16th World Conference on Earthquake, 16WCEE 2017

Santiago Chile, January 9th to 13th 2017

Paper N° 2850 (Abstract ID)

Registration Code: S-G1460583766

MAKING EARTHQUAKE EARLY WARNING FASTER AND MORE ACCURATE USING ETAS SEISMICITY MODELS AS BAYESIAN PRIOR

L. Yin[(1)], M.-A. Meier[(2)], T. Heaton[(3)]

[(1)] *PhD Candidate, California Institute of Technology, lyin@caltech.edu*
[(2)] *Post Doctoral Scholar, California Institute of Technology, mmeier@gps.caltech.edu*
[(2)] *Professor, California Institute of Technology, heaton@ caltech.edu*

Abstract

Conventional Earthquake Early Warning (EEW) algorithms focus on time-series analysis of waveform data, which unfortunately requires delay between data collection and alert delivery. In order to provide reliable warning as quickly as possible before the arrival of damaging ground shaking, we propose a new algorithm that uses Bayesian probabilistic inference to provide optimally fast estimates of earthquake location; in many cases, the earthquake location is available as soon as the first P-wave arrival at the station located closest to the epicenter. The algorithm uses Epidemic-Type Aftershock Sequence (ETAS) seismicity forecast model as Bayesian prior to provide a intuitive initial approximation; the analysis of seismic waveform information is incorporated into the solution as data becomes available over time. We have evaluated the algorithm for all 504 M4+ earthquakes in Southern California between 1990 and 2005. For the earliest epicentral location estimation at 0.5 seconds after P-wave detection, the median location error using seismicity forecast with waveform analysis improved by 58% relative to results using waveform analysis only. We also present location estimations of a M5.2 Lone Pine Earthquake and a M5.4 Chino Hills Earthquake in detail which highlight the importance of Bayesian seismic prior and waveform likelihood interaction. Our new strategy has shown promising results and implementation of this methodology should significantly enhance the performance of EEW systems.

Keywords: Earthquake Early Warning; ETAS seismicity models; Bayesian Probabilistic inference; rapid parameter estimation

1. Introduction

Due to the rapid advancement of digital seismic networks, Earthquake Early Warning (EEW) currently uses real-time waveform information to rapidly estimate source parameters such as the magnitude and location of an earthquake [1, 2]. The source parameters are then used to predict ground motion characteristics at various sites. This waveform analysis approach for source parameter characterization typically requires colleting a few seconds of data after the first detection of P-waves. Since the arrival of S-waves follows immediately after the arrival of P-waves in the epicentral region, processing delays must be minimized if we are to have any hope of providing warnings of potentially damaging S-waves near an earthquake's epicenter.

Cua(2005) and Cua and Heaton (2007) described a framework for deriving the probabilities of predicted ground shaking based on information that is available at any given instant for an EEW system [3, 4]. They suggested that Bayesian probabilistic approach could be used to simulate the type of common-sense analysis that is performed by human experts if they had the time to intervene in EEW. Bayesian probabilistic inference provides a natural framework to quantify uncertainties in combining heterogenous sources of information. In additional to waveform analysis in the conventional EEW algorithms, they suggested of using the fact that many earthquakes occur close to the locations of past earthquakes that occurred in prior minutes to days. If a seismic station detects shaking, then it is natural to inquire whether that station is close to previous activity. In order to use this concept in a working EEW system, we must develop this simple idea into a practical and reliable algorithm that estimates the probabilities of potential earthquake locations, given the seismic data available from the network and the eathquake catalog records, for the time immediately preceding the detected shaking [5].

We begin with Bayes' theorem which states that the probability magnitude *M* and station-epicenter-distances *R*, given seismic data *S(t)* available at time t is formulated:

$$P(M,R|S(t)) \propto P(S(t)|M,R)P(M,R) \tag{1}$$

$P(M,R|S(t))$ is usually refered as the Bayesian posterior function, $P(S(t)|M,R)$ is called the likelihood function and $P(M,R)$ is called the Bayesian prior. Most existing EEW algorithms focus on the likelihood function; that is, what magnitude and station-epicenter-distances produce seismic data similar to the data just recorded? For the purposes of this study, we assume that the likelihood function is provided by the Gutenberg Algorithm [6]. The Gutenberg Algorithm (GbA) uses a filterbank to decompose a real-time waveform into different frequency bands, and then efficiently searches an extensive waveform database to identify past records with similar time-frequency characteristics and uses them to compute probabilistic source parameter estimates for the real-time waveform. The GbA is designed to be optimally fast, first estimates are available using only 0.5 s of data from the closest station. However, earliest estimations could be poorly resolved due to trade-offs between earthquake magnitude and epicenter distance. This issue can be reduced by introducing prior information to assign relative probability based on establised empirical relationships. There are several choices for Bayesian prior, perhaps the most general choice is to assume that seismic activity obeys the Gutenberg-Ritcher frequency-magnitude law (GR law), in which case the prior probability can be approximated by $prob(M,R) = 10^{a-bM}R$, where a, b are GR law parameters and R is the scalar distance amplitude. This proposed prior simply states that the earthquake frequency decreases exponentially with magnitude and the area of a ring at distance R grows linearly with R.

In this study, we focus on earthquake location parameter estimation. We propose a Bayesian prior that is based on the principle that earthquake sequences tend to cluster in time and space[7]. In particular, we use Epidemic-Type Aftershock Sequence (ETAS) Models to derive a location distribution with far more spatial-temporal information than the simple prior mentioned previously[8]. We show how an ETAS prior can provide more accurate source-parameter estimates at short warning times with minimal seismic waveform data. In fact, the use of an ETAS seismicity forecast model as a Bayesian prior often provides accurate estimates of the epicenter location that are available simultaneously with the detection of an event. Even in the cases with a highly uncertain ETAS prior, the accurate location estimation is converged immediately with sufficient data processed by GbA.

We collected all 504 M4+ earthquakes in southern California between 1990 an 2015. We evaluated the location estimation performance of Bayesian analysis techniques making use of the GbA as the likelihood function combined with an ETAS model for the Bayesian prior. We also present the earliest estimations for examples of a M5.2 Lone Pine Earthquake and a M5.4 Chino Hills Earthquake in detail. The two examples demonstrate the importance of Bayesian prior and likelihood interaction to ensure the optimum results in different seismic environment.

2. Method

2.1 Bayesian Inference in EEW Location Estimation

To focus on location estimation only, Bayes' Theorem from Eq (1) can be simplified to:

$$P(Lat, Lon|S(t)) \propto P(S(t)|Lat, Lon)P(Lat, Lon) \tag{2}$$

where *Lat*, *Lon* are the epicenter location of the earthquake. In this formulation, $P(Lat, Lon|S(t))$ is the Bayesian posterior function, $P(S(t)|Lat, Lon)$ is the likelihood function and $P(Lat, Lon)$ is the Bayesian prior. The following describse how each of the terms are derived.

2.2 Prior Information – ETAS seismicity model

The Bayesian prior information is a spatial distribution of earthquake probability produced by ETAS models, where each of the observed earthquakes stochastically generates potential earthquakes in the future. This statistical method quantitatively describes the clustering property in earthquake sequence processes and the generated earthquakes have probability of generating secondary earthquakes, which differentiates this approach from other aftershock simulation methods [9].

The earthquake probability forecast map is created based on the premise that the location of the future earthquake is significantly influenced by the accumulation of time, location, and magnitude of previous observed earthquakes. We model earthquake sequences following Omori's Law in time and GR law in magnitude [10, 11]. The rate of earthquakes above magnitude M_{min} triggered by the ith observed earthquake, $\lambda_i(t)$, at time t is expressed:

$$\lambda_i(t) = \frac{K_0 10^{\alpha(M_i - M_{min})}}{(t - t_i + c)^p} \tag{3}$$

where K_0, α, c, p are ETAS model parameters, while M_i and t_i are the magnitude and occurrence time of the ith observed earthquake [12]. The rate of earthquake is then distributed spatially following Felzer and Brodsky distance relation to create the spacio-temporal rate of earthquakes, $\lambda_i(t, Lat, Lon)$, at a given time t and location Lat, Lon [13].

$$\lambda_i(t, Lat, Lon) = \frac{\lambda_i(t)}{r(Lat, Lon)^n} \tag{4}$$

where $r(Lat, Lon)$ is the epicenter distance between the point of interest to the observed earthquake, and n is Felzer and Brodsky coefficient. The total rate earthquake at a given time t and location Lat, Lon is the sum of the rate earthquake from all of the observed earthquakes combined with the background seismicity $\mu(Lat, Lon)$.

$$\lambda(t, Lat, Lon) = \mu(Lat, Lon) + \sum_i \lambda_i(t, Lat, Lon) \tag{5}$$

We model the earthquake occurrence process as a nonhomogeneous Poisson process in time. The probability, P, of one or more earthquakes occurring above M_{min} at location Lat, Lon and in the time range Δt is:

$$P(Lat, Lon) = 1 - \exp[-\int_t^{t+\Delta t} \lambda(t, Lat, Lon)dt] \tag{6}$$

Eq (6) is the input for the Bayesian prior information suggested in Eq (2). The probabilistic approach using ETAS method is a quantitative interpretation of the seismic forecast.

2.3 Likelihood Function– The Gutenberg Algorithm

The GbA is a probabilistic approach to estimate EEW source parameters using real-time waveform information. During an ongoing earthquake, GbA performs real-time time-frequency analysis on the collected waveform using minimum-phase-frequency filter banks, then efficiently search within a catalog of events for similar characteristics. At every time increment after the P-wave arrival, each triggered station computes the relative probability of magnitude and epicenter distance estimations. The focus is to explore maximum available information during the EEW process. Details of the algorithm can be referred to Meier (2015) [6].

In single station location inference, we convert the station-to-source distance probability density function from GbA onto a 2-dimensional spatial distribution that is most likely to produce the recorded waveform $S(t)$, $P^j(S(t)|Lat, Lon)$, according to the distance between the station j to the location Lat, Lon. Combining location estimation from multiple K stations is straightforward as the probabilistic formulation of the algorithm, simply requires that we multiply the single-station spatial probability functions:

$$P(S(t)|Lat, Lon) = \prod_{j=1}^{K} P^j(S(t)|Lat, Lon) \tag{7}$$

Eq (7) describes the likelihood function as suggested in Eq (2).

3. Data

We collected all 506 M4.0+ earthquakes in Southern California from 1990 to 2015; from which the catalog locations of the events are then compared to the estimated location parameters at every half-second interval after the event detection. The catalog location of all the events are shown in Fig. 1. The data set includes 03 October, 2009 M5.2 Lone Pine Earthquake, located at (-117.86, 36.39); and 29 August, 2008 M5.4 Chino Hills Earthquake, located at (-117.76, 33.95). Details of the two events are provided in Table 1. The two events represents two characteristic setting: the Lone Pine earthquake demonstrates the accuracy and speed estimations due to Bayesian prior information during a seismic sequence, which the Chino Hill earthquake demonstrates the importance of Bayesian likelihood function to reduce prior uncertainty during a seismic dominant period.

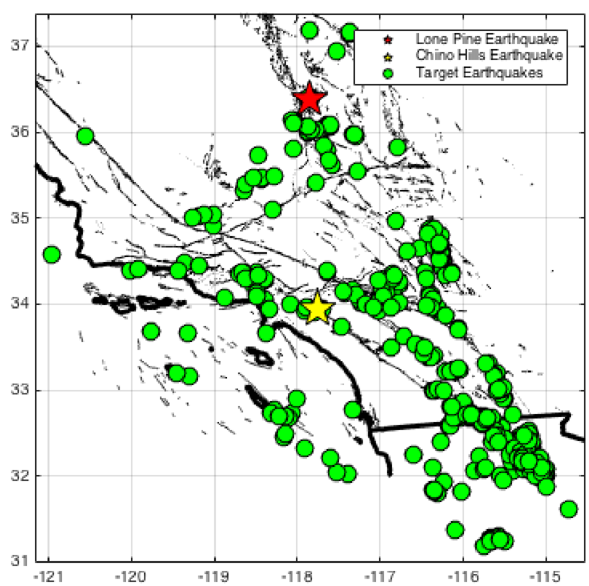


Fig.1 - Catalog location of the 506 target M4.0+ earthquakes in Southern California from 1990-2015. Including 2009 Lone Pine M5.2 Earthquake in red star and 2008 Chino Hills M5.4 Earthquake in yellow star.

Table 1 - Detail Information on Lone Pine Earthquake and Chino Hills Earthquake

Name	Time	Magnitude	Latitude	Longitude
Lone Pine Earthquake	2009/10/03 01:16:00	5.2	-117.86	36.39
Chino Hills Earthquake	2008/08/29 18:42:15	5.4	-117.76	33.95

The catalog data used in Bayesian prior information was downloaded from the Southern California Earthquake Data Center (http://data.scec.org/). It includes source parameter information, such as origin time, hypocenter location, and magnitude, of 567258 historic seismic events in Southern California from 1981 to 2015.

The waveform data set is collected from the global database of waveforms compiled in Meier (2015)[6]. All the waveforms have been preprocessed to eliminate poor quality records, including missing channel records, low signal-to-noise ratios, low sampling rates, clipped records. The subset of events used in this study includes 50750 three-component waveform records from a total of 3523 events.

4. Results

4.1 M5.2 Lone Pine Earthquake

At October 03, 2009 01:16:00, the M5.2 Lone Pine earthquake first triggered the vertical channel of station CI.CGO 3.5 seconds after the origin time. The station is about 20km northeast of the catalog event location. The location estimation in GbA resulted in a high initial uncertainty which decreased with time; the location error reaches 18km at 14 seconds after the origin time, as shown in Fig 3a)c)e).

At the moment of the first P-wave arrival at CI.CGO, the ETAS seismicity forecast map developed an earthquake probability with peak distribution at 15km southwest of the triggered station, shown in Fig 2. This was a consequence of the foreshock series recently accumulated in the area. By including the ETAS seismicity map as a Bayesian prior, the maximum posterior location estimation immediately converged to location error of 2km, as shown in Fig 3b)d)f).

At every time increment during the ongoing event, the location error estimated from GbA fluctuates around 10 to 20km; with the ETAS seismicity map included as the Bayesian prior, the location error reduced to less than 3km, as shown in Fig 4.

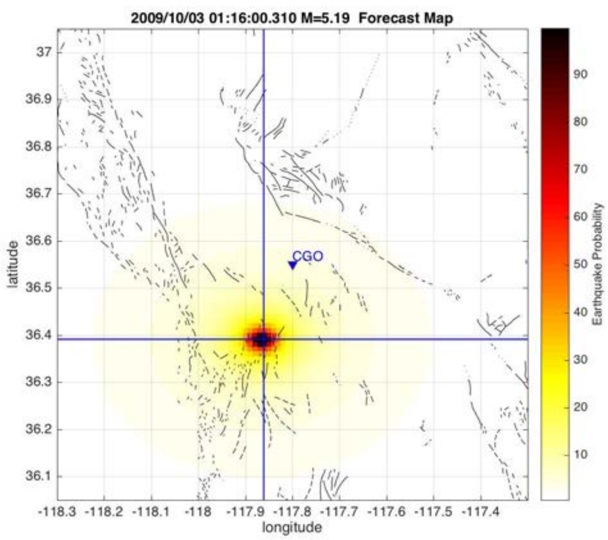


Fig 3 - Seismicity Forecast Map for Lone Pine M 5.2 Earthquake. It was produced immediately after the first station trigger at CI.CGO. The intersection of the two blue lines is the catalog location.

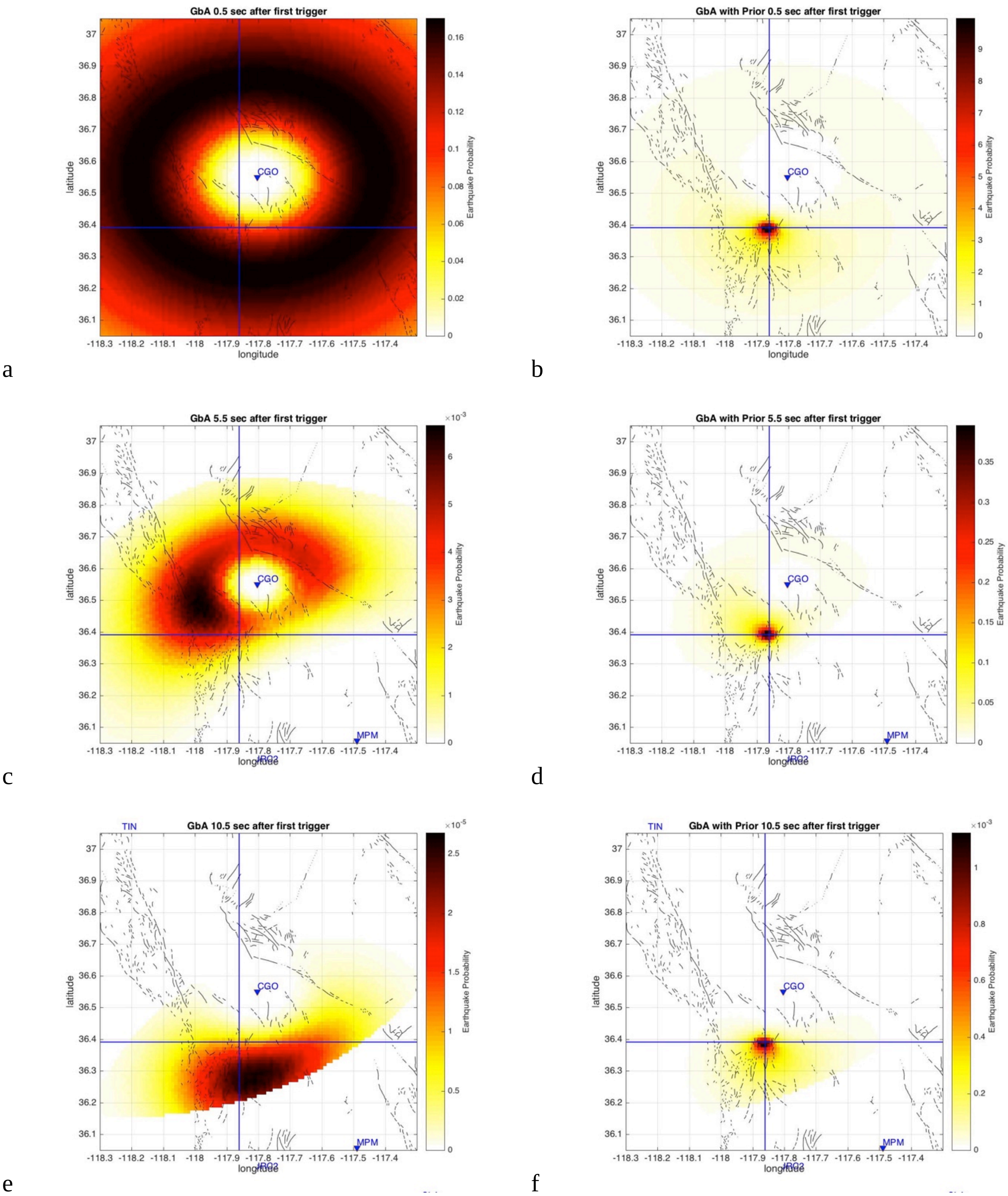


Fig 4 - Probabilistic location estimation map of the M5.2 Lone Pine Earthquake at various times after the first station trigger. a) c) e) are results of Gutenberg Algorithm at 0.5 sec, 5.5 sec, and 10.5 sec after the first trigger, respectively. b) d) f) are posterior results of Gutenberg Algorithm with Prior at 0.5 sec, 5.5 sec, and 10.5 sec after the first trigger, respectively. The intersection of the two blue lines is the catalog location.

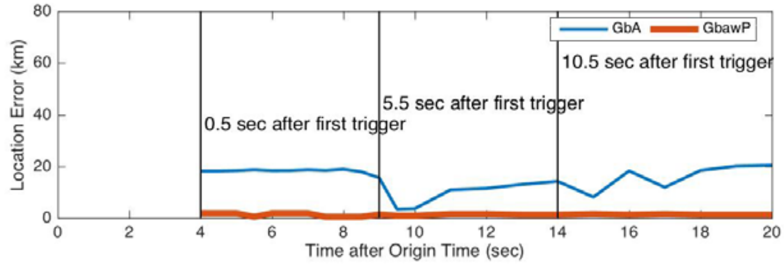


Fig 5 - M5.2 Lone Pine Earthquake location error as a function of time after the origin time. The blue and red lines are the location error results of the Gutenberg Algorithm, and the Gutenberg Algorithm with ETAS Prior, respectively.

4.2 M5.4 Chino Hill Earthquake

At August 29, 2008 18:42:15, the M5.4 Chino Hills earthquake first triggered the vertical channel of station CI.CHN 3 seconds after the origin time. The station is 5 km northeast of the catalog event location. Due to high station density in the area, sufficient waveform data was quickly collected, and GbA initial location estimation shows high accuracywith only 12km error, as shown in in Fig 7a)c)e).

Because there was no observed seismic cluster in the region, background seismicity greatly influenced the ETAS forecast, shown in Fig 6. Although the seismicity prior indicates relatively high earthquake probabilities around 40km east of the station, the hypothesis immediately updated by the GbA results with the incoming waveforms, shown in Fig 7b)d)f). At 1.0 sec after the first trigger, with 5 triggered stations data, the spatial distribution shows a clear shift from the ETAS to the GbA results. And a half second after that, with 9 stations triggered, the posterior probability is dominated by the GbA results.

Fig 8 shows that although the initial location error with the Bayesian prior is 42km, it quickly reduces to 8km within the following 1.5seconds, and the error remains low there after. At every time step, the posterior results update with the available waveform data.

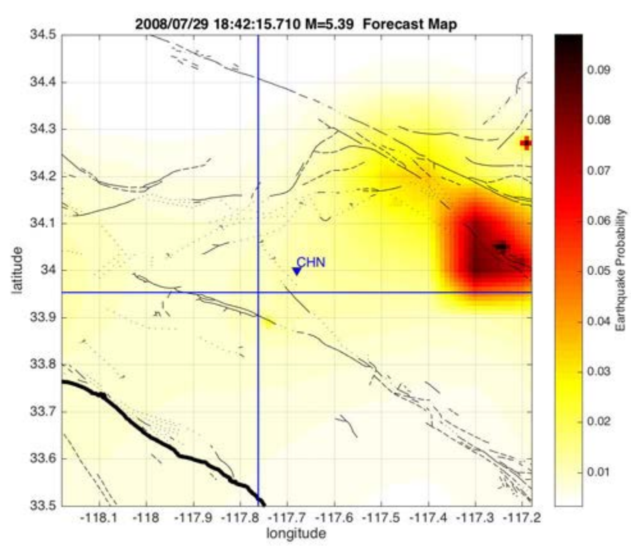


Fig 6 - Seismicity Forecast Map for Chino Hills M 5.4 Earthquake. It was produced immediately after the first station trigger at CI.CHN. The intersection of the two blue lines is the catalog location.

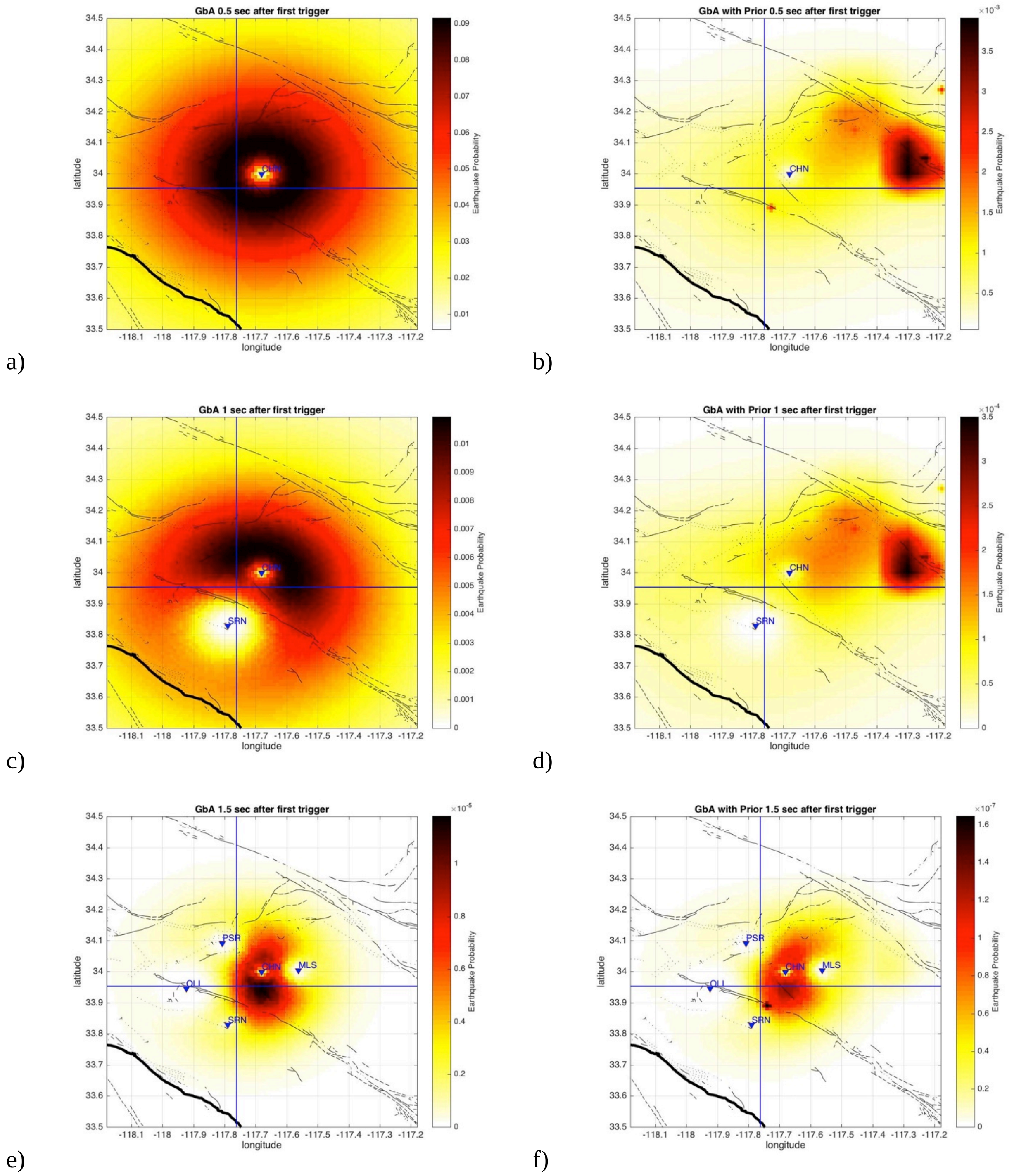


Fig 7 - Probabilistic location estimation map of the M5.4 Chino Hills Earthquake at various times after the first station trigger. a) c) e) are likelihood probabilities, results of GA at 0.5 sec, 1.0 sec, and 1.5 sec after the first trigger, respectively. b) d) f) are posterior probabilities, results of GA with Prior at 0.5 sec, 1.0 sec, and 1.5 sec after the first trigger, respectively. The intersection of the two blue lines is the catalog location.

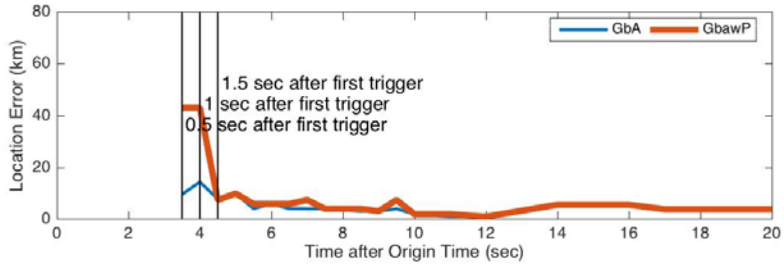
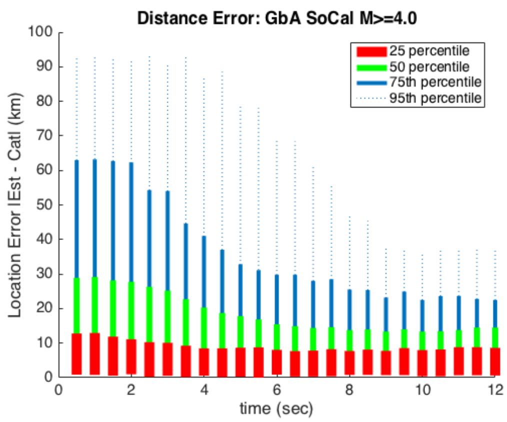
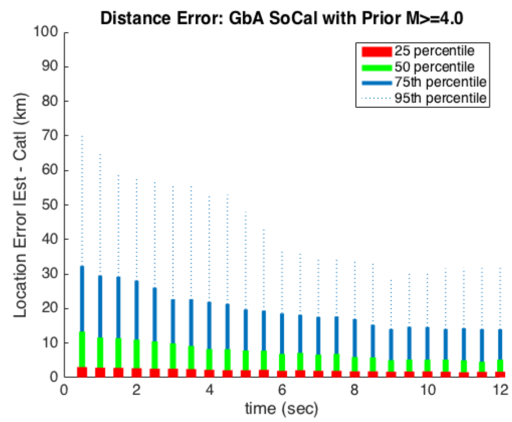


Fig 8 - M5.4 Chino Hills Earthquake location error as a function of time after the origin time. The blue and red lines are the location error results of the GbA, and GbA with ETAS Prior, respectively.

4.3 Overall Performance



a)



b)

Fig 9 - Location Error as a function of time after first trigger for 506 M4+ earthquakes in Southern California 1990-2015 a) likelihood performance: GA results b) posterior performance: GA with Prior results. The errors are specified at the 25th, 50th, 75th, and 95th percentile.

We evaluated the location error, the distance between the catalog location and the location of maximum probability, as a function of time after the first trigger for all 506 M4+ earthquakes in Southern California from 1990-2015, as shown in Fig 9. In Fig 9a), with only the waveform information is considered, the GbA median location error is 28km, 20km and 17km at 0.5sec, 5sec, and 10secs after the first P-wave detection. As expected, the error is initially large and reduced with time when more waveform data is collected. In Fig 9b), location error is calculated using Bayesian Inference by combining waveform and catalog information; the median location error is 12km, 8km, and 5km at 0.5 sec, 5sec, and 10secs after the first trigger. The median error reduction has improved substantially, especially in the first few seconds reaching a 58% improvement. The distance error at all percentile levels consistently decreased at every time increments.

5. Discussion

Many of the previous studies have shown that earthquakes tend to occur in places where there has been observed seismic activities; especially so with incidence of foreshocks often observed preceeding large earthquakes [14]. While most of the current EEW algorithms focusing on time-series analysis of waveforms, they fail to acknowledge that premeditary events are clearly related to subsequent larger earthquakes. ETAS Bayesian priors exploit the seismicity catalog information to provide optimally fast location approximation; logically using the spatial-temporal clustering property of earthquakes to ensures a higher the level of accuracy.

More significant positive impacts of ETAS Bayesian prior are reflected during aftershock and swarm earthquake sequences. During aftershock sequences, the repetitive ground shaking continuously deteriorates already weakened infrastructure components. Seismic damage can be even more significant if the aftershocks occur close to a populated urban area. In these cases, location estimations that uses the ETAS Bayesian prior guarantees the delivery of fast and accurate alerts immediately after the first P-wave detection, allowing EEW to offer more alerting time to rescue teams and residences for evacuation during aftershocks.

For the events with no obvious prior seismic activity in the proximity, ETAS produces a smooth spatial distribution of earthquake probability. In such case, the estimations are quickly dominated by the likelihood function of the waveform analysis, as demonstrated in the Chino Hills earthquake.

The Bayesian probabilistic approach in this study mimics human behavior in the decision making process during an ongoing earthquake. It first uses scientific intuition of the seismic knowledge to make a quick and rational approximation, and then analyzes real-time waveforms with the assistance of powerful computational tools. The Bayesian framework conveniently combines results from any independent probabilistic algorithm, such as the GbA with ETAS prior. For future development, additional algorithms can be incorporated in this ensemble framework to further enhance the posterior results.

Station topology constraints can be imposed as an additional Bayesian prior in an ideal network with no malfunctioning stations. The concept of the voronoi diagram of a network distribution implies that an earthquake must occur within the voronoi cell of the first triggered station, as the P-wave travel time from any of the points in this voronoi cell to the station is minimized [Rosenberger, 2009] [Cua, 2005]. However, in our offline study, due missing records and inconsistency station performance, station topology concept lead to poor results.

6. Conclusion

We proposed a probabilistic approach to obtain faster and more accurate EEW location estimations. We investigated EEW performance combining the GbA and ETAS seismicity models under Bayesian inference. Our results show that Bayesian inference with seismicity priors can reduce overall median location error by 58% for the first few seconds after P-wave arrival at the closest station to the epicenter. In most of the cases evaluated, accurate location estimation is available immediately after the first P-wave detection.

In the current technology, scientists are investigating sophisticated methods to exploit waveform information to estimate source parameters, while neglecting the most fundamental clustering phenomenon of seismic sequences. In this study, we demonstrated that both the Bayesian seismic prior and waveform likelihood are essential in EEW; only by analyzing both heterogeneous information, the location estimation in EEW could potentially achieve fast results with high confidence.

7. References

[1] Heaton TH (1985): A model for a seismic computerized alert network, *Science* 228, no. 4702, 987–990.

[2] Allen RM, Kanamori H (2003): The potential for earthquake early warning in southern California, *Science* 300, no. 5620, 786–789.

[3] Cua GB (2005): Creating the virtual seismologist: Developments in ground motion characterization and seismic early warning, Ph.D. Thesis, California Institute of Technology.

[4] Cua GB, Heaton TH (2007): The Virtual Seismologist (VS) method: A Bayesian approach to earthquake early warning, *Earthquake early warning systems*, Gasparini P, Manfredi G, Zschau J (editors), Springer-Verlag, Berlin, Germany 97-132

[5] Gerstenberger MC, Wiemer S, Jones LM, Reasenberg P (2005): Real-time forecast of tomorrow's earthquakes in California, *Nature* 435, 328 - 331

[6] Meier MA, Heaton TH, Clinton J (2015): The Gutenberg Algorithm: Evolutionary Bayesian Magnitude Estimates for Earthquake Early Warning with a Filter Bank, *Bulletin Seismological Society of America*. 105, no. 5.

[7] Abercrombie RE, Mori J (1996): Occurrence patterns of foreshocks to large earthquakes in the western United States, *Nature*, 381, 303-307.

[8] Ogata Y (1998): Space-time point-process models for earthquake occurrences, *Annals of Statistics* 50, 805–810.

[9] Felzer KR, Abercrombie RE, Ekström G (2003): Secondary aftershocks and their importance for aftershock prediction, *Bulletin of the Seismological Society of America* 93, 1,433–1,448.

[10] Utsu T (1961): A statistical study on the occurrence of aftershock, *Geophysical Magazine* 30, 521–605.

[11] Gutenberg B, Richter CF (1944): Frequency of earthquakes in California, *Bulletin of the Seismological Society of America* 4, 185–188.

[12] Felzer KR (2008): Simulated aftershock sequences for a M 7.8 earthquake on the Southern San Andreas Fault, *Seismology Research Letter* 80, 21-25.

[13] Felzer KR, Brodsky EE (2006): Decay of aftershock density with distance indicates triggering by dynamic stress, *Nature* 441, 735– 738.

[14] Reasenberg PA (1999): Foreshock occurrence before large earthquakes, *Journal of Geophysical Research* 104, 4755-4768.