Accuracy and Uncertainty Analysis of Selected Methodological Approaches to Earthquake Early Warning in Europe

Gemma Cremen,* Elisa Zuccolo, and Carmine Galasso

January 2021

Declaration of Competing Interests

The authors acknowledge there are no conflicts of interest recorded.

Abstract

Earthquake early warning (EEW) is becoming an increasingly attractive real-time strategy for mitigating the threats posed by potentially devastating incoming seismic events. As efforts accelerate to develop practical EEW-based solutions for earthquake-prone countries in Europe, it is important to understand and quantify the level of performance that can be achieved by the underlying seismological algorithms. We conduct a conceptual study on EEW performance in Europe, which explicitly focuses on the accuracy and associated uncertainties of selected methodological approaches. 23 events from four diverse European testbeds are used to compare the quality of EEW predictions produced by the Virtual Seismologist and PRESTo algorithms. We first examine the location and magnitude estimates of the algorithms, accounting for both bias and uncertainty in the resulting predictions. We then investigate the ground-shaking prediction capabilities of the source-parameter estimates, using an error metric that can explicitly capture the propagation of

*Corresponding Author. Mailing address: Chadwick Building, Gower Street, London, W1CE 6DE, United Kingdom. Email address: g.cremen@ucl.ac.uk
uncertainties in these estimates. Our work highlights the importance of accounting for EEW parameter uncertainties, which are often neglected in studies of EEW performance. Our findings can be used to inform current and future implementations of EEW systems in Europe. In addition, the evaluation metrics presented in this work can be used to determine EEW accuracy in any worldwide setting.

Introduction

Earthquake early warning (EEW) systems are becoming increasingly popular tools for mitigating seismic risk in urban areas (e.g., Allen and Melgar 2019). It is therefore important to understand the extent to which these systems perform as intended. The performance of EEW systems mainly depends on a trade-off between: (1) the accuracy of the source parameter (i.e., magnitude, location) and/or ground motion estimates of the underlying EEW algorithm; and (2) the speed at which the system issues an alert (e.g., Behr et al. 2015). This paper specifically focuses on aspect (1) for EEW in Europe, which has been the subject of major research efforts in recent years (Clinton et al. 2016).

We study the theoretical (offline) accuracy of Virtual Seismologist (Cua 2005; Cua and Heaton 2007; Cua et al. 2009) and PRobabilistic and Evolutionary early warning SysTem (PRESTo) (Satriano et al. 2008a; Lancieri and Zollo 2008; Satriano et al. 2011), which have been the most widely applied regional EEW algorithms in Europe to date (Cremen and Galasso 2020). We specifically focus on the location, magnitude, and ground-shaking estimation capabilities of the algorithms across four European testbeds that capture a diverse range of seismicity, seismotectonics, and seismic network densities. Note that aspect (2) of earthquake performance is assessed for real-time versions of the algorithms and the same regions in a recent companion study (Zuccolo et al. 2020). (The Vrancea testbed of the companion study is ignored in this paper, since the Virtual Seismologist algorithm has not been calibrated for the large depths of its associated events).

Our accuracy assessments explicitly incorporate uncertainties associated with the earthquake parameter estimations of the algorithms, which are crucial for well-informed decision-making on alert triggering (Meier et al. 2015; Cremen and Galasso 2020). From an engineering perspective for example, these uncertainties play a central role in the real-time probabilistic seismic hazard analysis framework that can be used
to determine whether issuing an EEW alarm would reduce the losses associated with an incoming event (Iervolino, 2011). Our work therefore represents a significant advancement over many previous studies of context-specific EEW accuracy (e.g., Hsu et al., 2018; Xu et al., 2017; Kodera et al., 2016; Böse et al., 2012; Colombelli et al., 2012; Hsu et al., 2016; Böse et al., 2012; Colombelli et al., 2015; Hoshiba and Aoki, 2015; Böse et al., 2014; Doi, 2011; Hartog et al., 2016; Mittal et al., 2019; Chen et al., 2019; Chung et al., 2020; Minson et al., 2020; Zollo et al., 2009; Cochran et al., 2018; Festa et al., 2018; Auclair et al., 2015) - including those that examine Virtual Seismologist (Behr et al., 2016) and PRESTo (Picozzi et al., 2015) - which focus exclusively on the performance of point-estimate predictions (i.e., that only consider mean or modal values of the estimates rather than their probability distributions) from EEW algorithms. Some work has examined uncertainty propagation for EEW (i.e., the effect of uncertain source-parameter estimates on the final ground-shaking predictions), but this has so far been limited to the context of hypothetical algorithms (Meier, 2017), simplistic simulated events (Iervolino et al., 2009), or empirical error models of parameter estimates (Brown et al., 2011).

To facilitate our calculations, we code the complete location and magnitude modules of the Virtual Seismologist and PRESTo algorithms in the MATLAB language, including the Bayesian priors and other probabilistic details. This setup provides maximum flexibility to produce rigorous statistical comparisons between the accuracy of both algorithms.

This paper is structured as follows. Examined Testbeds introduces the testbeds and associated seismic data that form the basis of our accuracy assessments. We then provide the details of the algorithms to be examined in Examined Algorithms. The first part of Results examines the quality of the algorithms’ location and magnitude estimates. The second part determines the capability of the source-parameter estimates to accurately predict the corresponding ground-motion amplitude, using a novel evaluation metric that captures source-parameter uncertainties and does not require knowledge on the ground-shaking threshold used for triggering alerts in the EEW system. We end with a discussion of the results in Conclusions.
Examined Testbeds

We consider four European testbeds: Southern Italy (ITA), Pyrenees (PYR), Southwest Iceland (ICE), and Western Greece (GRE). These regions are chosen to capture a diverse range of European seismic hazard levels, seismic network densities, and seismotectonic settings: collisional/subduction complex with a complicated back-arc/fore-arc/trench system (ITA), continent-continent collision with evolution of an orogenic belt (PYR), oceanic crust interplate transform faulting (ICE), and ocean-continent subduction (GRE).

Considered Events

Our study examines 23 events in total across the four testbeds. (See Figure 1 for earthquake and seismic station locations and Table 1 for event characteristics). For ICE and GRE, we use real recordings from moderate-to-large (i.e., magnitude > 5.5) events that occurred in the last 20 years, for which data were recorded on at least eight seismic stations. Recordings for GRE are obtained from the European Integrated Data Archive (see Data and Resources). We prioritise strong-motion data and use broadband data in its absence, discarding saturated velocigrams. Parameters of the GRE events are obtained from the earthquake catalogues of the National Observatory of Athens (GRE). Strong-motion recordings for ICE are accessed through the Internet Site for European Strong-motion Data (ISESD, see Data and Resources), and corresponding event parameters are retrieved from the catalogue of Panzera et al. (2016).

We generate synthetic seismograms for our assessments of ITA and PYR, due to a lack of available moderate-to-large earthquake recordings in these regions. Physics-based numerical simulations are used to compute the seismograms, according to the broadband ground-motion simulation method described in Crempien and Archuleta (2015), which uses an extended kinematic model of the seismic source (subdivided into point sources) with correlated random source parameters (rupture time, peak time, rise time and final slip) that are based on more than 300 simulations of dynamic rupture models (Schmedes et al. 2013). Green’s functions are computed using a 1-D layered Earth model and a frequency-wavenumber (FK) code Zhu and Rivera (2002). This code, which is efficient in obtaining high-frequency synthetics, is coupled with a random perturbation of the point-source focal mechanism that partially accounts for scattering effects at high frequencies. We consider one scenario earthquake per active fault in both testbeds, and use fault
parameters from the European Database of Seismogenic Faults (ESDF; Basili et al., 2013) to determine the characteristics of these events according to the following procedures/assumptions: (1) We assign the magnitude of an event as a random sample from a uniform distribution between 5.5 and the maximum magnitude of the fault; (2) We assume that an event’s hypocentre is located at a depth equivalent to the minimum fault depth plus 2/3 of the vertical width; (3) We use the average values of the strike, dip, and rake angles to define the focal mechanism; (4) We randomly sample the average rupture velocity for each scenario event from a uniform distribution between 65% and 85% of the fault plane’s shear wave velocity. We use the Wells and Coppersmith (1994) relationships to define the rupture fault dimensions, and the Causse et al. (2008) distributions to determine hypocentral position along the fault plane. The stress drop, which is assumed to be 3 MPa for both testbeds (Caporali et al. 2011) is used to determine the corner frequency (Allmann and Shearer, 2009). We use the Barberi et al. (2004) crustal velocity model for ITA and the Theunissen et al. (2018) crustal velocity model for PYR. Broadband seismograms (0-25 Hz) are calculated at the locations of all currently operating permanent seismic stations of the IRIS database (see Data and Resources) are positioned within 100 km of each epicentre. White noise is finally added to each generated seismogram, to facilitate the automatic detection of P-wave arrivals based on the Short Time Average over Long Time Average STA/LTA algorithm (Allen, 1982).

Note that the synthetic seismograms for ITA were carefully validated before use. The validation procedure consisted of comparing the synthetics with: 1) recordings (i.e., a recording on rock of the Mw 6.0 1978 Patti Gulf earthquake); (2) synthetics from other authors generated for the same area with a different methodology (retrieved from the Synthesis portal, http://synthesis.mi.ingv.it/); and (3) European and Italian ground-motion models (GMMs). The quality of the computed seismograms was deemed to be high, particularly in terms of characteristics that are relevant for EEW algorithms (i.e., peak displacement and frequency), based on visual inspections and quantitative examinations with the goodness-of-fit test proposed by Olsen and Mayhew (2010), which measures the misfit between waveforms according to commonly used metrics that characterise their time series. A discussion on other testing/rating methodologies for validating simulated ground motions to be used in engineering applications can be found in Galasso et al. (2012, 2013), for instance.
Examined Algorithms

We examine the theoretical performance of the Virtual Seismologist (VS) (Cua, 2005; Cua and Heaton, 2007; Cua et al., 2009) and the PRobabilistic and Evolutionary early warning SysTem (PRESTo) (Lancieri and Zollo, 2008; Satriano et al., 2008b, 2011) regional EEW algorithms across all testbeds. The similar (Bayesian) structure of both algorithms enables direct comparisons to be made.

Virtual Seismologist (VS) operates within a Bayesian framework, in which the set of possible epicentral location and magnitude values are jointly conditioned on the ground-motion amplitude measurements (associated with P- and/or S-waves) at triggered stations and the prior PDF incorporates an existing state of knowledge on relative earthquake probability. The magnitude and epicentral location estimates are subsequently translated to peak ground-shaking predictions, using envelope attenuation relationships documented in Cua and Heaton (2007). VS was originally part of the ShakeAlert® EEW system in California, but the slow operational performance of the algorithm resulted in its removal in 2016 (Chung and Allen, 2019). A version of VS is operating in Switzerland and has been tested for use in Greece, Turkey, Romania and Iceland (Behr et al., 2016).

PRESTo estimates location using the $RTLoc$ method proposed by Satriano et al. (2008b) and predicts magnitude according to the $RTMag$ procedure developed by Lancieri and Zollo (2008). $RTLoc$ produces multivariate normal probability density functions of hypocentral locations, based on P-wave arrival times and a velocity model. $RTMag$ uses a Bayesian framework for estimating magnitude, in which the likelihood function depends on initial peak displacement measurements and $RTLoc$ outputs. The prior PDF for a given time step is the posterior distribution obtained at the previous time step, and the prior for the first time step is optionally set as the Gutenberg-Richter distribution. Peak ground-motion parameters are computed based on the location and magnitude estimates, using a GMM. PRESTo is currently operating in real-time in Southern Italy, Turkey, Romania, and South Korea (Picozzi et al., 2015), and has also been tested for application in Austria and Slovenia (Picozzi et al., 2015), as well as the Iberian Peninsula (Pazos et al., 2015).
Algorithm Inputs

Location Inputs

We use a common (neutral) method to determine event arrival times for both algorithms, given that P-wave picking accuracy is not the focus of our evaluation. We leverage the SeisComP seismological software (see Data and Resources) and specifically use the picks associated with its preferred origin for a given event.

This origin is automatically selected using the scevent module of the software, according to a number of predefined rules (e.g., an origin is preferred to the previous one if it is computed using a greater number of picks and/or produces lower travel time residuals, etc). Velocity models input to the PRESTo algorithm are region-specific. The velocity models used for ITA and PYR are the same as those adopted for the generation of synthetic seismograms in both testbeds (see Considered Events). We use the Tryggvason et al. (2002) model for ICE and the Rigo et al. (1996) model for GRE. We use the Poisson’s solid approximation to derive undefined P-wave velocities from associated S-wave velocities (and vice versa), and we compute corresponding 3D travel-time grids using the NonLinLoc software (Lomax et al., 2000).

Magnitude Inputs

We estimate magnitudes based on seismogram data from stations that are associated with the preferred origin location estimated by SeisComP. The seismograms are first processed as follows (Boore and Bommer, 2005):

1. We apply instrument correction to GRE recordings;
2. We apply baseline correction;
3. We differentiate velocigrams to retrieve accelerograms;
4. We apply a high-pass filter with 3 s corner frequency for VS (Cua and Heaton, 2007) and a band-pass filter (0.075 - 3 Hz) to accelerograms for PRESTo (Satriano et al., 2011);
5. We integrate the accelerograms once to retrieve velocities, and twice to obtain displacements. For VS, we then extract maximum envelope values of vertical acceleration ($Z_A$), velocity ($Z_V$), displacement ($Z_D$), and root mean square horizontal acceleration ($H_A$), velocity ($H_V$) and displacement ($H_V$), computed within 1 s intervals starting from the P-wave arrival time. For PRESTo, we extract values of peak displacement ($P_d$) in three different time windows, accounting for the vector modulus of the three-component seismogram (Satriano et al., 2011). These time windows are (1) 2 s following the P-wave arrival if the P and S arrivals are at least 2 s apart; (2) 4 s after the P-wave arrival if the P and S arrivals are at least 4 s apart; and (3) 2
s following the S-wave arrival.

**Bayesian Priors**

Parameterisation of the Bayesian location prior for the VS algorithm depends on the number of stations triggered at a given instant, and spatial constraints provided by data associated with not-yet arrived P-waves (Cua and Heaton, 2007). If only one station has triggered, the location is constrained to the area of the associated Voronoi cell that is geometrically consistent with the surrounding non-triggered stations. For two triggers, the location is assumed to lie on the hyperbola passing between both stations, in line with the methodology described by Rydelek and Pujol (2004). For three triggers, the location is constrained to one point, i.e. the intersection of the two hyperbolae that pass between all triggered stations. All possible locations included in the prior are assigned equal weighting (i.e., a uniform distribution), and every other spatial point is assigned zero probability. The P-wave velocity used to determine P-wave arrivals at stations (for computing the location constraints) is taken as the average value within a 10 km depth, according to the appropriate velocity model provided in [Location Inputs](#).

The maximum magnitude and scaling ($b$) parameter of the Gutenberg-Richter distribution required for defining the VS Bayesian magnitude prior are retrieved for each event from the nearest point on a 0.1 degree by 0.1 degree grid of the European Seismic Hazard Model (ESHM13) (Woessner et al., 2015) model. A minimum magnitude of 4 is assumed in all cases. We use the same distribution for the Bayesian prior of the PRESTo algorithm.

**Results**

**Location and Magnitude Accuracy**

We quantify the accuracy of the algorithms’ location and magnitude components independently, to determine the accuracy of the estimates for different levels of algorithmic uncertainty in the source parameters. Since the quality of estimates should increase in time while the uncertainty decreases, this assessment is designed to capture various accuracy/speed trade-off thresholds that may be of interest to stakeholders for guiding
decision-making and EEW alert issuance. Location and magnitude accuracy are quantified for each algorithm using the root mean square error (RMSE) metric \cite{Hyndman2006}.

We compare the location estimates in terms of their epicentral distance to the following selected target sites in each region (values in parentheses respectively indicate longitudes and latitudes): (1) the port of Gioia Tauro (15.91°, 38.46°) in ITA, (2) Andorra (1.60°, 42.54°) in PYR, (3) Reykjavik (-21.94°, 64.15°) in ICE, and (4) Patras (21.73°, 38.25°) in GRE. The uncertainty levels considered for the estimates in this case are expressed in the form of coefficients of variation (CV, i.e., the ratio of the standard deviation to the mean) rather than standard deviations. This is because CV provides a measure of relative uncertainty, which is more appropriate for the large range of source-to-target distances used in the study. We specifically examine the mean distance prediction of each algorithm for the first estimate that has uncertainty lower than CV = 0.3, CV = 0.2, and CV = 0.1 (Figure 2).

It can be seen that PRESTo yields the best distance predictions across all uncertainty levels investigated except CV = 0.3. Its associated RMSE value is over 35% lower than that of VS for both CV = 0.2 and CV = 0.1, whereas the RMSE value for VS is 28% smaller in the largest uncertainty case (note that no consistent performance differences are observed between real and simulated events). If we consider a hypothetical $M_w$ 6 normal-faulting earthquake with $V_s30 = 800$ m/s and use the epicentral distance version of the Akkar et al. \cite{Akkar2014} GMM for the observed RMSE values, the median PGA predictions obtained for both algorithms are noticeably different, with the discrepancies ranging between 29% and 64% across the three cases. It is interesting to note that the VS RMSE values for epicentral distance increase as uncertainty decreases, which is the opposite of what is intuitively expected \cite{Cochran2018}. This observation is due to the effect of the algorithm’s Bayesian prior, which significantly narrows the range of location estimates (and therefore their uncertainty) after only two P-wave arrivals, thereby preventing any significant further improvements that may be achieved using information from additional stations.

We compare the mean magnitude predictions of both algorithms, using the first estimates with standard deviations ($\sigma_M$) below the following thresholds: 0.1, 0.2, and 0.3 (Figure 3). It can be observed that the results of the PRESTo algorithm are most accurate for all levels of uncertainty investigated. The PRESTo RMSE value is approximately 15% lower than that of VS for $\sigma_M = 0.2$, and over 20% lower in both other
cases (note that there are no distinct differences between the performance trends for real and simulated events). If we take a hypothetical $M_w$ 6 + RMSE normal-faulting earthquake with $V_{s30} = 800$ m/s and use the epicentral distance version of the Akkar et al. (2014) GMM at 30 km, we obtain non-negligible differences between the resulting median PGA predictions for the RMSE values of both algorithms. These differences range between 9% and 14% across the three uncertainty thresholds.

Propagation of Uncertainties and their Effect on Ground Motion

Finally, we investigate the impact of the algorithms’ location and magnitude estimation accuracy on the quality of resulting ground-shaking predictions, using the epicentral distance version of the Akkar et al. (2014) GMM. We simply assume rock ground conditions ($V_{s30} = 800$ m/s) in all cases (for both true and predicted ground shaking), given that site class does not influence the assessment of ground-motion accuracy related to location and magnitude, and use the style-of-faulting information provided in Table 1. We specifically focus on peak ground acceleration (PGA) predictions in this investigation.

PGA prediction accuracy is quantified with the $MD$ metric for sensitivity analyses (Chun et al., 2000), which has been used to examine the performance of GMMs in previous work (Cremen et al., 2020) and to determine the ground-shaking prediction accuracy of EEW algorithms in our companion paper (Zuccolo et al., 2020). For our application, $MD$ measures the difference between the cumulative distribution function (CDF) of PGA produced when the true source parameters are used in the GMM and the CDF obtained when an algorithm’s estimated source parameters are input to the model.

This type of comparison is useful if EEW alerts are issued based on a given probability of exceeding a prescribed value of PGA (Iervolino, 2011). This is because discrepancies in the CDFs indicate the potential for incorrect decisions to be made on whether or not to trigger an alarm (Iervolino, 2011). For example, a false alert may occur if the predicted PGA is greater than the actual PGA value at the exceedance threshold. On the other hand, an alert may be missed if this prediction is less than the true value. Our evaluation offers a significant advantage over many previous studies of EEW ground-shaking or intensity accuracy (e.g., Meier, 2017; Cochran et al., 2019; Minson et al., 2019; Meier et al., 2020), since it does not require the a priori definition of a subjective alert threshold.
As an advancement over our companion paper, we use a version of the $MD$ metric that can explicitly account for the propagated uncertainty of the EEW source-parameter estimates in the resulting PGA CDF. We use Monte Carlo sampling of the underlying distributions to capture all uncertainties, and calculate $MD$ according to the following equation:

$$MD = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left[ \frac{y_{n/N}^{i}}{y_{n/N}^{o}} - \frac{y_{n/N}^{o}}{y_{n/N}^{o}} \right]^2}$$

where $N$ is the number of Monte Carlo samples used (= 5,000 in this case), $n$ is the sample index, $y_{n/N}^{o}$ is the $(n/N)$th quantile of the true GMM CDF (0 < n < N), $y_{n/N}^{i}$ is the equivalent quantile for the predicted GMM distribution, and the denominator represents the mean of the true CDF. A lower value indicates a higher similarity between the predicted and true distributions of PGA. Figure 4 demonstrates the evaluation procedure, displaying the $MD$ values and corresponding PGA CDFs obtained for one scenario event.

We examine ground-shaking prediction accuracy by combining the estimations of location and magnitude obtained in Location and Magnitude Accuracy. We specifically consider the source-parameter estimates associated with the following thresholds of algorithmic uncertainties in location ($CV_R$) and magnitude ($\sigma_M$):

$CV_R = \sigma_M = 0.3$, $CV_R = \sigma_M = 0.2$, and $CV_R = \sigma_M = 0.1$. Our first investigation uses only the relevant point (i.e., mean) estimates for each algorithm. The corresponding $MD$ values obtained for both algorithms and each event are displayed in Figures 5a to 5c. It can be observed that PRESTo produces the lowest average $MD$ value across two considered uncertainty levels - this value is 38% lower for $CV_R = \sigma_M = 0.3$, and 23% lower for $CV_R = \sigma_M = 0.1$ - and approximately equivalent results are obtained for both algorithms in the case $CV_R = \sigma_M = 0.2$. Thus, we can generally conclude that PRESTo is the best algorithm in terms of ground-shaking prediction accuracy, when only point estimates of source parameter measurements are considered. (Note that there are no consistent differences between the $MD$ values observed for real and simulated events, in any examined case.)

We then determine the ground-shaking accuracy obtained when the uncertainty of the source-parameter estimates is propagated through to the PGA CDF. Corresponding $MD$ values for both algorithms and each event are displayed in Figures 5d to 5f. Firstly, this analysis produces larger average $MD$ values than those obtained when only point estimates of source parameters are considered, in almost all cases. This implies that
neglecting the underlying uncertainty of the EEW source parameter measurements tends to overestimate the
accuracy of the resulting ground-shaking predictions. PRESTo still produces more accurate PGA estimations
for the $CV_R = \sigma_M = 0.3$ and $CV_R = \sigma_M = 0.1$ uncertainty levels, for which the average $MD$ values for
PRESTo are respectively 22% and 7% lower than those of VS. However, the performance of VS is noticeably
better that that of PRESTo for $CV_R = \sigma_M = 0.2$; the average VS $MD$ value in this case is 12% lower than
the corresponding PRESTo value. Thus, the optimal algorithm for ground-shaking prediction in this case
depends on the level of uncertainty in the source parameters.

To provide some context on the practical consequences of the observed differences in $MD$ values, we
consider their implications for the specific case of predicting the median value of a hypothetical, realistic
(lognormal) GMM distribution with true median $= 0.5g$, known dispersion $= 0.7$, and predicted dispersion
$= 0.9$. Table 2 provides the median values of the predicted distributions that would lead to the $MD$ values
observed in Figure 5. It can be seen from the table that the median predictions associated with unique $MD$
values vary noticeably, such that they could feasibly lead to different decision outcomes (i.e., trigger/don’t
trigger alert) for an EEW alarm based on the median PGA prediction exceeding a particular threshold. For
example, based on the average $MD$ values of Figure 5a, a median prediction alert threshold of 0.7 g would
cause a false alert to be issued using the VS algorithm but would correctly result in no alarm being triggered
for the PRESTo algorithm prediction. We can thus conclude that the differences observed between $MD$
values have practical implications on the accuracy of ground-shaking estimates and EEW triggering.

Conclusions

This study has conceptually examined the offline accuracy of the VS and PRESTO regional EEW algo-
rithms, using seismic waveforms of 23 real and simulated events across four geographically disperse testbeds
in Europe. Our analyses have explicitly accounted for uncertainty in the algorithms’ source parameter mea-
surements, which represents a significant advancement over many previous studies of EEW performance that
only consider algorithmic point estimates.

We first assessed the algorithms’ mean source-parameter estimates, which corresponded to various un-
certainty thresholds that stakeholders may use to guide decision-making on EEW alert triggering. We found
that PRESTo was almost consistently the best-performing algorithm in terms of both location and magnitude estimation. PRESTo location estimates were over 35% more accurate than those of VS (except in the case of relatively large source-parameter uncertainty, i.e. CV$_R$=0.3), and its magnitude estimates were approximately 15 to 20% better. We therefore conclude that PRESTo should be used for EEW purposes that require mean estimates of location and magnitude, which is consistent with the recommendations of our companion paper (Zuccolo et al., 2020) that compared the real-time operational performance of both algorithms.

We also compared the capabilities of both algorithms in terms of ground-shaking (i.e., PGA) prediction, using a well-known European GMM. Accuracy at this stage of EEW is crucial if alerts are issued based on a given probability of exceeding a prescribed value of ground-motion amplitude or intensity. We used a technique leveraged from sensitivity analysis to quantify the quality of GMM predictions for a given set of location and magnitude estimates, which can also account for the propagation of their underlying uncertainties.

We found that PRESTo was the best algorithm for ground-shaking prediction, if only point estimates of the source parameters were used to determine the resulting distribution of PGA values. However, our conclusion changed when the uncertainty of the source parameters was also accounted for in the CDF of ground-motion amplitude. In this case, the performance of the VS algorithm was superior for the middle level of source-parameter uncertainty considered. We ultimately conclude that the best-performing algorithm in terms of ground-shaking prediction depends on the uncertainties introduced by the underlying source measurements. In addition, the accuracy of the predicted PGA distribution decreases when the source-parameter uncertainties are propagated through the calculations in almost all examined cases, which implies that neglecting this uncertainty tends to result in an overestimation of algorithm performance. In summary, our analyses clearly highlight the importance of explicitly accounting for source-parameter uncertainties when measuring the accuracy of final EEW predictions.

It is important to note that there are some limitations associated with this work. Firstly, the calibration of phase detection and association parameters was only carried out for the events examined in this study, and may not reflect the overall seismicity and network geometry of each area. Secondly, we did not explore the
sensitivity of the algorithms’ Bayesian priors. For example, magnitude priors retrieved from regional hazard studies (instead of the European ESHM13 model) may have resulted in the better performance of a given algorithm. In addition, a location prior with less severe constraints may have improved the accuracy of VS estimates. However, average $MD$ values obtained using the VS magnitude estimates and correct distance measurements yield the same conclusions on the relative performance of the algorithms as those presented in Propagation of Uncertainties and their Effect on Ground Motion tentatively suggesting that location accuracy (and therefore the choice of location prior) does not have a significant effect on the quality of ground-shaking estimates associated with VS (note that a more concrete conclusion on the effect of the VS prior chosen would also require an investigation of its influence on the accuracy of VS magnitude estimates). Furthermore, the empirical scaling relationships used to estimate magnitude in the algorithms may not be appropriate for all considered regions. Finally, our results may not reflect the performance of the algorithms across all European sites or regions, however the evaluation procedures presented in this paper could be used to conduct more detailed accuracy analyses for specific case studies. Despite the aforementioned constraints, our study nevertheless provides some notable insights on the accuracy and uncertainty of EEW estimates for European seismicity.

**Data and Resources**

No new data were created as part of this study. The European Integrated Data Archive was retrieved via the ORFEUS Data Center WebDC3 Web Interface at https://www.orfeus-eu.org/data/eida/ (last accessed April 2020). The Internet Site for European Strong-motion Data used was http://www.isesd.hi.is/ (last accessed March 2020). Station metadata were obtained from The International Federation of Digital Seismograph Networks, available at http://www.fdsn.org/ (last accessed May 2020). The IRIS (Incorporated Research Institutions for Seismology) station database was consulted at https://ds.iris.edu/gmap (last accessed February 2020). The National Observatory of Athens (NOA) earthquake catalogue was obtained at https://bbnet.gein.noa.gr/HL (last accessed May 2020). The European Database of Seismogenic Faults (ESDF) was retrieved from http://diss.rm.ingv.it/share-edsf/ (last accessed January 2020). Figures for this manuscript were produced using the Matplotlib Python library [Hunter 2007] and MATLAB®. The Seis-
Acknowledgements

We appreciate the thoughtful comments of two anonymous reviewers that significantly improved the article. This research is funded by the European Union’s Horizon 2020 research and innovation programme, specifically grant agreement number 821046: TURNkey - “Towards more Earthquake-resilient Urban Societies through a Multi-sensor-based Information System enabling Earthquake Forecasting, Early Warning and Rapid Response actions”. We thank Nikos Melis and Stratos Liadopoulos (National Observatory of Athens, Greece), Benedikt Halldorsson and Atefe Darzi (University of Iceland, Iceland), and Pierre Gehl (Bureau de Recherches Géologiques et Minières, France) for providing regional input data. We also acknowledge Bernd Weber, Dirk Roessler (gempa GmbH, Germany) and their wider team for providing SeiscomP and assisting with its running.

References

Akkar, S., M. Sandıkkaya, and J. Bommer (2014). Empirical ground-motion models for point-and extended-source crustal earthquake scenarios in Europe and the Middle East, Bull. Earthquake Eng. 12(1), 359–387.

Akkar, S., M. Sandıkkaya, M. Şenyurt, A. A. Sisi, B. Ay, P. Traversa, J. Douglas, F. Cotton, L. Luzi, B. Hernandez, et al. (2014). Reference database for seismic ground-motion in Europe (RESORCE), Bull. Earthquake Eng. 12(1), 311–339.

Allen, R. (1982). Automatic phase pickers: Their present use and future prospects, Bulletin of the Seismological Society of America 72(6B), S225–S242.

Allen, R. M. and D. Melgar (2019). Earthquake Early Warning: Advances, Scientific Challenges, and Societal Needs, Annu. Rev. Earth Planet Sci. 47.
Allmann, B. P. and P. M. Shearer (2009). Global variations of stress drop for moderate to large earthquakes, *J. Geophys. Res. Solid Earth* 114(1).

Auclair, S., X. Goula, J.-A. Jara, and Y. Colom (2015). Feasibility and interest in earthquake early warning systems for areas of moderate seismicity: Case study for the Pyrenees, *Pure Appl. Geophys.* 172(9), 2449–2465.

Barberi, G., M. Cosentino, A. Gervasi, I. Guerra, G. Neri, and B. Orecchio (2004). Crustal seismic tomography in the Calabrian Arc region, south Italy, *Phys. Earth Planet. Inter.* 147(4), 297–314.

Basili, R., M. M. Tiberti, V. Kastelic, F. Romano, A. Piatanesi, J. Selva, and S. Lorito (2013). Integrating geologic fault data into tsunami hazard studies, *Nat. Hazards Earth Syst. Sci.* 13(4), 1025–1050.

Behr, Y., J. Clinton, P. Kästli, C. Cauzzi, R. Racine, and M.-A. Meier (2015). Anatomy of an earthquake early warning (EEW) alert: Predicting time delays for an end-to-end EEW system, *Seismol. Res. Lett.* 86(3), 830–840.

Behr, Y., J. F. Clinton, C. Cauzzi, E. Hauksson, K. Jónsdóttir, C. G. Marius, A. Pinar, J. Salichon, and E. Sokos (2016). The Virtual Seismologist in SeisComP3: A new implementation strategy for earthquake early warning algorithms, *Seismol. Res. Lett.* 87(2A), 363–373.

Boore, D. M. and J. J. Bommer (2005). Processing of strong-motion accelerograms: Needs, options and consequences, *Soil Dyn. Earthq. Eng.* 25(2), 93–115.

Böse, M., R. Allen, H. Brown, G. Gua, M. Fischer, E. Hauksson, T. Heaten, M. Hellweg, M. Liukis, D. Neuhauser, et al. (2014). CISN ShakeAlert: An earthquake early warning demonstration system for California. In *Early warning for geological disasters*, pp. 49–69. Springer.

Böse, M., T. H. Heaton, and E. Hauksson (2012). Real-time finite fault rupture detector (FinDer) for large earthquakes, *Geophys. J. Int.* 191(2), 803–812.

Brown, H. M., R. M. Allen, M. Hellweg, O. Khainovski, D. Neuhauser, and A. Souf (2011). Development of the ElarmS methodology for earthquake early warning: Realtime application in California and offline testing in Japan, *Soil Dyn. Earthq. Eng.* 31(2), 188–200.
Caporali, A., S. Barba, M. M. Carafa, R. Devoti, G. Pietrantonio, and F. Riguzzi (2011). Static stress drop as determined from geodetic strain rates and statistical seismicity, *J. Geophys. Res. Solid Earth* 116(2).

Causse, M., F. Cotton, C. Cornou, and P. Y. Bard (2008). Calibrating median and uncertainty estimates for a practical use of empirical Green’s functions technique, *Bull. Seismol. Soc. Am.* 98(1), 344–353.

Chen, D.-Y., T.-L. Lin, H.-C. Hsu, Y.-C. Hsu, and N.-C. Hsiao (2019). An approach to improve the performance of the earthquake early warning system for the 2018 Hualien earthquake in Taiwan, *Terr. Atmos. Ocean. Sci.* 30, 423–433.

Chun, M.-H., S.-J. Han, and N.-I. Tak (2000). An uncertainty importance measure using a distance metric for the change in a cumulative distribution function, *Reliab. Eng. Syst. Saf.* 70, 313–321.

Chung, A. I. and R. M. Allen (2019). Optimizing Earthquake Early Warning Performance: ElarmS-3 ElarmS View project Building Response Using Ambient Noise View project, *Seismol. Res. Lett.* 90(2A).

Chung, A. I., M.-A. Meier, J. Andrews, M. Böse, B. W. Crowell, J. J. McGuire, and D. E. Smith (2020). ShakeAlert earthquake early warning system performance during the 2019 Ridgecrest earthquake sequence, *Bull. Seismol. Soc. Am.* 110(4), 1904–1923.

Clinton, J., A. Zollo, A. Marmureanu, C. Zulfikar, and S. Parolai (2016). State-of-the-art and future of earthquake early warning in the European region, *Bull. Earthq. Eng.* 14(9), 2441–2458.

Cochran, E. S., J. Bunn, S. E. Minson, A. S. Baltay, D. L. Kilb, Y. Kodera, and M. Hoshiba (2019). Event Detection Performance of the PLUM Earthquake Early Warning Algorithm in Southern California, *Bull. Seismol. Soc. Am.* 109(4), 1524–1541.

Cochran, E. S., M. D. Kohler, D. D. Given, S. Guiwits, J. Andrews, M.-A. Meier, M. Ahmad, I. Henson, R. Hartog, and D. Smith (2018). Earthquake early warning ShakeAlert system: Testing and certification platform, *Seismol. Res. Lett.* 89(1), 108–117.

Colombelli, S., O. Amoroso, A. Zollo, and H. Kanamori (2012). Test of a threshold-based earthquake early-warning method using Japanese data, *Bull. Seismol. Soc. Am.* 102(3), 1266–1275.
Colombelli, S., A. Caruso, A. Zollo, G. Festa, and H. Kanamori (2015). A P-wave-based, on-site method for earthquake early warning, *Geophys. Res. Lett.* **42**(5), 1390–1398.

Cremen, G. and C. Galasso (2020). Earthquake Early Warning: Recent Advances and Perspectives, *Earth-Sci. Rev.*, 103184.

Cremen, G., M. J. Werner, and B. Baptie (2020). A New Procedure for Evaluating Ground-Motion Models, with Application to Hydraulic-Fracture-Induced Seismicity in the United Kingdom, *Bull. Seismol. Soc. Am.*

Crempien, J. G. and R. J. Archuleta (2015). UCSB method for simulation of broadband ground motion from kinematic earthquake sources, *Seismol. Res. Lett.* **86**(1), 61–67.

Cua, G., M. Fischer, T. Heaton, and S. Wiemer (2009). Real-time performance of the virtual seismologist earthquake early warning algorithm in Southern California, *Seismol. Res. Lett.* **80**(5), 740–747.

Cua, G. and T. Heaton (2007). The Virtual Seismologist (VS) method: a Bayesian approach to earthquake early warning. In *Earthquake Early Warning Systems*, pp. 97–132.

Cua, G. B. (2005). *Creating the Virtual Seismologist: Developments in Ground Motion Characterization and Seismic Early Warning*. Ph. D. thesis, California Institute of Technology.

Doi, K. (2011). The operation and performance of earthquake early warnings by the Japan Meteorological Agency, *Soil Dyn. Earthq. Eng.* **31**(2), 119–126.

Festa, G., M. Picozzi, A. Caruso, S. Colombelli, M. Cattaneo, L. Chiaramale, L. Elia, C. Martino, S. Marzorati, M. Supino, et al. (2018). Performance of earthquake early warning systems during the 2016–2017 Mw 5–6.5 Central Italy sequence, *Seismol. Res. Lett.* **89**(1), 1–12.

Galasso, C., F. Zareian, I. Iervolino, and R. Graves (2012). Validation of ground-motion simulations for historical events using SDoF systems, *Bull. Seismol. Soc. Am.* **102**(6), 2727–2740.

Galasso, C., P. Zhong, F. Zareian, I. Iervolino, and R. W. Graves (2013). Validation of ground-motion simulations for historical events using MDoF systems, *Earthq. Eng. Struct. Dyn.* **42**(9), 1395–1412.
Hartog, J. R., V. C. Kress, S. D. Malone, P. Bodin, J. E. Vidale, and B. W. Crowell (2016). Earthquake Early Warning: ShakeAlert in the Pacific Northwest, *Bull. Seismol. Soc. Am.* 106(4), 1875–1886.

Hoshiba, M. and S. Aoki (2015). Numerical shake prediction for earthquake early warning: Data assimilation, real-time shake mapping, and simulation of wave propagation, *Bull. Seismol. Soc. Am.* 105(3), 1324–1338.

Hsu, T., P. Lin, H. Wang, H. Chiang, Y. Chang, C. Kuo, C. Lin, and K. Wen (2018). Comparing the performance of the NEEWS earthquake early warning system against the CWB system during the 6 February 2018 Mw 6.2 Hualien earthquake, *Geophys. Res. Lett.* 45(12), 6001–6007.

Hsu, T.-Y., H.-H. Wang, P.-Y. Lin, C.-M. Lin, C.-H. Kuo, and K.-L. Wen (2016). Performance of the NCREE’s on-site warning system during the 5 February 2016 Mw 6.53 Meinong earthquake, *Geophys. Res. Lett.* 43(17), 8954–8959.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment, *Comput. Sci. Eng.* 9(3), 90–95.

Hyndman, R. J. and A. B. Koehler (2006). Another look at measures of forecast accuracy, *Int. J. Forecast.* 22(4), 679–688.

Iervolino, I. (2011). Performance-based earthquake early warning, *Soil Dyn. Earthq. Eng.* 31(2), 209–222.

Iervolino, I., M. Giorgio, C. Galasso, and G. Manfredi (2009). Uncertainty in early warning predictions of engineering ground motion parameters: What really matters?, *Geophys. Res. Lett.* 36(4), L00B06.

Kodera, Y., J. Saitou, N. Hayashimoto, S. Adachi, M. Morimoto, Y. Nishimae, and M. Hoshiba (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(1), 202.

Lancieri, M. and A. Zollo (2008). A Bayesian approach to the real-time estimation of magnitude from the early P and S wave displacement peaks, *J. Geophys. Res. Solid Earth* 113(12).

Lomax, A., J. Vireux, P. Volant, and C. Berge-Thierry (2000). Probabilistic Earthquake Location in 3D and Layered Models. In *Advances in Seismic Event Location*. Springer, Dordrecht.
Meier, M.-A. (2017). How “good” are real-time ground motion predictions from earthquake early warning systems?, *J. Geophys. Res. Solid Earth* 122(7), 5561–5577.

Meier, M.-A., T. Heaton, and J. Clinton (2015). The Gutenberg algorithm: Evolutionary Bayesian magnitude estimates for earthquake early warning with a filter bank, *Bull. Seismol. Soc. Am.* 105(5), 2774–2786.

Meier, M. A., Y. Kodera, M. Böse, A. Chung, M. Hoshiba, E. Cochran, S. Minson, E. Hauksson, and T. Heaton (2020). How Often Can Earthquake Early Warning Systems Alert Sites With High-Intensity Ground Motion?, *J. Geophys. Res. Solid Earth* 125(2).

Minson, S. E., A. S. Baltay, E. S. Cochran, T. C. Hanks, M. T. Page, S. K. McBride, K. R. Milner, and M. A. Meier (2019). The Limits of Earthquake Early Warning Accuracy and Best Alerting Strategy, *Sci. Rep.* 9(1).

Minson, S. E., J. K. Saunders, J. J. Bunn, E. S. Cochran, A. S. Baltay, D. L. Kilb, M. Hoshiba, and Y. Kodera (2020). Real-Time Performance of the PLUM Earthquake Early Warning Method during the 2019 M 6.4 and 7.1 Ridgecrest, California, Earthquakes, *Bull. Seismol. Soc. Am.* 110(4), 1887–1903.

Mittal, H., Y.-M. Wu, M. L. Sharma, B. M. Yang, and S. Gupta (2019). Testing the performance of earthquake early warning system in northern India, *Acta Geophys.* 67(1), 59–75.

Olsen, K. B. and J. E. Mayhew (2010). Goodness-of-fit criteria for broadband synthetic seismograms, with application to the 2008 Mw 5.4 Chino Hills, California, earthquake, *Seismol. Res. Lett.* 81(5), 715–723.

Panzera, F., J. D. Zechar, K. S. Vogfjörd, and D. A. Eberhard (2016). A Revised Earthquake Catalogue for South Iceland, *Pure Appl. Geophys.* 173(1), 97–116.

Pazos, A., N. Romeu, L. Lozano, Y. Colom, M. López Mesa, X. Goula, J. A. Jara, J. V. Cantavella, A. Zollo, W. Hanka, and F. Carrilho (2015). A regional approach for earthquake early warning in South West Iberia: A feasibility study, *Bull. Seismol. Soc. Am.* 105(2), 560–567.

Picozzi, M., L. Elia, D. Pesaresi, A. Zollo, M. Mucciarelli, A. Gosar, W. Lenhardt, and M. Živic (2015). Trans-national earthquake early warning (EEW) in north-eastern Italy, Slovenia and Austria: first experience with PRESTo at the CE3RN network, *Adv. Geosci* 40, 51–61.
Picozzi, M., A. Zollo, P. Brondi, S. Colombelli, L. Elia, and C. Martino (2015). Exploring the feasibility of a nationwide earthquake early warning system in Italy, *J. Geophys. Res. Solid Earth* **120**(4), 2446–2465.

Rigo, A., H. Lyon-Caen, R. Armijo, A. Deschamps, D. Hatzfeld, K. Makropoulos, P. Papadimitriou, and I. Kassaras (1996). A microseismic study in the western part of the Gulf of Corinth (Greece): implications for large-scale normal faulting mechanisms, *Geophys. J. Int.* **126**(3), 663–688.

Rydelek, P. and J. Pujol (2004). Real-Time Seismic Warning with a Two-Station Subarray, *Bull. Seismol. Soc. Am.* **94**(4), 1546–1550.

Satriano, C., L. Elia, C. Martino, M. Lancieri, A. Zollo, and G. Iannaccone (2011). PRESTo, the earthquake early warning system for Southern Italy: Concepts, capabilities and future perspectives, *Soil Dyn. Earthq. Eng.* **31**(2), 137–153.

Satriano, C., A. Lomax, and A. Zollo (2008a). Real-time evolutionary earthquake location for seismic early warning, *Bull. Seismol. Soc. Am.* **98**(3), 1482–1494.

Satriano, C., A. Lomax, and A. Zollo (2008b). Real-Time Evolutionary Earthquake Location for Seismic Early Warning, *Bull. Seismol. Soc. Am.* **98**(3), 1482–94.

Satriano, C., Y. M. Wu, A. Zollo, and H. Kanamori (2011). Earthquake early warning: Concepts, methods and physical grounds, *Soil Dyn. Earthq. Eng.* **31**(2), 106–118.

Schmedes, J., R. J. Archuleta, and D. Lavallée (2013). A kinematic rupture model generator incorporating spatial interdependency of earthquake source parameters, *Geophys. J. Int.* **192**(3), 1116–1131.

Theunissen, T., S. Chevrot, M. Sylvander, V. Monteiller, M. Calvet, A. Villaseñor, S. Benahmed, H. Pauchet, and F. Grimaud (2018). Absolute earthquake locations using 3-D versus 1-D velocity models below a local seismic network: example from the Pyrenees, *Geophys. J. Int.* **212**(3), 1806–1828.

Tryggvason, A., S. T. Rögnvaldsson, and O. G. Flóvenz (2002). Three-dimensional imaging of the P-and S-wave velocity structure and earthquake locations beneath Southwest Iceland, *Geophys. J. Int.* **151**(3), 848–866.
Wells, D. L. and K. J. Coppersmith (1994). New Empirical Relationships among Magnitude, Rupture Length, Rupture Width, Rupture Area, and Surface Displacement, *Bull. Seismol. Soc. Am.* **84**(4), 974–1002.

Woessner, J., D. Laurentiu, D. Giardini, H. Crowley, F. Cotton, G. Grünthal, G. Valensise, R. Arvidsson, R. Basili, M. B. Demircioglu, S. Hiemer, C. Meletti, R. W. Musson, A. N. Rovida, K. Sesetyan, and M. Stucchi (2015). The 2013 European Seismic Hazard Model: key components and results, *Bull. of Earthquake Eng.* **13**(12), 3553–3596.

Xu, Y., J. Wang, Y.-M. Wu, and H. Kuo-Chen (2017). Reliability assessment on earthquake early warning: A case study from Taiwan, *Soil Dyn. Earthq. Eng.* **92**, 397–407.

Zhu, L. and L. A. Rivera (2002). A note on the dynamic and static displacements from a point source in multilayered media, *Geophys. J. Int.* **148**(3), 619–627.

Zollo, A., G. Iannaccone, M. Lancieri, L. Cantore, V. Convertito, A. Emolo, G. Festa, F. Gallović, M. Vassallo, C. Martino, et al. (2009). Earthquake early warning system in southern Italy: Methodologies and performance evaluation, *Geophys. Res. Lett.* **36**(5).

Zuccolo, E., G. Cremen, C. Galasso, and D. Roessler (2020). Comparing the performance of regional earthquake early warning algorithms in Europe 2, *Earth and Space Science Open Archive*.

**Full mailing address for each author**

Dr. Gemma Cremen

Chadwick Building

Gower Street

London

W1CE 6DE

United Kingdom
Table 1: Characteristics of the earthquakes considered in this study.

| Testbed | Earthquake ID | Magnitude ($M_w$) | Longitude (°) | Latitude (°) | Depth (km) | Style of Faulting |
|---------|---------------|-------------------|---------------|--------------|------------|-------------------|
| ITA     | ITCS042       | 5.6               | 15.03         | 38.35        | 17.0       | Strike-slip       |
|         | ITCS016       | 6.9               | 15.60         | 38.03        | 9.3        | Normal            |
|         | ITCS053       | 6.2               | 16.19         | 38.63        | 8.3        | Normal            |
|         | ITCS055       | 5.9               | 15.91         | 38.23        | 9.0        | Normal            |
|         | ITCS068       | 6.4               | 16.49         | 38.87        | 11.0       | Strike-slip       |
|         | ITCS080       | 5.6               | 16.18         | 38.42        | 9.0        | Normal            |
|         | ITCS082       | 6.3               | 16.02         | 38.37        | 8.3        | Normal            |
| PYR     | ESCS071       | 5.6               | 2.47          | 42.10        | 6.8        | Normal            |
|         | ESCS112       | 6.0               | 3.26          | 42.04        | 6.8        | Normal            |
|         | FRCS007       | 6.2               | 2.07          | 42.48        | 10.3       | Normal            |
|         | ESCS126       | 5.7               | 0.64          | 42.64        | 6.3        | Normal            |
|         | FRCS002       | 6.0               | 2.77          | 42.51        | 10.3       | Normal            |
|         | ESCS125       | 6.5               | 0.89          | 42.67        | 6.7        | Normal            |
| ICE     | 1998-06-04    | 5.5               | -21.29        | 64.04        | 5.9        | Strike-Slip       |
|         | 2000-06-17(1) | 6.4               | -20.37        | 63.97        | 6.4        | Strike-slip       |
|         | 2000-06-17(2) | 5.7               | -20.45        | 63.95        | 5.4        | Strike-slip       |
|         | 2000-06-21    | 6.5               | -20.71        | 63.97        | 5.0        | Strike-slip       |
|         | 2008-05-29    | 6.3               | -21.07        | 63.97        | 5.1        | Strike-slip       |
| GRE     | 2014-01-26    | 6.0               | 20.53         | 38.22        | 16.4       | Strike-Slip       |
|         | 2013-02-03    | 5.9               | 20.40         | 38.25        | 11.3       | Strike-slip       |
|         | 2015-11-17    | 6.4               | 20.60         | 38.67        | 10.7       | Strike-slip       |
|         | 2018-10-15    | 6.7               | 20.51         | 37.34        | 9.9        | Strike-slip       |
|         | 2018-10-30    | 5.8               | 20.45         | 37.46        | 5.5        | Reverse           |

Table 2: Practical consequences of the $MD$ values observed in Figure 5, considering the median prediction for a hypothetical lognormal GMM distribution with true median = 0.5g, known dispersion = 0.7, and predicted dispersion = 0.9.

| Figure Reference for $MD$ value | True Median (g) | Predicted Median (VS) (g) | Predicted Median (PRESTo) (g) |
|---------------------------------|-----------------|---------------------------|-------------------------------|
| Figure 5a                       | 0.5             | 0.86                      | 0.66                          |
| Figure 5b                       | 0.5             | 0.71                      | 0.71                          |
| Figure 5c                       | 0.5             | 0.61                      | 0.55                          |
| Figure 5d                       | 0.5             | 0.89                      | 0.77                          |
| Figure 5e                       | 0.5             | 0.72                      | 0.77                          |
| Figure 5f                       | 0.5             | 0.61                      | 0.59                          |

List of Figure Captions

1. Figure 1. Map of the testbeds examined in this study. Each inset shows considered earthquakes (red circles), target sites (green squares), and seismic stations (blue triangles).
2. Figure 2. Comparing the accuracy of mean epicentral distance estimates, for three different threshold
levels of epicentral distance estimate uncertainty (i.e., coefficient of variation) threshold: (a) CV\(_R\)=0.3, (b) CV\(_R\)=0.2, and (c) CV\(_R\)=0.1. Filled and empty markers respectively indicate real and simulated events.

3. Figure 3. Comparing the accuracy of mean magnitude estimates, for three different threshold levels of magnitude estimate uncertainty (i.e., standard deviation): (a) \(\sigma_M = 0.3\), (b) \(\sigma_M = 0.2\), and (c) \(\sigma_M = 0.1\). Filled and empty markers respectively indicate real and simulated events.

4. Figure 4. Demonstrating the GMM evaluation procedure for the ITCS068 event in the ITA testbed. The \(MD\) metric measures the difference between the true and estimated PGA CDF for a given set of location and magnitude predictions by an EEW algorithm. Evaluations are shown for the following threshold levels of uncertainty in the underlying EEW epicentral distance and magnitude measurements: (a) \(CV_R = \sigma_M = 0.3\), (b) \(CV_R = \sigma_M = 0.2\), and (c) \(CV_R = \sigma_M = 0.1\). In this case, all estimate uncertainties are propagated through to the PGA CDF.

5. Figure 5. Comparing the accuracy of ground-shaking estimates, for three different threshold levels of epicentral distance (CV\(_R\)) and magnitude (\(\sigma_M\)) uncertainty: (a,d) CV\(_R\) = \(\sigma_M\) = 0.3, (b,e) CV\(_R\) = \(\sigma_M\) = 0.2, and (c,f) CV\(_R\) = \(\sigma_M\) = 0.1. Each (filled and empty markers respectively indicate real and simulated events). The top panel (a,b,c) compares \(MD\) values obtained using point estimates of location and magnitude, and the bottom panel (d,e,f) compares the values obtained when location and magnitude estimate uncertainties are propagated through to the PGA CDF.
Figure 1: Map of the testbeds examined in this study. Each inset shows considered earthquakes (red circles), target sites (green squares), and seismic stations (blue triangles).

Figure 2: Comparing the accuracy of mean epicentral distance estimates, for three different threshold levels of epicentral distance estimate uncertainty (i.e., coefficient of variation) threshold: (a) $CV_R=0.3$, (b) $CV_R=0.2$, and (c) $CV_R=0.1$. Filled and empty markers respectively indicate real and simulated events.
Figure 3: Comparing the accuracy of mean magnitude estimates, for three different threshold levels of magnitude estimate uncertainty (i.e., standard deviation): (a) $\sigma_M = 0.3$, (b) $\sigma_M = 0.2$, and (c) $\sigma_M = 0.1$. Filled and empty markers respectively indicate real and simulated events.

Figure 4: Demonstrating the GMM evaluation procedure for the ITCS068 event in the ITA testbed. The $MD$ metric measures the difference between the true and estimated PGA CDF for a given set of location and magnitude predictions by an EEW algorithm. Evaluations are shown for the following threshold levels of uncertainty in the underlying EEW epicentral distance and magnitude measurements: (a) $CV_R = \sigma_M = 0.3$, (b) $CV_R = \sigma_M = 0.2$, and (c) $CV_R = \sigma_M = 0.1$. In this case, all estimate uncertainties are propagated through to the PGA CDF.
Figure 5: Comparing the accuracy of ground-shaking estimates, for three different threshold levels of epicentral distance ($CV_R$) and magnitude ($\sigma_M$) uncertainty: (a,d) $CV_R = \sigma_M = 0.3$, (b,e) $CV_R = \sigma_M = 0.2$, and (c,f) $CV_R = \sigma_M = 0.1$. Each data point corresponds to the resulting $MD$ values for one event (filled and empty markers respectively indicate real and simulated events). The top panel (a,b,c) compares $MD$ values obtained using point estimates of location and magnitude, and the bottom panel (d,e,f) compares the values obtained when location and magnitude estimate uncertainties are propagated through to the PGA CDF.