Bayesian Approach for Identification of Multiple Events in an Early Warning System

Annie Liu* Masumi Yamada†

Abstract

The 2011 Tohoku earthquake (Mw9.0) was followed by a large number of aftershocks that resulted in 70 early warning messages in the first month after the main shock. Of these warnings, a non-negligible fraction (63%) were false warnings where the largest expected seismic intensities were overestimated by at least two intensities or larger. These errors can be largely attributed to multiple concurrent aftershocks from distant origins that occur within a short period of time. Based on a Bayesian formulation that considers the possibility of having more than one event present at any given time, we propose a novel likelihood function suitable for classifying multiple concurrent earthquakes, which uses amplitude information. We use a sequential Monte Carlo heuristic whose complexity grows linearly with the number of events. We further provide a particle filter implementation and empirically verify its performance in with the aftershock records after the Tohoku earthquake. The initial case

*aliu@cms.caltech.edu, Computer Science, California Institute of Technology, Pasadena, CA, USA
†masumi@eqh.dpri.kyoto-u.ac.jp, Disaster Prevention Research Institute, Kyoto University, Uji, Japan

studies suggest promising performance of this method in classifying multiple seismic events that occur closely in time.

Introduction

During the highly seismically active period after a major earthquake, multiple earthquakes can occur almost concurrently at different locations. In this case, the seismic waves measured by the ground sensors contain mixed signals from more than one source. If the detection algorithm assumes only one quake, the estimated quake parameters (e.g. location and magnitude) will not be accurate. These inaccurate estimates can lead to false warnings that are often observed after large earthquakes.

The 2011 off the Pacific Coast of Tohoku Earthquake (hereafter called Tohoku earthquake) caused significant damage over a large area of northeastern Honshu. An earthquake early warning (EEW) was issued to the public in the Tohoku region about 8 s after the first P-arrival, which is 31 s after the origin time (Hoshiba et al., 2011; Sagiya et al., 2011; Hoshiba and Iwakiri, 2011). There was no blind zone, i.e., warnings were received at all locations before the S-arrivals, since the earthquake was fairly far offshore.

The main earthquake was followed by a large number of aftershocks that resulted in 70 early warnings issued in the first month after the main shock (JMA, 2011). Among these, 63% of the warnings contained significant errors where the estimated seismic intensities were at least two scales larger than the observed ones. As a comparison, only 29% of the warnings contain such errors prior to the Tohoku earthquake. Post-event analysis revealed that 73% of these errors could be attributed to failure to classify multiple concurrent quakes either from the same hypocenter separated by a short amount of time or from spatially distant origins (JMA, 2011). One of the main reasons of these false alarms is that the current approach uses mainly P-wave arrival time to estimate the hypocenter.

In this paper, we propose a novel approach to detect and classify multiple concurrent earthquakes in the current Japan Meteorological Agency (JMA) system framework. We introduce an approximate Bayesian method that estimates the location, magnitude, and origin time of multiple concurrent aftershocks. In contrast to the current JMA system, this approach produces multiple sets of estimation for earthquakes that occur closely in time. The experimental results from several case studies suggest that this approach can successfully detect and estimate the parameters of multiple concurrent earthquakes.

Data and Processing

This paper includes strong motion data observed by the JMA seismic stations during and after the Tohoku earthquake. We evaluate the new classification approach on three sections of these records as summarized below. For each record, the values included in the JMA EEW are compared to the values that appear in the JMA catalog in Table 1.

Data set

Case 1: 15 March 2011, 1:36:00 - 1:38:00 (two small earthquakes). Early warnings were issued to the public based on an estimated JMA magnitude of 5.9 at 21 s after the first P-wave detection (see Data and Resources). However, the largest observed seismic intensity was only 2 in the JMA seismic intensity scale. As shown in Table 1, at least two events about 200 km apart of magnitude 2.5 and 3.3 occurred within 15 s. Since the second event started close in time to the wave arrivals of the first event, the EEW system treated these separate events as one single earthquake and as a result, overestimated the magnitude.

Case 2: 20 March 2011, 14:19:00 - 14:21:00 (two small earthquakes). The JMA EEW system estimated a magnitude of 7.6 at 6.6 s after the initial detection of P-wave and issued a warning to the public (see Data and Resources). However, the largest observed seismic intensity was only 3. Again, as shown in Table 1, the overestimation can possibly be attributed to mistaking two smaller quakes about 150 km apart (M3.0 and M4.7) that occurred within 5 s for one large quake, since the occurrence of the second event was close in time to the wave arrival of the first event.

Case 3: 11 March 2011, 14:46:00 - 14:49:00 (Tohoku earthquake). To demonstrate that the method can also handle the classification of a single event, we also include the analysis of the Tohoku earthquake (Mw9.0). An early warning was issued to the public in the Tohoku region about 8 s after the first P-arrival, which is 31 s after the origin time (see Data and Resources).

Processing

This paper uses the three component acceleration data with a sampling rate of 100 Hz from about 200 stations. The acceleration data was first converted to SAC format and decimated by a factor of 100, reducing the sampling frequency to 1 Hz. The decimation was not necessary but was used to reduce computation time. Each component of the decimated acceleration $k(t)$ was then converted to displacement $A(t)$. The conversion was done by twice integration of $k(t)$ using a recursive digital filter with the frequency response of a mechanical seismometer (Katsumata, 2008).

$$A(t) = gn \times [k(t) + h_0 \cdot k(t-1) + h_1 \cdot k(t-2)] - h_2 \cdot A(t-1) - h_3 \cdot A(t-2), \qquad (1)$$

where the function gain factor gn and filter constants h_0, h_1, h_2, h_3 depend on the sampling frequency, damping constant, and natural period of the seismometer. For a JMA seismometer with 100Hz sampling, 0.55 damping constant, and 6 s natural period, the values correspond to:

$$gn = 0.0000248691025,\ h_0 = 1.0,\ h_1 = 1.0,\ h_2 = -1.9889474,\ h_3 = 0.9895828. \quad (2)$$

The following approach to classification uses both the vector sum of the three component displacement $A(t)$ as well as the vertical component of acceleration $k(t)$. The picking is done with STA/LTA of $k(t)$ with a short term window of 1 s and long term window of 10 s. The method also computes expected P- and S-wave arrival times (t_p and t_s) to determine whether a station should have observed P-wave or S-wave or neither. These arrival times are computed with the JMA 1D layered velocity structure (Ueno et al., 2002).

Bayesian Method

The problem of continuous parameter estimation for multiple events can be formulated as a Bayesian inference problem. Let θ be the vector of parameters that characterizes an event and Θ be a set of events that are parametrized by θ's, $\Theta = \{\o, \{\theta_1\}, \ldots, \{\theta_1, \theta_2, \ldots\}\}$. Suppose $\mathbf{z}_{1:t}$ is the complete history of observations from all the stations till the current time t, the posterior $P(\Theta_t|\mathbf{z}_{1:t})$ reveals the distribution of information of current ongoing events at time t given the evidence and prior information.

$$P(\Theta_t|\mathbf{z}_{1:t}) = \frac{P(\mathbf{z}_t|\Theta_t)\ P(\Theta_t|\mathbf{z}_{1:t-1})}{P(\mathbf{z}_t|\mathbf{z}_{1:t-1})}, \quad (3)$$

where $P(\mathbf{z}_t|\Theta_t)$ is the likelihood function and is typically denoted as L, $L(\mathbf{z}_t|\Theta_t) = P(\mathbf{z}_t|\Theta_t)$. $P(\Theta_t|\mathbf{z}_{1:t-1})$ is the updated prior at time t,

$$P(\Theta_t|\mathbf{z}_{1:t-1}) = \int P(\Theta_t|\Theta_{t-1})P(\Theta_{t-1}|\mathbf{z}_{1:t-1})\,\mathrm{d}\Theta_{t-1}, \tag{4}$$

and $P(\Theta_0|\mathbf{z}_0) \equiv P(\Theta_0)$ is the prior distribution of Θ.

Particle Filter

In general, Equation (3) does not have a closed-form solution, and there exists several sub-optimal solutions to approximate the posterior distribution (Arulampalam et al., 2002), one of which is grid search. Grid search, though simple to implement, suffers a few problems. First of all, when the parameters are continuous and not sufficiently restricted, the method cannot cover the complete parameter space since there can only be a finite number of grids. Secondly, the grid size is predefined, and as a result, it requires a large number of grids to achieve good coverage at a desired resolution.

Another solution is the Particle Filter (PF), which is a sequential Monte Carlo method that approximates the posterior distribution with a set of weighted particles (Doucet et al., 2001). As the number of particles goes to infinity, the solution from PF approaches the optimal solution. There is a rich literature on PF and its variation (Doucet et al., 2001; Arulampalam et al., 2002; Liu and Chen, 1998). The basic procedure is summarized below for reference.

Sampling. At the beginning of each iteration, the value of each particle is drawn from an *important density* function $q(\Theta_t^i|\Theta_{t-1}^i, \mathbf{z}_t)$. For $i = 1, \ldots, N$

$$\Theta_t^i \sim q(\Theta_t^i|\Theta_{t-1}^i, \mathbf{z}_t). \tag{5}$$

where $\sim$ denotes that the sample Θ_t^i is drawn according to the distribution $q(\cdot)$.

Weight update. PF approximates the posterior with a collection of weighted particles.

$$P(\Theta_t|\mathbf{z}_{1:t}) \approx \sum_{i=1}^{N} w_t^i \cdot \delta(\Theta_t - \Theta_t^i), \tag{6}$$

where w_t^i is the weight for particle i at time t. The sum of total weights are normalized to 1.

$$\sum_{i=1}^{N} w_t^i = 1. \tag{7}$$

The weights for all particles are updated as new evidence $\mathbf{z}_t$ comes in and renormalized at the end of each update.

$$w_t^i \propto w_{t-1}^i \frac{L(\mathbf{z}_t|\Theta_t^i)P(\Theta_t^i|\Theta_{t-1}^i)}{q(\Theta_t^i|\Theta_{t-1}^i, \mathbf{z}_t)}, \tag{8}$$

where $q(\cdot)$ is the same important density that appears in the sampling step. To simplify the calculation, $q(\cdot)$ is often chosen to be the transition prior $P(\Theta_t^i|\Theta_{t-1}^i)$. Since the terms cancel out in the right hand side, the new weight is directly proportional to the likelihood $L(\mathbf{z}_t|\Theta_t^i)$.

Resampling. Because the posterior is approximated with discrete particles, the system suffers *sample degeneracy* after a few update iterations when the weight is concentrated on a very small number of particles. The decrease in weight variance determines the degree of degeneracy that can be approximated with $\widehat{N_{eff}}$ (Arulampalam et al., 2002),

$$\widehat{N_{eff}} = \frac{1}{\sum_{i=1}^{N}(w_t^i)^2}. \tag{9}$$

Small $\widehat{N_{eff}}$ indicates severe degeneracy in which case resampling is required. Resampling essentially eliminates particles with negligible weight by generating a new set of N equally weighted particles according to current distribution $P(\Theta_t|\mathbf{z}_{1:t})$. There exists many methods for sampling from a discrete distribution, which we will not discuss here.

Each iteration typically involves one *sampling* and one *weight update*. *Resampling* only happens when $\widehat{N_{eff}}$ drops below a certain threshold.

Model

In the rest of the section, we discuss the practical implementation details of a PF-based real-time parameter estimation system for multiple earthquakes. The parameters we would like to estimate are $\theta = [x, y, D, M, t_0]$, where $x =$ longitude (deg), $y =$ latitude (deg), $D =$ depth (km), M = JMA magnitude, and t_0 = origin time. Complete pseudo code (Algorithm 1) is included in the end of this section.

Prior distribution. The prior $P(\theta)$ determines how the particles are initialized. A good prior encodes geographical information such as the location of nearby fault lines to the station that first triggered, and the most common magnitudes generated at the fault lines. This information can be compiled from historical earthquake catalog for each station and used in real time when initializing the PF. If prior information is absent, then a flat prior can be used instead. The choice of prior distribution affect the quality of the estimates and the convergence rate. Prior distribution of large coverage may cause the initial estimates to be unstable because little evidence is present. Priors of small coverage may result in slow convergence or false convergence (converging at the wrong values). These tradeoffs can be evaluated empirically. In this paper, we use a uniform flat prior of ± 100 km for location, ± 10 km for depth, ± 1 magnitude for event magnitude, and ± 10 s for event origin time.

Likelihood function. The performance of the particle filter for parameter estimation depends largely upon the design of the likelihood function. In addition to the arrival time and measured amplitude from the *triggered* stations that current JMA approach uses, our likelihood function also utilize the same information from *non-triggered* stations as well because they also convey important information about the event.

In this paper, we use the attenuation relationship developed by JMA for magnitude estimation. The relationship is stated as follows (Hoshiba and Ozaki, 2013). Let A_{max} be the maximum displacement measured by a seismometer after the onset of an event. The earthquake P-wave and S-wave magnitude M_p and M_s can be expressed as a function of the linear distance from the station to the hypocenter (R), the depth of the hypocenter (D), and the maximum displacement for P-wave (A^p_{max}) or the maximum displacement of the entire duration (A^{p+s}_{max}).

$$0.72M_p = \log A^p_{max} + 1.2\log R + 5\times 10^{-4}R - 5.0\times 10^{-3}D + 0.46, \tag{10}$$

$$0.87M_s = \log A^{p+s}_{max} + \log R + 1.9\times 10^{-3}R - 5.0\times 10^{-3}D + 0.98. \tag{11}$$

The relationship between the parameters is illustrated in Figure 1. These formulae are specifically tailored for the geological compositions in Japan (see Data and Resources). The P-wave and S-wave magnitudes are expressed in terms of the maximum displacement A_{max} rather than the maximum acceleration or velocity because the scatter of displacement is smaller.

Given Equation (10) and Equation (11) and that the displacement is log-normally distributed $A \sim \ln\mathcal{N}(\mu, \sigma^2)$, we propose the following likelihood function for a single station,

$$L(z|x, y, D, M, t_0) = \frac{\exp \frac{-(\log A_{max} - \log A_{exp})^2}{2\sigma^2}}{A_{max} \cdot \sigma\sqrt{2\pi}}. \tag{12}$$

Here A_{exp} is the expected A_{max} and σ is the standard deviation of displacement measurement. Depending on whether the station has observed P-wave, S-wave, or neither, the expected maximum displacement and its standard deviation are different. For convenience, by rearranging Equation (10) and Equation (11), we can compute A_{exp} and σ for the following three cases.

Note that Equation (12) is based on amplitude which departs from standard arrival-time based methods. The main reason for adopting this approach is the observation that the information of no shaking is critical in separating and classifying multiple earthquakes that occur close in space and time. This will be further discussed in Discussion.

- Has not observed any seismic wave:

$$\log A_{exp} = \log A_{noise},\ \sigma = \sigma_{noise}. \tag{13}$$

- Has observed P-wave:

$$\log A_{exp} = 0.72M_p - 1.2\log R - 5 \times 10^{-4}R + 5.0 \times 10^{-3}D - 0.46,\ \sigma = \sigma_p. \tag{14}$$

- Has observed S-wave:

$$\log A_{exp} = 0.87M_s - \log R - 1.9 \times 10^{-3}R + 5.0 \times 10^{-3}D - 0.98,\ \sigma = \sigma_s. \tag{15}$$

A_{noise} and σ_{noise} are the noise in displacement measurement due to recent environmental noise and can be computed independently for each station by keeping a running window. σ_p and σ_s can be precomputed from historical earthquake data.

The decision of which A_{exp} to compute for a station depends on whether P-wave, S-wave, or neither has arrived at the station. The expected travel time of P-wave and S-wave (t_p and t_s) can be computed with ray theory, given the relative location of the station to a hypocenter (x, y, D). Comparison between t_p, t_s, the absolute current time t, and the absolute event start time t_0 gives direct estimation of which A_{exp} to compute for a station. Figure 2 provides a illustrative summary of these design ideas.

This design of the likelihood function is based on the maximum displacement A_{max} that a seismometer observes during the shaking of P- or S-wave. However, a seismometer may not observe the maximum displacement immediately after the wave arrival. In this case, the initial estimates can be highly incorrect using this likelihood function. A simple delay function $g(\cdot)$ can be included to approximate the instantaneous displacement before the maximum is observed,

$$A_{exp} = g(t - t_0 - t_p)A_{max},\ 0 \leq g(\cdot) \leq 1. \tag{16}$$

where t and t_0 are the absolute current time and the absolute event origin time. t_p is the expected P-wave travel time. An example of $g(t)$ is a left shifted sigmoid function.

The likelihood $L(\cdot|\cdot)$ is applied in each time step to update the weight of each particle. Assuming that each station makes independent observation and the collection of observations from all stations is $\mathbf{z}$, the complete likelihood function becomes

$$L(\mathbf{z}|x, y, D, M, t_0) = \prod_{i=1}^{n} L(z_i|x, y, D, M, t_0), \tag{17}$$

where n is the number of stations. Note that the independence assumption is a minor simplification since nearby stations may have correlated observations.

Generalized Particle Filter

Particles are initialized according to a prior distribution on the parameters. Since we are approximating an unbounded and continuous 5-dimensional space with a bounded and discrete one, care must be taken to ensure that the particles have sufficient coverage and the number of required particles stays bounded. This is especially important for the seismic application since both the number of parameters and the range of values they can take are large. One way to ensure particle diversity with a limited number of particles is to adopt the *Regularized Particle Filter (RPF)* approach (Arulampalam et al., 2002).

RPF differs from common particle filter only in the resampling stage. Rather than sampling from a discrete approximation of the posterior density $P(\cdot|\mathbf{z})$ as in Equation (6), RPF samples from a *continuous* approximation (Musso et al., 2001). More specifically, RPF draws samples from the approximation,

$$P(\theta|\mathbf{z}) \approx \sum_{i=1}^{N} w_i \cdot K_h(\theta - \theta_i), \tag{18}$$

where $K_h(\theta) = \frac{1}{h}K(\theta/h), h > 0$ is the rescaled kernel density of $K(\cdot)$. h is the bandwidth, and w_i is the normalized weight for particle i. As a comparison, $K_h(\theta)$ is the Dirac delta function $\delta(\theta)$ in the regular particle filter. Special care is given to the design of kernels to minimize the error between approximated and actual distribution. Under the assumption that all particles are equally-weighted and the density is Gaussian, the optimal kernel is the Epanechnikov kernel (Musso et al., 2001).

$$K_{opt}(x) = \begin{cases} \frac{n_x+2}{2C_{n_x}}(1 - \|x\|) & \|x\| < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{19}$$

where n_x is the dimension of the parameter space, C_{n_x} is the volume of the unit hypersphere in $\mathbb{R}^{n_x}$. Figure 3 lists a few popular kernels in the literature.

The bandwidth vector h can be chosen proportionally to the variance in the particle population by computing the Cholesky decomposition of the empirical covariance matrix (Bickel and Levina, 2008).

Approximate Method for Multiple Concurrent Quakes

PF allows for solving the Bayesian inference problem when exact inference is intractable; however, for the estimates to approach the optimal solution, the number of required particles must grow exponentially with the number of events.

Fortunately, as shown in historical records, the probability of having n concurrent earthquakes within a time window of 60 seconds is exponentially small for large n ($n > 3$). Incorporating this information into the prior distribution can significantly reduce the size of the state space. But the state space may still be too large for efficient real-time computation even with this information. For example, suppose that the quake can be parameterized by a 5-parameter vector θ, $\theta = [x\ y\ D\ M\ t_0]^T$ where $[x\ y\ D]^T$ is the [longitude, latitude, depth] coordinate, M is the event magnitude, and t_0 is the event starting time. In the presence of $n = 3$ quakes, the states to be searched reside in a $5 \times 3 = 15$-dimensional space.

This amount of computation may be executable in reasonable time on a supercomputer or a networked system of computers with parallel implementation of particle filter (Durham and Geweke, 2013; Miao et al., 2010). In this paper, however, we propose a simple heuristics to keep track of multiple quakes. The heuristics has the desired property such that the complexity grows linearly with the number of the events.

As a first approximation, the heurisitcs initializes separate particle filters $pf_1(\theta_1), pf_2(\theta_2), \ldots$

for all possible quakes rather than keeping track of all events within one particle filter $pf(\Theta = \{\theta_1, \theta_2, \ldots\})$. Each particle filter communicates its current estimate $\hat{\theta}$ at *the end* of each update step to all other particle filters. Specifically, each particle filter pf_i computes the following posterior at time t,

$$P(\theta_i^t | \mathbf{z}, \{\hat{\theta_j^{t-1}}, j \neq i\}). \tag{20}$$

This approximation breaks down the $5n$ state space where n is the number of concurrent quakes, and dramatically reduces the required computations to keep all events estimation up to date. It is suboptimal, however, since all the particles from $pf_1, pf_2, \ldots$ combined only cover a small fraction of the complete parameter space.

The heuristic initializes a new particle filter with each single station P-wave pick, using a high enough threshold such that noisy detections are filtered out. Since local detection can be due to an existing event that is being tracked by another particle filter, it is necessary to condition new initialization on a separate metric. A natural choice of metric is $\mathbb{P}\left[z|\hat{\theta}_1, \hat{\theta}_2, \ldots\right]$, i.e., the probability that the triggered measurement can be explained by existing events. Computation of this metric can follow directly from the single station likelihood calculation as in Equation (12); however, determining A_{exp} is nontrivial in this case since it involves computing the additive effect of the interference of multiple wavefronts. We propose an alternative metric which allows for rapid computation; the metric is the probability of shaking due to *any* of the existing events and threshold on the highest probability:

$$\max_i \mathbb{P}\left[z|\hat{\theta}_i\right] = \max_i L(z|\hat{\theta}_i) \begin{cases} < \tau, & \text{initialize new pf} \\ \geq \tau, & \text{do nothing} \end{cases}. \tag{21}$$

By tuning the threshold τ, we adjust how conservative the system is in declaring new events. The complete algorithm is outlined in Algorithm 1 in the appendix for reference.

Results

We carried out the particle filter parameter estimation approach on the data described in Data and Processing, using a flat prior around the first triggered station and 1,000 particles for each particle filter. The algorithm updates at a one-second interval and all experiments were run in simulated real-time.

Case 1: 15 March 2011, 1:36:00 - 1:38:00 (two small earthquakes). 20 trials were performed during this period of time. Snapshots of the particle distribution for one of the runs are shown in Figure 4. The averaged time histories of the estimated parameters across all 20 runs were compared against the JMA unified catalog (marked as dotted lines) in Figure 5. The standard deviations across all runs are included as the error bars. The labeled x-axis corresponds to seconds since the first detection of the first event. As the results demonstrate, the first particle filter was initiated at the first P-wave arrivals, and 15 s later, another particle filter was created. This approach successfully identified the two separate events. In addition, all estimates converge within 10 s after the initializations. On average, the method is able to localize the epicenters to within 20 km and produce magnitude estimates with an error of ± 1, relative to the JMA unified hypocenter catalog (Table 1).

Case 2: 20 March 2011, 14:19:00 - 14:21:00 (two small earthquakes). We repeated the analyses for the dataset of Case 2, where two small earthquakes occurred 5 s apart. The snapshots of particle distributions and time series of estimated parameters are included in Figure 6 and Figure 7. Note that in this example, because the first event occurred offshore and there were fewer near-source recordings, localization and estimation of other parameters are more challenging than for Case 1. Indeed, the results showed that the estimates converge slower (about 30 s for event A), and the averaged localization error was relatively large (about 80 km for event A, relative to the JMA unified hypocenter catalog (Table 1). However, the algorithm was still able to identify and separate the two events and provide accurate estimates of their magnitudes to within ± 0.5.

Case 3: 11 March 2011, 14:46:00 - 14:49:00 (Tohoku earthquake). We used the dataset of the Tohoku earthquake to show that the approach also works for a single event. The snapshots of particle distributions and time series of estimated parameters are included in Figure 8 and Figure 9. Since the event was originated offshore, there was substantial localization error in the initial estimates. However, the averaged error decreased with time and converged at less than 40 km at 40 s after the initial P-wave arrival. The magnitude estimate grew from 6.0 to 8.4 as the earthquake rupture propagated, which is consistent with the earthquake rupture physics. At convergence, all five estimated parameters were close to the values in the JMA catalog.

Discussion

Current JMA methods to detecting and associating multiple quakes perform well when the events are far apart in space or time. However, they have been shown to generate many false alarms when events are close in space or time (Sagiya et al., 2011). The empirical studies suggest that the particle based heuristic can successfully separate multiple concurrent seismic events and provide reasonable estimates of their parameters. And the speed of convergence may be improved by incorporating P-wave arrival time in the likelihood, i.e., the residual between observed and predicted P-wave arrival times. The results show that estimated parameters converge in less than 10 s for inland earthquakes. For offshore earthquakes, the estimates converge in 20-30 s. In terms of localization error, we observed less than 20 km for inland earthquakes, and 20-80 km for offshore events.

In order to classify multiple concurrent earthquakes, the use of non-triggered stations is important. The current JMA EEW system uses arrival times of waves at only the triggered stations in the hypocenter calculation. As a result, when multiple earthquakes occur around the same time and the later event occurs close to the wave arrival times of the earlier event, the EEW system treats these events as one single earthquake. If this is the case and the stations around the later event observes non-negligible shakings, the current system may overestimate the magnitude because these stations are far away from the estimated hypocenter (i.e., the location of the earlier event). In our approach, the likelihood function uses information from not only the triggered stations but also the non-triggered ones. This design together with the adaptive measure of A_{noise} allow the algorithm to identify unaffected regions between events and is therefore crucial in separating multiple concurrent earthquakes.

Another advantage of our approach is the use of regularized particle filter to circumvent the need for intensive computation that traditional grid search requires. Although a prior distribution is still required as mentioned in Model, such a distribution can be compiled from historical records. Alternatively, initial measurements can be used to "select" the appropriate priors to achieve better performance (Liu et al., 2011).

This approach is also subject to several weaknesses. For example, the algorithm is sensitive to the choice of prior distribution, the number of particles, the values of A_{noise}, σ_{noise}, σ_p and σ_s. While these values can be adjusted and adapted in real time, it requires extensive empirical studies and analyses of historical records for the algorithm to be robust. Some of the slow convergence and high variance results in Results may be attributed to suboptimal choices in these parameters.

In this paper we use only three cases to test the proposed method, so we are currently carrying out more extensive evaluations of our method using the many examples of multiple earthquake sequences that have occurred over the last several years.

As a side note, the performance of parameter estimation for multiple seismic events is limited by how well one can model the ground motion when multiple wavefronts overlap. In the algorithm proposed in Model, this model is not considered. While the omission makes little difference in the case studies where the events are spatially far apart (greater than 100 km), if we want to apply the same technique to separate aftershocks from mainshock that occur close in time, then such model should be considered.

Conclusion

In the seismically active period, multiple earthquakes of similar or distant origins can take place at almost the same time. Failure to identify them as separate events leads to poor estimates of their parameters. The error in estimates can in turn cause false warnings. In this paper, we study the problem of detecting and classifying multiple earthquakes that occur close in time. Based on a Bayesian formulation that considers the possibility of having more than one event present at any given time, we propose a novel likelihood function suitable for classifying multiple concurrent earthquakes and present a sequential Monte Carlo heuristic whose complexity grows linearly with the number of events. The performance of the heuristic is empirically validated with three sets of JMA seismic records after the 2011 Tohoku earthquake. The initial studies show that the approach is able to successfully separate multiple events that occur close in space and time and estimate their parameters in realtime to a reasonable degree of precision in comparison to official values determined by JMA in the post event analyses. Although complete validation and characterization are required before this method applied in realtime detection, the initial results show that our approach can reduce the chance of overestimation of earthquake magnitude and, as a result, contribute to the design of a better EEW system.

Data and Resources

Waveform data used in the present study were extracted from continuous recordings of the stations within the JMA strong motion network. The JMA EEW performance of three cases is available at; `http://www.seisvol.kishou.go.jp/eq/EEW/kaisetsu/joho/20110315013605/content/contentout.html`, `http://www.seisvol.kishou.go.jp/eq/EEW/kaisetsu/joho/20110320141959/content/contentout.html`, and `http://www.seisvol.kishou.go.jp/eq/EEW/kaisetsu/joho/20110311144640/content/contentout.html` (last accessed July 2013). We use Seismic Analysis Code (`http://www.iris.washington.edu/`

software/sac/manual/fileformat.html, last accessed July 2013) for the data processing. The JMA attenuation relationship are available in the report of the second JMA EEW evaluation committee (http://www.seisvol.kishou.go.jp/eq/EEW/MeetingHYOUKA/t02/shiryou.pdf, last accessed July 2013).

Acknowledgements

The first author would like to thank professor Jim Mori at the Kyoto University for the generous hosting and the warm welcome she received from all the group members. The authors would also like to thank JMA for providing the strong motion seismic data. This work was generously funded by the 2011 National Science Foundation (NSF) East Asian and Pacific Summer Institute (EAPSI) Fellowship and the Funding Program for Next Generation World-Leading Researchers (NEXT Program).

References

Arulampalam, M., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *arXiv.org*.

Doucet, A., de Freitas, N., Gordon, N., and Smith, A. (2001). *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer.

Durham, G. and Geweke, J. (2013). Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments. *arXiv.org*.

Hoshiba, M. and Iwakiri, K. (2011). Initial 30 seconds of the 2011 off the Pacific coast of Tohoku Earthquake(M (w) 9. 0)-amplitude and tau(c) for magnitude estimation for Earthquake Early Warning -. *Earth Planets Space*, 63(7):553–557.

Hoshiba, M., Iwakiri, K., and Hayashimoto, N. (2011). Outline of the 2011 off the Pacific coast of Tohoku Earthquake (M_w 9.0)—Seismicity: foreshocks, mainshock, aftershocks, and induced activity—. *Earth Planets Space*, 63(7):513–518.

Hoshiba, M. and Ozaki, T. (2013). Earthquake early warning and tsunami warning of the japan meteorological agency, and theirperformance in the 2011 off the pacific coast of tohoku earthquake (mw9.0). In Wenzel, F. and Zschau, J., editors, *Early Warning for Geological Disasters*. Springer-Verlag Berlin Heidelberg.

JMA (2011). Earthquake early warning report for tohoku 2011. `http://www.jma.go.jp/jma/press/1104/28b/eew_hyouka_2.pdf`.

Katsumata, A. (2008). Recursive digital filter with frequency response of a mechanical seismograph. Technical report, Seismology and Volcanology Research Department, Meteorological Research Institute.

Liu, A. H., Bunn, J., and Chandy, K. M. (2011). Sensor networks for the detection and tracking of radiation and other threats in cities. In *The 10th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 1–12.

Liu, J. S. and Chen, R. (1998). Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–1044.

Miao, L., Zhang, J. J., Chakrabarti, C., and Papandreou-Suppappola, A. (2010). A new parallel implementation for particle filters and its application to adaptive waveform design. *Signal Processing Systems (SIPS), 2010 IEEE Workshop on*, pages 19–24.

Musso, C., Oudjane, N., and Legland, F. (2001). Improving regularized particle filters. In Doucet, A., de Freitas, N., and Gordon, N., editors, *Sequential Monte Carlo Methods in Practice. New York*, number 12, pages 247–271. Statistics for Engineering and Information Science.

Sagiya, T., Kanamori, H., Yagi, Y., Yamada, M., and Mori, J. (2011). Rebuilding seismology. *Nature*, (473):146–148.

Ueno, H., Hatakeyama, S., Aketagawa, T., Funasaki, J., and Hamada, N. (2002). Improvement of hypocenter determination procedures in the japan meteorological agency. *Quarterly Journal of Seismology*, 65(1-4):123–134.

Table 1: Summary of the earthquake information studied in this paper.

	M_{est}	Lon	Lat	Dep	Date	Time	M	Lon	Lat	Dep
Case 1	5.9	138.6	36.9	10	03-15	01:35:57.35	2.5	138.610	36.938	3.4
					03-15	01:36:12.72	3.3	139.879	35.526	20.5
Case 2	7.6	142.1	38.2	30	03-20	14:19:38.27	3.0	141.935	38.286	42.3
					03-20	14:19:58.06	4.7	140.794	37.082	7.2
Case 3	8.6	142.7	38.2	10	03-11	14:46:48.08	9.0	142.861	38.103	23.7

The first four columns correspond to the real-time JMA EEW records. The last six columns are the values documented in the JMA unified hypocenter catalog. Both Case 1 and 2 contain two events.

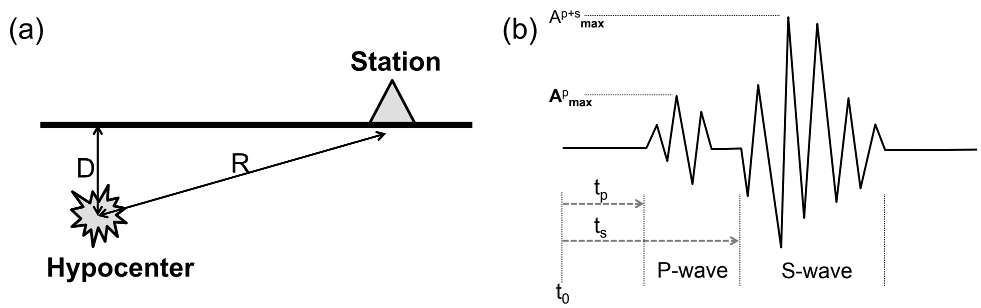

Figure 1: Illustrations of the parameters used in Model. (a) Hypocenter and seismic station and (b) amplitude and arrival times. t_p and t_s mark the arrival time of the P-wave and S-wave since the start of the earthquake at t_0, $t_p \leq t_s$.

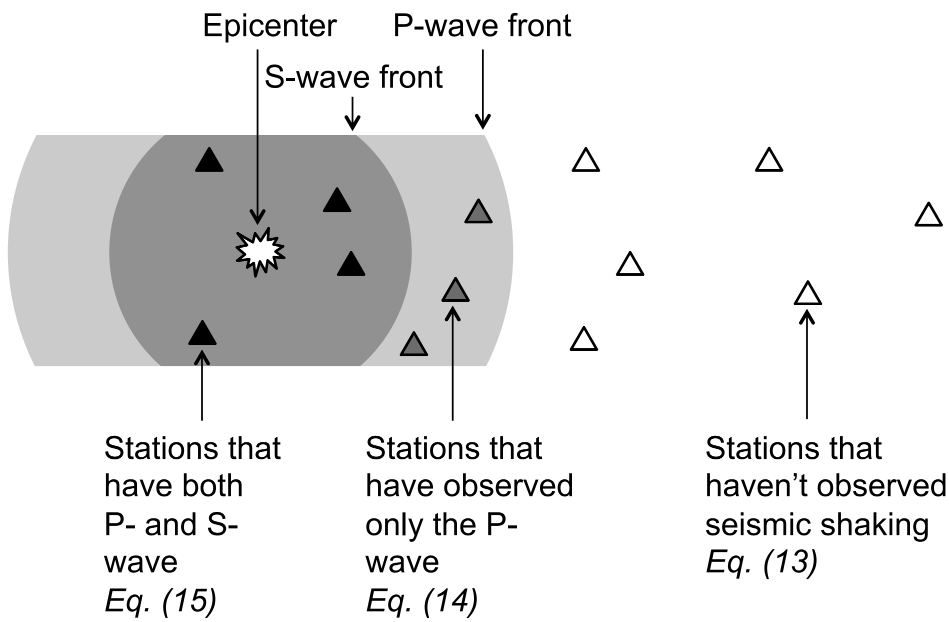

Figure 2: Illustrative summary of the design of a single station likelihood function. The expected observation made by a station depends on whether it should have observed P-wave, S-wave, or neither, given an hypocenter estimate.

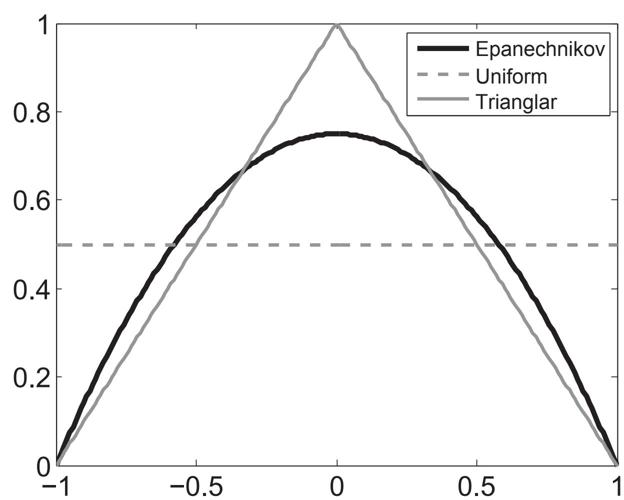

Figure 3: Some popular smoothing kernels used in regularized particle filter. Each kernel integrates to 1 to ensure that the resulting density is still a probability density function.

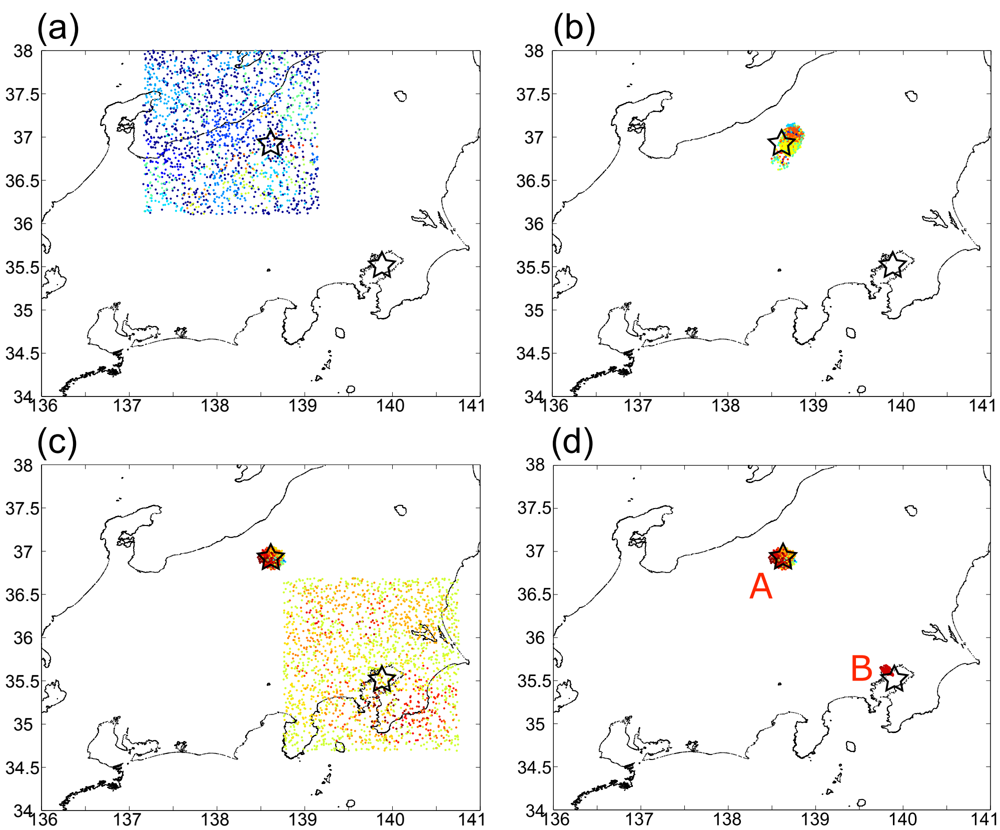

Figure 4: Distributions of 2,000 Particles visualized on the map at (a)1 s, (b)2 s, (c)14 s, and (d)17 s after 1:36:07 on 15 March, 2011. The time correspond to seconds elapsed since the first P-wave detection. The official epicenters for the two events as appeared in the JMA catalog are marked as stars and labeled in (d) for reference.

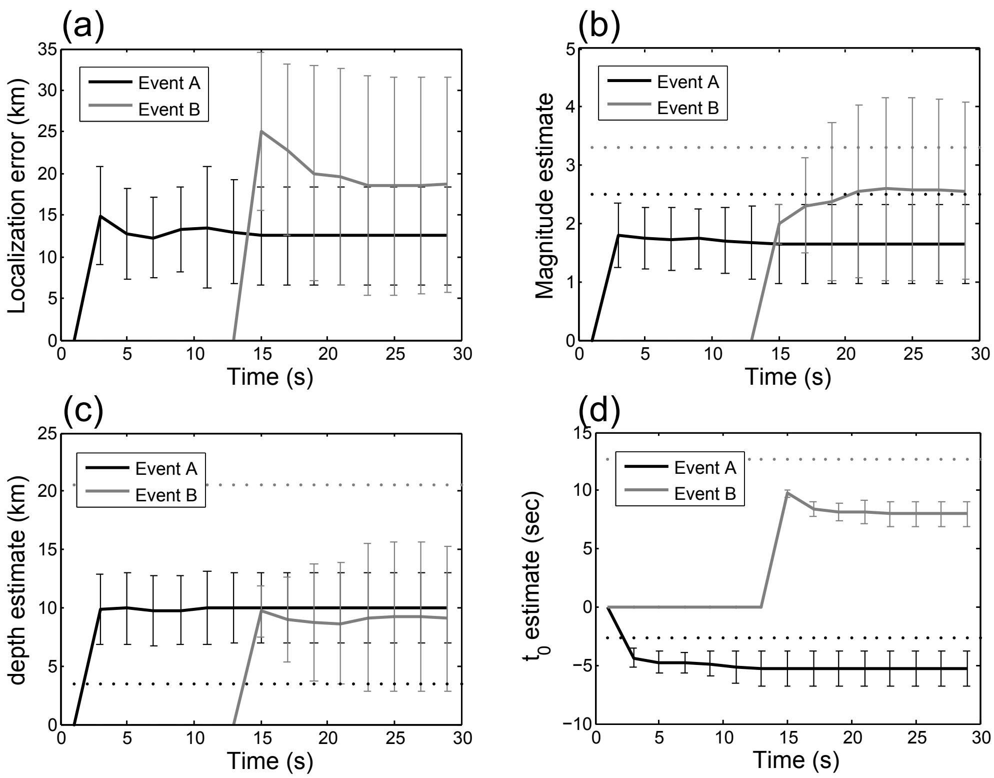

Figure 5: Results compiled from 20 independent runs for the period between 1:36:07 and 1:36:37 on 15 March, 2011. Time histories of the (a) localization error, (b) magnitude, (c) depth of the hypocenter and (d) origin time of the event. The two events are labeled according to Figure 4(d). Averaged time histories across all 20 runs are marked as solid lines, and the official values that appear in the JMA catalog are marked as dashed lines. The standard deviations across all runs are shown as error bars. The time displayed on the x-axis is relative to the first pick from the earliest event.

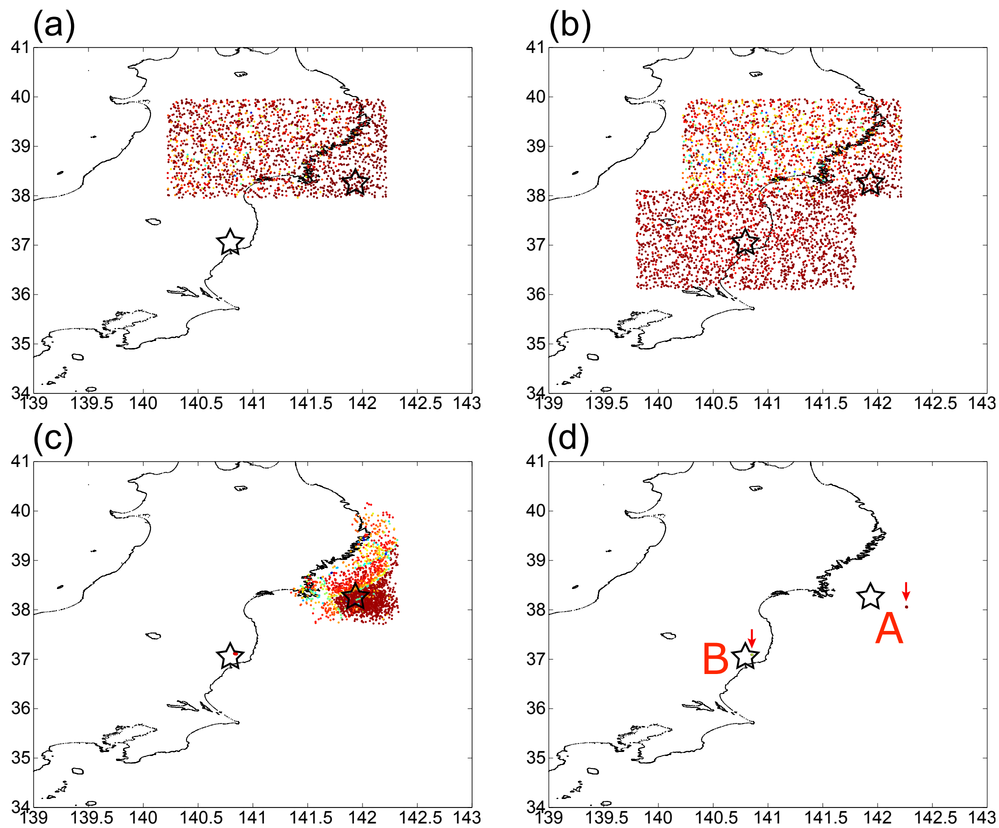

Figure 6: Distributions of 2,000 Particles visualized on the map at (a)2 s, (b)7 s, (c)17 s, and (d)37 s after 14:19:56 on 20 March, 2011. The symbols are defined the same way as in Figure 4.

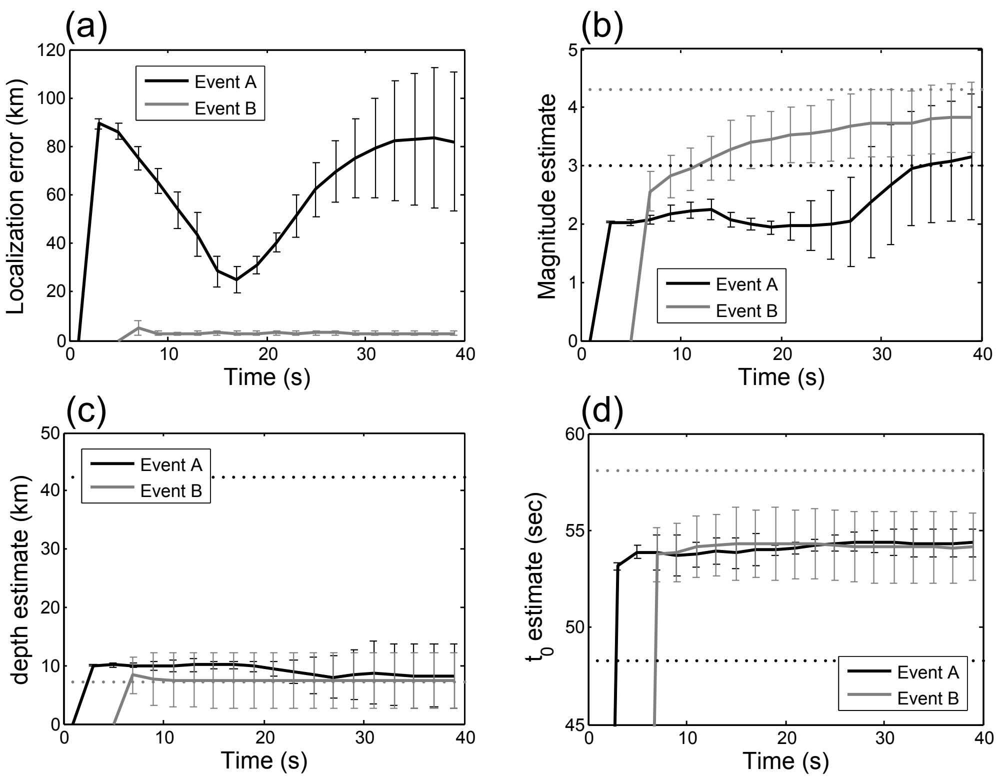

Figure 7: Results compiled from 20 independent runs in the period between 14:19:56 and 14:20:36 on 20 March, 2011. The subfigures and included symbols are defined the same way as in Figure 5.

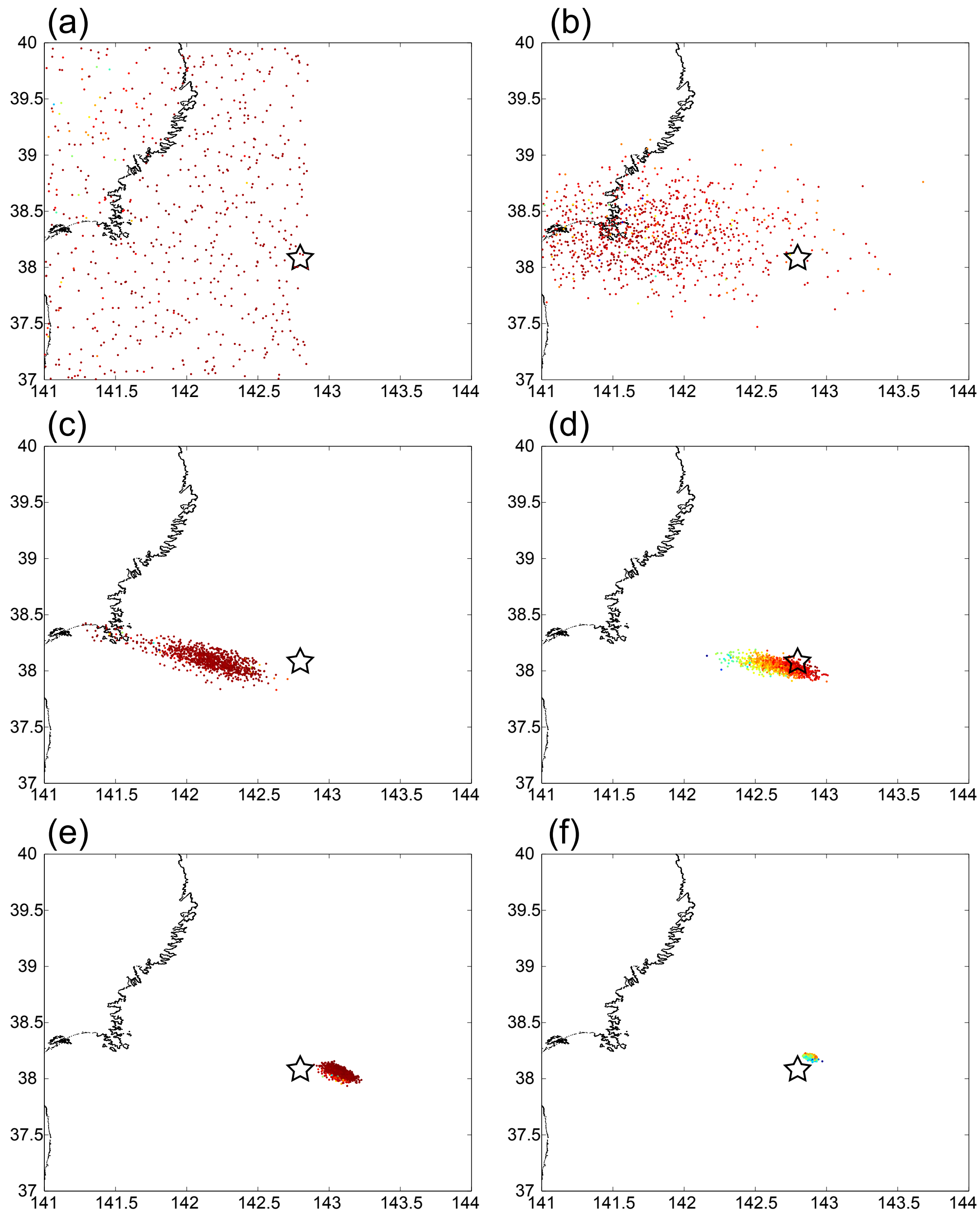

Figure 8: Distributions of 1,000 Particles visualized on the map at (a)2 s, (b)7 s, (c)13 s, (d)22 s, (e)32 s, and (f)62 s after 14:46:46 on 11 March, 2011. The symbols are defined the same way as in Figure 4.

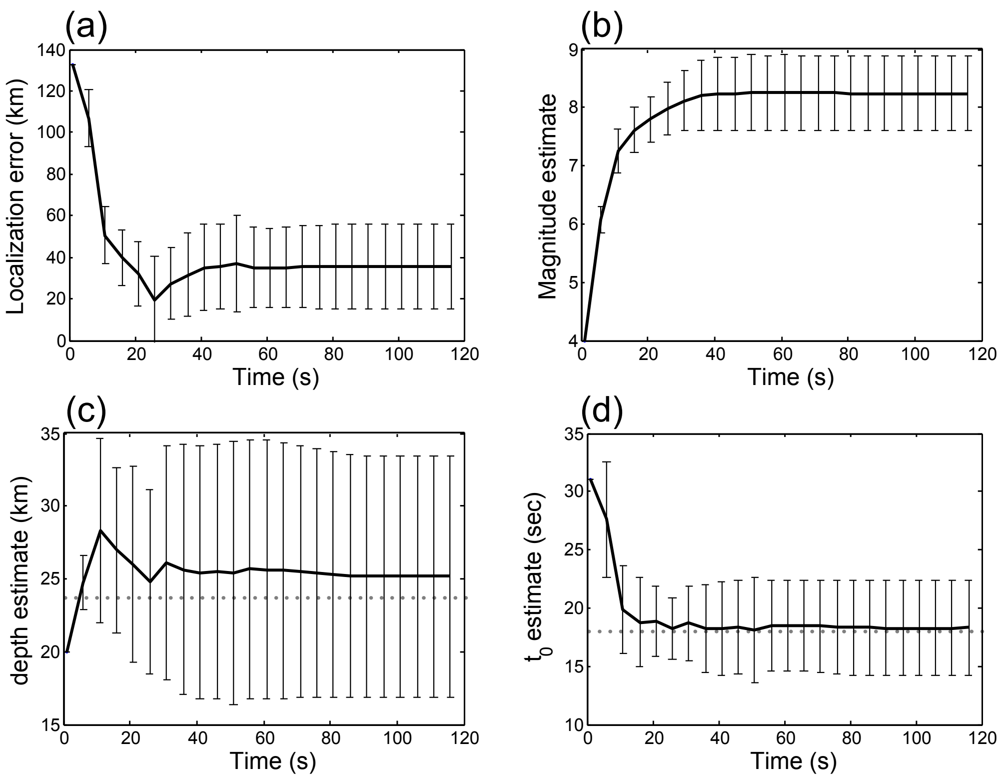

Figure 9: Results compiled from 15 independent runs for the period between 14:46:46 and 14:48:46 on 11 March, 2011. The subfigures and included symbols are defined the same way as in Figure 5.

A Appendix

Algorithm 1: Outline of regularized Particle Filter for multiple seismic event detection. The "CONVERGED" criteria can be substituted with desired conditions, e.g. change in estimates $\left\|\hat{\theta}_{t-10} - \hat{\theta}_{t-1}\right\| < \delta$.

$PF \leftarrow \{\}$
Initialize thresholds τ, α
Initialize bandwidth vector $h' \in \mathbb{R}^{n_x}$
while *not end* **do**
    ⋆ Check for new event
    $Z \leftarrow$ list of station measurements that triggered
    for $z \in Z$ **do**
        $pr \leftarrow \max_k L(z|\theta_k)$
        if $pr < \tau$ **then**
            $\left[\{\theta_i, w_i\}_{i=1}^N\right] \leftarrow RPF\left[\{\theta_i, w_i\}_{i=1}^N, z\right]$
            for $i = 1 \rightarrow N$ **do**
                Draw $\theta_i \sim P(\theta, z)$
                Assign weights based on prior and z, $w_i \sim P(\theta, z)$
            $pf \leftarrow \left[\{\theta_i, w_i\}_{i=1}^N, z\right]$
            $PF \leftarrow PF \cup pf$
    ⋆ Update weight, resample if needed
    for $pf \in PF$ **do**
        $\{\theta_i, w_i\}_{i=1}^N \leftarrow pf$
        for $i = 1 \rightarrow N$ **do**
            $w_i \leftarrow w_i L(\mathbf{z}|\theta_i)$
        $\left[\{\theta_i, w_i\}_{i=1}^N\right] \leftarrow NORMALIZE\left[\{\theta_i, w_i\}_{i=1}^N\right]$
        Compute $\widehat{N_{eff}} \leftarrow \frac{1}{\sum_{i=1}^N w_i^2}$
        if $\widehat{N_{eff}} < \alpha$ **then**
            $\left[\{\theta_i, w_i\}_{i=1}^N\right] \leftarrow RESAMPLE\left[\{\theta_i, w_i\}_{i=1}^N, z\right]$
            for $i = 1 \rightarrow N$ **do**
                Draw $\epsilon \sim K$ from the Epanechnikov Kernel
                Compute weighted empirical covariance matrix S_k of $\{\theta_i, w_i\}_{i=1}^N$
                Compute lower triangle $D_k = chol(S_k)$, $D_k D_k^T = S_k$
                $\theta_i \leftarrow \theta_i + h' D_k \epsilon$
    ⋆ Check for termination
    for $pf \in PF$ **do**
        $\{\theta_i, w_i\}_{i=1}^N \leftarrow pf$
        if $CONVERGED\left[\{\theta_i, w_i\}_{i=1}^N\right]$ **then**
            $PF = PF - pf$