Geophysical Journal International

Geophys. J. Int. (2012) **191,** 803–812 doi: 10.1111/j.1365-246X.2012.05657.x

Real-time Finite Fault Rupture Detector (FinDer) for large earthquakes

Maren Böse, Thomas H. Heaton and Egill Hauksson

Seismological Laboratory, California Institute of Technology, 1200 *E. California Blvd., Mail Code* 252–21*, Pasadena,* CA 91125*, USA.*
E-mail: mboese@caltech.edu

Accepted 2012 August 20. Received 2012 June 29; in original form 2012 January 10

SUMMARY

To provide rapid estimates of fault rupture extent during large earthquakes, we have developed the *Fin*ite Fault Rupture *De*tecto*r* algorithm, 'FinDer'. FinDer uses image recognition techniques to detect automatically surface-projected fault ruptures in real-time (assuming a line source) by estimating their current centroid position, length L, and strike θ. The approach is based on a rapid high-frequency near/far-source classification of ground motion amplitudes in a dense seismic network (station spacing $<$50 km), and comparison with a set of pre-calculated templates using 'Matching by Correlation'. To increase computational efficiency, we perform the correlation in the wavenumber domain. FinDer keeps track of the current dimensions of a rupture in progress. Errors in L are typically on the same order as station spacing in the network. The continuously updated estimates of source geometries as provided by FinDer make predicted shaking intensities more accurate and thus more useful for earthquake early warning, ShakeMaps, and related products. The applicability of the algorithm is demonstrated for several recorded and simulated earthquakes with different focal mechanisms, including the 2009 M_w 6.3 L'Aquila (Italy), the 1999 M_w 7.6 ChiChi (Taiwan) and the M_w 7.8 ShakeOut scenario earthquake on the southern San Andreas Fault (California).

Key words: Image processing; Spatial analysis; Earthquake ground motions; Seismic monitoring and test-ban treaty verification; Earthquake interaction, forecasting and prediction; Early warning.

INTRODUCTION

Using current data acquisition, telemetry and processing technologies, seismologists are able to determine within seconds the magnitude and location of a local earthquake using dense networks of seismic sensors with real-time communication to a central processing site. However, the current processing algorithms are insufficient in the case of large earthquakes ($M_w > 6$); large earthquakes rupture for long distances along seismic faults (M_w 7 $\sim$ 60 km, M_w 7.5 $\sim$ 130 km, and M_w 8 $\sim$ 300 km; Wells & Coppersmith 1994); it is the distance to the closest fault rupture that causes the peak strength of shaking at a given site. To estimate the distribution of shaking and potential damage, for example, for earthquake early warning (EEW) in the seconds before seismic waves hit or for the coordination of emergency responders afterwards, knowledge on the approximate location, length and orientation of the earthquake rupture is crucial. Estimating fault rupture extent in real-time though is difficult; the temporal and spatial evolution of extended fault ruptures can be highly complex.

The majority of EEW systems estimate the expected shaking intensity at a given user site from the magnitude and epicentral distance R_{epi} of the earthquake applying empirical ground motion prediction equations (e.g. Allen *et al.* 2009; Böse *et al.* 2012a). For larger earthquakes ($M_w > 6$), however, it would be more appropriate to use rupture-to-site distances, such as Joyner–Boore distance R_{jb}, which required knowledge on source finiteness (Böse & Heaton 2010). Neglecting the rupture extent can underestimate shaking intensities and result in warnings not being issued because shaking predictions do not meet predefined thresholds.

A prominent example is from the recent 2011 M_w 9 Tohoku-Oki earthquake in Japan; while ground motion predictions by the Japanese Meteorological Agency (JMA) were fairly accurate in the Sendai area, which is close to the point of rupture nucleation, a warning was not transmitted to people and users in the Kanto region around Tokyo. Even though Kanto shaking was strong, it was so far from the epicentre (R_{epi} $>$350 km) that it was seriously under predicted by the JMA EEW system (Hoshiba & Iwakiri 2011; Sagiya *et al.* 2011). The recognition of fault rupture extent would likely have significantly improved the system performance during the Tohoku-Oki earthquake (Kurahashi & Irikura 2011).

Information on earthquake source finiteness is also important for the computation of rapid response ShakeMaps, which are maps of spatially interpolated recorded ground motions. These maps provide critical information to search and rescue teams and other

stakeholders in the aftermath of strong earthquakes (Wald *et al.* 1999). Generally, fault finiteness is not explicitly considered in ShakeMaps within the first hour(s) following the earthquake. In the case of the 2011 Tohoku-Oki earthquake, for instance, a first ShakeMap considering source finiteness was available 2–3 hr after the event (Hayes *et al.* 2011). Currently, fault dimensions in ShakeMaps are estimated from near-fault strong motion data, surface offset observations, geodetic displacements, regional and local waveforms and teleseismic data, which need to be processed and interpreted (Wald *et al.* 2005). Recently, Convertito *et al.* (2012) proposed a Bayesian approach to estimate fault rupture extent within minutes using the residuals of observed and predicted ground-motion quantities.

To provide estimates of fault rupture extent during large earthquakes while still in progress, we have developed the *Fin*ite Fault Rupture *De*tecto*r* (FinDer) algorithm. FinDer uses image recognition techniques to detect and map finite fault ruptures automatically and in real-time based on a rapid near/far-source classification of seismic observations recorded by a dense seismic network (station spacing <50 km), and comparison of these observations with a set of pre-calculated templates. The continuously updated estimates of source dimensions as provided by FinDer make predicted ground motions more accurate (since they can be predicted from R_{jb} rather than R_{epi}) and thus more useful for EEW, ShakeMaps and related products.

FINITE FAULT RUPTURE DETECTOR 'FinDer'

The Finite Fault Rupture Detector algorithm FinDer detects and maps finite source earthquake ruptures ($M_w > 6$) in real-time. Given data available at time t_i, the algorithm estimates the current centroid position, rupture length L and strike θ of the assumed line source (without weighting for slip). The algorithm accomplishes this by comparing recorded ground motion amplitudes with a set of pre-calculated templates. The processing flow is shown in Fig. 1.

FinDer presumes the existence of a dense seismic network (station spacing <50 km) with real-time data transmission to a central processing site. Let us assume that at a given time t_i a set of ground motion observations at all or at a subset of seismic sensors is available. Whenever an earthquake is detected [for example, using the Virtual Seismologist (Cua & Heaton 2007), PreSEIS (Böse *et al.* 2008; Böse *et al.* 2012b) or any other algorithm for EEW (Allen *et al.* 2009)], ground motion amplitudes are soil-to-rock corrected,

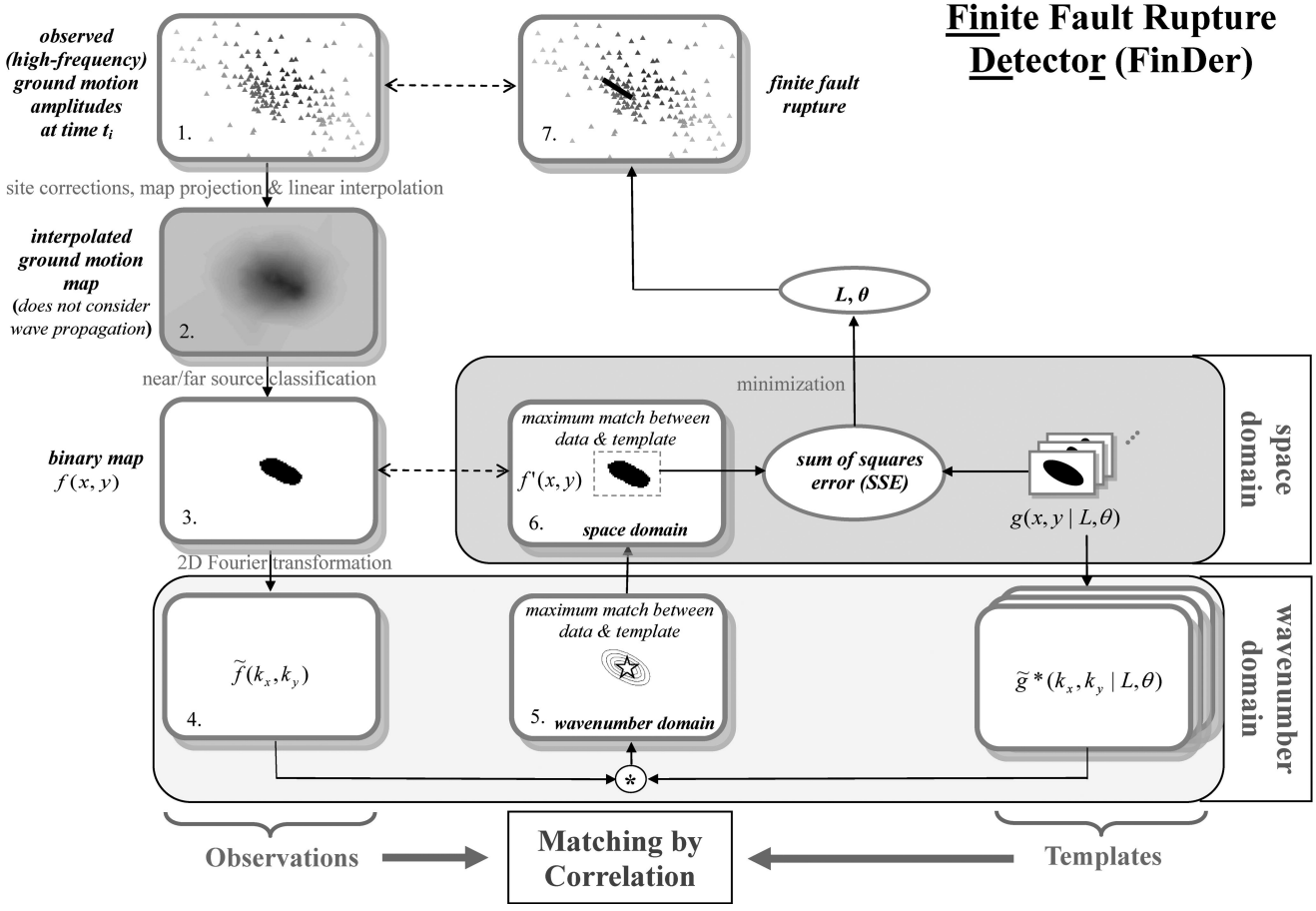

Figure 1. Processing steps of the Finite Fault Rupture Detector (FinDer) algorithm to determine the current centroid position, length L, and strike θ of an ongoing earthquake rupture. Step 1: gathering of maximum ground motion amplitudes in a seismic network at time t_i. Step 2: empirical site-corrections, map projection, and spatial interpolation of amplitudes using zero-amplitude boundary condition. Step 3: near/far source classification, for example, based on acceleration thresholds. Steps 4–6: comparison of binary map $f(x,y)$ obtained from classification with a set of pre-calculated templates for various L and θ using 'Matching by Correlation'. The best match is achieved at the point of maximum correlation (Step 6). For increased speed, correlation is performed in the wavenumber domain. Step 7: the template (and thus L and θ) with the best fit is obtained by minimizing the misfit (SSE error) between template and ground motion observations using a Direct Pattern Search algorithm. Steps 1–7 are repeated regularly, for example, every $\Delta t = 1$ to 2 s.

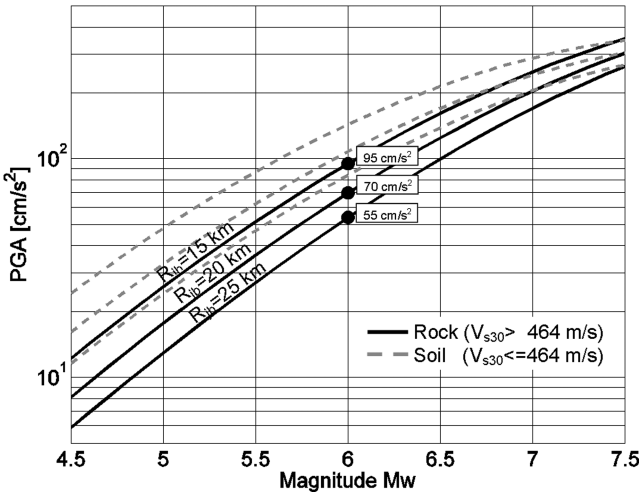

Figure 2. Saturation of peak ground acceleration (PGA) as a function of moment magnitude M_w and Joyner–Boore distance R_{jb} using relations by Cua (2005). A simple acceleration threshold is effective for near/far-source classification, because large PGA values are typically expected for stations close to the rupture, and these large PGA values occur for all events $M_w > 6$. In this study, we apply three acceleration thresholds at 95, 70 and 55 cm s^{-2}, to identify near-source stations for $M_w > 6$ at $R_{jb} \leq 15$, $\leq$20 and $\leq$25 km, respectively.

projected onto a map, and interpolated spatially onto a fine grid assuming a zero-amplitude boundary condition (Fig. 1, Steps 1 and 2). Shortly after rupture nucleation the seismic *P* and *S* waves will have reached only a few sensors in the network. The spatial interpolation of maximum amplitudes is carried out over all stations, independent from the type of observation (*P* wave, *S* wave, surface wave, background noise); that is, independent from wave propagation.

In the next step, FinDer assigns a unit less value of '+1' to all sites in the interpolated map that appear to be close to the rupture, and a value of '0' to those further away. This near/far-source classification gives a binary image $f(x,y)$ with Cartesian coordinates x and y (Fig. 1, Step 3). In this study, we use simple high-frequency (acceleration) thresholds for near/far-source classification, motivated by the observation that high acceleration values are typically observed very close to the rupture only (Fig. 2). We apply acceleration thresholds at 95, 70 and 55 cm s^{-2}, to identify near-source stations for earthquakes with $M_w \geq 6$ at Joyner–Boore distances of $R_{jb} \leq 15$, $\leq$20 km and $\leq$25 km, respectively (Fig. 2). Principally, however, we could also apply more sophisticated algorithms for classification that make use, for example, of additional information from mid- and low-frequency observations (Yamada *et al.* 2007; Böse *et al.* 2012b).

To estimate the current centroid position, length L, and strike θ of the underlying rupture, FinDer compares the binary *image* $f(x,y)$ obtained from near/far-source classification with a set of around 12,500 pre-calculated *templates* $g(x,y|L,\theta)$ for various L and θ (here: 5 km $\leq L \leq$ 350 km with increments of $\Delta L = 5$ km, and $0^\circ \leq \theta < 180^\circ$ with $\Delta\theta = 1^\circ$). We assume that the spatial coverage of each template is much smaller than that of the interpolated map. For a systematic comparison of the impact of different near/far-source distances in this paper, we generated three sets of templates. Sites close to the rupture ($R_{jb} \leq 15$, $\leq$20 and $\leq$25 km, respectively) were assigned a unit less value of '+1', and a value of '0' to those further away. To determine the optimum solutions of L and θ at time t_i, we need to find (1) for a given L–θ pair, the place in $f(x,y)$ that shows the best match with template $g(x,y|L,\theta)$, and (2) the template with the best overall fit, that is smallest error.

To solve requirements (1) and (2) in real-time, we use the templates $g(x,y|L,\theta)$ as spatial filters and compute the sum of products for each location of g in f. The best match(es) is (are) the location(s) of the maximum value(s) in the resulting correlation image. This method is known as 'Matching by Correlation' (e.g. Gonzales *et al.* 2004). However, since this approach is generally computationally intensive and therefore not well suited for real-time procedures such as EEW, we implement correlation in the wavenumber domain, using the correlation theorem:

$$f(x,y) \star g(x,y\,|L,\theta) \Leftrightarrow \tilde{f}\left(k_x,k_y\right)\tilde{g}^*\left(k_x,k_y\,|L,\theta\right). \quad (1)$$

Eq. (1) relates the spatial correlation to the product of the Fourier transforms $\tilde{f}(k_x,k_y)$ and $\tilde{g}^*(k_x,k_y|L,\theta.)$ in the wavenumber domain (k_x,k_y), where '$\star$' denotes correlation and '*' the complex conjugate. In other words, spatial correlation is obtained as the inverse Fourier transform of the product of the transform of one function times the conjugate of the transform of the other (e.g. Gonzales *et al.* 2004). This approach is very efficient and can be parallelized for additional speed if needed.

The maximum of the correlation image $\tilde{f}(k_x,k_y)\tilde{g}^*(k_x,k_y|L,\theta)$ gives the location of the best match of *subimage* $f'(x,y)$ and template $g(x,y|L,\theta)$ for a given L–θ pair (Fig. 1, Steps 4–6). If we suspect that multiple earthquakes occur simultaneously, we can analyse further (local) maxima in the correlation image.

To determine the optimum solutions of L and θ, we need to find the template with the smallest misfit, that is, the minimum sum of squared errors (SSE):

$$\mathrm{SSE} = \sum_x \sum_y \left[f'(x,y) - g(x,y\,|L,\theta)\right]. \quad (2)$$

We use a Direct Pattern Search optimization algorithm (Mathworks Matlab Global Optimization Toolbox), which usually finds a minimum of eq. (2) after less than 15 iterations. Thus we can quickly estimate the current location and dimensions of the ongoing rupture, and determine L and θ in real-time (Fig. 1, Step 7). Once the current rupture length L is determined, we can also estimate magnitude M_w from empirical L–M_w relations, such as proposed by Wells & Coppersmith (1994). Note, however, that these relations depend on the type of focal mechanism and on whether the rupture is along the surface or subsurface, involving additional uncertainties of $\sim\pm$0.1–0.2 M_w units in the L–M_w relationship.

Steps 1–7 are repeated every Δt seconds using the currently available ground motion amplitudes. The total processing time depends on the desired resolution and spatial coverage of the interpolated maps and templates. In this study, we used a resolution of 0.025° for $L \leq 50$ km and a resolution of 0.05° for $L > 50$ km. For the examples shown in this paper, the computation of L and θ at each time t_i takes on the order of around 1 s on a modern desktop computer, that is, we can update the estimates every $\Delta t = 1$–2 s. A higher resolution would increase the computational effort without gaining much additional information about the rupture of a large earthquake.

DATA

In this study, we apply FinDer to several recorded and simulated earthquakes with different focal mechanisms, including the 2009 M_w 6.3 L'Aquila (Italy) normal fault earthquake (e.g. Ameri *et al.* 2012; Gallovič & Zahradník 2012), the 1999 M_w 7.6 ChiChi

Table 1. Observed and estimated rupture length L_{obs} and L_{est}, and azimuth θ_{obs} and θ_{est} at six time steps after origin. Errors in L are on the same order as station spacing. Map resolution is 0.05° (∼5 km), that is L_{est} can change in these examples in increments of $\Delta L = 5$ km only. $L_{est,15}$ and $\theta_{est,15}$ were estimated using a near/far-source classification threshold of 95 cm s^{-2} ($R_{jb} \leq 15$ km); $L_{est,20}$ and $\theta_{est,20}$ refer to a near/far-source classification threshold of 70 cm s^{-2} ($R_{jb} \leq 20$ km), and $L_{est,25}$ and $\theta_{est,25}$ refer to 55 cm s^{-2} ($R_{jb} \leq 25$ km). See Figs 4 and 5 for plotted data. Note that the rupture duration of the M7.8 ShakeOut scenario earthquake exceeds 60 s.

Event	10 s	20 s	30 s	40 s	50 s	60 s	Final
*M*6.3 L'Aquila (RAN)							
L_{obs}	20 km	20 km	20 km	20 km	20 km	20 km	20 km
$L_{est,15}$	5 km	10 km	15 km	20 km	20 km	20 km	20 km
$L_{est,20}$	5 km	10 km	10 km	10 km	10 km	10 km	10 km
$L_{est,25}$	5 km	15 km	20 km	30 km	30 km	30 km	30 km
θ_{obs}	140°	140°	140°	140°	140°	140°	140°
$\theta_{est,15}$	45°	66°	84°	91°	110°	110°	110°
$\theta_{est,20}$	135°	133°	136°	136°	139°	139°	139°
$\theta_{est,25}$	133°	127°	120°	117°	110°	110°	110°
*M*7.6 ChiChi (TSMIP)							
L_{obs}	40 km	80 km	100 km	100 km	100 km	100 km	100 km
$L_{est,15}$	5 km	45 km	80 km	110 km	120 km	120 km	120 km
$L_{est,20}$	10 km	45 km	80 km	110 km	125 km	135 km	135 km
$L_{est,25}$	5 km	35 km	75 km	105 km	125 km	130 km	130 km
θ_{obs}	23°	23°	23°	23°	23°	23°	23°
$\theta_{est,15}$	45°	37°	37°	37°	34°	36°	36°
$\theta_{est,20}$	132°	73°	51°	42°	23°	18°	18°
$\theta_{est,25}$	43°	20°	28°	33°	37°	46°	46°
*M*7.8 ShakeOut (CISN)							
L_{obs}	50 km	80 km	110 km	135 km	165 km	195 km	300 km
$L_{est,15}$	5 km	40 km	70 km	110 km	145 km	225 km	335 km
$L_{est,20}$	5 km	45 km	85 km	115 km	160 km	225 km	330 km
$L_{est,25}$	10 km	50 km	90 km	125 km	165 km	240 km	350 km
θ_{obs}	121°	121°	121°	121°	121°	121°	121°
$\theta_{est,15}$	101°	120°	120°	120°	120°	120°	120°
$\theta_{est,20}$	146°	125°	123°	122°	119°	119°	119°
$\theta_{est,25}$	139°	127°	124°	122°	119°	119°	119°

(Taiwan) thrust fault earthquake (e.g. Ji *et al.* 2001) and the 2008 M_w 7.8 ShakeOut scenario (southern California) strike-slip fault earthquake (Jones *et al.* 2008).

On 2009 April 6, a M_w 6.3 earthquake hit the city of L'Aquila (Central Italy), located at only a few kilometres northeast of the epicentre, and causing more than 300 casualties and large damage to the city and neighboured villages (Ameri *et al.* 2012). The earthquake occurred along a NW–SE trending normal fault, with an estimated rupture length of $L_{obs} \approx 20$ km and strike of $\theta_{obs} \approx 140°$ (Ameri *et al.* 2012). We downloaded the corrected strong-motion waveforms recorded by the Italian Strong motion Network (*Rete Accelerometrica Nazionale*, RAN) from the Italian strong-motion database (ITACA, http://itaca.mi.ingv.it); we then determined the maximum absolute ground motion amplitudes (largest value of all three components) of the corrected records in time windows of 1 s on each record to obtain acceleration envelopes, such as described in Cua & Heaton (2007). About 14 of the waveforms were recorded within 50 km distance from the earthquake epicentre (Ameri *et al.* 2012).

The 1999 M_w 7.6 ChiChi earthquake (Taiwan) was one of the largest inland events in the 20th century and produced a relatively dense set of strong motion recordings. It produced an almost 100-km-long surface rupture along the primarily NS striking Chelungpu fault with offsets of up to 8 m near the northern end of the fault (Ji *et al.* 2001). The Harvard CMT solution shows a shallow (27°) dipping thrust movement towards east. Due to this small dip angle we do not expect the surface-projected rupture (as predicted by FinDer) to agree with the surface trace of the Chelungpu fault, but to be shifted to the east. From a fault model proposed by Ji *et al.* (2001) we estimate the rupture length, azimuth, and velocity of the ChiChi earthquake as $L_{obs} \approx 100$ km, $\theta_{obs} \approx 23°$, and $v_r = 2.0$ km s^{-1}, respectively (Table 1).

Seismic waveforms of the ChiChi earthquake were recorded at 441 free-field strong-motion stations of the Taiwan Strong-Motion Instrumentation Program (TSMIP) operated by the Central Weather Bureau (CWB) with an average station spacing of ∼5 km. We downloaded the acceleration records from the COSMOS Virtual Datacenter (http://db.cosmos-eq.org), and determined the maximum absolute ground motion amplitudes of the corrected records in time windows of 1 s of each record to obtain acceleration envelopes (Fig. 3a). We apply Vs_{30} values from the National Center for Research on Earthquake Engineering (NCREE) and CWB (http://geo.ncree.org.tw) for empirical soil-to-rock corrections (explanation follows).

The ShakeOut scenario earthquake is a best-estimate scenario rupture of a M_w 7.8 earthquake on the southernmost portion of the San Andreas Fault (Jones *et al.* 2008). The rupture of this right-lateral strike slip event starts at the Salton Sea in the south and travels over a distance of $L_{obs} = 300$ km in a northwesterly direction towards Lake Hughes. The ShakeOut scenario earthquake was used to study

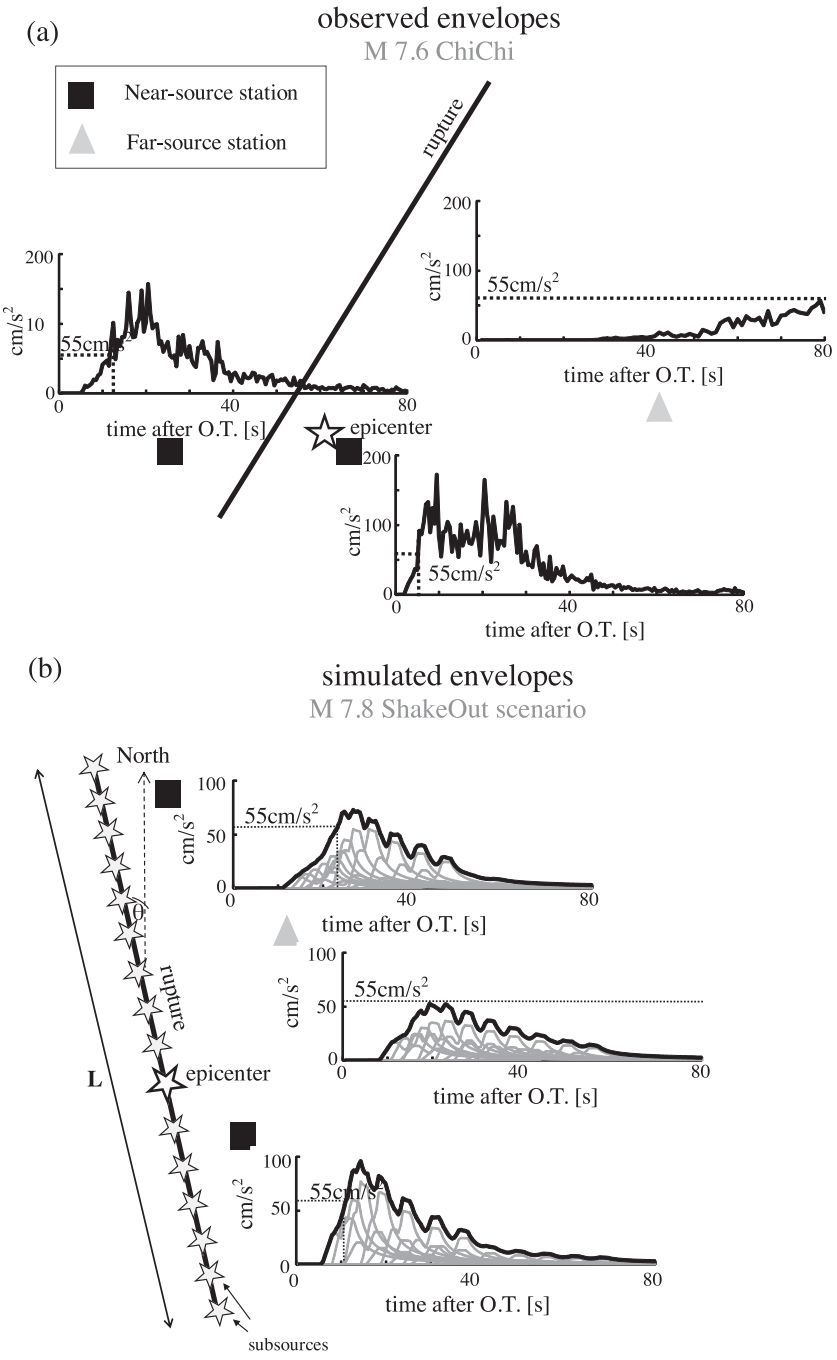

Figure 3. Examples for (a) observed and (b) simulated acceleration envelope functions for the M_w 7.6 ChiChi and the M_w 7.8 ShakeOut scenario earthquake rupture (black lines). The simulations in (b) were obtained by summing the envelopes of multiple smaller subsource events of M_w 6 (grey stars). We use high-frequency thresholds at 95, 70 and 55 cm s^{-2}, to identify near-source stations for earthquakes with $M_w > 6$ at Joyner–Boore distances of $R_{jb} \leq 15$, ≤ 20 and ≤ 25 km, respectively (here we show classification results for 55 cm s^{-2} only). The envelopes are used to determine if and when these thresholds at a given station are reached.

the potential impact of a major earthquake on southern California and also formed the bases for the 2008 November 13, ShakeOut Earthquake Drill with over five million participants (Jones *et al.* 2008).

Graves *et al.* (2010) simulated seismic ground-motions for the ShakeOut scenario earthquake in the low- to mid-frequency band using a finite difference method. However, these waveforms cannot be used for the testing of FinDer, since our near/far-source classification is based on high-frequency motions. We therefore apply a method proposed by Yamada & Heaton (2008) to simulate the envelope time functions of high-frequency ground motions during large earthquakes.

Simulated envelope time series

To calculate the envelope time functions of ground motions for the ShakeOut scenario earthquake, we apply a simple 2D source-model

approach (Yamada & Heaton 2008) that extends earlier ground-motion models by Cua (2005) and Cua & Heaton (2007). In this approach the fault surface is divided into multiple (not overlapping) subfaults, each represented by a single point (sub-) source that radiates *P* and *S* waves when the rupture front arrives. The envelope of the *i*th subsource is described by

$$E_i\left(t\,|M,\,R_{\rm epi}\right) = E\left[t - t_i\,|M_i,\,R_{{\rm epi},i}\right], \tag{3}$$

where M_i is the magnitude of the subsource i, $R_{{\rm epi},i}$ is the (epicentral) subsource-to-site distance, and t_i is the time delay due to the rupture propagation. Eq. (3) can be solved analytically for point source earthquakes ($M \leq 6$) from ground-motion models proposed by Cua (2005) and Cua & Heaton (2007).

For larger finite-source earthquakes ($M > 6$), the total envelope of ground motion $E_{\rm total}$ at a given site is modelled as the combination of the responses of each subsource i (Fig. 3b). For high-frequency motions with roughly random phase, the square root of the sum of the squares of the envelope amplitudes from each subsource gives a good approximation of $E_{\rm total}$:

$$E_{\rm total}\left(t\,|M,\,R_{\rm jb}\right) = \sqrt{\sum\nolimits_{i=0}^{N} E_i^2\left(t\,|M,\,R_{\rm epi}\right)}, \tag{4}$$

where N is the number of the point sources. This approach is only suited for the modelling of high-frequency ground motions that seem to be fairly insensitive to the effects of source radiation and directivity (Yamada & Heaton 2008).

We assume that the dimensions of all subsources are uniform and that each subsource magnitude is $M_i = 6$, corresponding to a subfault length of $l \approx 10$ km (Wells & Coppersmith 1994). The number of subsources is determined from $N \approx (L/l)$. We use a 1-D seismic velocity model for southern California to determine the *P* and *S* wave arrivals for the ShakeOut scenario (Hutton *et al.* 2010). Fig. 3(b) shows three example simulations.

Cua (2005) and Cua & Heaton (2007) determined ground motion models for both rock and soil conditions. They classified sites with $V_{s30} \leq 434$ m s^{-1} as 'soil' and sites with $V_{s30} > 434$ m s^{-1} as 'rock', where V_{s30} is the average shear wave velocity taken over the uppermost 30 m. We simulate a set of envelope functions for both rock and soil condition for M 6 and distances of up to 100 km, and obtain a mean amplification factor of peak amplitudes in the order of 1.35. We take this empirical value for soil-to-rock corrections of simulated and observed ground motion envelopes in this study.

We simulate the rupture of the ShakeOut scenario earthquake propagating with a rupture speed of $v_r = 2.9$ km s^{-1}, using the current distribution of seismic real-time stations in the southern California Integrated Seismic Network (CISN/SCSN; www.cisn.org and www.scsn.org) with an average station spacing of ∼20 km. The rupture of the $M_{\rm w}$ 7.8 ShakeOut scenario earthquake extends across the Big Bend of the San Andreas Fault. We approximate this trace by a linear rupture with $\theta_{\rm obs} = 121^\circ$, which requires shifting the epicentre by 0.15° (∼17 km) to the east (Table 1). We apply V_{s30} values determined by Wills *et al.* (2000) for the simulation of envelope functions at rock and soil stations.

Alike for the recorded L'Aquila and ChiChi earthquakes, we apply thresholds of 95, 70 and 55 cm s^{-2} to the simulated ground-motion envelopes to estimate, whether a site is close to the rupture (≤15, ≤20 or ≤25 km, respectively) or further away. The simulations help us to identify if and when these thresholds at a given station are exceeded.

RESULTS

We replay the records of each earthquake as simulated real-time data streams and emulate the temporal evolution of predictions by FinDer at regular time steps t_i after origin time (O.T.). We assume that the transmission delays of seismic waveform data from the stations to the central processing facility are insignificantly small and negligible.

Table 1 and Fig. 4 compare the observed and estimated rupture lengths and azimuths, $L_{\rm obs}$, $L_{\rm est}$, $\theta_{\rm obs}$ and $\theta_{\rm est}$, for each of the three earthquakes. Values for $L_{\rm obs}$ and $\theta_{\rm obs}$ are taken from the literature and were discussed earlier; $L_{\rm est,15}$ and $\theta_{\rm est,15}$ were estimated using a near/far-source classification threshold of 95 cm s^{-2} ($R_{\rm jb} \leq 15$ km); $L_{\rm est,20}$ and $\theta_{\rm est,20}$ refer to a near/far-source classification threshold of 70 cm s^{-2} ($R_{\rm jb} \leq 20$ km), and $L_{\rm est,25}$ and $\theta_{\rm est,25}$ refer to 55 cm s^{-2} ($R_{\rm jb} \leq 25$ km). Fig. 5 shows screenshots of the estimated near/far source station classifications (using an acceleration threshold of 70 cm s^{-2}) and estimated rupture parameters in map view at four time steps after origin.

For all events we observe a good convergence towards the observed values $L_{\rm obs}$ and $\theta_{\rm obs}$ with increasing time. The smaller the station spacing, the quicker the detection of a large earthquake, and the more accurate are the estimated parameters (Table 1; Figs 4 and 5). The map resolution in these examples is 0.05° (∼5 km), that is $L_{\rm est}$ increases in increments of 5 km only. Note that the earthquake ruptures are usually still in progress when FinDer provides the first estimates of their dimensions and azimuth. Usually, the predictions of L are smaller or equal to the current rupture length; that is, FinDer keeps track of the current rupture parameters without predicting their future evolution (Table 1; Figs 4 and 5). Generally, errors in L are on the same order as the station spacing. There are no systematic differences in the performance depending on the three near/far-source classification thresholds.

The rupture length of the ChiChi earthquake appears to be overestimated by ∼30 per cent around 1 min after rupture nucleation (Table 1, Fig. 5b). Shin & Teng (2001), however, suggested the occurrence of a secondary rupture along the Shihtan fault shortly after the ChiChi main shock. This rupture is consistent with an extended zone of near-source classified stations, north of the main rupture, as was also observed by Yamada *et al.* (2007). Both ruptures occur very close to each other in space and time and cannot be distinguished by FinDer. Also note that during the ChiChi earthquake some stations along the Taiwanese east coast were incorrectly classified as near-source sites (Fig. 5b); the predictions of L and θ, however, are robust and not affected by these misclassifications.

Estimated shaking intensity

Most EEW systems, including the current version of the CISN ShakeAlert Earthquake Early Warning Demonstration System for California (Böse *et al.* 2012a), estimate the shaking intensity at a given user site from predicted magnitudes and epicentral distances $R_{\rm epi}$ using empirical ground motion prediction equations. More accurate predictions are obtained, if $R_{\rm epi}$ is replaced by Joyner–Boore distance $R_{\rm jb}$; that is, the closest distance of the user site to the surface-projected earthquake rupture (Böse & Heaton 2010). This procedure, however, requires the knowledge or estimation of rupture dimensions, such as those provided by FinDer.

In the following, we will analyse the estimated and observed Modified Mercalli Intensity (MMI) intensities for the $M_{\rm w}$ 7.8 ShakeOut scenario earthquake in three cities in southern California: Palm Springs ($R_{\rm epi} = 80$ km), Riverside ($R_{\rm epi} = 160$ km) and Los Angeles

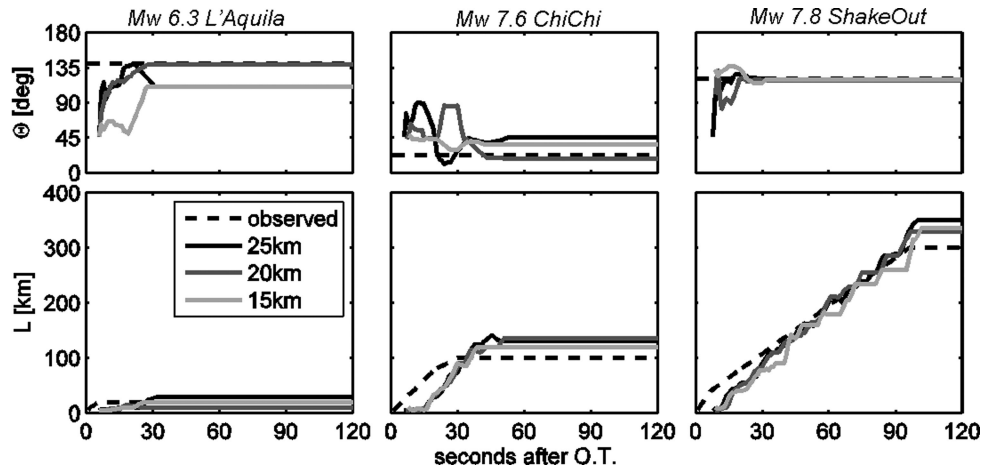

Figure 4. Comparison of observed and estimated rupture length L and strike θ for the M_w 6.3 L'Aquila, M_w 7.6 ChiChi and M_w 7.8 ShakeOut scenario earthquakes within the first 2 min after origin. The dashed lines show the current rupture length assuming constant rupture velocities of $v_r = 2.0$ km s^{-1} for the ChiChi earthquake, and $v_r = 2.9$ km s^{-1} for the two other events. FinDer keeps track of the current rupture parameters; estimates in L are usually below or equal to the current rupture dimensions. The plots show the results for three templates with near/far-source classification thresholds at 95 cm s^{-2} ($R_{jb} \leq$ 15 km), 70 cm s^{-2} ($R_{jb} \leq 20$ km) and 55 cm s^{-2} ($R_{jb} \leq 25$ km).

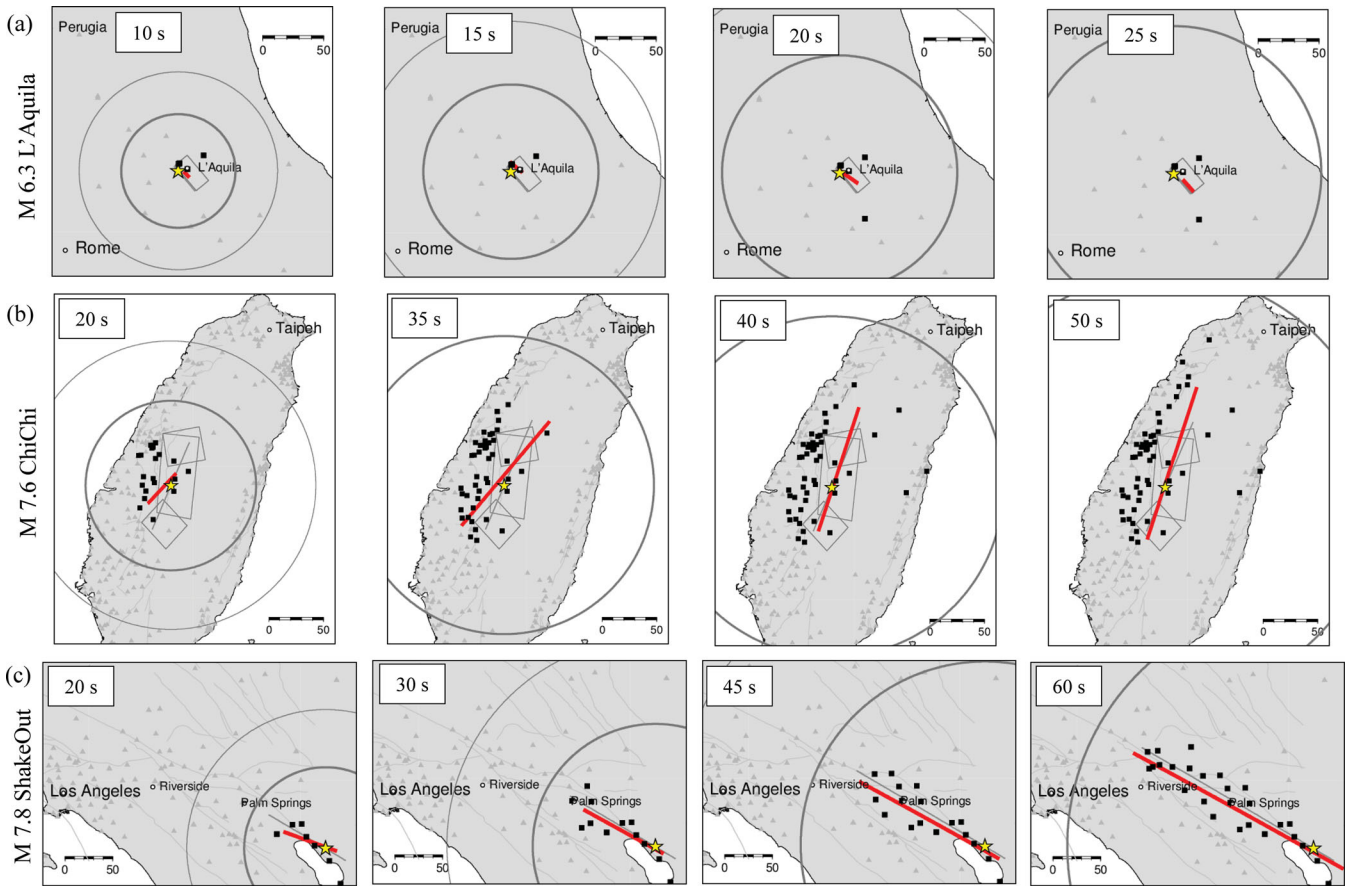

Figure 5. Results of near/far-source classification and predicted (surface-projected) ruptures for the (a) M_w 6.3 L'Aquila (Italy), (b) M_w 7.6 ChiChi (Taiwan), and (c) M_w 7.8 ShakeOut scenario (southern California) earthquakes at four time steps after rupture nucleation. Black squares mark near-source, grey triangles far-source classified stations determined from high-frequency thresholds (here: 70 cm s^{-2}). Circles show the P and S wave fronts; grey lines show evolution of L_{obs}. The grey rectangles in (b) show a three-plane fault geometry with shallow dip towards east determined from static inversion (Ji *et al.* 2001). The estimated ruptures are robust and not strongly affected by misclassified stations in (b). The final fault rupture extent of the ChiChi earthquake is overestimated by ∼30 per cent, caused by a secondary rupture along the Shihtan fault in the north shortly after the nucleation of the main shock.

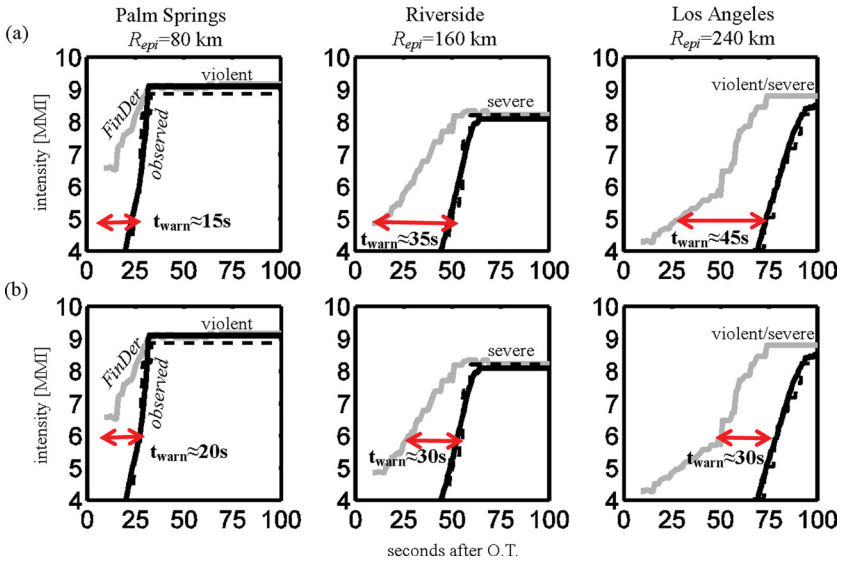

Figure 6. Estimated and observed MMI intensities for the M_w 7.8 ShakeOut scenario earthquake at three cities in southern California (for rock condition). The temporal evolution of MMI is determined from empirical peak-value-to-MMI relations (Wald *et al.* 1999) using simulations by Graves *et al.* (2010; solid black lines) and envelope simulations in the work (dashed black lines). The grey lines show the evolution of MMI using Joyner–Boore distances R_{jb} derived from the FinDer predicted rupture dimensions. An EEW system could provide (a) 15–45 s of warning if configured to issue a warning if MMI ≥ 5 ('moderate' shaking), and (b) 20–30 s of warning if MMI ≥ 6 ('strong' shaking).

Downtown (R_{epi} = 240 km). Fig. 6 shows the temporal evolution of observed (simulated) MMI values at these sites determined by Aagaard (personal communication, 2012) from the ShakeOut simulations (Graves *et al.* 2010) using empirical peak-value-MMI relations (Fig. 6, solid black lines). These intensities agree well with the intensity evolution determined from the envelope times series simulated in this work (Fig. 6, dashed black lines); for the latter we assumed that the MMI values increase by one full intensity unit in Los Angeles Downtown due to basin effects.

The observed (simulated) intensities in the three cities are $MMI_{obs} \approx 9$ (Palm Springs), $MMI_{obs} \approx 8$ (Riverside), and $MMI_{obs} \approx$ 8–9 (Los Angeles), corresponding to 'violent', 'severe' and 'violent/severe' shaking, respectively (Fig. 6, solid and dashed black lines). If we estimate these intensities from magnitude M_w (which is assumed to be known from another (point-source) algorithm for EEW) and epicentral distances R_{epi}, and apply empirical relations by Cua & Heaton (2007) and Wald *et al.* (1999), we obtain $MMI_{Repi} \approx$ 6 (Palm Springs), $MMI_{Repi} \approx 5$ (Riverside) and $MMI_{Repi} \approx 4$ (Los Angeles), corresponding to 'strong', 'moderate' and 'light' shaking, respectively. This means that the intensities in this example were underestimated by up to around three to four units, if rupture dimensions were neglected, that is if MMI was estimated from R_{epi} rather than from R_{jb}. An operational EEW system would likely not have issued a warning to users in Riverside and Los Angeles although shaking at these sites could be very strong and damaging.

We can improve these estimates of shaking intensities, if we consider rupture-to-site distances from the FinDer estimated rupture dimensions (Fig. 6, solid grey lines). Strong shaking (MMI ≥ 5) in the three cities in this example starts at ∼25, ∼55 and ∼90 s after O.T. (Fig. 6). An EEW system set up to issue a warning whenever the expected MMI intensity exceeds $MMI_{warn} = 5$ (*moderate* shaking), could provide warnings of ∼15 s (Palm Springs), ∼35 s (Riverside) and ∼45 s (Los Angeles), if rupture dimensions were considered (Fig. 6a); a system set up to issue a warning whenever the expected MMI intensity exceeds $MMI_{warn} = 6$ (*strong* shaking), could provide warnings of ∼20 s (Palm Springs), ∼30 s (Riverside), and ∼30 s (Los Angeles; Fig. 6b).

DISCUSSIONS

The FinDer algorithm keeps track of an ongoing earthquake rupture and gives estimates of its location, length, and orientation in real-time (Figs 4 and 5). FinDer, however, does not predict the future evolution of this rupture; that is, it does not attempt to predict whether the rupture will keep growing or not, as is implicitly done in other EEW (point source) algorithms. The probability for an earthquake to evolve into a large event is largely controlled by the characteristics of the rupturing fault (Böse & Heaton 2010). An EEW system for large earthquakes will thus benefit from the real-time recognition of the fault, along which rupture occurs. FinDer is suited for such rapid fault association: if the predicted location and azimuth of the rupture agrees well with the parameters of a known active fault, for example, of the San Andreas Fault, the probability that rupture is occurring along this fault is high. Also, the usage of two types of templates, for *specific* ruptures (for a given fault) and for *generic* ruptures (as used in this study) is possible. If rupture is occurring along a smooth (mature) fault, a warning should be issued immediately, because the probability for a large earthquake is high (Böse & Heaton 2010).

For software such as ShakeMap, it would be necessary to provide estimates of how the rupture length L is distributed relative to the epicentre. Thus, rather than L and θ, it would be beneficial to provide $L+$, $L-$ and θ, where $L+$ is defined as the length in the direction of θ, and L- the opposite direction. While the current FinDer algorithm does not determine the location of the earthquake epicentre, we anticipate that by combining our algorithm with other (point-source) methods, such as the 'Virtual Seismologist' (Cua & Heaton 2007), estimates of $L+$ and $L-$ could be provided in real-time.

The application of image recognition techniques, such as in FinDer, requires the availability of seismic near- and far-source

observations from a dense seismic network. We estimate that the required station spacing should be <50 km, but further study on the trade-off of station spacing and solution convergence is required. We have shown in this paper that the errors in estimated rupture dimensions are typically on the same order as the station spacing in the network. The application of FinDer to dense seismic real-time networks as are currently operated in Japan, California, Taiwan, Italy or New Zealand, is thus straightforward. However, if additional constraints, for example information on known active faults, are used to assign *a priori* probabilities to L and θ or to reduce the space of possible solutions, FinDer may also have applications in less dense networks.

We have adopted simple high-frequency thresholds for near/far source classification of ground motion observations. Site corrections seem to play a secondary role only. While the recognition of finite fault ruptures in FinDer clearly depends on a good separation of near- and far-source stations, the selection of classification thresholds is not too critical for the overall performance, as long as the general shape of the near-source ellipse remains largely unchanged. We recommend adapting the classification threshold to the station density of the seismic network to which FinDer is applied; in a dense network we expect a number of stations to be very close to the rupture, while in a sparse network there is a chance that no station is classified as near-source if the threshold is too high. The higher the station density, the larger thresholds can be selected to reduce the uncertainty in the predicted rupture strike. However, note that a high threshold value generally requires longer time to be exceeded, that is warning times will likely be reduced if thresholds are too high.

The FinDer algorithm is also suited to process multiple earthquake ruptures that occur simultaneously in the area of interest. This can be done by analysing multiple maxima in the correlation image (Fig. 1). However, if the earthquakes occur very close to each other in space and time, for example, in complex foreshock or aftershock sequences, the separation of single events becomes more and more difficult (e.g. Karakus & Heaton 2011); such rupture complexity can be seen in Fig. 5b for the ChiChi earthquake and its secondary rupture in the north along the Shihtan fault (Fig. 5b).

CONCLUSIONS AND OUTLOOK

Although large earthquakes are infrequent, many people could benefit from an EEW system capable of processing data from a finite source (Heaton 1985; Allen 2006). Large earthquakes ($M_w > 6$), with rupture lengths of several tens to hundreds of kilometres, cause damaging ground shaking over large areas. Warning times in advance of the onset of strong shaking to these areas may exceed more than 1 min, because seismic ruptures usually propagate with ∼80 per cent of the seismic S-wave speed only.

We have developed a FinDer algorithm that uses image recognition techniques to rapidly detect and map large (finite source) earthquake ruptures in real-time. The predicted rupture dimensions help improving source and ground motion estimates for EEW, Shakemaps and related products, since they can be predicted from R_{jb} rather than R_{epi}. Using three recorded and simulated earthquakes, we have shown in this study that the real-time detection and mapping of finite faults from image recognition of seismic observations is feasible, and will improve ground motion predictions.

The proposed algorithm is suited for the finite-fault detection within seismic networks with dense station spacing, where stations are fairly evenly distributed. Great subduction-zone earthquakes, such as the 2011 M_w 9 Tohoku-Oki event, will require some modifications to the algorithm and/or templates, to handle the one-sided station coverage. Also, subduction-zone earthquakes might need other methods for near/far source classification, in particular because closest stations are typically several tens to more than a hundred kilometres away from the rupture.

ACKNOWLEDGMENTS

Strong-motion records of the 2009 M_w 6.3 L'Aquila earthquake recorded by RAN were downloaded from the Italian strong-motion database ITACA (http://itaca.mi.ingv.it; last visited 2012 April); strong-motion records of the 1999 M_w 7.6 ChiChi earthquake recorded by TSMIP were downloaded from the COSMOS Virtual Datacenter (http://db.cosmos-eq.org, last visited 2011 December). We would like to thank Dr. Georgia Cua for providing codes to simulate ground motion envelopes and amplitudes used in this study. Most figures in this paper were generated with Matlab 7.8; maps in Fig. 5 were made with the Generic Mapping Tools version 4.2.1 (www.soest.hawaii.edu/gmt; Wessel & Smith 1998). We would like to thank Brad Aagaard and Robert Graves for data and discussions on the ShakeOut scenario earthquake. We are very grateful for the helpful comments and suggestions by two anonymous reviewers and the Editor, Dr. Frank Krüger. This work is funded through contract G09AC00258 from USGS/ANSS to the California Institute of Technology (Caltech). Funding was also provided by a grant from the Gordon and Betty Moore Foundation to Caltech. This is contribution #10068 of the Seismological Laboratory, Geological and Planetary Sciences at Caltech.

REFERENCES

Allen, R.M., 2006. Probabilistic warning times for earthquake ground shaking in the San Francisco Bay Area, *Seismol. Res. Lett.,* **77**(3), 371–376.

Allen, R.M., Gasparini, P., Kamigaichi, O. & Böse, M., 2009. The status of earthquake early warning around the world: an introductory overview, *Seismol. Res. Lett.,* **80**(5), 682–693, doi:10.1785/gssrl.80.5.682.

Ameri, G., Gallovič, F. & Pacor, F., 2012. Complexity of the M_w 6.3 2009 L'Aquila (central Italy) earthquake: 2. Broadband strong motion modeling, *J. geophys. Res.,* **117,** B04308, doi:10.1029/2011JB008729.

Böse, M. & Heaton, T.H., 2010. Probabilistic prediction of rupture length, slip and seismic ground motions for an ongoing rupture: implications for early warning for large earthquakes, *Geophys. J. Int.,* **183**(2), 1014–1030, doi:10.1111/j.1365-246X.2010.04774.x.

Böse, M., Wenzel, F. & Erdik, M., 2008. PreSEIS: a neural network based approach to earthquake early warning for finite faults, *Bull. seism. Soc. Am.,* **98**(1), 366–382, doi:10.1785/0120070002.

Böse, M. *et al*., 2012a. CISN ShakeAlert—an earthquake early warning demonstration system for California, in *Early Warning for Geological Disasters – Scientific Concepts and Current Practice,* eds Wenzel, F. & Zschau, J., Springer, Heidelberg, in press.

Böse, M., Heaton, T. & Hauksson, E., 2012b. Rapid estimation of earthquake source and ground-motion parameters for earthquake early warning using data from single three-component broadband or strong-motion sensor, *Bull. seism. Soc. Am.,* **102**(2), 738–750, doi:10.1785/0120110152.

Convertito, V., Caccavale, M., De Matteis, R., Emolo, A., Wald, D. & Zollo, A., 2012. Fault extent estimation for near-real-time ground-shaking map computation purposes, *Bull. seism. Soc. Am.,* **102**(2), 661–679, doi:10.1785/0120100306.

Cua, G., 2005. Creating the Virtual Seismologist: developments in earthquake early warning and ground motion characterization, *PhD thesis*, Department of Civil Engineering, California Institute of Technology.

Cua, G. & Heaton, T., 2007. The Virtual Seismologist (VS) method: a Bayesian approach to earthquake early warning, in *Earthquake Early Warning Systems,* pp. 85–132, eds Gasparini, P., Manfredi, G. & Zschau, J., Springer, Heidelberg.

Gallovič, F. & Zahradník, J., 2012. Complexity of the M_w 6.3 2009 L'Aquila (central Italy) earthquake: 1. Multiple finite-extent source inversion, *J. geophys. Res.,* **117,** B04307, doi:10.1029/2011JB008709.

Gonzales, R.C., Woods, R.F. & Eddins, S.L., 2004. *Digital Image Processing using MATLAB,* Pearson Prentice Hall, Upper Saddle River, NJ.

Graves, R.W., Aagaard, B.T. & Hudnut, K.W., 2010. The ShakeOut earthquake source and ground motion simulations, *Earthq. Spectra,* **27**(2), 273–291, doi:10.1193/1.3570677.

Hayes, G.P., Earle, P.S., Benz, H.M., Wald, D.J., Briggs, R.W., the USGS/NEIC Earthquake Response Team, 2011. 88 hours: the U.S. Geological Survey National Earthquake Information Center response to the 11 March 2011 M_w 9.0 Tohoku earthquake, *Seism. Res. Lett.,* **82**(4), 481–493.

Heaton, T., 1985. A model for a seismic computerized alert network, *Science,* **228**(4702), 987–990.

Hoshiba, M. & Iwakiri, K., 2011. Initial 30 seconds of the 2011 off the Pacific coast Tohoku earthquake (M_w 9.0)—amplitude and τ_c for magnitude estimation for earthquake early warning, *Earth Planets Space,* **30,** 553–557.

Hutton, L.K., Woessner, J. & Hauksson, E., 2010. Seventy-Seven Years (1932 – 2009) of earthquake monitoring in southern California, *Bull. seism. Soc. Am.,* **100**(2), 423–446, doi:10.1785/0120090130.

Ji, C., Helmberger, D.V., Song, T., Ma, K.-F. & Wald, D.J., 2001. Co-seismic slip distribution and tectonic implication of 1999 Chi-Chi earthquake, *Geophys. Res. Lett.,* **28**(23), 4379–4382.

Jones, L.M. *et al*., 2008. The ShakeOut Scenario, U.S. Geological Survey Open File Report 2008–1150, California Geological Survey.

Karakus, G. & Heaton, T., 2011. Methodologies to automatically recognize foreshock/mainshock pairs, *Proc. 2011 AGU Fall Meeting,* abstract S53A-2268.

Kurahashi, S. & Irikura, K., 2011. Source model for generating strong ground motions during the 2011 off the Pacific coast Tohoku earthquake, *Earth Planets Space,* **63,** 571–576.

Sagiya, T., Kanamori, H., Yagi, Y., Yamada, M. & Mori, J., 2011. Rebuilding seismology, *Nature,* **473,** 146–148, doi:10.1038/473146a.

Shin, T.-C. & Teng, T.-L., 2001. An overview of the 1999 Chi-Chi, Taiwan, earthquake, *Bull. seism. Soc. Am.,* **91,** 895–913.

Wald, D., Quitoriano, V., Heaton, T., Kanamori, H., Scrivner, C. & Worden, C., 1999. TriNet ShakeMaps: rapid generation of instrumental ground motion and intensity maps for earthquakes in southern California, *Earthq. Spectr.,* **15,** 537–556.

Wald, D.J., Worden, B.C., Quitoriano, V. & Pankow, K.L., 2005. ShakeMap Manual: Users Guide, Technical Manual, and Software Guide, USGS Techniques and Methods 12–A1, 128 pp.

Wells, D.L. & Coppersmith, K.J., 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area and surface displacement, *Bull. seism. Soc. Am.,* **84,** 974–1002.

Wessel, P. & Smith, W.H.F., 1998. New, improved version of the Generic Mapping Tools Released, *EOS, Trans. Am. geophys. Un.,* **79,** 579.

Wills, C.J., Petersen, M.D., Bryant, W.A., Reichle, M.S., Saucedo, G.J., Tan, S.S., Taylor, G.C. & Treiman, J.A., 2000. A site conditions map for California based on geology and shear wave velocity, *Bull. seism. Soc. Am.,* **90**(6b), 187–208.

Yamada, M. & Heaton, T., 2008. Real-time estimation of fault rupture extent using envelopes of acceleration, *Bull. seism. Soc. Am.* **98**(2), 607–619, doi:10.1785/0120060218.

Yamada, M., Heaton, T. & Beck, J., 2007. Early warning systems for large earthquakes: near-source versus far-source classification, *Bull. seism. Soc. Am.,* **97,** 1890–1910.