Rapid Source Parameter Estimations of Southern California Earthquakes Using PreSEIS

Nina Köhler,1 Georgia Cua,2 Friedemann Wenzel,1 and Maren Böse3

INTRODUCTION

Earthquake early warning (EEW) systems provide real-time estimates of earthquake source and ground motion parameters to users before strong ground shaking occurs at sites of interest (Kanamori et al. 1997; Kanamori 2005). They make use of the fact that the most destructive ground shaking during an earthquake is caused by S- and surface waves, which travel much slower than P waves and also slower than electromagnetic signals carrying warnings to potential users. Real-time information systems can minimize loss of life and property damage and are therefore an important tool in short-term seismic hazard mitigation and disaster management (Wenzel et al. 2001). If an alarm can be issued seconds before the onset of the strong ground motions, automatic emergency actions can be initiated such as slowing down high speed trains or shutting down computers or gas distribution, for instance (Goltz 2002).

EEW systems are of two main types, regional and on-site. The former uses a dense network of seismic stations to locate the earthquake, determine its magnitude, and estimate the ground motion at given sites of interest. The latter uses the observations at a single sensor to estimate the ensuing ground motion at the same site (Kanamori 2005). While regional systems work more accurately, they need more time to estimate earthquake source parameters.

EEW systems are currently operated in Japan (Nakamura 1989; Kamigaichi 2004; Horiuchi et al. 2005; Hoshiba et al. 2008), Taiwan (Wu and Teng 2002), Mexico (Espinosa-Aranda et al. 1995), Turkey (Erdik et al. 2003; Alcik et al. 2009), and Romania (Wenzel et al. 1999; Böse et al. 2007). New algorithms for EEW are being developed and tested in California (Allen and Kanamori 2003; Wurman et al. 2007; Allen et al. 2009; Wu et al. 2007; Böse et al. 2009; Cua and Heaton 2007), Italy (Zollo et al. 2006; Satriano et al. 2008; Zollo et al. 2009), and Turkey (Böse et al. 2008).

PreSEIS (Pre-SEISmic shaking) is a neural network-based approach to EEW that takes advantage of both regional and on-site early warning (Böse 2006; Böse et al. 2008). It inverts time-dependent seismic attributes derived from ground motion observations at different stations in a seismic network as soon as the first station is triggered by the arriving P wave. Starting at this point in time, PreSEIS estimates the most likely hypocentral location and magnitude of the earthquake and updates these estimates at regular time steps. Using the example of the Turkish megacity Istanbul, Böse et al. (2008) tested PreSEIS for a large suite of simulated earthquake scenarios along the main Marmara fault. The study showed a robust performance of the algorithm and demonstrated a clear and fast convergence of source parameter estimates toward correct solutions. However, the use of synthetic data is of limited meaning, since aspects such as site effects or noisy records, for example, can only be accounted for to a certain extent, although they are of major importance for a potential implementation of the method.

This study presents the first performance test of PreSEIS using real earthquake data combined with empirical relations from southern California (Cua 2005; Cua and Heaton 2007). It aims to analyze the functionality of PreSEIS in terms of 1) its capability to handle real data, 2) its operational suitability, and 3) its ability to define the problems remaining before a possible future implementation.

METHOD AND DATA

PreSEIS determines the most likely hypocentral location (latitude, longitude, depth) and moment magnitude Mw using the ground motion information available at regular time steps from a network of seismic stations. Two types of input information are used: first, the P-wave arrival time differences of the various stations relative to the first triggered station, and second, amplitude information. While Böse et al. (2008) used the cumulative absolute velocity of seismic records as amplitude information, we use the ground motion envelope, defined as the maximum absolute value of ground motion on a given channel over a one-second time window (Cua 2005; Cua and Heaton 2007). The time step at which the source parameter estimates are updated is set to 0.5 s.

At 0.5 s after the P-wave detection at the first station, PreSEIS starts estimating the hypocentral location and Mw. When more stations trigger and longer ground motion time series become available, the additional information contributes to the estimates. Thus, PreSEIS is able to use the full waveforms recorded at each station to infer information about the source parameters rather than using the early P phase only.

1. Geophysical Institute, Karlsruhe University, Germany
2. Swiss Seismological Service, ETH Zurich, Switzerland
3. Seismological Laboratory, California Institute of Technology, U.S.A.
During the first seconds of fault rupture, the $P$ wave has usually only arrived at a few stations of the network, leading to an underdetermined inversion problem for locating the hypocenter. By including the information on not-yet-triggered stations, however, the range of possible solutions can be confined, as, for example, Horiuchi et al. (2005) have done in Japan. PreSEIS requires data availability from all stations at any time; a nonfunctional station will be interpreted as one where no signal has yet arrived.

For the inversion of seismic source parameters, PreSEIS makes use of artificial neural networks (ANNs). Their high tolerance for noisy input data turns ANNs into attractive tools for EEW. ANNs consist of large numbers of simple, interconnected processing units (neurons). The importance of each connection is controlled by a weight parameter. The weight parameters of an ANN are iteratively adapted to the inversion problem by learning from a set of examples with known input and known output values. Regarding our earthquake source parameters, the ANNs learn from a training set of events with known hypocenter locations and magnitudes and the required input information from their ground motion records. A priori information about likely source locations, directivity effects, site conditions, etc. is therefore included automatically.

The outputs of the ANNs are explicit functions of the input values and the weight parameters. For each time step, two neural networks are designed. The first one estimates the hypocenter location using the arrival time differences, and the second one uses the ground motion envelopes combined with the outputs of the first network to estimate $M_w$. Once the training of the ANNs is finished, PreSEIS is able to process unknown data that follow the same statistical patterns as the training examples. For further details about PreSEIS, see Böse (2006) and Böse et al. (2008).

In order to evaluate the functionality of PreSEIS, we use a set of southern California earthquakes recorded by the Southern California Seismic Network (SCSN). The data is a subset of the data used by Cua (2005) to derive the Virtual Seismologist (VS) method, a probabilistic approach to EEW based on Bayes’s theorem. The VS uses ratios and envelope attenuation relations of seismic ground motion to determine the posterior probabilities of earthquake locations and magnitudes (Cua 2005; Cua and Heaton 2007).

The data subset consists of 70 earthquakes plus an additional set of four events that occurred in the same source region. The earthquakes have source depths ranging between 0.01 and 31.39 km and $2.0 \leq M_w \leq 7.3$. In our study, we limit the number of stations to 15. However, if enough training examples are available, the possible number of stations in PreSEIS is not limited (Böse 2006; Böse et al. 2008). Figure 1 displays the locations of the earthquakes and the seismic stations.

![Figure 1. Map of southern California with the 15 SCSN station sites (black inverted triangles), the 74 training events (gray circles), and the 1999 $M_w$ 7.1 Hector Mine (HM), the 2002 $M_w$ 4.8 Yorba Linda (YL), and the 2004 $M_w$ 6.0 Parkfield (PA) earthquakes (white circles). The black squares give the locations of larger cities within the study area.](image)
The 74 events are used as a training set for the ANNs. The source parameter estimations are then tested on three example events, the 3 September 2002 $M_w$ 4.8 Yorba Linda earthquake, located at 33.92°N and 177.78°W at a depth of 12.9 km (Hauksson et al. 2002); the 16 October 1999 $M_w$ 7.1 Hector Mine earthquake, located at 34.59°N and 116.27°W at a depth of 5 km (Graizer et al. 2002); and the 8 September 2004 $M_w$ 6.0 Parkfield earthquake, located by SCSN at 35.82°N and 120.37°W at a depth of 7.9 km (Figure 1). These events represent characteristic settings: the Hector Mine and Yorba Linda earthquakes are located within the network, the Hector Mine earthquake is located near one station but with some distance to the next stations, and the Parkfield earthquake is located remotely.

The earthquake records were downloaded from the Southern California Earthquake Data Center (http://www.data.scec.org/). When possible, the 100 samples per second, high gain, broadband (HH) channel was taken. A baseline correction was applied and the data were corrected for the instrument gain to obtain ground motion velocity. The velocity records were differentiated once to obtain ground motion acceleration. When the HH channel was clipped, the 100 samples per second, low gain accelerometer channel was downloaded instead. A baseline correction was applied and the data was again corrected for the instrument gain to obtain ground motion acceleration (Cua 2005; Cua and Heaton 2007).

Due to missing records or poor signal-to-noise ratios, only 40% of the records at the 15 stations were available. The missing ones were replaced by synthetic envelopes, predicted by applying envelope attenuation relationships established by Cua (2005). These relationships are used to calculate the expected ground motion envelope as a function of time, given the magnitude and epicentral distance of the earthquake. They were inferred from the observed envelopes by parameterizing them as a function of $P$- and $S$-wave envelopes and ambient noise at the station. The relationships predict envelopes for peak vertical and root mean square of the peak horizontal acceleration, velocity, and displacement data for both rock and soil sites. In this study, the $P$- and $S$-wave onset times at the 15 stations were determined using constant seismic velocities of $v_p = 6.8$ km s$^{-1}$ and $v_s = 3.9$ km s$^{-1}$. Since the current version of PreSEIS is limited to complete datasets, the data from Cua (2005) is highly appropriate, because it allows us to replace missing records with suitable synthetics.

Figure 2 shows an example of the observed and predicted envelopes of vertical acceleration from the Yorba Linda earthquake. The predicted envelopes were used in this study only when no observations were available.

The following section presents the source parameter estimates obtained by using vertical acceleration data.

**RESULTS**

As it can be seen from Figure 1, the epicentral distances between the 74 training earthquakes and the 15 seismic stations vary widely. This is reflected in the average time that is needed until additional ground motion information becomes available by subsequent triggered stations. The upper plot in Figure 3 demonstrates that ground motion information from a second station is, on average, available 5.5 s later than from the first station, and, on average, 8.5 s later from a third station. To trigger all 15 stations takes, on average, 64.5 s.

The middle plot in Figure 3 displays the absolute errors in hypocenter locations, derived from training the ANNs with the 74 events. The localization errors are defined as the differences between the true, SCSN locations and the estimated hypocenter locations. The errors are classified as the 25th, 50th, 75th, and 95th percentiles, with the 50th percentile marking the median. The localization errors clearly decrease with ongoing time, with a median error of 21.0 km only 0.5 s after the first $P$ wave arrival. After 17.0 s, the median localization error reaches a roughly constant level, showing an average error of 9.2 km. The average localization error obtained after 50 s is 10.5 km. The bottom plot in Figure 3 displays the mean errors and standard deviations of moment magnitude estimates of all 74 training events, defined as differences between the estimated magnitudes and their true, SCSN values. The largest average error reaches a deviation of -0.37 magnitude units at 2.0 s after the first $P$-wave trigger, while the largest standard deviation of the distribution is ±1.55 magnitude units obtained at the first time step, i.e., 0.5 s after the first $P$-wave arrival. As time goes on, the standard deviations clearly decrease.

Figure 4 shows the temporal distribution of triggered stations and source parameter estimates obtained by PreSEIS for the Hector Mine, Yorba Linda, and Parkfield earthquakes. The localization and magnitude curves are smoothed over 13 time steps, which is 6.5 s, using a moving average filter. Applying a smoothing average procedure to the outputs of neural networks is a legitimate and common procedure; the outputs can show outliers, e.g., caused by unfavorable weight initialization in the beginning of the training stage (Bishop 1995). Again, the localization errors are the absolute errors in hypocenter location, i.e., including source depths.

Although the $M_w$ 7.1 Hector Mine earthquake occurred close to station Hector, it takes 9.0 s until the $P$ wave arrives at the second-nearest station. Another seven stations trigger within the following 3.5 s. After 32.0 s, the $P$ wave has arrived at all 15 stations. The localization error shows a roughly constant level with proceeding time. The error has an initial value of 11.2 km with a maximum error of 23.6 km at 28.0 s. Averaged over the analyzed 50 s, a mean localization error of 14.1 km is obtained, which is 3.6 km larger than the mean error derived from training of the ANNs (Figure 3). The initial moment magnitude estimate of $M_w = 5.7 ± 1.6$ is 1.4 magnitude units smaller than the SCSN magnitude. With triggering of the second-nearest station, the estimate has improved to $M_w = 6.7 ± 0.5$. PreSEIS needs 13.5 s until the correct magnitude is estimated, which correlates with ground motion information from nine stations.

The epicenter of the $M_w$ 4.8 Yorba Linda earthquake is closely located to two of the seismic stations; the second station triggers only 1.8 s after the first one. The third and fourth
stations trigger within the following 7.7 s. It takes 39.5 s until ground motion information from all 15 stations is available. The initial localization error of the Yorba Linda earthquake is high, with a value of 104.1 km. After 4 s (two stations), it can be reduced to about 40 km. After 22 s, the localization error shows an average level of 22.3 km. The initial moment magnitude estimate of the Yorba Linda earthquake is \( M_w = 4.4 \pm 1.6 \), only 0.4 magnitude units smaller than the SCSN magnitude. After 5.0 s, the correct value of \( M_w = 4.9 \pm 0.6 \) is obtained, which corresponds to two triggered stations.

For the \( M_w 6.0 \) Parkfield earthquake, the first three stations trigger within 2.5 s. It takes another 8.5 s until the \( P \) wave reaches the fourth station. All 15 stations are triggered after 41 s. These longer times are because the epicenter is located at a significantly larger distance away from the network compared to the other two example earthquakes. As is also the case with the Yorba Linda earthquake, the initial localization error is high, showing a value of more than 120 km. Within 5 to 10 s, the error is reduced to about 60 km. After approximately 20 s, an average level of 19.5 km is obtained. The initial magni-
The magnitude estimate is $M_w = 5.3 \pm 1.6$, which is 0.7 magnitude units smaller than the SCSN magnitude. After 2.5 s (two stations), the estimate has improved to $M_w = 5.4 \pm 0.9$, and the correct moment magnitude of $M_w = 6.0 \pm 0.6$ is reached at 5.0 s (two stations). It is obvious, however, that in the subsequent seconds the magnitude is slightly overestimated and later slightly underestimated.

We repeated the analyses presented in this study by replacing all observed ground motion envelopes by predicted ones and also by adding a constant Gaussian noise signal to all predicted envelopes to give them a more realistic shape. Both variations did not significantly influence the results, proving the great flexibility of ANNs.

**DISCUSSION AND CONCLUSIONS**

This study represents the first application of the early warning approach PreSEIS to real data cases, using 74 earthquakes from southern California recorded at stations of the SCSN, combined with empirical relations derived by Cua (2005). PreSEIS is based on artificial neural networks and uses the $P$-wave arrival time differences at a network of seismic stations as well as the ground motion envelopes to estimate the most likely hypocentral location and magnitude of the earthquake. The estimates are continuously updated at time steps of 0.5 s. The objective of this study was to investigate whether PreSEIS is able to estimate the source parameters from real earthquake observations instead of from synthetics, as shown by Böse et al. (2008). The functionality is tested by analyzing the performances for three example events representing different geometrical settings, the 1999 $M_w$ 7.1 Hector Mine earthquake, the 2002 $M_w$ 4.8 Yorba Linda earthquake, and the 2004 $M_w$ 6.0 Parkfield earthquake.

For the Yorba Linda and Parkfield earthquakes, the study reveals unusually high initial localization errors of 104 km and more than 120 km, respectively. Once the $P$ waves arrive at two to three stations, these errors can be reduced by more than 50% and show continuous decrease as time goes on. Despite large localization errors, the moment magnitude estimates are of remarkable quality, considering the fact that the estimated hypocenter locations at each time step contribute as inputs for the magnitude estimations. Once ground motion information from two stations is available, the magnitude estimates are of 90.0 % (Parkfield earthquake), 94.4 % (Hector Mine earthquake), and 100% (Yorba Linda earthquake) accuracy, respectively. The time until the magnitudes are estimated correctly depends strongly on the station density around the epicenter and varies between 5.0 and 13.5 s.

The standard deviations of the source parameter estimates are derived from the training of the neural networks with all 74 events. They show highest values within the first few seconds and clearly decrease with progressing time, demonstrating the inverse relationship between the reliability of estimates and remaining warning time in an EEW system. This stresses the importance of updating the source parameter estimates with ongoing time.

**Figure 3.** Top: Mean temporal distribution of the number of triggered stations of all 74 earthquakes (solid line) and their minimum and maximum distributions (dashed lines). Middle: Absolute errors of hypocenter locations for all 74 training events. The errors are specified by the 25th, 50th, 75th, and 95th percentiles of the error distributions. Bottom: Mean moment magnitude estimates (circles) with standard deviations (error bars) derived from all 74 training events. The time axes in the three figures are relative to the time when each $P$ wave triggers the first respective station. All source parameter estimates are updated at 0.5-s intervals.
Rather than presenting a fully operational EEW algorithm, the major objective of this study was to investigate the functionality of PreSEIS in terms of 1) its capability to handle real data, 2) its operational suitability, and 3) its ability to define the problems remaining before a possible future implementation.

1. The application of PreSEIS to real data cases combined with synthetics shows a stable and robust performance. The quality of outputs depends neither on the number of missing observations that had to be replaced by predictions nor on added noise. Indeed, repeating the test using only predictions did not change the quality of results. The major remaining problem of the application of PreSEIS to real data will therefore be the existence of nonfunctional stations. We are aware that failing stations appear quite frequently in a seismic network. However, we are confident that suitable training epochs including various configurations of failing stations can reduce the impact of missing observations or even completely eliminate the effect.

2. Our study confirms that ANNs can indeed be suitable and attractive tools for earthquake early warning applications. The stability of results discussed above proves the tolerance of ANNs toward various types of input data, which is of great advantage in transferring the method to different regions. Our study also shows that the time necessary to obtain reliable source parameter estimates can be reduced by more appropriate geometrical settings and shorter interstation distances, which makes the method generally suitable for dense seismic early warning networks. A precondition for application of the method to a new region would be the availability of a suitable training dataset, which should be as extensive and various as possible to allow for a vast amount of a priori information.

3. In summary, we conclude that the major remaining problems regarding an implementation of PreSEIS are the

\[ \text{Figure 4. Number of triggered stations and source parameter estimates of the } M_w 7.1 \text{ Hector Mine (top), } M_w 4.8 \text{ Yorba Linda (middle), and } M_w 6.0 \text{ Parkfield earthquakes (bottom) with ongoing time after triggering of the first station. Left: Temporal distribution of the number of triggered stations. Middle: Absolute errors of hypocenter locations (solid curves) with mean errors (dotted curves) obtained from the training process in Figure 3 (50th percentile). Right: Moment magnitude estimates (solid curves) with standard deviations derived from the training process (gray-shaded areas) in Figure 3. All curves are smoothed over 6.5 seconds.} \]
nonfunctional stations, as well as the P-wave detection and, in some study areas, the availability of suitable training examples. Despite these obstacles, these questions are potentially resolvable and we will address them in future work.

ACKNOWLEDGMENTS

This work was funded by the EU project SAFER and the EDIM-project of the BMFB GEOTECHNOLOGIEN program of the German Federal Ministry of Education and Research. We would also like to thank the ETH Zurich and the NERIES project for funding the scientific exchange. The seismological data was recorded at the Southern California Seismic Network and downloaded from the Southern California Earthquake Data Center (http://www.data.scec.org/). The topography data used for the Generic Mapping Tools (GMT) map was taken from Satellite Geodesy, Scripps Institution of Oceanography, University of California San Diego (http://topex.ucsd.edu/).

REFERENCES


Geophysical Institute
Karlsruhe University
Hertzstrasse 16
76187 Karlsruhe
nina.koehler@gpi.uni-karlsruhe.de
(N. K.)