

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

Gemma Cremen*, Elisa Zuccolo, and Carmine Galasso

January 2021

1 Declaration of Competing Interests

2 The authors acknowledge there are no conflicts of interest recorded.

3 Abstract

4 Earthquake early warning (EEW) is becoming an increasingly attractive real-time strategy for mitigating the
5 threats posed by potentially devastating incoming seismic events. As efforts accelerate to develop practical
6 EEW-based solutions for earthquake-prone countries in Europe, it is important to understand and quantify
7 the level of performance that can be achieved by the underlying seismological algorithms. We conduct a
8 conceptual study on EEW performance in Europe, which explicitly focuses on the accuracy and associated
9 uncertainties of selected methodological approaches. 23 events from four diverse European testbeds are used
10 to compare the quality of EEW predictions produced by the Virtual Seismologist and PRESTo algorithms.
11 We first examine the location and magnitude estimates of the algorithms, accounting for both bias and
12 uncertainty in the resulting predictions. We then investigate the ground-shaking prediction capabilities
13 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

14 uncertainties in these estimates. Our work highlights the importance of accounting for EEW parameter
15 uncertainties, which are often neglected in studies of EEW performance. Our findings can be used to inform
16 current and future implementations of EEW systems in Europe. In addition, the evaluation metrics presented
17 in this work can be used to determine EEW accuracy in any worldwide setting.

18 **Introduction**

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

26 We study the theoretical (offline) accuracy of Virtual Seismologist (Cua, 2005; Cua and Heaton, 2007;
27 Cua et al., 2009) and PRobabilistic and Evolutionary early warning SysTem (PRESTo) (Satriano et al.,
28 2008a; Lancieri and Zollo, 2008; Satriano et al., 2011), which have been the most widely applied regional
29 EEW algorithms in Europe to date (Cremen and Galasso, 2020). We specifically focus on the location,
30 magnitude, and ground-shaking estimation capabilities of the algorithms across four European testbeds that
31 capture a diverse range of seismicity, seismotectonics, and seismic network densities. Note that aspect (2) of
32 earthquake performance is assessed for real-time versions of the algorithms and the same regions in a recent
33 companion study (Zuccolo et al., 2020). (The Vrancea testbed of the companion study is ignored in this
34 paper, since the Virtual Seismologist algorithm has not been calibrated for the large depths of its associated
35 events).

36 Our accuracy assessments explicitly incorporate uncertainties associated with the earthquake parameter
37 estimations of the algorithms, which are crucial for well-informed decision-making on alert triggering (Meier
38 et al., 2015; Cremen and Galasso, 2020). From an engineering perspective for example, these uncertain-
39 ties play a central role in the real-time probabilistic seismic hazard analysis framework that can be used

40 to determine whether issuing an EEW alarm would reduce the losses associated with an incoming event
41 (Iervolino, 2011). Our work therefore represents a significant advancement over many previous studies of
42 context-specific EEW accuracy (e.g., Hsu et al., 2018; Xu et al., 2017; Kodera et al., 2016; Böse et al., 2012;
43 Colombelli et al., 2012; Hsu et al., 2016; Böse et al., 2012; Colombelli et al., 2015; Hoshiha and Aoki, 2015;
44 Böse et al., 2014; Doi, 2011; Hartog et al., 2016; Mittal et al., 2019; Chen et al., 2019; Chung et al., 2020;
45 Minson et al., 2020; Zollo et al., 2009; Cochran et al., 2018; Festa et al., 2018; Auclair et al., 2015) - includ-
46 ing those that examine Virtual Seismologist (Behr et al., 2016) and PRESTo (Picozzi et al., 2015) - which
47 focus exclusively on the performance of point-estimate predictions (i.e., that only consider mean or modal
48 values of the estimates rather than their probability distributions) from EEW algorithms. Some work has
49 examined uncertainty propagation for EEW (i.e., the effect of uncertain source-parameter estimates on the
50 final ground-shaking predictions), but this has so far been limited to the context of hypothetical algorithms
51 (Meier, 2017), simplistic simulated events (Iervolino et al., 2009), or empirical error models of parameter
52 estimates (Brown et al., 2011).

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

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

64 **Examined Testbeds**

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

70 **Considered Events**

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

80 We generate synthetic seismograms for our assessments of ITA and PYR, due to a lack of available
81 moderate-to-large earthquake recordings in these regions. Physics-based numerical simulations are used
82 to compute the seismograms, according to the broadband ground-motion simulation method described in
83 Crempien and Archuleta (2015), which uses an extended kinematic model of the seismic source (subdivided
84 into point sources) with correlated random source parameters (rupture time, peak time, rise time and final
85 slip) that are based on more than 300 simulations of dynamic rupture models (Schmedes et al., 2013).
86 Green's functions are computed using a 1-D layered Earth model and a frequency-wavenumber (FK) code
87 (Zhu and Rivera, 2002). This code, which is efficient in obtaining high-frequency synthetics, is coupled with
88 a random perturbation of the point-source focal mechanism that partially accounts for scattering effects
89 at high frequencies. We consider one scenario earthquake per active fault in both testbeds, and use fault

90 parameters from the European Database of Seismogenic Faults (ESDF; Basili et al., 2013) to determine
91 the characteristics of these events according to the following procedures/assumptions: (1) We assign the
92 magnitude of an event as a random sample from a uniform distribution between 5.5 and the maximum
93 magnitude of the fault; (2) We assume that an event’s hypocentre is located at a depth equivalent to the
94 minimum fault depth plus 2/3 of the vertical width; (3) We use the average values of the strike, dip, and
95 rake angles to define the focal mechanism; (4) We randomly sample the average rupture velocity for each
96 scenario event from a uniform distribution between 65% and 85% of the fault plane’s shear wave velocity. We
97 use the Wells and Coppersmith (1994) relationships to define the rupture fault dimensions, and the Causse
98 et al. (2008) distributions to determine hypocentral position along the fault plane. The stress drop, which
99 is assumed to be 3 MPa for both testbeds (Caporali et al., 2011) is used to determine the corner frequency
100 (Allmann and Shearer, 2009). We use the Barberi et al. (2004) crustal velocity model for ITA and the
101 Theunissen et al. (2018) crustal velocity model for PYR. Broadband seismograms (0-25 Hz) are calculated
102 at the locations of all currently operating permanent seismic stations of the IRIS database (see **Data and**
103 **Resources**) are positioned within 100 km of each epicentre. White noise is finally added to each generated
104 seismogram, to facilitate the automatic detection of P-wave arrivals based on the Short Time Average over
105 Long Time Average STA/LTA algorithm (Allen, 1982).

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

117 **Examined Algorithms**

118 We examine the theoretical performance of the Virtual Seismologist (VS) (Cua, 2005; Cua and Heaton,
119 2007; Cua et al., 2009) and the PRobabilistic and Evolutionary early warning SysTem (PRESTo) (Lancieri
120 and Zollo, 2008; Satriano et al., 2008b, 2011) regional EEW algorithms across all testbeds. The similar
121 (Bayesian) structure of both algorithms enables direct comparisons to be made.

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

131 PRESTo estimates location using the *RTL*oc method proposed by Satriano et al. (2008b) and predicts
132 magnitude according to the *RTM*ag procedure developed by Lancieri and Zollo (2008). *RTL*oc produces
133 multivariate normal probability density functions of hypocentral locations, based on P-wave arrival times
134 and a velocity model. *RTM*ag uses a Bayesian framework for estimating magnitude, in which the likelihood
135 function depends on initial peak displacement measurements and *RTL*oc outputs. The prior PDF for a given
136 time step is the posterior distribution obtained at the previous time step, and the prior for the first time
137 step is optionally set as the Gutenberg-Richter distribution. Peak ground-motion parameters are computed
138 based on the location and magnitude estimates, using a GMM. PRESTo is currently operating in real-time
139 in Southern Italy, Turkey, Romania, and South Korea (Picozzi et al., 2015), and has also been tested for
140 application in Austria and Slovenia (Picozzi et al., 2015), as well as the Iberian Peninsula (Pazos et al.,
141 2015).

142 **Algorithm Inputs**

143 **Location Inputs**

144 We use a common (neutral) method to determine event arrival times for both algorithms, given that P-wave
145 picking accuracy is not the focus of our evaluation. We leverage the SeisComP seismological software (see
146 **Data and Resources**) and specifically use the picks associated with its preferred origin for a given event.
147 This origin is automatically selected using the *scevent* module of the software, according to a number of
148 predefined rules (e.g., an origin is preferred to the previous one if it is computed using a greater number of
149 picks and/or produces lower travel time residuals, etc). Velocity models input to the PRESTo algorithm are
150 region-specific. The velocity models used for ITA and PYR are the same as those adopted for the generation
151 of synthetic seismograms in both testbeds (see **Considered Events**). We use the Tryggvason et al. (2002)
152 model for ICE and the Rigo et al. (1996) model for GRE. We use the Poisson’s solid approximation to derive
153 undefined P-wave velocities from associated S-wave velocities (and vice versa), and we compute corresponding
154 3D travel-time grids using the *NonLinLoc* software (Lomax et al., 2000).

155 **Magnitude Inputs**

156 We estimate magnitudes based on seismogram data from stations that are associated with the preferred origin
157 location estimated by SeisComP. The seismograms are first processed as follows (Boore and Bommer, 2005):
158 (1) We apply instrument correction to GRE recordings; (2) We apply baseline correction; (3) We differentiate
159 velocigrams to retrieve accelerograms; (4) We apply a high-pass filter with 3 s corner frequency for VS (Cua
160 and Heaton, 2007) and a band-pass filter (0.075 - 3 Hz) to accelerograms for PRESTo (Satriano et al., 2011);
161 (5) We integrate the accelerograms once to retrieve velocities, and twice to obtain displacements. For VS, we
162 then extract maximum envelope values of vertical acceleration (ZA), velocity (ZV), displacement (ZD), and
163 root mean square horizontal acceleration (HA), velocity (HV) and displacement (HV), computed within
164 1 s intervals starting from the P-wave arrival time. For PRESTo, we extract values of peak displacement
165 (P_d) in three different time windows, accounting for the vector modulus of the three-component seismogram
166 (Satriano et al., 2011). These time windows are (1) 2 s following the P-wave arrival if the P and S arrivals
167 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

168 s following the S-wave arrival.

169 **Bayesian Priors**

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

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

186 **Results**

187 **Location and Magnitude Accuracy**

188 We quantify the accuracy of the algorithms' location and magnitude components independently, to determine
189 the accuracy of the estimates for different levels of algorithmic uncertainty in the source parameters. Since
190 the quality of estimates should increase in time while the uncertainty decreases, this assessment is designed
191 to capture various accuracy/speed trade-off thresholds that may be of interest to stakeholders for guiding

192 decision-making and EEW alert issuance. Location and magnitude accuracy are quantified for each algorithm
193 using the root mean square error (RMSE) metric (Hyndman and Koehler, 2006).

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

203 It can be seen that PRESTo yields the best distance predictions across all uncertainty levels investigated
204 except $CV_R = 0.3$. Its associated RMSE value is over 35% lower than that of VS for both $CV_R = 0.2$
205 and $CV_R = 0.1$, whereas the RMSE value for VS is 28% smaller in the largest uncertainty case (note that
206 no consistent performance differences are observed between real and simulated events). If we consider a
207 hypothetical M_w 6 normal-faulting earthquake with $V_{s30} = 800$ m/s and use the epicentral distance version
208 of the Akkar et al. (2014) GMM for the observed RMSE values, the median PGA predictions obtained for
209 both algorithms are noticeably different, with the discrepancies ranging between 29% and 64% across the
210 three cases. It is interesting to note that the VS RMSE values for epicentral distance increase as uncertainty
211 decreases, which is the opposite of what is intuitively expected (Cochran et al., 2018). This observation is
212 due to the effect of the algorithm’s Bayesian prior, which significantly narrows the range of location estimates
213 (and therefore their uncertainty) after only two P-wave arrivals, thereby preventing any significant further
214 improvements that may be achieved using information from additional stations.

215 We compare the mean magnitude predictions of both algorithms, using the first estimates with standard
216 deviations (σ_M) below the following thresholds: 0.1, 0.2, and 0.3 (Figure 3). It can be observed that the
217 results of the PRESTo algorithm are most accurate for all levels of uncertainty investigated. The PRESTo
218 RMSE value is approximately 15% lower than that of VS for $\sigma_M = 0.2$, and over 20% lower in both other

219 cases (note that there are no distinct differences between the performance trends for real and simulated
220 events). If we take a hypothetical M_w 6 + RMSE normal-faulting earthquake with $V_{s30} = 800$ m/s and use
221 the epicentral distance version of the Akkar et al. (2014) GMM at 30 km, we obtain non-negligible differences
222 between the resulting median PGA predictions for the RMSE values of both algorithms. These differences
223 range between 9% and 14% across the three uncertainty thresholds.

224 **Propagation of Uncertainties and their Effect on Ground Motion**

225 Finally, we investigate the impact of the algorithms' location and magnitude estimation accuracy on the
226 quality of resulting ground-shaking predictions, using the epicentral distance version of the Akkar et al.
227 (2014) GMM. We simply assume rock ground conditions ($V_{s30} = 800$ m/s) in all cases (for both true
228 and predicted ground shaking), given that site class does not influence the assessment of ground-motion
229 accuracy related to location and magnitude, and use the style-of-faulting information provided in Table 1.
230 We specifically focus on peak ground acceleration (PGA) predictions in this investigation.

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

237 This type of comparison is useful if EEW alerts are issued based on a given probability of exceeding a
238 prescribed value of PGA (Iervolino, 2011). This is because discrepancies in the CDFs indicate the potential
239 for incorrect decisions to be made on whether or not to trigger an alarm (Iervolino, 2011). For example, a
240 false alert may occur if the predicted PGA is greater than the actual PGA value at the exceedance threshold.
241 On the other hand, an alert may be missed if this prediction is less than the true value. Our evaluation
242 offers a significant advantage over many previous studies of EEW ground-shaking or intensity accuracy (e.g.,
243 Meier, 2017; Cochran et al., 2019; Minson et al., 2019; Meier et al., 2020), since it does not require the a
244 priori definition of a subjective alert threshold.

245 As an advancement over our companion paper, we use a version of the MD metric that can explicitly
 246 account for the propagated uncertainty of the EEW source-parameter estimates in the resulting PGA CDF.
 247 We use Monte Carlo sampling of the underlying distributions to capture all uncertainties, and calculate MD
 248 according to the following equation:

$$MD = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N [y_{n/N}^i - y_{n/N}^o]^2}}{\frac{1}{N} \sum_{n=1}^N y_n^o} \quad (1)$$

249 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
 250 the (n/N) th quantile of the true GMM CDF ($0 < n < N$), $y_{n/N}^i$ is the equivalent quantile for the predicted
 251 GMM distribution, and the denominator represents the mean of the true CDF. A lower value indicates a
 252 higher similarity between the predicted and true distributions of PGA. Figure 4 demonstrates the evaluation
 253 procedure, displaying the MD values and corresponding PGA CDFs obtained for one scenario event.

254 We examine ground-shaking prediction accuracy by combining the estimations of location and magnitude
 255 obtained in **Location and Magnitude Accuracy**. We specifically consider the source-parameter estimates
 256 associated with the following thresholds of algorithmic uncertainties in location (CV_R) and magnitude (σ_M):
 257 $CV_R = \sigma_M = 0.3$, $CV_R = \sigma_M = 0.2$, and $CV_R = \sigma_M = 0.1$. Our first investigation uses only the relevant
 258 point (i.e., mean) estimates for each algorithm. The corresponding MD values obtained for both algorithms
 259 and each event are displayed in Figures 5a to 5c. It can be observed that PRESTo produces the lowest
 260 average MD value across two considered uncertainty levels - this value is 38% lower for $CV_R = \sigma_M = 0.3$,
 261 and 23% lower for $CV_R = \sigma_M = 0.1$ - and approximately equivalent results are obtained for both algorithms
 262 in the case $CV_R = \sigma_M = 0.2$. Thus, we can generally conclude that PRESTo is the best algorithm in terms
 263 of ground-shaking prediction accuracy, when only point estimates of source parameter measurements are
 264 considered. (Note that there are no consistent differences between the MD values observed for real and
 265 simulated events, in any examined case.)

266 We then determine the ground-shaking accuracy obtained when the uncertainty of the source-parameter
 267 estimates is propagated through to the PGA CDF. Corresponding MD values for both algorithms and each
 268 event are displayed in Figures 5d to 5f. Firstly, this analysis produces larger average MD values than those
 269 obtained when only point estimates of source parameters are considered, in almost all cases. This implies that

270 neglecting the underlying uncertainty of the EEW source parameter measurements tends to overestimate the
271 accuracy of the resulting ground-shaking predictions. PRESTo still produces more accurate PGA estimations
272 for the $CV_R = \sigma_M = 0.3$ and $CV_R = \sigma_M = 0.1$ uncertainty levels, for which the average MD values for
273 PRESTo are respectively 22% and 7% lower than those of VS. However, the performance of VS is noticeably
274 better than that of PRESTo for $CV_R = \sigma_M = 0.2$; the average VS MD value in this case is 12% lower than
275 the corresponding PRESTo value. Thus, the optimal algorithm for ground-shaking prediction in this case
276 depends on the level of uncertainty in the source parameters.

277 To provide some context on the practical consequences of the observed differences in MD values, we
278 consider their implications for the specific case of predicting the median value of a hypothetical, realistic
279 (lognormal) GMM distribution with true median = 0.5g, known dispersion = 0.7, and predicted dispersion
280 = 0.9. Table 2 provides the median values of the predicted distributions that would lead to the MD values
281 observed in Figure 5. It can be seen from the table that the median predictions associated with unique MD
282 values vary noticeably, such that they could feasibly lead to different decision outcomes (i.e., trigger/don't
283 trigger alert) for an EEW alarm based on the median PGA prediction exceeding a particular threshold. For
284 example, based on the average MD values of Figure 5a, a median prediction alert threshold of 0.7 g would
285 cause a false alert to be issued using the VS algorithm but would correctly result in no alarm being triggered
286 for the PRESTo algorithm prediction. We can thus conclude that the differences observed between MD
287 values have practical implications on the accuracy of ground-shaking estimates and EEW triggering.

288 Conclusions

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

294 We first assessed the algorithms' mean source-parameter estimates, which corresponded to various un-
295 certainty thresholds that stakeholders may use to guide decision-making on EEW alert triggering. We found

296 that PRESTo was almost consistently the best-performing algorithm in terms of both location and mag-
297 nitude estimation. PRESTo location estimates were over 35% more accurate than those of VS (except in
298 the case of relatively large source-parameter uncertainty, i.e. $CV_R=0.3$), and its magnitude estimates were
299 approximately 15 to 20% better. We therefore conclude that PRESTo should be used for EEW purposes
300 that require mean estimates of location and magnitude, which is consistent with the recommendations of
301 our companion paper (Zuccolo et al., 2020) that compared the real-time operational performance of both
302 algorithms.

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

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

320 It is important to note that there are some limitations associated with this work. Firstly, the calibration
321 of phase detection and association parameters was only carried out for the events examined in this study, and
322 may not reflect the overall seismicity and network geometry of each area. Secondly, we did not explore the

323 sensitivity of the algorithms' Bayesian priors. For example, magnitude priors retrieved from regional hazard
324 studies (instead of the European ESHM13 model) may have resulted in the better performance of a given
325 algorithm. In addition, a location prior with less severe constraints may have improved the accuracy of VS
326 estimates. However, average MD values obtained using the VS magnitude estimates and correct distance
327 measurements yield the same conclusions on the relative performance of the algorithms as those presented
328 **in Propagation of Uncertainties and their Effect on Ground Motion**, tentatively suggesting that
329 location accuracy (and therefore the choice of location prior) does not have a significant effect on the quality
330 of ground-shaking estimates associated with VS (note that a more concrete conclusion on the effect of the VS
331 prior chosen would also require an investigation of its influence on the accuracy of VS magnitude estimates).
332 Furthermore, the empirical scaling relationships used to estimate magnitude in the algorithms may not be
333 appropriate for all considered regions. Finally, our results may not reflect the performance of the algorithms
334 across all European sites or regions, however the evaluation procedures presented in this paper could be used
335 to conduct more detailed accuracy analyses for specific case studies. Despite the aforementioned constraints,
336 our study nevertheless provides some notable insights on the accuracy and uncertainty of EEW estimates
337 for European seismicity.

338 Data and Resources

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

349 ComP platform used is available at <https://www.seiscomp.de>, and related professional software support is
350 provided by gempa GmbH (<http://www.gempa.de>).

351 Acknowledgements

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

361 References

- 362 Akkar, S., M. Sandikkaya, and J. Bommer (2014). Empirical ground-motion models for point-and extended-
363 source crustal earthquake scenarios in Europe and the Middle East, *Bull. Earthquake Eng.* **12**(1), 359–387.
- 364 Akkar, S., M. Sandikkaya, M. Şenyurt, A. A. Sisi, B. Ay, P. Traversa, J. Douglas, F. Cotton, L. Luzi,
365 B. Hernandez, et al. (2014). Reference database for seismic ground-motion in Europe (RESORCE), *Bull.*
366 *Earthquake Eng.* **12**(1), 311–339.
- 367 Allen, R. (1982). Automatic phase pickers: Their present use and future prospects, *Bulletin of the Seismo-*
368 *logical Society of America* **72**(6B), S225–S242.
- 369 Allen, R. M. and D. Melgar (2019). Earthquake Early Warning: Advances, Scientific Challenges, and Societal
370 Needs, *Annu. Rev. Earth Planet Sci.* **47**.

371 Allmann, B. P. and P. M. Shearer (2009). Global variations of stress drop for moderate to large earthquakes,
372 *J. Geophys. Res. Solid Earth* **114**(1).

373 Auclair, S., X. Goula, J.-A. Jara, and Y. Colom (2015). Feasibility and interest in earthquake early warning
374 systems for areas of moderate seismicity: Case study for the Pyrenees, *Pure Appl. Geophys.* **172**(9),
375 2449–2465.

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

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

380 Behr, Y., J. Clinton, P. Kästli, C. Cauzzi, R. Racine, and M.-A. Meier (2015). Anatomy of an earthquake early
381 warning (EEW) alert: Predicting time delays for an end-to-end EEW system, *Seismol. Res. Lett.* **86**(3),
382 830–840.

383 Behr, Y., J. F. Clinton, C. Cauzzi, E. Hauksson, K. Jónsdóttir, C. G. Marius, A. Pinar, J. Salichon, and
384 E. Sokos (2016). The Virtual Seismologist in SeisComp3: A new implementation strategy for earthquake
385 early warning algorithms, *Seismol. Res. Lett.* **87**(2A), 363–373.

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

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

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

393 Brown, H. M., R. M. Allen, M. Hellweg, O. Khainovski, D. Neuhauser, and A. Souf (2011). Development
394 of the ElarmS methodology for earthquake early warning: Realtime application in California and offline
395 testing in Japan, *Soil Dyn. Earthq. Eng.* **31**(2), 188–200.

396 Caporali, A., S. Barba, M. M. Carafa, R. Devoti, G. Pietrantonio, and F. Riguzzi (2011). Static stress drop
397 as determined from geodetic strain rates and statistical seismicity, *J. Geophys. Res. Solid Earth* **116**(2).

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

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

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

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

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

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

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

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

418 Colombelli, S., O. Amoroso, A. Zollo, and H. Kanamori (2012). Test of a threshold-based earthquake early-
419 warning method using Japanese data, *Bull. Seismol. Soc. Am.* **102**(3), 1266–1275.

420 Colombelli, S., A. Caruso, A. Zollo, G. Festa, and H. Kanamori (2015). A P-wave-based, on-site method for
421 earthquake early warning, *Geophys. Res. Lett.* **42**(5), 1390–1398.

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

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

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

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

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

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

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

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

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

442 Galasso, C., P. Zhong, F. Zareian, I. Iervolino, and R. W. Graves (2013). Validation of ground-motion
443 simulations for historical events using MDoF systems, *Earthq. Eng. Struct. Dyn.* **42**(9), 1395–1412.

444 Hartog, J. R., V. C. Kress, S. D. Malone, P. Bodin, J. E. Vidale, and B. W. Crowell (2016). Earthquake
445 Early Warning: ShakeAlert in the Pacific Northwest Earthquake Early Warning: ShakeAlert in the Pacific
446 Northwest, *Bull. Seismol. Soc. Am.* **106**(4), 1875–1886.

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

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

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

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

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

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

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

461 Kodera, Y., J. Saitou, N. Hayashimoto, S. Adachi, M. Morimoto, Y. Nishimae, and M. Hoshiaba (2016).
462 Earthquake early warning for the 2016 Kumamoto earthquake: Performance evaluation of the current
463 system and the next-generation methods of the Japan Meteorological Agency, *Earth Planets Space* **68**(1),
464 202.

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

467 Lomax, A., J. Vireux, P. Volant, and C. Berge-Thierry (2000). Probabilistic Earthquake Location in 3D and
468 Layered Models. In *Advances in Seismic Event Location*. Springer, Dordrecht.

469 Meier, M.-A. (2017). How “good” are real-time ground motion predictions from earthquake early warning
470 systems?, *J. Geophys. Res. Solid Earth* **122**(7), 5561–5577.

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

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

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

479 Minson, S. E., J. K. Saunders, J. J. Bunn, E. S. Cochran, A. S. Baltay, D. L. Kilb, M. Hoshihara, and Y. Kodera
480 (2020). Real-Time Performance of the PLUM Earthquake Early Warning Method during the 2019 M 6.4
481 and 7.1 Ridgecrest, California, Earthquakes, *Bull. Seismol. Soc. Am.* **110**(4), 1887–1903.

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

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

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

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

491 Picozzi, M., L. Elia, D. Pesaresi, A. Zollo, M. Mucciarelli, A. Gosar, W. Lenhardt, and M. Živčić (2015).
492 Trans-national earthquake early warning (EEW) in north-eastern Italy, Slovenia and Austria: first expe-
493 rience with PRESTo at the CE3RN network, *Adv. Geosci* **40**, 51–61.

494 Picozzi, M., A. Zollo, P. Brondi, S. Colombelli, L. Elia, and C. Martino (2015). Exploring the feasibility of
495 a nationwide earthquake early warning system in Italy, *J. Geophys. Res. Solid Earth* **120**(4), 2446–2465.

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

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

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

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

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

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

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

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

515 Tryggvason, A., S. T. Rögnvaldsson, and O. G. Flóvenz (2002). Three-dimensional imaging of the P-and
516 S-wave velocity structure and earthquake locations beneath Southwest Iceland, *Geophys. J. Int.* **151**(3),
517 848–866.

518 Wells, D. L. and K. J. Coppersmith (1994). New Empirical Relationships among Magnitude, Rupture Length,
519 Rupture Width, Rupture Area, and Surface Displacement, *Bull. Seismol. Soc. Am.* **84**(4), 974–1002.

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

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

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

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

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

533 **Full mailing address for each author**

534 Dr. Gemma Cremen
535 Chadwick Building
536 Gower Street
537 London
538 W1CE 6DE
539 United Kingdom

540

541

542 Dr. Elisa Zuccolo
543 European Centre for Training and Research in Earthquake Engineering
544 Via A. Ferrata, 1
545 Pavia
546 Italy
547
548
549 Dr. Carmine Galasso
550 Chadwick Building
551 Gower Street
552 London
553 W1CE 6DE
554 United Kingdom
555

Tables

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

Testbed	Earthquake ID	Magnitude (M_w)	Longitude ($^\circ$)	Latitude ($^\circ$)	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

561 levels of epicentral distance estimate uncertainty (i.e., coefficient of variation) threshold: (a) $CV_R=0.3$,
562 (b) $CV_R=0.2$, and (c) $CV_R=0.1$. Filled and empty markers respectively indicate real and simulated
563 events.

564 3. Figure 3. Comparing the accuracy of mean magnitude estimates, for three different threshold levels
565 of magnitude estimate uncertainty (i.e., standard deviation): (a) $\sigma_M = 0.3$, (b) $\sigma_M = 0.2$, and (c)
566 $\sigma_M = 0.1$. Filled and empty markers respectively indicate real and simulated events.

567 4. Figure 4. Demonstrating the GMM evaluation procedure for the ITCS068 event in the ITA testbed.
568 The MD metric measures the difference between the true and estimated PGA CDF for a given set of
569 location and magnitude predictions by an EEW algorithm. Evaluations are shown for the following
570 threshold levels of uncertainty in the underlying EEW epicentral distance and magnitude measure-
571 ments: (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
572 estimate uncertainties are propagated through to the PGA CDF.

573 5. Figure 5. Comparing the accuracy of ground-shaking estimates, for three different threshold levels
574 of epicentral distance (CV_R) and magnitude (σ_M) uncertainty: (a,d) $CV_R = \sigma_M = 0.3$, (b,e) CV_R
575 $= \sigma_M = 0.2$, and (c,f) $CV_R = \sigma_M = 0.1$. Each (filled and empty markers respectively indicate real
576 and simulated events). The top panel (a,b,c) compares MD values obtained using point estimates of
577 location and magnitude, and the bottom panel (d,e,f) compares the values obtained when location and
578 magnitude estimate uncertainties are propagated through to the PGA CDF.

Figures

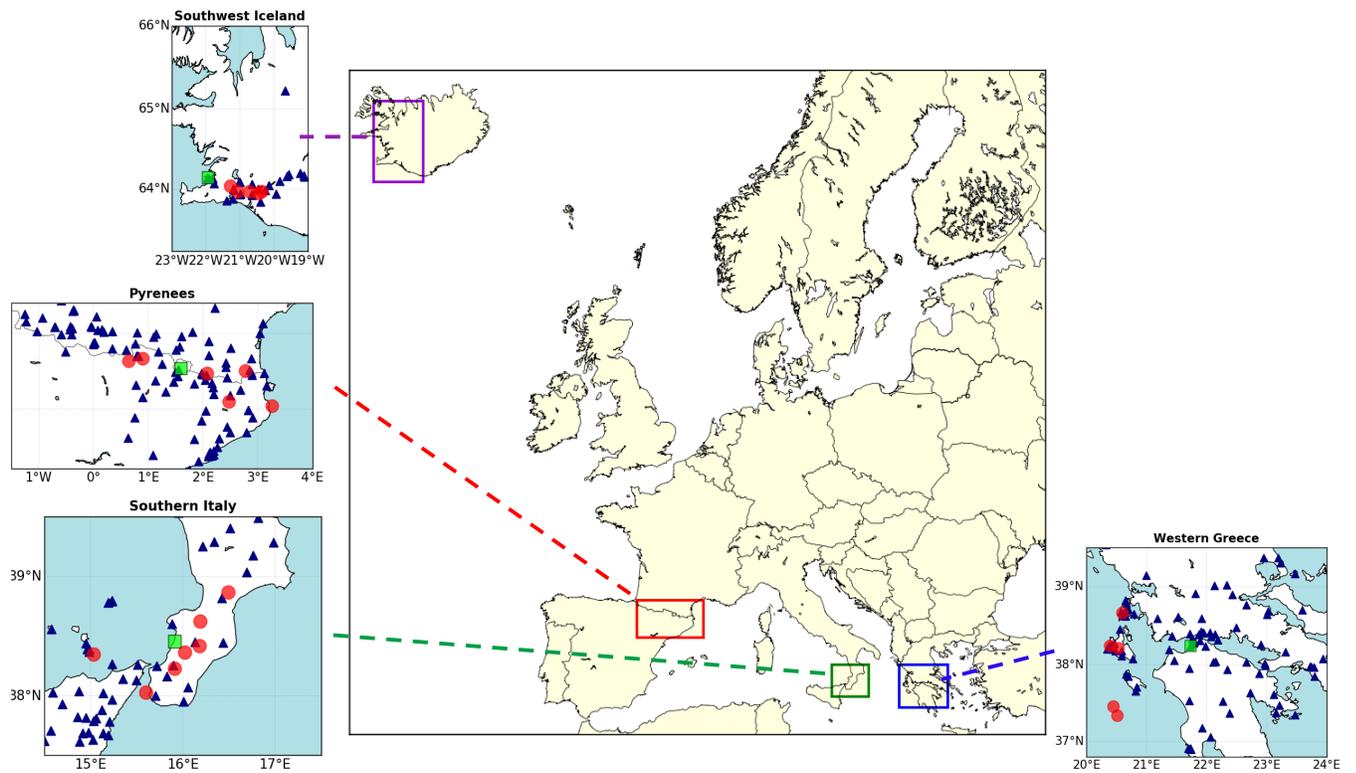


Figure 1: Map of the testbeds examined in this study. Each inset shows considered earthquakes (red circles), target sites (green squares), and seismic stations (blue triangles).

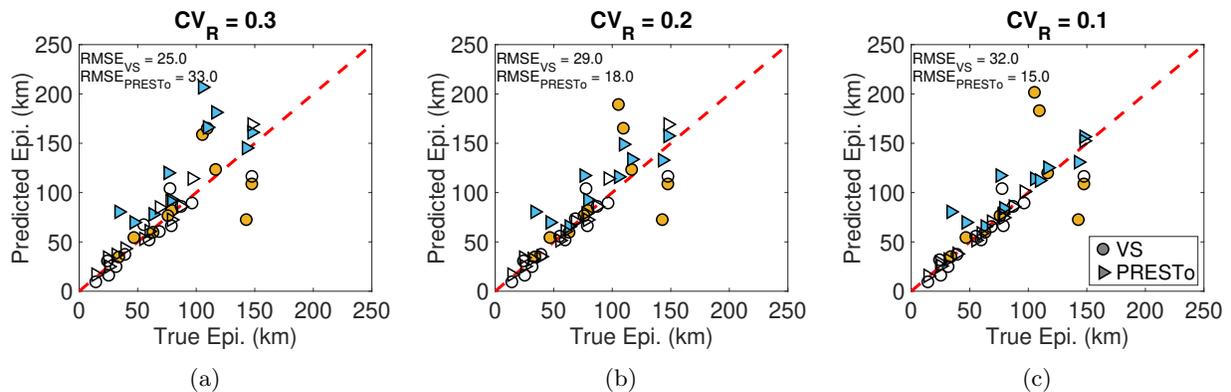


Figure 2: Comparing the accuracy of mean epicentral distance estimates, for three different threshold levels of epicentral distance estimate uncertainty (i.e., coefficient of variation) threshold: (a) $CV_R=0.3$, (b) $CV_R=0.2$, and (c) $CV_R=0.1$. Filled and empty markers respectively indicate real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

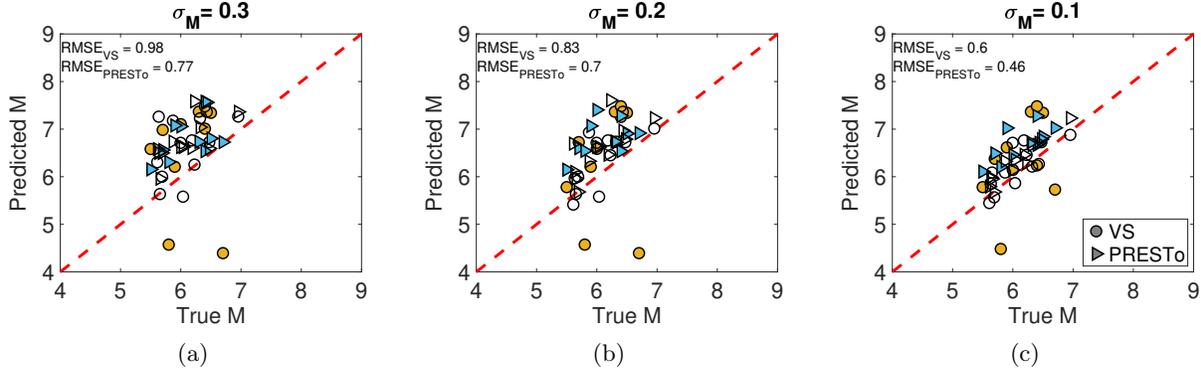


Figure 3: Comparing the accuracy of mean magnitude estimates, for three different threshold levels of magnitude estimate uncertainty (i.e., standard deviation): (a) $\sigma_M = 0.3$, (b) $\sigma_M = 0.2$, and (c) $\sigma_M = 0.1$. Filled and empty markers respectively indicate real and simulated events.

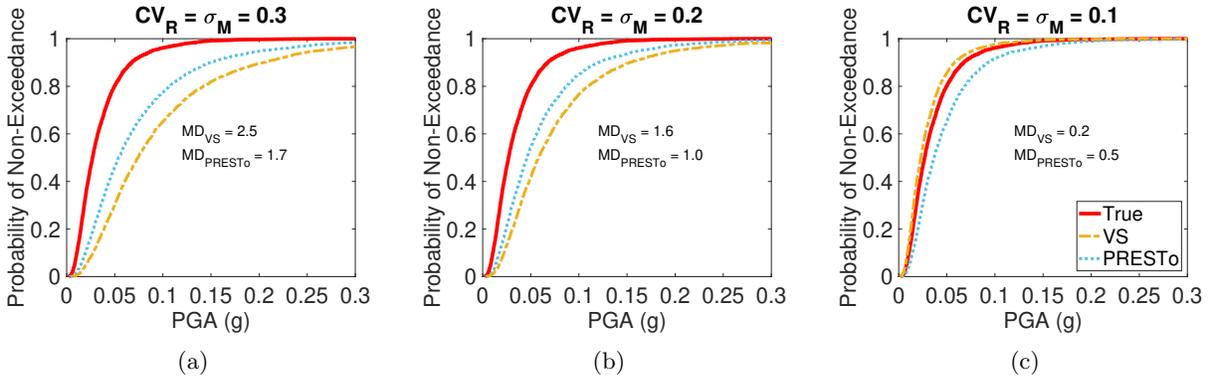


Figure 4: Demonstrating the GMM evaluation procedure for the ITCS068 event in the ITA testbed. The MD metric measures the difference between the true and estimated PGA CDF for a given set of location and magnitude predictions by an EEW algorithm. Evaluations are shown for the following threshold levels 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.

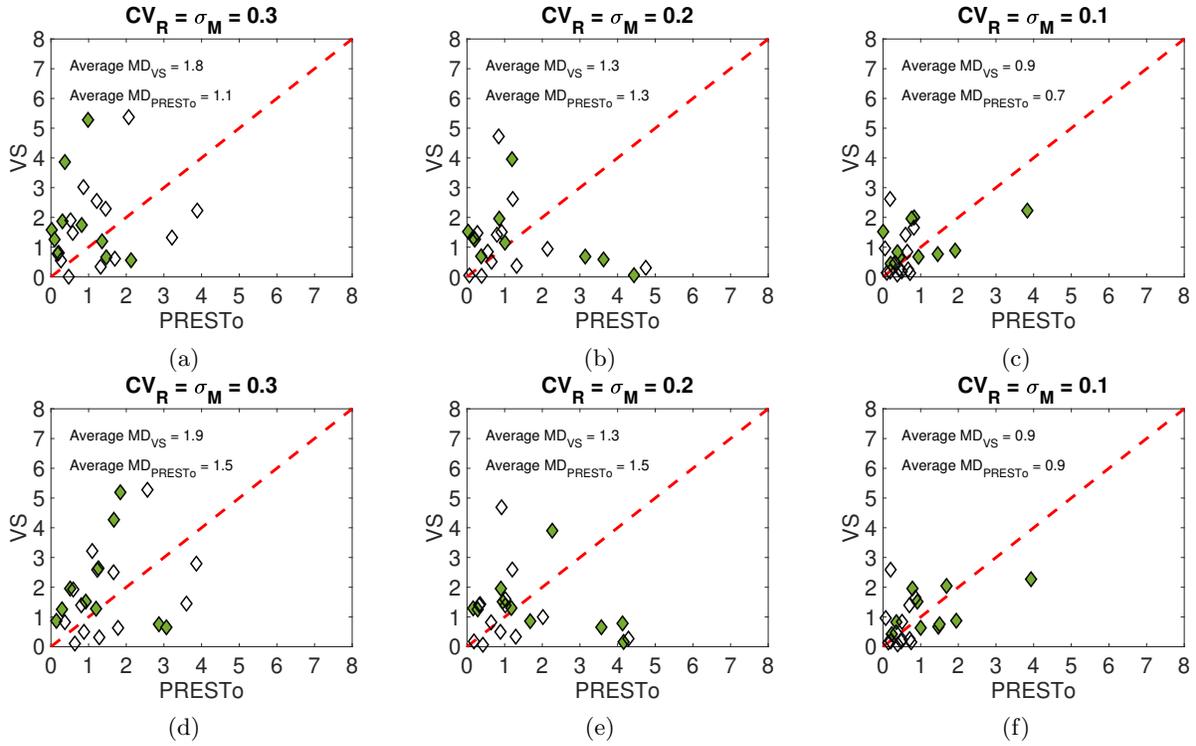


Figure 5: Comparing the accuracy of ground-shaking estimates, for three different threshold levels of epicentral distance (CV_R) and magnitude (σ_M) uncertainty: (a,d) $CV_R = \sigma_M = 0.3$, (b,e) $CV_R = \sigma_M = 0.2$, and (c,f) $CV_R = \sigma_M = 0.1$. Each data point corresponds to the resulting 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.