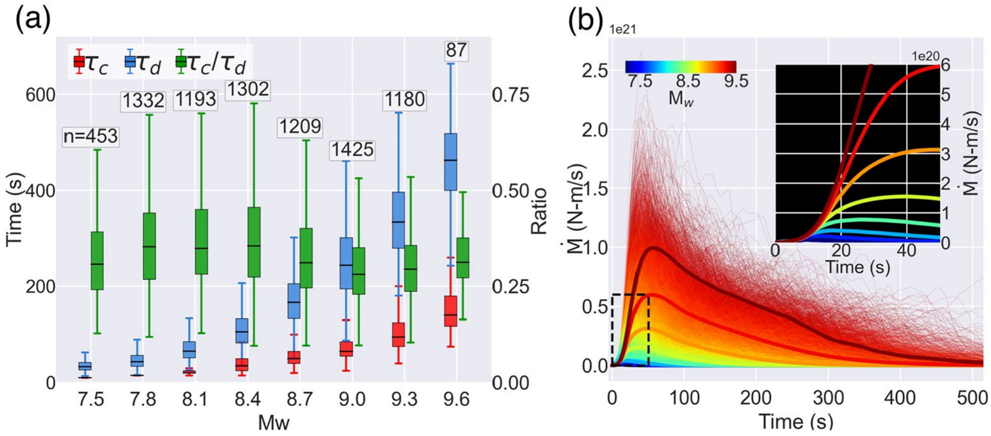

Figure 7. Warning time ratios and STF analysis. Panel (a) shows the duration (τ_d), time to the correct prediction (τ_c), and the ratio between these two for each magnitude bin. Texts indicate the number of samples for each bin. Panel (b) shows the STF of 36,800 rupture scenarios color coded by Mw. Thick lines denote the averaged STF of different magnitude bins. Inset shows the zoom-in view of the averaged source time functions.

metric flat and long tail Dreger-STF, which have growth patterns that can frequently be seen in worldwide databases (Figure 7b) (Mena et al., 2010). Thus, our model can provide practical earlier warning while updating its magnitude as time progresses; this is only possible when the real-time STF can be accurately measured.

4.3. Model Uncertainty

The uncertainty in magnitude prediction is important for practical EEW systems. A probabilistic output layer could potentially be an estimator of the model confidence; however, in our regression-type model, such an output layer is not straightforward. Here, we analyze the model performance on the testing data set to estimate uncertainty. Assuming that the distribution of the testing data set is complete, we calculate the model accuracy as a function of time (i.e., length of PGD data used) and its magnitude (Figure 8). Figure 8a shows the prediction accuracies with respect to their final magnitudes. We find that generally high accuracies occurred at the right-hand side of the estimated duration curve, suggesting a final magnitude is more likely to be correctly determined after the source termination. For example, the predicted magnitude of Mw9.0 has an accuracy of 77% at 100 s, and this rises to 98% when estimating the same magnitude at 200 s. On the other hand, the low accuracies at the beginning of the prediction suggest that the initial rupture signals are not good indicators of final magnitude, which is consistent with previous source studies (Goldberg et al., 2018; Ide, 2019; Meier et al., 2017; Melgar & Hayes, 2019; Rydelek & Horiuchi, 2006). This is also demonstrated in Figure S9, where the same current STF shapes may lead to different final magnitudes. We also note that for very large events (Mw9.2+), high accuracies can occur at the very early stage, prior to the source duration. This is due to a large slip influencing the beginning of the STF (e.g., Figure 6b), and since the largest possible magnitude is limited by the finite fault geometry (i.e., in our case, Mw9.6), the possibility that an Mw9.2+ event grows into a larger event is limited.

We further show the model accuracies with respect to their current or time-dependent magnitude (Figure 8b). This is also the actual label that the model learns during the training step. However, Figure 8b shows that the misfits are not evenly distributed, large misfits occur at the beginning of the prediction. The mismatches at the beginning of the data-label pairs are mainly due to the travel time delay introduced in Section 2.3. In addition to this, GNSS noise can also dominate the beginning of the records when only few stations have true rupture signals, contributing to large uncertainty at this time. The effect of GNSS noise can also be seen in both Figures 8a and 8b, where larger misfits occur constantly throughout the time window for smaller events (e.g., Mw below 8.5).

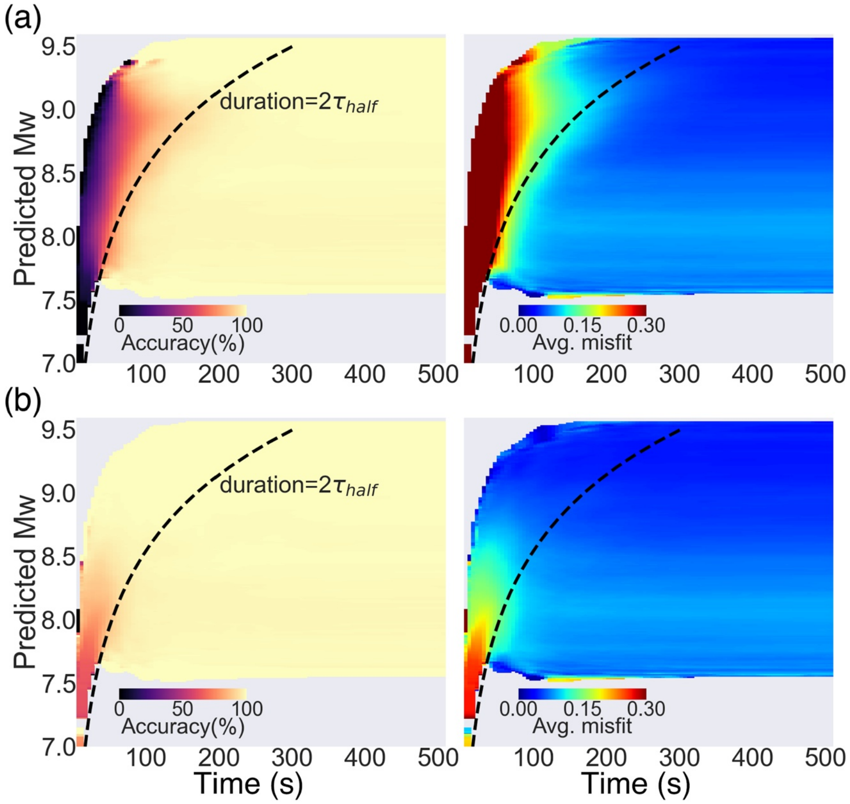

Figure 8. Model performance on the testing data set. Panel (a) shows the prediction accuracy (i.e., number of successful prediction/total samples) as a function of peak ground displacement time and Mw, where a successful prediction is defined as when the predicted and final Mw misfit is smaller than 0.3. Dashed line shows the estimated duration from Duputel et al. (2013). Panel (b) same as (a), but define a successful prediction is when the predicted and time-dependent Mw misfit is smaller than 0.3. Note that the time-dependent Mw is the integration from the STF at the current time.

4.4. Limitations and Future Work

We have shown that M-LARGE has the ability to learn complex rupture patterns from the crustal deformation data. Also, it significantly outperforms other HR-GNSS EEW algorithms. However, we note that it still has some limitations, and these should be targets for potential improvements in the future. First, once M-LARGE is trained, the model is not global in scope, it is presently limited by the simulated earthquakes, waveforms, and network geometry for a specific region. So at present, the model can only be applied to the specific configuration of GNSS sites at a specific subduction zone, Chile. In machine learning, whether a model can be applied to other data not seen in training, such as that from a different subduction zone, is called generalization. It is evident that the approach we have followed here is tied to the specific network geometry and subduction zone and will not immediately generalize to other tectonic settings. For M-LARGE to be useful elsewhere, it will need to be retrained to another specific geometry and perceived possible set of ruptures for that tectonic environment. For example, consider the Cascadia Subduction Zone where an operational EEW system exists (Kohler et al., 2020) and already uses simple PGD scaling from GNSS for magnitude calculation. Steps to implement M-LARGE in this new setting would involve taking the known 3D geometry of the megathrust and simulating an adequate set of ruptures in a desired magnitude range (~M7-M9.2), waveforms would be synthesized for the ~1000 GNSS station network (Murray et al., 2018), polluted with real recorded GNSS real-time noise, and used to train another version of M-LARGE. For such a real-world implementation, careful thought would have to go into other non-subduction zone sources common to the region, such as offshore strike slip faults and inslab events. These kinds of events could be simulated as well and included in the training. Validation would be challenging given the dearth of large megathrust ruptures but could be done on other real-recorded events such as M7 events in the Mendocino triple junction or the M6.7 Nisqually earthquake as shown by Crowell et al. (2016). This discussion shows how it is nontrivial to apply the method to a new region, so future avenues of research will include whether the model can perhaps be generalized by other means such as introducing feature engineering (e.g., extracting the hypocentral distance used by GFAST). We have not yet attempted this because regional heterogeneity such as site effects, tectonic environments, fault geometries, and station distribution vary, so global model generalization is nontrivial. However, synthesis of the ruptures and GNSS waveforms is fast enough that the algorithm could be adapted to other environments.

M-LARGE, like any geodesy-based technique, is only useful for large magnitude events typically in the M7+ range (e.g., Crowell et al., 2013; Melgar et al., 2015). Damaging shaking during earthquakes can also occur at significantly smaller magnitudes (e.g., Minson et al., 2021). The approach proposed here is not meant to replace seismic methods but rather to work in tandem with them. Saturation is a persistent concern for EEW and by combining networks, data types, and algorithms EEW systems can respond to a wider variety of events.

We note that for the 2010 Mw8.8 Maule earthquake example, there is a 17 s gap without recording due to the lack of near-field stations. This performance could be sped up by ~10 s if the information delay introduced by the travel times could be reduced, that is if station coverage were expanded offshore. The model performance is strongly reliant on the training data set behaving according to what is seen in world databases. As a result, an outlier event with a unique rupture may still prove challenging. More simulated events that incorporate rupture variability would improve M-LARGE's performance by making it resilient to complex rupture scenarios.

The model architecture and hyperparameters are selected arbitrarily; however, the scale of hyperparameters is comparable to those in similar studies (e.g., Ross et al., 2018; Zhu & Beroza, 2019). We do not find significant model improvement from tuning the hyperparameters we used (Figure 2), probably because the model has already reached its accuracy limit (i.e., 99%) based on the current architecture. Any further improvement will require a different model design. We find that the logarithmic scaling function of PGD features has better performance against the commonly adopted linear scaling. This is consistent with the existence of log-linear PGD and magnitude relationships (Crowell et al., 2016; Melgar et al., 2015; Ruhl et al., 2019) making the input and output pairs less complicated during model training.

We trained the model to learn the evolution of STF directly, which is based on the assumption that earthquake sources are not strongly deterministic (Ide, 2019; Rydelek & Horiuchi, 2006), this is also the source characteristic for our simulations (Figure 7b), where initial rupture signals (i.e., <5 s) do not provide the information of final magnitude, as opposed to strongly deterministic scenarios (Olson & Allen, 2005; Wu & Zhao, 2006) where the initial rupture signal for small or large magnitude events are fundamentally different. Even though the final magnitude can be made earlier than the source termination, which has been discussed in Section 4.1, earthquake determinism is not the main exploration in this study. However, source observation studies with moderate or weakly deterministic behavior (Goldberg et al., 2018; Meier et al., 2017; Melgar & Hayes, 2019) may be incorporated into future models to speed up the warning time. Also, as shown in Figure S9, the possibility of the final magnitude can be inferred by tracking the STF's shape when rupture proceeds, providing another way to potentially speed up the warning time.

Lastly, the earthquake magnitude is not the only important factor for EEW. In fact, the source location, rupture length, width, and slip are equally important for an accurate ground motion prediction or tsunami amplitude forecasts. In this paper, we have successfully demonstrated that M-LARGE is capable of learning Mw directly from raw observations. This is a starting point for new types of EEW algorithm, and we anticipate that, given this success, M-LARGE could be expanded to directly forecast earthquake hazards.

5. Conclusion

Developing frameworks to provide timely warning during the largest magnitude earthquakes remains an outstanding scientific and technological challenge. EEW systems continue to expand and have proliferated to many countries across the globe (Allen & Melgar, 2019). Despite this, how these systems will perform in rare but high consequence, large magnitude earthquakes is uncertain. Here, we have combined the knowledge of where great earthquakes will occur, their average expected rupture characteristics, state of the art sensor technology, and deep learning to rapidly characterize large magnitude earthquakes from their crustal deformation patterns. In simulated real-time testing, the resulting EEW algorithm, M-LARGE, has a significantly better performance than current GNSS algorithms and can be retrained to apply to any specific region capable of generating large events. As such, M-LARGE represents a new approach to EEW that, if made operational, can work in tandem with the current technologies and provide accurate, unsaturated Mw estimation and fast alerts that will lead to increased resilience. This is the first demonstration of our approach, future work to achieve generalization of the method is still needed and will include more training and testing data, interacting with existing EEW methods, and creating new data for different tectonic environments and GNSS network configurations.

Data Availability Statement

The rupture simulations and waveforms can be found on Zenodo: https://doi.org/10.5281/zenodo.5015610 (Lin et al., 2020). The code of M-LARGE can be obtained at https://doi.org/10.5281/zenodo.4527253 (Lin, 2021).

Acknowledgments
We thank the editor, Rachel Abercrombie, the associate editor, Ronni Grapenthin, and an anonymous reviewer for their careful reviews, which have helped us to greatly improve the manuscript. This work is funded by the National Aeronautics and Space Administration NESSF grant 80NSSC18K1420 and National Science Foundation grants 1663834 and 1835661. This work is also partially funded by NASA grants 80NSSC19K0360 and 80NSSC19K1104. We thank the Centro Sismologico Nacional of Chile for operation of the HR-GNSS network and access to their data. We would also like to thank Amy Williamson and Brendan Crowell for discussions on the performance of GFAST on simulated events.

References

Allen, R. M., & Melgar, D. (2019). Earthquake early warning: Advances, scientific challenges, and societal needs. *Annual Review of Earth and Planetary Sciences*, *47*, 361–388. https://doi.org/10.1146/annurev-earth-053018-060457

Báez, J. C., Leyton, F., Troncoso, C., del Campo, F., Bevis, M., Vigny, C., et al. (2018). The Chilean GNSS network: Current status and progress toward early warning applications. *Seismological Research Letters*, *89*(4), 1546–1554. https://doi.org/10.1785/0220180011

Blaser, L., Krüger, F., Ohrnberger, M., & Scherbaum, F. (2010). Scaling relations of earthquake source parameter estimates with special focus on subduction environment. *Bulletin of the Seismological Society of America*, *100*(6), 2914–2926. https://doi.org/10.1785/0120100111

Bock, Y., & Melgar, D. (2016). Physical applications of GPS geodesy: A review. *Reports on Progress in Physics*, *79*(10), 106801. https://doi.org/10.1088/0034-4885/79/10/106801

Boore, D. M., & Bommer, J. J. (2005). Processing of strong-motion accelerograms: Needs, options and consequences. *Soil Dynamics and Earthquake Engineering*, *25*(2), 93–115. https://doi.org/10.1016/j.soildyn.2004.10.007

Böse, M., Heaton, T. H., & Hauksson, E. (2012). Real-time finite fault rupture detector (FinDer) for large earthquakes. *Geophysical Journal International*, *191*(2), 803–812. https://doi.org/10.1111/j.1365-246x.2012.05657.x

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. *Bulletin of the Seismological Society of America*, *109*(4), 1524–1541. https://doi.org/10.1785/0120180326

Colombelli, S., Allen, R. M., & Zollo, A. (2013). Application of real-time GPS to earthquake early warning in subduction and strike-slip environments. *Journal of Geophysical Research: Solid Earth*, *118*(7), 3448–3461. https://doi.org/10.1002/jgrb.50242

Crowell, B. W., Melgar, D., Bock, Y., Haase, J. S., & Geng, J. (2013). Earthquake magnitude scaling using seismogeodetic data. *Geophysical Research Letters*, *40*(23), 6089–6094. https://doi.org/10.1002/2013gl058391

Crowell, B. W., Schmidt, D. A., Bodin, P., Vidale, J. E., Baker, B., Barrientos, S., & Geng, J. (2018). G-FAST earthquake early warning potential for great earthquakes in Chile. *Seismological Research Letters*, *89*(2A), 542–556. https://doi.org/10.1785/0220170180

Crowell, B. W., Schmidt, D. A., Bodin, P., Vidale, J. E., Gomberg, J., Renate Hartog, J., et al. (2016). Demonstration of the Cascadia G-FAST geodetic earthquake early warning system for the Nisqually, Washington, earthquake. *Seismological Research Letters*, *87*(4), 930–943. https://doi.org/10.1785/0220150255

DeMets, C., Gordon, R. G., & Argus, D. F. (2010). Geologically current plate motions. *Geophysical Journal International*, *181*(1), 1–80. https://doi.org/10.1111/j.1365-246x.2009.04491.x

Duputel, Z., Tsai, V. C., Rivera, L., & Kanamori, H. (2013). Using centroid time-delays to characterize source durations and identify earthquakes with unique characteristics. *Earth and Planetary Science Letters*, *374*, 92–100. https://doi.org/10.1016/j.epsl.2013.05.024

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. *Bulletin of the Seismological Society of America*, *108*(5A), 2347–2369. https://doi.org/10.1785/0120180034

Geller, R. J. (1976). Scaling relations for earthquake source parameters and magnitudes. *Bulletin of the Seismological Society of America*, *66*(5), 1501–1523. https://doi.org/10.1785/bssa0660061801

Geng, J., Pan, Y., Li, X., Guo, J., Liu, J., Chen, X., & Zhang, Y. (2018). Noise characteristics of high-rate multi-GNSS for subdaily crustal deformation monitoring. *Journal of Geophysical Research: Solid Earth*, *123*(2), 1987–2002. https://doi.org/10.1002/2018jb015527

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 315–323).

Goldberg, D. E., & Melgar, D. (2020). Generation and validation of broadband synthetic P waves in semistochastic models of large earthquakes. *Bulletin of the Seismological Society of America*, *110*(4), 1982–1995. https://doi.org/10.1785/0120200049

Goldberg, D. E., Melgar, D., & Bock, Y. (2019). Seismogeodetic P-wave amplitude: No evidence for strong determinism. *Geophysical Research Letters*, *46*(20), 11118–11126. https://doi.org/10.1029/2019gl083624

Goldberg, D. E., Melgar, D., Bock, Y., & Allen, R. M. (2018). Geodetic observations of weak determinism in rupture evolution of large earthquakes. *Journal of Geophysical Research: Solid Earth*, *123*(11), 9950–9962. https://doi.org/10.1029/2018jb015962

Grapenthin, R., Johanson, I. A., & Allen, R. M. (2014). Operational real-time GPS-enhanced earthquake early warning. *Journal of Geophysical Research: Solid Earth*, *119*(10), 7944–7965. https://doi.org/10.1002/2014jb011400

Graves, R. W., & Pitarka, A. (2010). Broadband ground-motion simulation using a hybrid approach. *Bulletin of the Seismological Society of America*, *100*(5A), 2095–2123. https://doi.org/10.1785/0120100057

Hayes, G. P. (2017). The finite, kinematic rupture properties of great-sized earthquakes since 1990. *Earth and Planetary Science Letters*, *468*, 94–100. https://doi.org/10.1016/j.epsl.2017.04.003

Hayes, G. P., Moore, G. L., Portner, D. E., Hearne, M., Flamme, H., Furtney, M., & Smoczyk, G. M. (2018). Slab2, a comprehensive subduction zone geometry model. *Science*, *362*(6410), 58–61. https://doi.org/10.1126/science.aat4723

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hoshiba, M., Iwakiri, K., Hayashimoto, N., Shimoyama, T., Hirano, K., Yamada, Y., & Kikuta, H. (2011). Outline of the 2011 off the Pacific coast of Tohoku Earthquake (M w 9.0)—Earthquake early warning and observed seismic intensity—. *Earth Planets and Space*, *63*(7), 547–551. https://doi.org/10.5047/eps.2011.05.031

Hoshiba, M., & Ozaki, T. (2014). Earthquake early warning and tsunami warning of the Japan Meteorological Agency, and their performance in the 2011 off the Pacific Coast of Tohoku Earthquake (Mw9.0). In *Early warning for geological disasters* (pp. 1–28). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-12233-0_1

Hutchison, A. A., Böse, M., & Manighetti, I. (2020). Improving early estimates of large earthquake's final fault lengths and magnitudes leveraging source fault structural maturity information. *Geophysical Research Letters*, *47*, e2020GL087539. https://doi.org/10.1029/2020gl087539

Ide, S. (2019). Frequent observations of identical onsets of large and small earthquakes. *Nature*, *573*(7772), 112–116. https://doi.org/10.1038/s41586-019-1508-5

Kawamoto, S., Hiyama, Y., Ohta, Y., & Nishimura, T. (2016). First result from the GEONET real-time analysis system (REGARD): The case of the 2016 Kumamoto earthquakes. *Earth Planets and Space*, *68*(1), 190. https://doi.org/10.1186/s40623-016-0564-4

Kodera, Y., Hayashimoto, N., Moriwaki, K., Noguchi, K., Saito, J., Akutagawa, J., et al. (2020). First-year performance of a nationwide earthquake early warning system using a wavefield-based ground-motion prediction algorithm in Japan. *Seismological Research Letters*, *91*(2A), 826–834. https://doi.org/10.1785/0220190263

Kodera, Y., Yamada, Y., Hirano, K., Tamaribuchi, K., Adachi, S., Hayashimoto, N., et al. (2018). The propagation of local undamped motion (PLUM) method: A simple and robust seismic wavefield estimation approach for earthquake early warning. *Bulletin of the Seismological Society of America*, *108*(2), 983–1003. https://doi.org/10.1785/0120170085

Kohler, M. D., Smith, D. E., Andrews, J., Chung, A. I., Hartog, R., Henson, I., et al. (2020). Earthquake early warning ShakeAlert 2.0: Public rollout. *Seismological Research Letters*, *91*(3), 1763–1775. https://doi.org/10.1785/0220190245

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, *90*(1), 3–14. https://doi.org/10.1785/0220180259

Larson, K. M. (2009). GPS seismology. *Journal of Geodesy*, *83*(3–4), 227–233. https://doi.org/10.1007/s00190-008-0233-x

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

LeVeque, R. J., Waagan, K., González, F. I., Rim, D., & Lin, G. (2016). Generating random earthquake events for probabilistic tsunami hazard assessment. *Global Tsunami Science: Past and Future* (Vol. I, pp. 3671–3692). Cham: Birkhäuser. https://doi.org/10.1007/978-3-319-55480-8_2

Lin, J. T. (2021). *jiunting/MLARGE: First release of MLARGE (Version v1.0.0)*. Zenodo. https://doi.org/10.5281/zenodo.4527253

Lin, J. T., Chang, W. L., Melgar, D., Thomas, A., & Chiu, C. Y. (2019). Quick determination of earthquake source parameters from GPS measurements: A study of suitability for Taiwan. *Geophysical Journal International*, *219*(2), 1148–1162. https://doi.org/10.1093/gji/ggz359

Lin, J. T., Melgar, D., Thomas, A., & Searcy, J. (2020). *Chilean subduction zone rupture scenarios and waveform data [Data set]*. Zenodo. https://doi.org/10.5281/zenodo.5015610

Lomax, A., Michelini, A., & Jozinović, D. (2019). An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. *Seismological Research Letters*, *90*(2A), 517–529. https://doi.org/10.1785/0220180311

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml* (Vol. 30, p. 3).

Meier, M. A., Ampuero, J. P., & Heaton, T. H. (2017). The hidden simplicity of subduction megathrust earthquakes. *Science*, *357*(6357), 1277–1281. https://doi.org/10.1126/science.aan5643

Meier, M. A., Heaton, T., & Clinton, J. (2016). Evidence for universal earthquake rupture initiation behavior. *Geophysical Research Letters*, *43*(15), 7991–7996. https://doi.org/10.1002/2016gl070081

Melgar, D., Bock, Y., & Crowell, B. W. (2012). Real-time centroid moment tensor determination for large earthquakes from local and regional displacement records. *Geophysical Journal International*, *188*(2), 703–718. https://doi.org/10.1111/j.1365-246x.2011.05297.x

Melgar, D., Crowell, B. W., Bock, Y., & Haase, J. S. (2013). Rapid modeling of the 2011 Mw 9.0 Tohoku-Oki earthquake with seismogeodesy. *Geophysical Research Letters*, *40*(12), 2963–2968. https://doi.org/10.1002/grl.50590

Melgar, D., Crowell, B. W., Geng, J., Allen, R. M., Bock, Y., Riquelme, S., et al. (2015). Earthquake magnitude calculation without saturation from the scaling of peak ground displacement. *Geophysical Research Letters*, *42*(13), 5197–5205. https://doi.org/10.1002/2015gl064278

Melgar, D., Crowell, B. W., Melbourne, T. I., Szeliga, W., Santillan, M., & Scrivner, C. (2020). Noise characteristics of operational real-time high-rate GNSS positions in a large aperture network. *Journal of Geophysical Research: Solid Earth*, *125*(7), e2019JB019197. https://doi.org/10.1029/2019jb019197

Melgar, D., & Hayes, G. P. (2017). Systematic observations of the slip pulse properties of large earthquake ruptures. *Geophysical Research Letters*, *44*(19), 9691–9698. https://doi.org/10.1002/2017gl074916

Melgar, D., & Hayes, G. P. (2019). Characterizing large earthquakes before rupture is complete. *Science Advances*, *5*(5), eaav2032. https://doi.org/10.1126/sciadv.aav2032

Melgar, D., LeVeque, R. J., Dreger, D. S., & Allen, R. M. (2016). Kinematic rupture scenarios and synthetic displacement data: An example application to the Cascadia subduction zone. *Journal of Geophysical Research: Solid Earth*, *121*(9), 6658–6674. https://doi.org/10.1002/2016jb013314

Mena, B., Mai, P. M., Olsen, K. B., Purvance, M. D., & Brune, J. N. (2010). Hybrid broadband ground-motion simulation using scattering Green's functions: Application to large-magnitude events. *Bulletin of the Seismological Society of America*, *100*(5A), 2143–2162. https://doi.org/10.1785/0120080318

Minson, S. E., Baltay, A. S., Cochran, E. S., McBride, S. K., & Milner, K. R. (2021). Shaking is almost always a surprise: The earthquakes that produce significant ground motion. *Seismological Society of America*, *92*(1), 460–468. https://doi.org/10.1785/0220200165

Minson, S. E., Murray, J. R., Langbein, J. O., & Gomberg, J. S. (2014). Real-time inversions for finite fault slip models and rupture geometry based on high-rate GPS data. *Journal of Geophysical Research: Solid Earth*, *119*(4), 3201–3231. https://doi.org/10.1002/2013jb010622

Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, *47*(1), e2019GL085976. https://doi.org/10.1029/2019gl085976

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-17591-w

Murray, J. R., Crowell, B. W., Grapenthin, R., Hodgkinson, K., Langbein, J. O., Melbourne, T., et al. (2018). Development of a geodetic component for the US West Coast earthquake early warning system. *Seismological Research Letters*, *89*(6), 2322–2336. https://doi.org/10.1785/0220180162

Olson, E. L., & Allen, R. M. (2005). The deterministic nature of earthquake rupture. *Nature*, *438*(7065), 212–215. https://doi.org/10.1038/nature04214

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

Pitarka, A., Graves, R., Irikura, K., Miyakoshi, K., & Rodgers, A. (2020). Kinematic rupture modeling of ground motion from the M7 Kumamoto, Japan earthquake. *Pure and Applied Geophysics*, *177*(5), 2199–2221. https://doi.org/10.1007/s00024-019-02220-5

Riquelme, S., Medina, M., Bravo, F., Barrientos, S., Campos, J., & Cisternas, A. (2018). W-phase real-time implementation and network expansion from 2012 to 2017: The experience in Chile. *Seismological Research Letters*, *89*(6), 2237–2248. https://doi.org/10.1785/0220180146

Ross, Z. E., Meier, M. A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, *108*(5A), 2894–2901. https://doi.org/10.1785/0120180080

Ruhl, C. J., Melgar, D., Geng, J., Goldberg, D. E., Crowell, B. W., Allen, R. M., et al. (2019). A global database of strong-motion displacement GNSS recordings and an example application to PGD scaling. *Seismological Research Letters*, *90*(1), 271–279. https://doi.org/10.1785/0220180177

Ruiz, S., & Madariaga, R. (2018). Historical and recent large megathrust earthquakes in Chile. *Tectonophysics*, *733*, 37–56. https://doi.org/10.1016/j.tecto.2018.01.015

Rydelek, P., & Horiuchi, S. (2006). Is earthquake rupture deterministic? *Nature*, *442*(7100), E5–E6. https://doi.org/10.1038/nature04963

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929–1958.

Vallée, M., & Douet, V. (2016). A new database of source time functions (STFs) extracted from the SCARDEC method. *Physics of the Earth and Planetary Interiors*, *257*, 149–157. https://doi.org/10.1016/j.pepi.2016.05.012

van den Ende, M. P., & Ampuero, J. P. (2020). Automated seismic source characterization using deep graph neural networks. *Geophysical Research Letters*, *47*(17), e2020GL088690. https://doi.org/10.1029/2020gl088690

Williamson, A. L., Melgar, D., Crowell, B. W., Arcas, D., Melbourne, T. I., Wei, Y., & Kwong, K. (2020). Toward near-field tsunami forecasting along the Cascadia subduction zone using rapid GNSS source models. *Journal of Geophysical Research: Solid Earth*, *125*, e2020JB019636. https://doi.org/10.1029/2020jb019636

Wu, Y. M., & Zhao, L. (2006). Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophysical Research Letters*, *33*(16), L16312. https://doi.org/10.1029/2006gl026871

Ye, L., Lay, T., Kanamori, H., & Rivera, L. (2016). Rupture characteristics of major and great (Mw ≥ 7.0) megathrust earthquakes from 1990 to 2015: 1. Source parameter scaling relationships. *Journal of Geophysical Research: Solid Earth*, *121*(2), 826–844. https://doi.org/10.1002/2015jb012426

Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–273.

References From the Supporting Information

Duputel, Z., Tsai, V. C., Rivera, L., & Kanamori, H. (2013). Using centroid time-delays to characterize source durations and identify earthquakes with unique characteristics. *Earth and Planetary Science Letters*, *374*, 92–100. https://doi.org/10.1016/j.epsl.2013.05.024

Goda, K., Yasuda, T., Mori, N., & Maruyama, T. (2016). New scaling relationships of earthquake source parameters for stochastic tsunami simulation. *Coastal Engineering Journal*, *58*(3), 1650010-1–1650010-40. https://doi.org/10.1142/s0578563416500108

Graves, R., & Pitarka, A. (2015). Refinements to the Graves and Pitarka (2010) broadband ground-motion simulation method. *Seismological Research Letters*, *86*(1), 75–80. https://doi.org/10.1785/0220140101

Graves, R. W., & Pitarka, A. (2010). Broadband ground-motion simulation using a hybrid approach. *Bulletin of the Seismological Society of America*, *100*(5A), 2095–2123. https://doi.org/10.1785/0120100057

LeVeque, R. J., Waagan, K., González, F. I., Rim, D., & Lin, G. (2016). Generating random earthquake events for probabilistic tsunami hazard assessment. In *Global Tsunami Science: Past and Future* (Vol. I, pp. 3671–3692). Cham: Birkhäuser. https://doi.org/10.1007/978-3-319-55480-8_2

Mai, P. M., & Beroza, G. C. (2002). A spatial random field model to characterize complexity in earthquake slip. *Journal of Geophysical Research*, *107*(B11), ESE-10-1–ESE-10-21. https://doi.org/10.1029/2001jb000588

Melgar, D., & Hayes, G. P. (2017). Systematic observations of the slip pulse properties of large earthquake ruptures. *Geophysical Research Letters*, *44*(19), 9691–9698. https://doi.org/10.1002/2017gl074916

Melgar, D., & Hayes, G. P. (2019). The correlation lengths and hypocentral positions of great earthquakes. *Bulletin of the Seismological Society of America*, *109*(6), 2582–2593. https://doi.org/10.1785/0120190164

Melgar, D., LeVeque, R. J., Dreger, D. S., & Allen, R. M. (2016). Kinematic rupture scenarios and synthetic displacement data: An example application to the Cascadia subduction zone. *Journal of Geophysical Research: Solid Earth*, *121*(9), 6658–6674. https://doi.org/10.1002/2016jb013314

Mena, B., Mai, P. M., Olsen, K. B., Purvance, M. D., & Brune, J. N. (2010). Hybrid broadband ground-motion simulation using scattering Green's functions: Application to large-magnitude events. *Bulletin of the Seismological Society of America*, *100*(5A), 2143–2162. https://doi.org/10.1785/0120080318

Pasyanos, M. E., Masters, T. G., Laske, G., & Ma, Z. (2014). LITHO1. 0: An updated crust and lithospheric model of the Earth. *Journal of Geophysical Research: Solid Earth*, *119*(3), 2153–2173. https://doi.org/10.1002/2013jb010626

Zhu, L., & Rivera, L. A. (2002). A note on the dynamic and static displacements from a point source in multilayered media. *Geophysical Journal International*, *148*(3), 619–627. https://doi.org/10.1046/j.1365-246x.2002.01610.x