submitted to *Geophys. J. Int.*

Multi-events Earthquake Early Warning algorithm using a Bayesian approach

S. Wu[1], M. Yamada[2], K. Tamaribuchi[3] and J.L. Beck[1]

[1] *California Institute of Technology, Pasadena, California, USA*

[2] *Disaster Prevention Research Institute, Kyoto University, Uji, Japan*

[3] *Japan Meteorological Agency, Japan*

SUMMARY

Current earthquake early warning (EEW) systems lack the ability to appropriately handle multiple concurrent earthquakes, which led to many false alarms during the 2011 Tohoku earthquake sequence in Japan. This paper uses a Bayesian probabilistic approach to handle multiple concurrent events for EEW. We implement the theory using a 2-step algorithm. First, an efficient approximate Bayesian model class selection scheme is used to estimate the number of concurrent events. Then, the Rao-Blackwellized Importance Sampling method with a sequential proposal probability density function is used to estimate the earthquake parameters, i.e., hypocenter location, origin time, magnitude and local seismic intensity. A real data example based on two months data (March 9 to April 30, 2011) around the time of the 2011 M9 Tohoku earthquake is studied to verify the proposed algorithm. Our algorithm results in over 90% reduction in the number of incorrect warnings compared to the existing EEW system operating in Japan.

Key words: Earthquake Early Warning; Bayesian model class selection; Rao-Blackwellized Importance Sampling; Sequential sampling; 2011 Tohoku earthquake.

1 INTRODUCTION

During the 2011 M9 Tohoku earthquake in Japan, an earthquake early warning (EEW) system operated by the Japan Meteorological Agency (JMA) provided the first early warning 5.4s after the first P-wave detection, which was before the S-wave arrival throughout all areas in Japan (Hoshiba et al. 2011). However, during the two months after the mainshock on March 11, because of the significantly increased seismicity around the region, 44 inappropriate warnings (missed and false alarms) were issued among the 70 early warnings broadcast to the public (JMA, 2013a,b). Over 70% of the inappropriate warnings were caused by multiple concurrent events (referred as the *multi-events* problem in this paper). Motivated by these warning errors, our study focuses on the improvement of EEW using a Bayesian probabilistic approach and proposes an algorithm to properly handle the multi-events problem. The proposed algorithm first finds the most probable number of earthquake given a set of seismic data based on the theory of Bayesian model class selection, and then estimates the earthquake parameters (e.g., hypocenter location, magnitude and origin time) of each event based on the method of Bayesian inference. The JMA EEW system is used as an example to demonstrate the details of the algorithm.

The JMA EEW system began operating in 2007; it is the first country-wide EEW system, broadcasting warnings to the public when the expected intensity measure (in the JMA seismic intensity scale) is greater than or equal to "5 lower" (denoted as 5−). The current system uses a continuous datastream from around 300 JMA seismic stations and 700 Hi-net stations located all around Japan (Hoshiba et al. 2008). The datasets from the JMA network and the Hi-net network are processed in separate algorithms to provide individual estimates on hypocenter location and origin time. Then these results are combined to provide the final warning. The magnitude is estimated from the JMA network based on the JMA magnitude estimation equation (Kamigaichi, 2004; JMA, 2010, 2012), which is given in Section 3.3.2. Finally, the seismic intensity at each site is estimated based on the estimated magnitude and hypocenter location using an attenuation relationship (Si & Midorikawa 2000).

During the period of active seismicity after the Tohoku earthquake, the JMA EEW system sometimes treated the data from multiple events as a single earthquake. As a result, some of the near source stations of the second event were treated as far source stations of the first event (Figure 1). This led to a significant overestimate of magnitude and caused false alarms. Recently, Yamada et al. (2012) and Liu & Yamada (2014) proposed a multi-events recognition algorithm using maximum displacement amplitude of triggered and non-triggered stations based on a particle filter approach. However, displacement amplitude is not sensi-

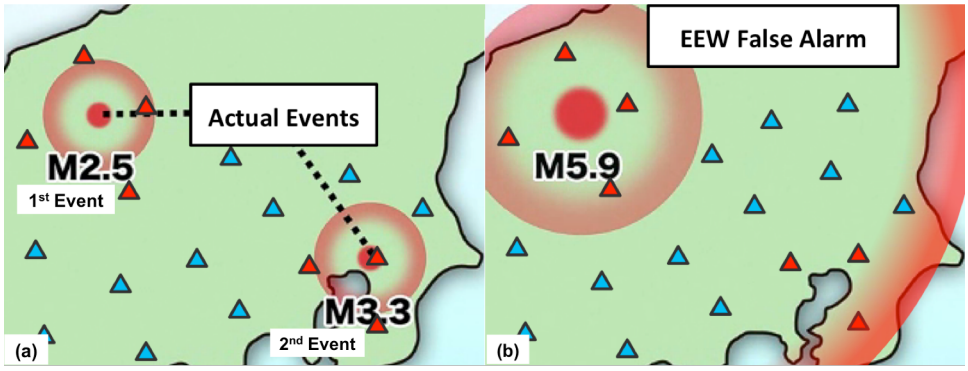

Figure 1. Mechanisms of false alarms: (a) two small earthquakes occurred concurrently being distant from each other; (b) EEW treated near source stations of one small event as far source stations from a large event. Triangle: seismic station (triggered stations in red).

tive enough for providing accurate hypocenter location estimate. One way to improve this problem is to include the P-wave picking time information, which is conventionally used in the hypocenter location. This paper integrates P-wave picking time and maximum displacement amplitude from both JMA and Hi-net seismic stations into a *single* algorithm under a Bayesian probability framework. We expect our approach to provide faster and more accurate warnings for EEW systems.

The outline of this paper is as follow: Section 2 states the details of the data used in the case study; Section 3 explains the theoretical foundation of the proposed algorithm; Section 4 illustrates the implementation details of the algorithm in practice; Section 5 shows the results of the case study followed by some discussions; Section 6 concludes the paper with suggested directions for future study.

2 DATA DESCRIPTION

We use the Japanese seismic data from March 9 to April 30, 2011 in this paper. This period includes the Tohoku foreshock that happened on March 9, the mainshock on March 11 and the major part of the aftershock sequence. Continuous waveform data from around 300 JMA stations and 700 Hi-net stations are used in this study. The JMA seismic network uses accelerometers, whereas the Hi-net uses short-period velocity meters. The velocity meters have a lower sensitivity for long-period waves and suffer from large amplitude saturation (Shiomi et al. 2005). Therefore, we remove instrumental response of the Hi-net data and adjust them to the JMA data (Yamada et al. 2014). Two features of the waveform data are used in this study: the maximum displacement amplitude and the P-wave picking time. The displacement

amplitude is computed using a 6s high-pass filter (Katsumata 2008). Then, the maximum displacement amplitude is obtained by continuously tracking the maximum value as a function of time after a station is triggered. The P-wave picking time is computed using an algorithm suggested in Yamada et al. (2014), which is based on a filter proposed in Allen (1978).

The performance for earthquake parameter estimation of the proposed algorithm is evaluated using the JMA unified catalogue obtained from the National Research Institute for Earth Science and Disaster Prevention (NIED) webpage (NIED, 2014). The results of the current JMA EEW system is obtained from the JMA webpage (JMA EEW, 2014). A total of 71 events are selected to evaluate the performance of EEW systems (see Appendix A).

The JMA magnitude used in this paper are based on the maximum displacement for large earthquakes, and maximum velocity for smaller earthquakes. The magnitude of the Tohoku mainshock is moment magnitude.

3 BAYESIAN THEORY FOR MULTI-EVENTS EEW ALGORITHM

In the case of multi-events, it is important for an EEW system to identify the number of concurrent events given the current data set, as well as their information, in order to broadcast accurate warnings to the users. This section describes the theory and method for identifying the number of concurrent earthquakes and estimating the source parameters. For the ease of derivation, this paper adopts the following notation:

- $\mathcal{D}_{1:t} = \{D_j(1:t)|j = 1, ..., N_{st}\}$—set of waveform data $D_j(1:t)$ for N_{st} stations from initial time step 1 to time step t
- $\mathcal{F}_t = \{F_j(t)|j = 1, ..., N_{st}\}$—set of vectors of data features $F_j(t)$ (e.g., maximum displacement amplitude and P-wave picking time in this study), which are extracted from the waveform data $\mathcal{D}_{1:t}$, for each of the N_{st} stations, and used for parameter estimation
- $\mathcal{M}_n$—Bayesian model class (Beck 2010) that assumes n concurrent events are captured within the current data set $\mathcal{D}_{1:t}$
- $\Theta_n = \{\theta_l|l = 1, ..., n\}$—set of vectors of earthquake parameters θ_l for each of the n events given by $\mathcal{M}_n$ (assume Θ_n is independent of time t, i.e., it is a static variable)
- $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$ - the standard Gaussian PDF (probability density function)
- $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{y^2}{2}\right)\mathrm{d}y$ - the standard Gaussian CDF (cumulative density function)
- $p(x|y)$ - conditional PDF of x given y (x is a continuous variable)
- $P(x|y)$ - probability of x given y (x is a discrete variable)

3.1 Bayesian probability approach for EEW

In general, an EEW system receives and processes seismic network data continuously to provide updated warning and earthquake information. In our approach, the first step is to introduce the Bayesian model class selection framework to find the most probable number of earthquakes. The next step is a numerical method for Bayesian inference, called Rao-Blackwellized Importance Sampling (RBIS) (Liu 2002), to estimate the earthquake parameters for each identified event. In this section, this two-step process is expressed in a mathematical form based on applying fundamental probability theory at every discrete time t.

The first step is to find the most probable number of events, $\hat{n}$, that explains the current data set at time t, i.e., find the model class $\mathcal{M}_{\hat{n}}$ that maximizes the posterior probability over all model classes. This optimization problem is often referred to as Bayesian model class selection in the literature (Beck 2010). Rather than using all data $\mathcal{D}_{1:t}$ at time t in the posterior probability for $\mathcal{M}_n$, we use only the set of feature vectors $\mathcal{F}_t$ extracted from $\mathcal{D}_{1:t}$. By Bayes' theorem:

$$P(\mathcal{M}_n|\mathcal{F}_t) = \frac{p(\mathcal{F}_t|\mathcal{M}_n)P(\mathcal{M}_n)}{p(\mathcal{F}_t)} \propto p(\mathcal{F}_t|\mathcal{M}_n)P(\mathcal{M}_n) \tag{1}$$

Assuming a non-informative prior $P(\mathcal{M}_n) = constant\ \forall n$ to avoid imposing any bias on any model class before the data is collected, the probability of model $\mathcal{M}_n$ can be expressed as:

$$\begin{aligned} &P(\mathcal{M}_n|\mathcal{F}_t) \propto p(\mathcal{F}_t|\mathcal{M}_n) \\ &\Rightarrow \hat{n} = \text{argmax}_n\{P(\mathcal{M}_n|\mathcal{F}_t)\} = \text{argmax}_n\{p(\mathcal{F}_t|\mathcal{M}_n)\} \end{aligned} \tag{2}$$

where by the Total Probability Theorem, the evidence for $\mathcal{M}_n$ given by data $\mathcal{F}_t$ is:

$$p(\mathcal{F}_t|\mathcal{M}_n) = \int p(\mathcal{F}_t|\Theta_n, \mathcal{M}_n)p(\Theta_n|\mathcal{M}_n)\,\mathrm{d}\Theta_n \tag{3}$$

The models for $p(\mathcal{F}_t|\Theta_n, \mathcal{M}_n)$ and $p(\Theta_n|\mathcal{M}_n)$ are introduced later.

The second step is to find the earthquake parameter values $\Theta_{\hat{n}}$ for all events given $\mathcal{M}_{\hat{n}}$, i.e., find the posterior PDF $p(\Theta_{\hat{n}}|\mathcal{M}_{\hat{n}}, \mathcal{F}_t)$. This is a Bayesian inference problem under the specified model class $\mathcal{M}_{\hat{n}}$. By Bayes' theorem:

$$\begin{aligned} p(\Theta_{\hat{n}}|\mathcal{M}_{\hat{n}}, \mathcal{F}_t) &= \frac{p(\mathcal{F}_t|\Theta_{\hat{n}}, \mathcal{M}_{\hat{n}})p(\Theta_{\hat{n}}|\mathcal{M}_{\hat{n}})}{p(\mathcal{F}_t|\mathcal{M}_{\hat{n}})} \\ &\propto p(\mathcal{F}_t|\Theta_{\hat{n}}, \mathcal{M}_{\hat{n}})p(\Theta_{\hat{n}}|\mathcal{M}_{\hat{n}}) \end{aligned} \tag{4}$$

Here, the evidence function $p(\mathcal{F}_t|\mathcal{M}_{\hat{n}})$ in Equation 4 is the same one as in Equation 3 that is used for finding $\hat{n}$.

3.2 Efficient approximate scheme for Bayesian model class selection

There is limited time to perform the full Bayesian model class selection scheme through calculating the evidence for all possible model classes $\mathcal{M}_n$. Existing methods to calculate or estimate the evidence function $p(\mathcal{F}_t|\mathcal{M}_n)$ are not fast enough for this purpose. This motivates the need for an efficient approximate, but suboptimal, model class selection scheme that is also robust.

We assume that $\hat{n}$ is a monotonically increasing function of time t because earthquakes happen in sequence. Exploiting this pattern, instead of searching for an optimal $\hat{n}$ at every second, one may start with $\hat{n} = 0$ and increase $\hat{n}$ by one when a fast calculated criterion is satisfied. Intuitively, $\hat{n}$ should be increased by one when the current data set $\mathcal{D}_{1:t}$ cannot be well explained by any of the identified events given the currently selected model class $\mathcal{M}_{\hat{n}}$. In other words, let $\mathcal{M}_{\hat{n}} = \{M_l|l = 1, ..., \hat{n}\}$ where M_l represents each event identified within $\mathcal{M}_{\hat{n}}$, one may increase $\hat{n}$ by one when the following criterion is met:

$$p(\mathcal{F}_t|M_l) < \tau_{new} \;\; \forall l = 1, ..., \hat{n} \tag{5}$$

Here, τ_{new} is some empirical threshold (possibly depending on $\hat{n}$) for how well the current data set features $\mathcal{F}_t$ are explained by an event M_l, and $p(\mathcal{F}_t|M_l)$ is calculated by an integral similar to Equation 3 except that the integration is over θ_l, the earthquake parameters for M_l.

Equation 5 involves calculations with the complete feature set $\mathcal{F}_t$. One can further simplify the criterion by calculating with only one set of features $F_j(t)$ that is extracted from a single station j. Because most earthquakes have only one first triggered station, each event can be represented by its first triggered station. One can continuously search for newly triggered stations that have a low probability to be caused by any of the existing events in $\mathcal{M}_{\hat{n}}$. Those stations are likely to be the first triggered stations of new events. As a result, a more efficient criterion is:

$$p(F_j(t)|M_l) < \tau_{new} \text{ for newly triggered station } j \;\&\; \forall l = 1, ..., \hat{n} \tag{6}$$

where $p(F_j(t)|M_l) = \int p(F_j(t)|\theta_l, M_l)p(\theta_l|M_l)\,\mathrm{d}\theta_l$

3.3 Bayesian inference for earthquake parameters using RBIS

With the efficient approximate scheme, θ_l is only dependent on M_l within $\mathcal{M}_{\hat{n}}$. Hence, the posterior of each $\theta_l \in \Theta_n$, $p(\theta_l|\mathcal{M}_{\hat{n}}, \mathcal{F}_t) = p(\theta_l|M_l, \mathcal{F}_t)$, can be found separately. For notational simplicity, M_l will be left as implicit whenever θ_l appears in the rest of this paper. Hence, the

posterior for the parameters for each event becomes:

$$p(\theta_l|\mathcal{F}_t) = \frac{p(\mathcal{F}_t|\theta_l)p(\theta_l)}{p(\mathcal{F}_t|M_l)} \propto p(\mathcal{F}_t|\theta_l)p(\theta_l) \tag{7}$$

Given a model class $\mathcal{M}_{\hat{n}}$, $p(\theta_l|\mathcal{F}_t)$ of each event $M_l \in \mathcal{M}_{\hat{n}}$ is needed for broadcasting an appropriate warning. In practice, a numerical scheme is often used to estimate the posterior PDF. As mentioned in Section 2, two data features, $F_j(t) = [F_{p,j}(t), F_{a,j}(t)]$, are chosen for each station j, where $F_{p,j}$ is the P-wave picking time and $F_{a,j}$ is the logarithm of maximum displacement amplitude. Also, five earthquake parameters, $\theta_l = [lat_l, lon_l, d_l, m_l, t0_l]$, are needed for each event M_l, which are the hypocenter latitude, hypocenter longitude, hypocenter depth, earthquake magnitude and origin time, respectively. In Section 3.3.1, we present the RBIS (Rao-Blackwellized Importance Sampling) method for estimating the posterior PDF. This involves analytically integrating over some parameters and only sampling the remaining ones. We use it for improving the efficiency of the Importance Sampling. The likelihood functions in the Bayesian approach are calculated based on the existing JMA EEW model, as explained in Section 3.3.2.

3.3.1 Rao-Blackwellized Importance Sampling of Posterior PDF

Figure 2 shows the variable dependency between the chosen data features and earthquake parameters. $F_{p,j}$ is a function of hypocenter location and origin time, and $F_{a,j}$ is a function of location, origin time and magnitude. Note that $F_{a,j}$ is a function of origin time because it is calculated from different attenuation relations depending on the state of station j (no wave arrived, P-wave arrived or S-wave arrived), which is controlled by the origin time of an earthquake. Based on this model and the assumption that the predictions of the station data are independent of each other, the likelihood function $p(\mathcal{F}_t|\theta_l)$ in Equation 7 is expressed as:

$$p(\mathcal{F}_t|\theta_l) = \prod_{j=1}^{N_{st}} [p(F_{p,j}(t)|\theta_l)p(F_{a,j}(t)|\theta_l)] \tag{8}$$

For the likelihood model given next in Section 3.3.2, there is no analytical solution for the posterior PDF, so a numerical method is required. Here, we use importance sampling because it is computationally efficient for real-time running of EEW. Based on the structure shown in Figure 2, however, we observe that magnitude m can be analytically treated without significantly increasing the complexity of the equations. Hence, θ_l is partitioned into $[m, \tilde{\theta}_l]$, where $\tilde{\theta}_l$ includes all the other parameters that will be sampled. A set of sequential proposal PDFs based on using the posterior PDF from the previous time step can be implemented for efficiency in the importance sampling scheme (Section 4.5 shows an example). We therefore

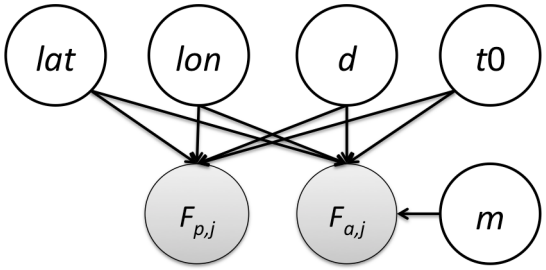

Figure 2. Bayesian network model of probabilistic dependency between the data features $\mathcal{F}_t$ and earthquake parameters Θ_n.

use the RBIS method with a sequential proposal PDF as follows:

$$p(\theta_l|\mathcal{F}_t) = p(m|\tilde{\theta}_l, \mathcal{F}_t)p(\tilde{\theta}_l|\mathcal{F}_t) \approx \sum_{i=1}^{N_s} w_i p(m|\tilde{\theta}_l^{(i)}, \mathcal{F}_t)\delta(\tilde{\theta}_l - \tilde{\theta}_l^{(i)}) \tag{9}$$

where N_s samples $\tilde{\theta}_l^{(i)}$ are drawn randomly from the proposal PDF $q(\tilde{\theta}_l)$ and the corresponding importance weights are:

$$w_i \propto \frac{p(\tilde{\theta}_l^{(i)}|\mathcal{F}_t)}{q(\tilde{\theta}_l^{(i)})} \text{ with } \sum_{i=1}^{N_s} w_i = 1 \tag{10}$$

In order to exploit the information from the previous time step and maintain computational efficiency, we choose $q(\tilde{\theta}_l^{(i)})$ to be a uniform distribution with a domain that contains the high probability regions in $p(\tilde{\theta}_l^{(i)}|\mathcal{F}_{t-1})$. By using a uniform distribution for both the prior $p(\theta_l)$ and the proposal PDF $q(\theta_l)$, the weights w_i simplify to:

$$w_i \propto p(\mathcal{F}_t|\tilde{\theta}_l^{(i)}) \tag{11}$$

$$\text{where } p(\mathcal{F}_t|\tilde{\theta}_l^{(i)}) = \int p(\mathcal{F}_t|\theta_l^{(i)})p(m)\,\mathrm{d}m,\ \theta_l^{(i)} = [\tilde{\theta}_l^{(i)}, m] \tag{12}$$

3.3.2 Gaussian likelihood implementation

A Gaussian model is chosen for predicting the features $F_{p,j}$ and $F_{a,j}$ because of the resulting simplicity in the EEW algorithm:

(a) Gaussian likelihood for pick time:

$$p(F_{p,j}(t)|\theta_l) = \frac{1}{\sigma_{p,j}(\theta_l)}\phi\left(\frac{F_{p,j}(t) - \mu_{p,j}(\theta_l)}{\sigma_{p,j}(\theta_l)}\right) \tag{13}$$

where $\mu_{p,j}(\theta_l)$ = theoretical P-wave arrival time at station j given θ_l calculated using a simple 1D velocity structure (Ueno et al. 2002) and $\sigma_{p,j}(\theta_l)$ = empirical standard deviation for the P-wave arrival time model

(b) Gaussian likelihood for log of maximum displacement amplitude:

$$p(F_{a,j}(t)|\theta_l) = \frac{1}{\sigma_{a,j}(\theta_l)}\phi\left(\frac{F_{a,j}(t) - \mu_{a,j}(\theta_l)}{\sigma_{a,j}(\theta_l)}\right) \tag{14}$$

where $\mu_{a,j}(\theta_l)$ = attenuation equation for $F_{a,j}$ at station j given θ_l and $\sigma_{a,j}(\theta_l)$ = empirical standard deviation for the attenuation model

In general, a Gaussian likelihood model with either a Gaussian prior or an uniform prior will result in a Gaussian posterior, allowing analytical Bayesian updating, if the mean of the likelihood function is a linear function of the model parameters. However, in the case of EEW, $\mu_{a,j}(\theta_l)$ is usually a nonlinear function of the parameters θ_l, so an analytical solution cannot be obtained. We therefore use the RBIS approximate numerical method. Here, we use the JMA ground motion prediction equations (GMPE) for the P- and S-waves (Kamigaichi, 2004; JMA, 2010, 2012) to define $\mu_{a,j}(\theta_l)$ and $\sigma_{a,j}(\theta_l)$. Let S_s be the set of indices of S-wave arrived stations, S_p be the set of indices of P-wave arrived stations, and S_n be the set of indices of no wave-arrival stations, then the mean of the log of the maximum displacement amplitude is expressed as:

$$\mu_{a,j} = \begin{cases} \mu_{noise,j} & j \in S_n \\ 0.72m - 1.2\log R_j - 0.0005R_j + 0.005d - 0.46 & j \in S_p \\ 0.87m - \log R_j - 0.0019R_j + 0.005d - 0.98 & j \in S_s \end{cases} \tag{15}$$

Here, R_j is the hypocenter-to-station distance (km) as a function of lat, lon and d; $\mu_{noise,j}$ and $\sigma^2_{a,j}(j \in S_n)$ are the mean and variance of the station noise, which are determined empirically based on the historical data for station j. Note that $t0$ plays a role in deciding the set that station j belongs to based on a travel-time model.

3.3.3 Analytical expressions for the algorithm

Although a fully analytical solution is not available, the chosen likelihood function has a mean that depends linearly on earthquake magnitude m given the other parameters $\tilde{\theta}_l$ (see Equation 15). Hence, one can apply the idea of RBIS to m to obtain a partial analytical solution that improves both the accuracy and efficiency of the numerical scheme.

EEW requires the mean (expected value) and standard deviation of the earthquake parameters given the current data set, i.e., $E[\theta_l|\mathcal{F}_t]$ and $Var[\theta_l|\mathcal{F}_t]$, in order to release appropriate warnings. Note that $Var[\theta_l] = E[\theta_l^2] - (E[\theta_l])^2$ and based on Equation 9, for any function

$f(\theta_l)$:

$$\begin{aligned} E[f(\theta_l)|\mathcal{F}_t] &= \int f(\theta_l) p(\theta_l|\mathcal{F}_t)\,\mathrm{d}\theta_l \\ &\approx \int f(\theta_l) \sum_{i=1}^{N_s} w_i p(m|\tilde{\theta}_l^{(i)}, \mathcal{F}_t)\delta(\tilde{\theta}_l - \tilde{\theta}_l^{(i)})\,\mathrm{d}\theta_l \\ &= \sum_{i=1}^{N_s} w_i \int f(m, \tilde{\theta}_l^{(i)}) p(m|\tilde{\theta}_l^{(i)}, \mathcal{F}_t)\,\mathrm{d}m \end{aligned} \tag{16}$$

We can therefore derive the following expressions:

$$\begin{aligned} E[\tilde{\theta}_l|\mathcal{F}_t] &= \sum_{i=1}^{N_s} w_i \tilde{\theta}_l^{(i)} \\ Var[\tilde{\theta}_l|\mathcal{F}_t] &= \sum_{i=1}^{N_s} w_i (\tilde{\theta}_l^{(i)})^2 - (E[\tilde{\theta}_l|\mathcal{F}_t])^2 \end{aligned} \tag{17}$$

$$\begin{aligned} E[m|\mathcal{F}_t] &= \sum_{i=1}^{N_s} w_i \int m\, p(m|\tilde{\theta}_l^{(i)}, \mathcal{F}_t)\,\mathrm{d}m \\ Var[m|\mathcal{F}_t] &= \sum_{i=1}^{N_s} w_i \int m^2 p(m|\tilde{\theta}_l^{(i)}, \mathcal{F}_t)\,\mathrm{d}m - (E[m|\mathcal{F}_t])^2 \end{aligned} \tag{18}$$

As a result, we need analytical expressions for $p(m|\tilde{\theta}_l^{(i)}, \mathcal{F}_t)$ and w_i (or $p(\mathcal{F}_t|\tilde{\theta}_l^{(i)})$). Then we can also derive an analytical expression for $p(\mathcal{F}_t|M_l)$ in Equation 5 or 6. For notational simplicity, l is omitted whenever (i) occurs since the sample index i is always linked with the earthquake index l:

(a) *Analytical form for* $p(m|\tilde{\theta}^{(i)}, \mathcal{F}_t)$

Note that m only depends on $F_{a,j}$ for $j = 1, ..., N_{st}$, i.e., $p(m|\tilde{\theta}^{(i)}, \mathcal{F}_t) = p(m|\tilde{\theta}^{(i)}, F_{a,1}, ..., F_{a,N_{st}})$. Again, by Bayes' Theorem and the independent data assumption:

$$p(m|\tilde{\theta}^{(i)}, F_{a,1}, ..., F_{a,N_{st}}) \propto p(F_{a,1}, ..., F_{a,N_{st}}|\theta^{(i)})p(m) = p(m)\prod_{j=1}^{N_{st}} p(F_{a,j}|\theta^{(i)}) \tag{19}$$

Applying Equation 14 and re-arranging Equation 15, we can rewrite the expression for $p(F_{a,j}|\theta^{(i)})$

to be:

$$p(F_{a,j}|\theta^{(i)}) = \begin{cases} \frac{1}{\sigma_{a,j}^{(i)}}\phi\left(\frac{F_{a,j}-\mu_{noise,j}}{\sigma_{a,j}^{(i)}}\right) & j \in S_n \\ \frac{1}{\sigma_{a,j}^{(i)}}\phi\left(\frac{m-f_{JMA,j}^{(i)}}{\alpha_j^{(i)}\sigma_{a,j}^{(i)}}\right) & j \in S_p \cup S_s \end{cases}$$

$$\text{where } f_{JMA,j}^{(i)} = \begin{cases} (F_{a,j} + 1.2\log R^{(i)} + 0.0005R^{(i)} - 0.005d^{(i)} + 0.46)/0.72 & j \in S_p \\ (F_{a,j} + \log R^{(i)} + 0.0019R^{(i)} - 0.005d^{(i)} + 0.98)/0.87 & j \in S_s \end{cases} \quad (20)$$

$$\text{and } \alpha_j^{(i)} = \begin{cases} 1/0.72 & j \in S_p \\ 1/0.87 & j \in S_s \end{cases}$$

Here, $R^{(i)}$ and $d^{(i)}$ are obtained from $\tilde{\theta}^{(i)}$ and $\sigma_{a,j}^{(i)}$ depends on sample $\tilde{\theta}^{(i)}$ and station j. Note that in actual implementation for the case study, a bias correction of 0.3 is added to $f_{JMA,j}^{(i)}$ if the term is calculated from Hi-net stations following Yamada et al. (2014).

By the well-known property of product of Gaussian functions, we can derive the following expression:

$$\prod_{j\in S_p\cup S_s} \frac{1}{\sigma_{a,j}^{(i)}}\phi\left(\frac{m-f_{JMA,j}^{(i)}}{\alpha_j^{(i)}\sigma_{a,j}^{(i)}}\right) = \frac{z^{(i)}}{\sigma_m^{(i)}}\phi\left(\frac{m-\mu_m^{(i)}}{\sigma_m^{(i)}}\right)$$

$$\text{where } \left(\sigma_m^{(i)}\right)^2 = \left(\sum_{j\in S_p\cup S_s} \frac{1}{\left(\alpha_j^{(i)}\sigma_{a,j}^{(i)}\right)^2}\right)^{-1}, \ \mu_m^{(i)} = \left(\sigma_m^{(i)}\right)^2 \left(\sum_{j\in S_p\cup S_s} \frac{f_{JMA,j}^{(i)}}{\left(\alpha_j^{(i)}\sigma_{a,j}^{(i)}\right)^2}\right) \quad (21)$$

$$\text{and } z^{(i)} = \frac{\sqrt{2\pi\left(\sigma_m^{(i)}\right)^2}}{\prod_{j\in S_p\cup S_s}\sqrt{2\pi\left(\sigma_{a,j}^{(i)}\right)^2}} \exp\left[\frac{1}{2}\left(\frac{\left(\mu_m^{(i)}\right)^2}{\left(\sigma_m^{(i)}\right)^2} - \sum_{j\in S_p\cup S_s}\frac{\left(f_{JMA,j}^{(i)}\right)^2}{\left(\alpha_j^{(i)}\sigma_{a,j}^{(i)}\right)^2}\right)\right]$$

As a result, we conclude that:

$$\prod_{j=1}^{N_{st}} p(F_{a,j}|\theta^{(i)}) = \frac{z^{(i)}}{\sigma_m^{(i)}}\phi\left(\frac{m-\mu_m^{(i)}}{\sigma_m^{(i)}}\right) \prod_{j\in S_n}\frac{1}{\sigma_{a,j}^{(i)}}\phi\left(\frac{F_{a,j}-\mu_{noise}}{\sigma_{a,j}^{(i)}}\right) \quad (22)$$

(b) *Analytical form for* $p(\mathcal{F}_t|\tilde{\theta}^{(i)})$

First, we find an analytical expression for $p(\mathcal{F}_t|\theta^{(i)})$ by combining Equation 8, 13 and 22:

$$p(\mathcal{F}_t|\theta^{(i)}) = \frac{z^{(i)}}{\sigma_m^{(i)}}\phi\left(\frac{m-\mu_m^{(i)}}{\sigma_m^{(i)}}\right) \prod_{j\in S_n}\frac{1}{\sigma_{a,j}^{(i)}}\phi\left(\frac{F_{a,j}-\mu_{noise,j}}{\sigma_{a,j}^{(i)}}\right) \prod_{j=1}^{N_{st}}\frac{1}{\sigma_{p,j}^{(i)}}\phi\left(\frac{F_{p,j}-\mu_{p,j}^{(i)}}{\sigma_{p,j}^{(i)}}\right) \quad (23)$$

where $\mu_{p,j}^{(i)}$ and $\sigma_{p,j}^{(i)}$ depend on sample $\tilde{\theta}^{(i)}$ and station j.

Combining Equation 12 and 23:

$$p(\mathcal{F}_t|\tilde{\theta}^{(i)}) = z^{(i)} \prod_{j\in S_n} \frac{1}{\sigma_{a,j}^{(i)}} \phi\left(\frac{F_{a,j} - \mu_{noise,j}}{\sigma_{a,j}^{(i)}}\right) \prod_{j=1}^{N_{st}} \frac{1}{\sigma_{p,j}^{(i)}} \phi\left(\frac{F_{p,j} - \mu_{p,j}^{(i)}}{\sigma_{p,j}^{(i)}}\right) \int \frac{p(m)}{\sigma_m^{(i)}} \phi\left(\frac{m - \mu_m^{(i)}}{\sigma_m^{(i)}}\right) \mathrm{d}m \tag{24}$$

(c) *Choice of prior $p(m)$*

The integral in Equation 24 can be solved analytically only for some specific choices of the prior $p(m)$, for example, a Gaussian prior or a uniform prior. One common choice is a prior that derives from the well-known Gutenberg-Richter law, which is used, for example, in the PRESTo approach (Satriano et al. 2011; Zollo et al. 2009). Based on our empirical study for Japan, this prior results in a consistent underestimation of the magnitude when warnings are necessary, and so a uniform prior is chosen instead:

$$p(m) = \begin{cases} 0 & m < a \;\&\; m > b \\ 1/(b-a) & a \le m \le b \end{cases} \tag{25}$$

where the parameters are chosen to be $a = 0$ and $b = 10$ because this is the likely range of earthquake magnitudes. As a result, the integral in Equation 24 can be expressed in terms of the Gaussian PDF and CDF to obtain:

$$\begin{aligned} p(\mathcal{F}_t|\tilde{\theta}^{(i)}) &= \frac{z^{(i)} c_0^{(i)}}{b-a} \prod_{j\in S_n} \frac{1}{\sigma_{a,j}^{(i)}} \phi\left(\frac{F_{a,j} - \mu_{noise,j}}{\sigma_{a,j}^{(i)}}\right) \prod_{j=1}^{N_{st}} \frac{1}{\sigma_{p,j}^{(i)}} \phi\left(\frac{F_{p,j} - \mu_{p,j}^{(i)}}{\sigma_{p,j}^{(i)}}\right) \\ \text{where } c_0^{(i)} &= \Phi\left(\frac{b - \mu_m^{(i)}}{\sigma_m^{(i)}}\right) - \Phi\left(\frac{a - \mu_m^{(i)}}{\sigma_m^{(i)}}\right) \end{aligned} \tag{26}$$

This constant $c_0^{(i)}$ is due to the prior $p(m)$ truncating the Gaussian PDF for m outside $[a, b]$. Note that in most cases, $c_0^{(i)} \approx 1$ because the Gaussian PDF with $\mu_m^{(i)}$ and $\sigma_m^{(i)}$ should have most of its density between the chosen prior values a and b. Also, if S_p and S_s are both a null set $\emptyset$, the factor $\frac{z^{(i)} c_0^{(i)}}{b-a}$ is deleted from Equation 26.

(d) *Analytical solution for $\int m\, p(m|\tilde{\theta}^{(i)}, \mathcal{F}_t)\, \mathrm{d}m$ and $\int m^2 p(m|\tilde{\theta}^{(i)}, \mathcal{F}_t)\, \mathrm{d}m$*

These two integrals in Equation 18 are simply the first and second moment of the uniform PDF if S_p and S_s are both a null set $\emptyset$, or the first and second moment of the truncated

Gaussian PDF otherwise. Their analytical expressions can be written as follow:

$$\begin{aligned}
&\int m\, p(m|\tilde{\theta}^{(i)}, \mathcal{F}_t)\, \mathrm{d}m \equiv \tilde{\mu}_m^{(i)} \\
&\qquad = \begin{cases} \frac{a+b}{2}, & S_p \cup S_s = \emptyset \\ \mu_m^{(i)} + \frac{\sigma_m^{(i)}}{c_0^{(i)}} \left[\phi\left(\frac{a-\mu_m^{(i)}}{\sigma_m^{(i)}} \right) - \phi\left(\frac{b-\mu_m^{(i)}}{\sigma_m^{(i)}} \right) \right], & \text{otherwise} \end{cases} \\
&\int m^2 p(m|\tilde{\theta}^{(i)}, \mathcal{F}_t)\, \mathrm{d}m = \left(\tilde{\mu}_m^{(i)}\right)^2 + \left(\tilde{\sigma}_m^{(i)}\right)^2 \\
&\text{where } \left(\tilde{\sigma}_m^{(i)}\right)^2 = \begin{cases} \frac{(b-a)^2}{12}, & S_p \cup S_s = \emptyset \\ \left(\frac{\sigma_m^{(i)}}{c_0^{(i)}}\right)^2 \tilde{c}^{(i)}, & \text{otherwise} \end{cases}
\end{aligned} \tag{27}$$

$$\begin{aligned}
\tilde{c}^{(i)} = &\left(c_0^{(i)}\right)^2 + \frac{c_0^{(i)}(a-\mu_m^{(i)})}{\sigma_m^{(i)}} \phi\left(\frac{a-\mu_m^{(i)}}{\sigma_m^{(i)}} \right) - \frac{c_0^{(i)}(b-\mu_m^{(i)})}{\sigma_m^{(i)}} \phi\left(\frac{b-\mu_m^{(i)}}{\sigma_m^{(i)}} \right) \\
&- \left(\phi\left(\frac{a-\mu_m^{(i)}}{\sigma_m^{(i)}} \right) - \phi\left(\frac{b-\mu_m^{(i)}}{\sigma_m^{(i)}} \right) \right)^2
\end{aligned} \tag{28}$$

Finally, we can obtain an analytical expression for the evidence for the earthquake event M_l, i.e., $p(\mathcal{F}_t|M_l)$ in Equation 5 or $p(F_j(t)|M_l)$ in Equation 6 (depending on the choice of criterion), based on the samples and weights from RBIS (Newton 1994). From our empirical study, we found that $p(F_j(t)|\hat{\theta}_l)$, where $\hat{\theta}_l$ is the optimal value of θ_l that maximizes $p(\theta_l|\mathcal{F}_t)$, is a good estimator for $p(F_j(t)|M_l)$. Hence, to further reduce computational effort, we used the following criterion in our case study:

$$p(F_j(t)|\hat{\theta}_l) < \tau_{new} \text{ for newly triggered station } j \;\&\; \forall l = 1, ..., \hat{n} \tag{29}$$

4 PRACTICAL IMPLEMENTATION FOR MULTI-EVENTS EEW ALGORITHM

In this section, the theory explained in the previous section is implemented for an actual EEW system. To improve computational efficiency, this practical implementation requires reduction of the sampling space, choice of stations to be used, controlling the number of events and sample updating. An outline of the implementation is summarized at the end of this section.

4.1 Reduction of the sampling space

The earthquake magnitude can be estimated by an analytical solution in the proposed RBIS-based EEW algorithm (see Section 3.3). Hence, the sampling space is shrunk from five to

four dimensions: latitude, longitude, depth and origin time of an earthquake. The number of samples required to accurately represent a PDF usually increases exponentially as the dimension of the sampling space increases. Therefore, reduction of the sampling space is always desired. In general, the uncertainty in origin time is very small given the hypocenter location. This means that the posterior PDF of origin time given the hypocenter location often peaks around one value. Instead of using many samples for the origin time as a separate sampling dimension, we can estimate the peaked PDF by only one sample calculated based on a deterministic model for each sample of hypocenter location. The single sample value is found by minimizing the residual between the P-wave picking time and the theoretical P-wave arrival time from all triggered stations in an event. As a result, we are able to reduce the sampling dimension from five to three (excluding magnitude and origin time).

4.2 Use of non-triggered data for P-wave picking time likelihood model

Most of the earthquake location algorithms only utilize the data from triggered stations and ignore stations without a recorded picking time. However, the fact that a P-wave has not yet arrived at a station is also an important piece of information. Based on the concept of using "not-yet-arrived data" from Horiuchi et al. (2005) and Satriano et al. (2008), we adopt the following rules for calculating $p(F_{p,j}(t)|\theta_l)$ in Equation 13:

(a) When station j has not yet been triggered, but the theoretical arrival time suggests the opposite: the Gaussian variable $F_{p,j}(t)$ equals current time to penalize the estimation. Additionally, $p(F_{p,j}(t)|\theta_l)$ has a lower limit, e.g., 0.004 (represents a three standard deviation error probability under a Gaussian model), in order to avoid a single malfunctioned station that will never be triggered driving the final likelihood value of an event to zero.

(b) When station j has not yet been triggered and the theoretical arrival time suggests the same: $p(F_{p,j}(t)|\theta_l) = 1$, which represents that no information is extracted from this station.

(c) When station j has been triggered: $F_{p,j}(t)$ equals the recorded picking time value and $p(F_{p,j}(t)|\theta_l)$ is directly calculated using Equation 13.

4.3 Choice of stations used in likelihood models

Seismic waves attenuate when traveling through the ground and the uncertainty of the attenuation model accumulates as the traveling distance increases. Hence, information from stations that are far from the hypocenter does not make a significant contribution to estimation of the earthquake parameters. Theoretically, in the absence of noise, four stations are enough to

pinpoint the location and origin time of an earthquake. In practice, the parameters of most earthquakes can be accurately identified based on data from around five to ten stations closest to the hypocenter. However, for offshore earthquakes, because the station distribution is often one-sided, wide azimuth coverage becomes an important factor to accurately estimate the hypocenter location and origin time.

To reduce the computation time, instead of using all stations for every event, $\mathcal{F}_t$ is reduced to include only information from a small subset of stations (e.g., 20 stations for our case study) that are selected based on the first triggered station of an event using the following two-step method:

(a) Select a subset of stations that have the closest distance to the triggered station. For example, in our case study, we pick 10 closest stations to the center of the Voronoi cell of the first triggered station based on the assumption that the hypocenter is most likely to be around that center point without any further information given.

(b) Select the remaining stations one-by-one that contribute to an increase of azimuth coverage from stations that are close to the center of the Voronoi cell of the first triggered station. For example, in our case study, the remaining 10 stations are selected such that each leads to an increase in azimuth coverage. If the first 10 stations have a complete azimuth coverage of 360^o already (often the case for an inland earthquake), the remaining 10 stations will simply be chosen from the closest stations.

Theoretically, exclusion of data from certain stations is equivalent to assuming $p(F_j(t)|\theta_l) \approx 1$ (no important information is extracted) for those stations. Figure 3 shows an example of the resulting station selection for a first triggered station along the coastline and an inland first triggered station.

4.4 Creating, merging and deleting events

To achieve a fast algorithm, we use the approximate model class selection, as shown in Equation 29, instead of the full Bayesian model class selection scheme. For improving robustness of the system, a merging or deleting event criterion that allows decreasing the number of events in $\mathcal{M}_{\hat{n}}$ is added. Each step is summarized as follow:

(a) *Creating based on approximate model class selection*: To choose an equivalent to threshold τ_{new} in the approximate scheme, we create a new event if the current data is outside three standard deviations of the optimal Gaussian likelihood model.

(b) *Deleting based on picking time alignment*: After the triggering of the first few stations

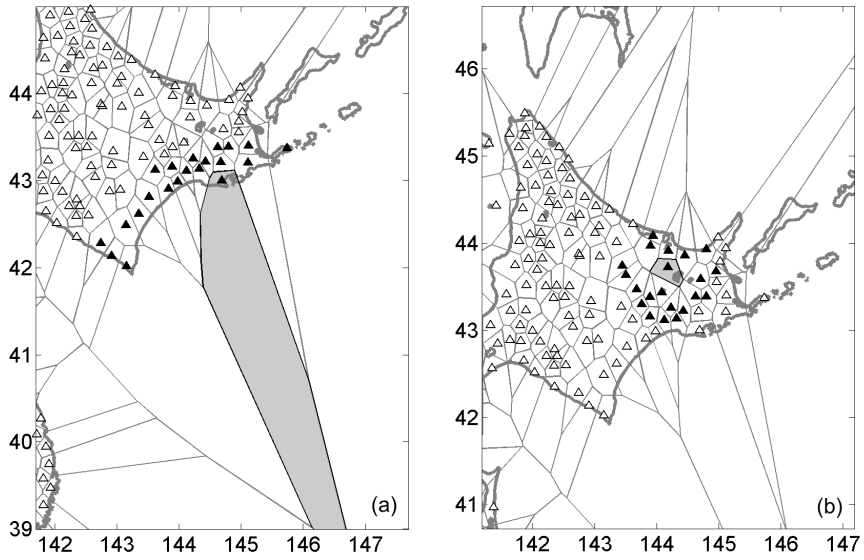

Figure 3. Example of station selection in Japan for (a) coastline and (b) inland first triggered station. Seismic stations are denoted in triangles. Selected stations are filled in black. The gray area represents the Voronoi cell of the first triggered station (the triangle inside the gray area).

of an event (e.g., 7 out of the 20 stations in our case study), the hypocenter and origin time estimates will start to converge. From this moment, if the observed picking times of the stations deviate from the theoretical P-wave arrival times based on the current parameter estimates (e.g., 4 standard deviation under a Gaussian model), this event is likely to be falsely identified and can be deleted from the algorithm.

(c) *Merging based on hypocenter and origin time estimate*: To avoid false alarms due to a duplicated event caused by noisy data, it is beneficial to maintain a unique set of event records in the algorithm. Two events are merged if their converged hypocenter and origin time estimates are reasonably close (e.g., two events are within 10km radius and the expected origin times are less than three seconds apart in our case study). In some cases, even if the two events are actually not identical events, it is beneficial to merge them within the EEW system in order to avoid issuing a confusing warning.

4.5 Prior, proposal PDF and sample updating

For the importance sampling, we draw initial samples from a uniform prior distribution in a three-dimensional hypocenter space (latitude, longitude and depth). The size of the sampled area is chosen to cover the Voronoi cell of the first triggered station, and the depth range is between 0km to 100km. In order to maintain the efficiency of the algorithm, the prior for both

latitude and longitude is limited to a maximum width of one degree from the first triggered station. However, for offshore earthquakes, the optimal solution may be outside this prior and it requires some treatment to achieve a fast and robust convergence to the optimal solution. The proposed algorithm therefore estimates earthquake parameters based on a set of samples drawn from a proposal PDF, which is constructed initially based on the prior PDF when an event is first created. After that, to construct the proposal PDF based on information used in the posterior PDF of the earthquake parameters from the previous time step, we divide the procedure into two cases:

(a) If the distance between the mean epicenter location at the current and the previous time step exceeds some predetermined ratio of the range of the prior on the latitude and longitude, the new proposal PDF for the epicenter is constructed by shifting the center of the uniform prior to the mean of the epicenter obtained from the current time step (Figure 4).

(b) Otherwise, a new proposal PDF is constructed based on a resampling scheme (Liu & Yamada 2014) for better convergence of the hypocenter estimate.

Case (b) is for reducing the variance of importance sampling, and so to improve the efficiency of the algorithm. Most of the inland earthquakes fall into this case. Case (a) is designed to cover the case of an offshore earthquake, since we use a smaller prior to maintain computational efficiency. The new proposal PDF allows fast convergence to the actual epicenter location when it is away from the area of the current samples. For the criterion in case (a), we choose a ratio of 50% for both latitude and longitude based on empirical experience considering a trade-off between the efficiency of importance sampling convergence and the variance of the expected value after new data is collected.

4.6 Algorithm summary

Actual implementation of the method can be summarized as a two-step algorithm at each given time step t (Figure 5). Starting from an initial step $t = 0$, first, earthquake parameters of each existing event are updated based on the newly received seismic data from the network. Second, the estimate of the number of concurrent events ($\hat{n}$) is updated by the approximate model class selection scheme with the predetermined creating, merging and deleting criteria. The process is repeated until the termination of the process.

Step 1: To update the earthquake parameters of each existing event n, we repeat the following steps for $n = 1$ to $\hat{n}$:

(a) Extract information from the waveform data to compute the features used in event n.

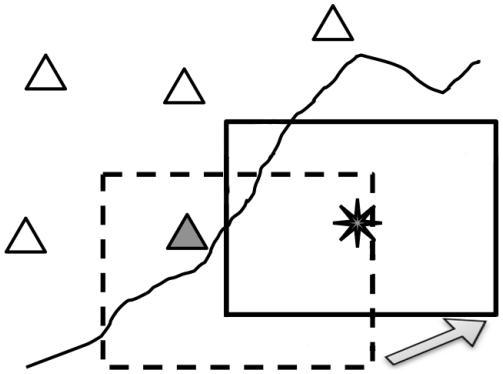

Figure 4. Example of shifting the proposal PDF in Case (a). Triangles are seismic stations. Solid triangle is the triggered station for a potential new event. Dotted rectangle is the uniform prior PDF with a mean epicenter estimated at the black star. The new uniform proposal PDF is shifted to the solid rectangle.

(b) Update weights for the new feature values as in Section 3.3.

(c) Construct a proposal PDF by the suggested method in Section 4.5 to update samples if necessary.

(d) Update all earthquake parameter estimates using the RBIS method, as explained in Section 3.3.

Step 2: After updating the estimates, we compute the most probable number of concurrent events by the approximate model class selection scheme (see Section 4.4):

(a) Merge existing events (reduce number of concurrent events) that have similar converged estimates of earthquake parameters.

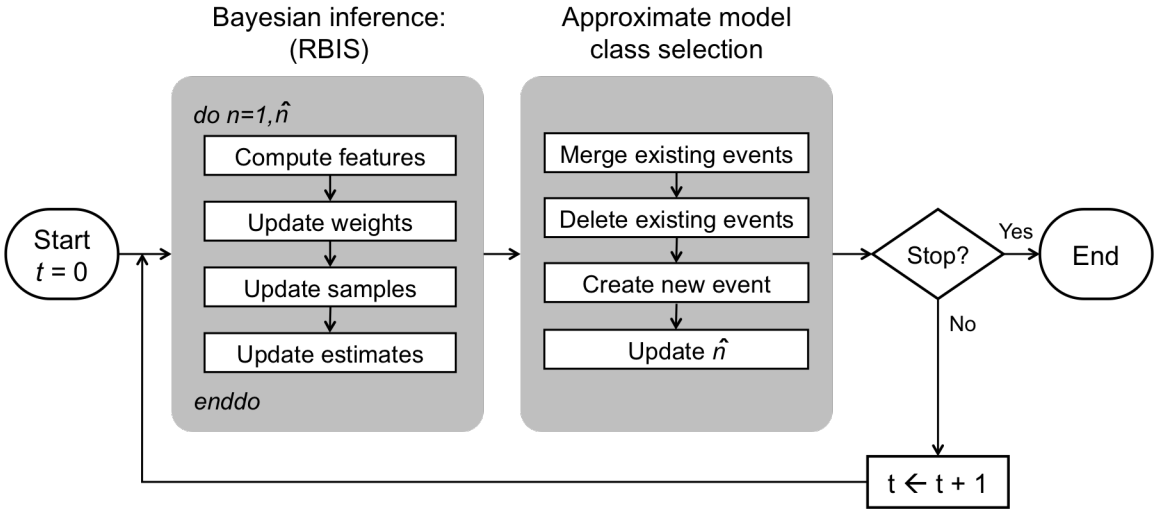

Figure 5. Flowchart of a two-step RBIS EEW algorithm at each time step t.

Table 1. Information of the Tohoku foreshock on March 9, 2011.

Date	Origin Time	Latitude	Longitude	Depth(km)	Magnitude
2011/3/9	11:45:13	38.33	143.28	8.28	M7.3

(b) Delete existing events (reduce number of concurrent events) that have inconsistent theoretical P-wave arrival times comparing to the observed P-wave picking times from the current data.

(c) Create a new event (increase number of concurrent events) with a newly triggered station as the theoretical first triggered station for the event when the criterion is met.

(d) Finally, we obtain the updated most probable number of concurrent events, $\hat{n}$.

5 CASE STUDY RESULTS AND DISCUSSION

5.1 Performance of a single event

First, we illustrate the performance of the proposed algorithm for a single large event on March 9, 2011. This event occurred in the Pacific ocean off the Tohoku region in Japan and it is considered as a foreshock of the Tohoku event on March 11. Table 1 shows the information of this event.

Figure 6 shows three snapshots of the distribution of weighted samples for this event. After the earthquake started at 11:45:13, its P-wave first arrived at the coastline 28s later and a new event was triggered in the system at 11:45:41. The first set of samples were created using the prior based on Voronoi cell information (Figure 6(a)). Then, the samples migrate toward the actual epicenter after 1s (Figure 6(b)), and the samples converge stably 5s after the event is triggered (Figure 6(c)). The posterior distribution represented by the weighted samples is actually very peaked around the actual epicenter.

Figure 7 shows the two components that sum up to the final weight in log-scale: the P-wave picking time (F_p) likelihood and the maximum displacement amplitude (F_a) likelihood. One can observe that when the event is first triggered, the scale of the picking time likelihood is smaller compared to a few seconds later. At this early stage, there is not enough picking time information from the stations to fully constrain the hypocenter location. The maximum displacement amplitude likelihood helps refine the convergence to a smaller region. After that, the picking time likelihood becomes reliable enough to dominate the weight contribution and the amplitude likelihood is no longer so important.

Figure 8 shows a time history of the five earthquake parameters (latitude and longitude

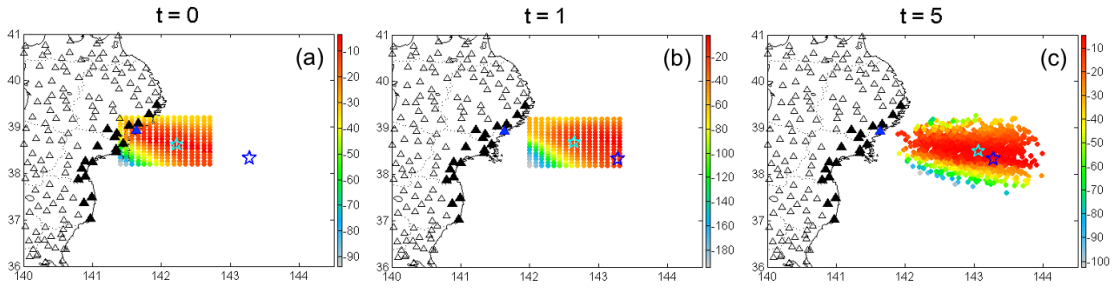

Figure 6. Weighted-sample distribution for the foreshock on March 9, 2011 at (a) $t = 0$s, (b) $t = 1$s and (c) $t = 5$s after the first trigger. Color of the samples is proportional to ln(weight) with higher value in red and lower value in blue. Dark and light blue stars indicate the actual and estimated epicenter location, respectively. Open black triangles are seismic stations. Solid ones are selected stations for this event, and the blue triangle is the first triggered station.

are combined to calculate the error of epicentral distance R). The algorithm takes 30s–40s to completely converge because it is an offshore event. The depth estimate is not as good as the other parameters, as expected, and its uncertainty is also significantly higher than the others.

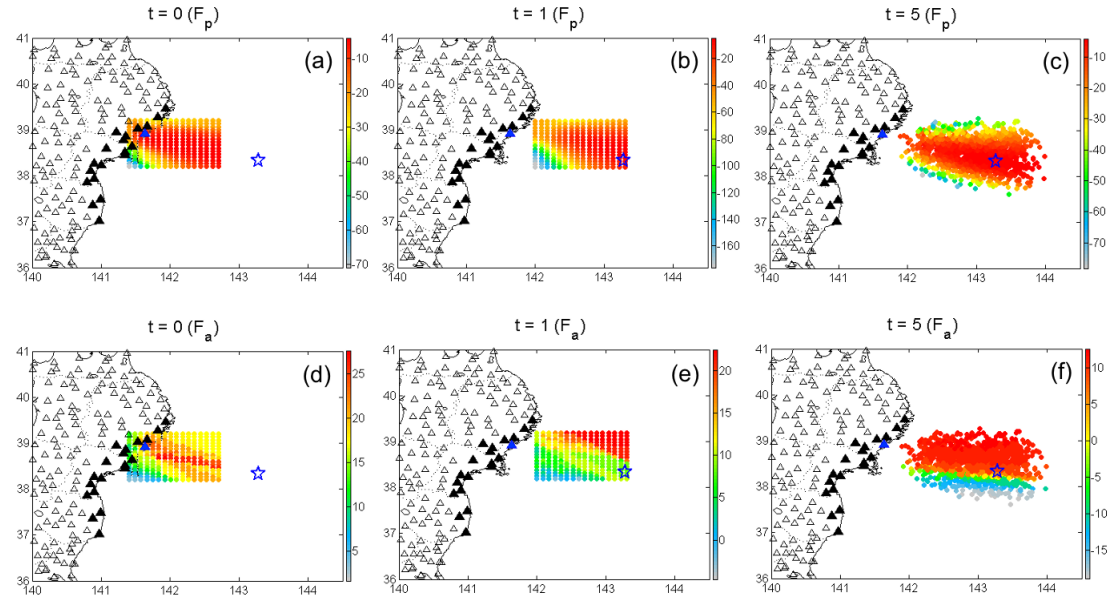

Figure 7. Distribution of the P-wave picking time (F_p) likelihood at (a) $t = 0$s, (b) $t = 1$s and (c) $t = 5$s and the maximum displacement amplitude (F_a) likelihood at (a) $t = 0$s, (b) $t = 1$s and (c) $t = 5$s of the sample weights for the foreshock on March 9, 2011. The symbols are in the same format as in Figure 6.

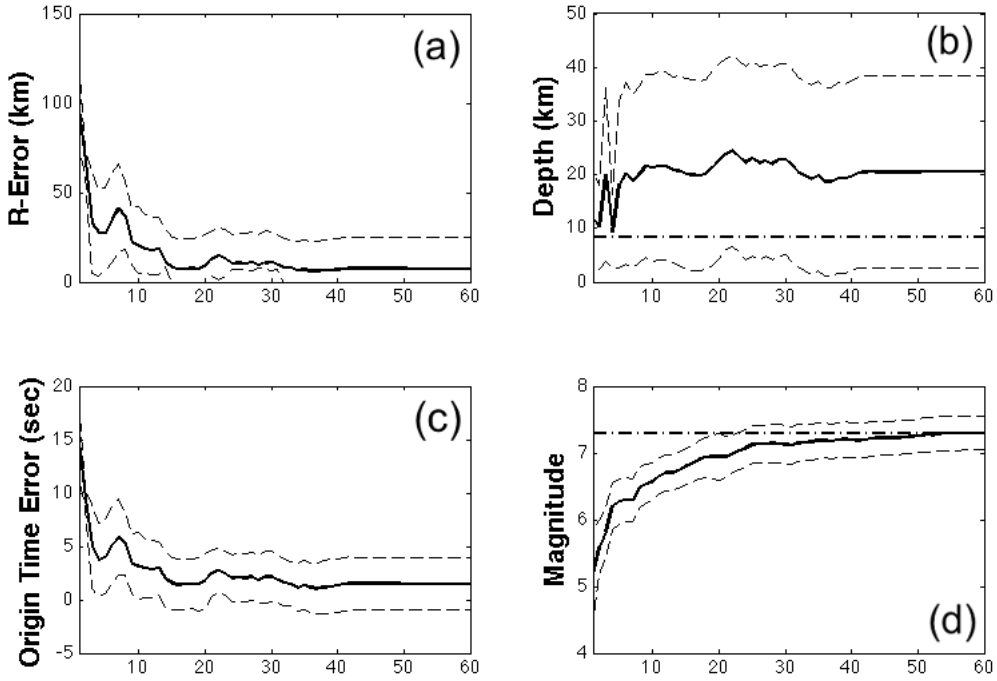

Figure 8. Convergence summary for the foreshock on March 9, 2011. (a) Time histories of the epicentral distance error between the estimated epicenter and the actual epicenter, (b) depth, (c) origin time error and (d) magnitude. The solid and dashed lines are the mean and ±1 std. values, respectively. The horizontal dash-dotted lines are actual parameter values of the earthquake. The x-axis shows the time (sec) after the event is triggered.

5.2 Performance of two overlapped events

Next, we show the results for two concurrent events that overlap with each other within a short period of time. The two events occurred on March 19, 2011 in the Ibaraki Prefecture. The two hypocenters were at the same location and the origin times of the two events were only 28s apart, which makes it very difficult to separate the two events for EEW. Table 2 shows the information of both events.

Figure 9 shows four snapshots of the distribution of weighted samples for the two overlap-

Table 2. Information of the two overlapped concurrent events on March 19, 2011.

Date	Origin Time	Latitude	Longitude	Depth(km)	Magnitude
2011/3/19	18:56:20	36.78	140.57	5.76	M4
2011/3/19	18:56:48	36.78	140.57	5.37	M6.1

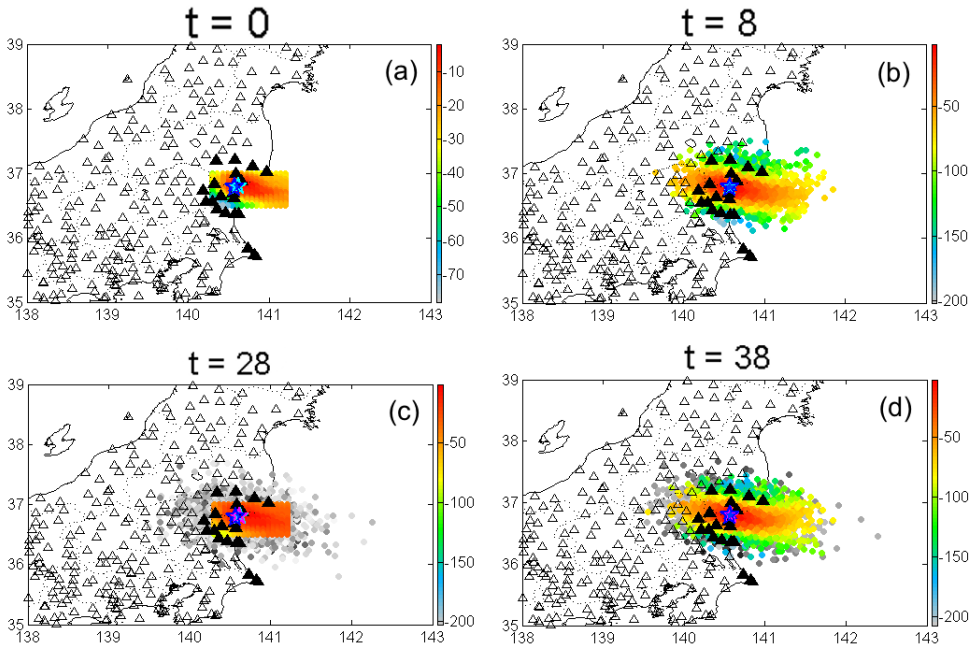

Figure 9. Weighted-sample distribution for the two concurrent events on March 19, 2011. Colored samples in the upper two plots represent the first event. Gray and colored samples in the lower two plots represent the first and second event, respectively. The symbols are in the same format as in Figure 6. The estimated epicenters in (c) and (d) overlap with the actual epicenter.

ping events. The first event started at 18:56:20 and the P-wave arrived at the closest station 3s later (Figure 9(a)). Because it was an inland earthquake, the initial sample set at the time of event trigger was already well converged to the actual epicenter. The samples were completely stable after a few seconds later (Figure 9(b)). The second event started at the exact same location at 18:56:48 and a new event was successfully triggered by the system 3s later. Again, the prior sample set was already well converged at the time the second event was triggered and the samples were completely stable a few seconds later.

Figure 10 shows a time history of the five earthquake parameters. Compared to the single offshore event on March 9 (Figure 8), the errors for all parameters are small when the events are first triggered. This is typical of the fast convergence that is achieved for the case of an inland earthquake. The only problem is that the magnitude estimate of the first event is affected by the seismic waves generated by the second event. After triggering from the second event, the magnitude estimate of the first event converges to the magnitude of the second event. However, no false alarm is triggered for the purpose of EEW because the two events are close enough in time and space. Although one can set a more sophisticated convergence criterion to appropriately stop estimating the parameters for an event 30s after it is triggered,

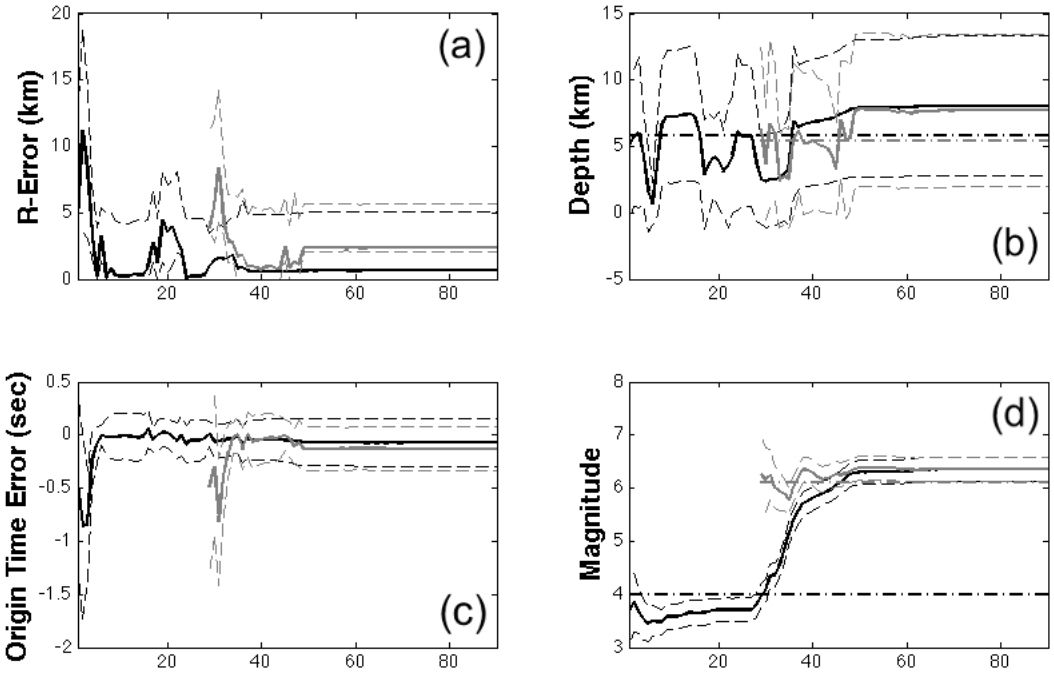

Figure 10. Convergence summary of the new algorithm for the two concurrent events on March 19, 2011. Black lines and gray lines correspond to the first and second event, respectively. The format is the same as in Figure 8.

this is likely to increase the error for many other events. Therefore, it may not be practical to do so for the purpose of EEW.

5.3 Summary of all identified events

After continuously running our proposed algorithm with the 50-day data (March 9 to April 30), a total of 895 earthquakes have been identified. However, the 10-hour data from 15:00 to 23:59 on March 11 was excluded from the comparison with the catalogue, because a significant increase of seismicity after the mainshock at 14:46 results in an incomplete catalogue which makes it difficult to compare. As a result, we use a total of 850 earthquakes for the comparison. Figure 11 shows the magnitude histogram of the 850 earthquakes. The number of earthquake with magnitude between 3 to 7 follows the Gutenberg-Richter relationship. Therefore, the proposed algorithm can reliably identify M3 or above earthquakes.

Figure 12 shows the histograms of the residuals of latitude, longitude, depth, origin time, magnitude and Japanese seismic intensity, calculated by subtracting the catalogue values from our EEW algorithm estimates. The relatively peaked histograms verify that the proposed

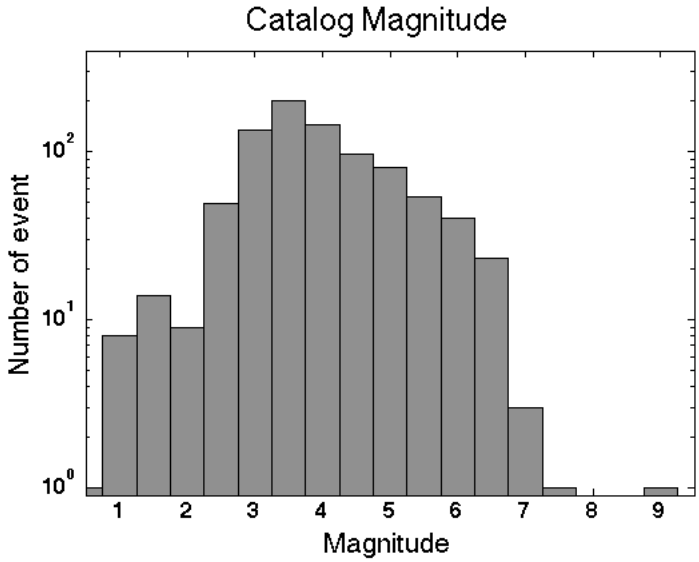

Figure 11. Histogram of the catalogue magnitudes for the 850 earthquakes between March 9 and April 30 (plotted in log-scale).

algorithm can predict the earthquake parameters accurately. A relatively higher variance is observed for longitude estimation of some events. Most of them are due to offshore events, where there is a lack of seismic station coverage along the longitudinal direction in the Tohoku region. Also, due to the low sensitivity of travel time to depth, the depth shows the largest variance among the three components of the hypocenter location estimate. It may appear in the histogram that the magnitude estimates have a slight bias towards overestimation. To check this, we show the comparison between catalogue and estimated magnitude in Figure 13. It shows that some of the earthquakes with magnitude $2.5-4.5$ are largely overestimated. This may be because the signal to noise ratio is low for smaller earthquakes, especially for the long period component. In terms of EEW seismic intensity, the proposed algorithm demonstrates a sufficient performance based on the JMA standard, where ± 1 unit of error for EEW seismic intensity estimates is expected (JMA, 2005; Kamigaichi et al., 2009).

5.4 Comparison with JMA EEW

There were 71 warnings released by the JMA EEW system within the period of March 11 to April 30 (Tamaribuchi et al. 2014). We compare the accuracy of the proposed algorithm with that of the existing JMA EEW system using Japanese seismic intensity. Warnings are released to the public if the expected intensity is greater than or equal to $5-$. Appendix A lists the details of all 71 events, including the results of the seismic intensity estimates from

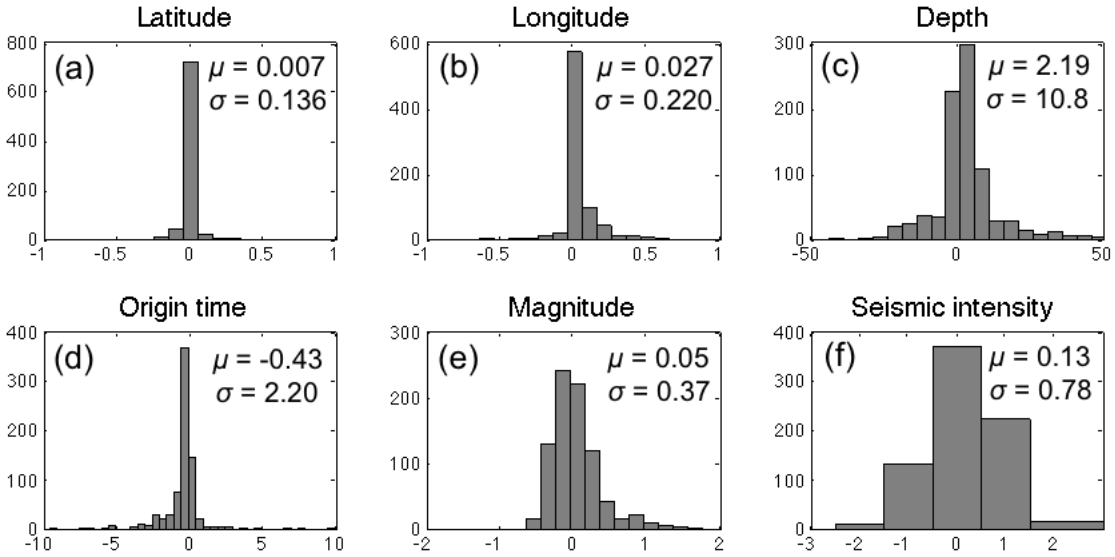

Figure 12. Residual histogram of (a) latitude (deg), (b) longitude (deg), (c) depth (km), (d) origin time (sec), (e) magnitude and (f) Japanese seismic intensity. X-axis shows the residual calculated by: new EEW algorithm estimates – Catalogue values. Y-axis shows the number of earthquakes. μ = mean and σ = standard deviation.

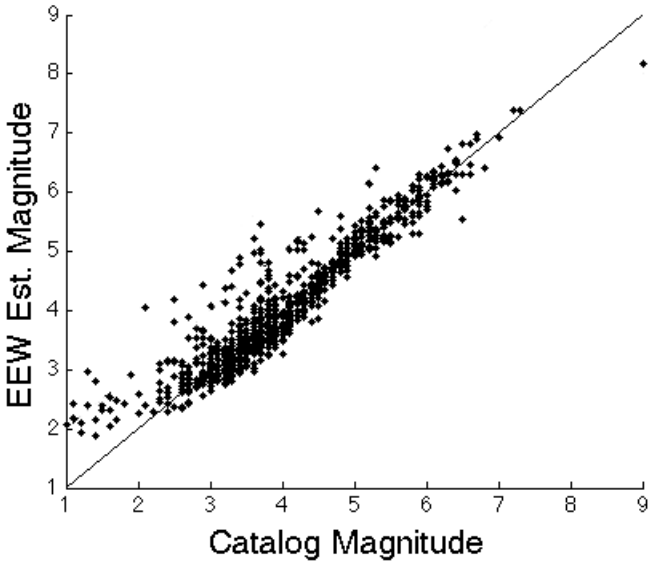

Figure 13. Plot of EEW estimated magnitudes (final) versus catalogue magnitudes. The solid black line indicates the case of perfect prediction.

the catalogue, JMA EEW system, an algorithm proposed by Tamaribuchi et al. (2014) and the proposed algorithm in this paper.

Our algorithm prevents over 90% of the false alarm cases produced by the existing JMA EEW. There are only three cases of incorrect warnings produced by our algorithm. In one case, the algorithm incorrectly deleted an actual event due to the interference of multiple aftershocks around that period (Event 2 had an accurate seismic intensity estimate, but the event was incorrectly deleted after a few seconds). This suggests a more sophisticated deleting criterion is desirable. In another case, there were two concurrent earthquakes that were too close in time and space, as explained in Section 5.2 (Event 34). In the final case, the algorithm produced accurate estimates of the earthquake parameters, but due to a deficiency in the attenuation relationship, an underestimate of the maximum seismic intensity was produced (Event 46).

Figure 14 shows the histograms of seismic intensity residual for the three EEW systems based on the 71 selected earthquakes. It verifies that the proposed algorithm improves the performance of EEW in over 90% of the cases based on seismic intensity compared to the existing JMA EEW system, which has a difficulty in handling multi-events. Note that Figure 14 shows that the algorithm proposed in this paper improves the seismic intensity estimate by around 30% compared with the one proposed by Tamaribuchi et al. (2014). In fact, both algorithms use a probabilistic approach to solve the multi-events cases. A major difference is that our algorithm includes both Hi-net and JMA data for estimating all earthquake parameters and we use only two features (P-wave picking time and maximum displacement amplitude), whereas Tamaribuchi et al. (2014) use a total of four features (the outputs of a B-Δ method and principal component analysis, P-wave picking time and maximum displacement amplitude) and JMA data only. Not only does the inclusion of Hi-net data improve the warning lead time (Yamada et al. (2014) shows an average increase of 3.6s for the EEW warning lead time), it is shown here that it also improves the warning accuracy. Although some treatment of the data is needed in order to integrate data from the two different seismic networks (Yamada et al. 2014), the resulting benefits make it worthwhile to do so.

6 CONCLUSION

In order to improve the accuracy of EEW, this study proposes a probability-based EEW algorithm to identify multiple concurrent earthquakes. An approximate method for Bayesian model class selection is applied to solve for the number of concurrent events in the EEW multi-events problem. Because of the short time limitations of EEW, a simple numerical method, the Rao-Blackwellized Importance Sampling with a set of sequential proposal PDFs, is used

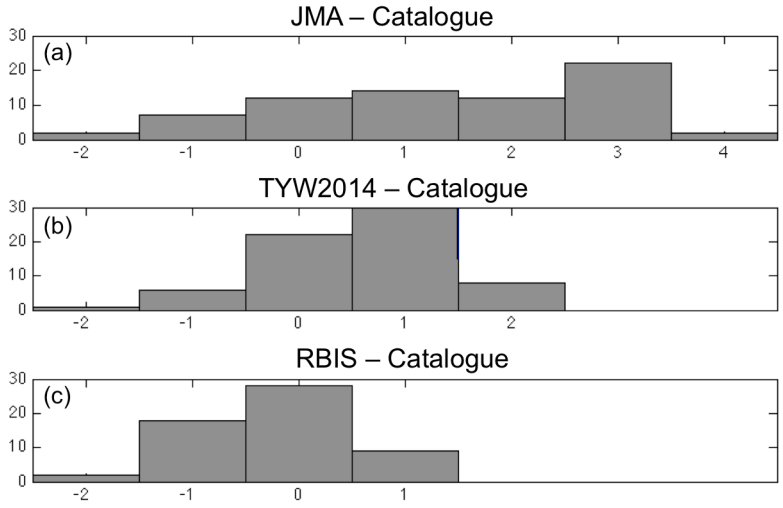

Figure 14. Histogram of the error on the seismic intensity estimate of (a) the existing JMA EEW system; (b) an algorithm proposed by Tamaribuchi et al. (2014) (TYW2014); and (c) our algorithm (RBIS) for the selected 71 earthquakes. The x-axis shows the error of the Japanese seismic intensity calculated by: EEW estimate - Catalogue value.

to estimate the earthquake parameters and the necessary equations are derived analytically as much as possible.

The existing JMA EEW system is used to demonstrate the process of applying the proposed probabilistic method to an existing deterministic EEW model. Two features, the P-wave picking time and the maximum displacement amplitude, are chosen for earthquake parameter estimation. A Gaussian model is used for the likelihood function of both features. A real example based on two months data (March 9 to April 30, 2011) around the time of the March 11, 2011 Tohoku earthquake is studied to verify the proposed algorithm. Over 90% of the false alarms are avoided by our algorithm, which accurately identifies multiple concurrent events in comparison with the existing JMA EEW system. Also, the importance of having a denser network is demonstrated by the examples. The integration of data from the JMA and Hi-net stations in a single algorithm greatly enhances the average warning lead time (around 3.6s) and the accuracy of the EEW predictions (around 30% in seismic intensity).

ACKNOWLEDGMENTS

We would like to thank the Japan Meteorological Agency and NIED for providing the seismic

waveform data during the period of 2011 M9 Tohoku earthquake. This research was supported by the funding program for Next Generation World-Leading Researchers in Japan.

REFERENCES

Allen, R., 1978. Automatic earthquake recognition and timing from single traces, *Bulletin of the Seismological Society of America*, **68**(5), 1521–1532.

Beck, J., 2010. Bayesian system identification based on probability logic, *Structural Control and Health Monitoring*, **17**(7), 825–847.

Horiuchi, S., Negishi, H., Abe, K., Kamimura, A., & Fujinawa, Y., 2005. An automatic processing system for broadcasting earthquake alarms, *Bulletin of the Seismological Society of America*, **95**(2), 708–718.

Hoshiba, M., Kamigaichi, O., Saito, M., Tsukada, S., & Hamada, N., 2008. Earthquake early warning starts nationwide in japan, *EOS Trans. AGU*, **89**, 73–74.

Hoshiba, M., Iwakiri, K., Hayashimoto, N., & Shimoyama, T., 2011. Outline of the 2011 off the pacific coast of tohoku earthquake (mw 9.0) - earthquake early warning and observed seismic intensity, *Earth Planets and Space*, **63**(7), 547–551.

Japan Meteorological Agency (JMA), 2005. Earthquake early warning accuracy evaluation, `http://www.data.jma.go.jp/svd/eew/data/nc/kako/seidohyoka_20051031.pdf`, [Accessed 5 June, 2014].

Japan Meteorological Agency (JMA), 2010. 2nd earthquake early warning meeting (technical), `http://www.seisvol.kishou.go.jp/eq/EEW/Meeting_HYOUKA/t02/index.html`, [Accessed 17 September, 2013].

Japan Meteorological Agency (JMA), 2012. 4th earthquake early warning meeting (technical), `http://www.seisvol.kishou.go.jp/eq/EEW/Meeting_HYOUKA/t04/index.html`, [Accessed 17 September, 2013].

Japan Meteorological Agency (JMA), 2013a. 4th earthquake early warning meeting, `http://www.seisvol.kishou.go.jp/eq/EEW/Meeting_HYOUKA/04/index.html`, [Accessed 30 August, 2013].

Japan Meteorological Agency (JMA), 2013b. Earthquake early warning report, `http://www.seisvol.kishou.go.jp/eq/EEW/kaisetsu/joho/joho.html`, [Accessed 16 October, 2013].

Japan Meteorological Agency (JMA) Earthquake Early Warning (EEW), 2014. Earthquake Early Warning historical broadcast, `http://www.data.jma.go.jp/svd/eew/data/nc/pub_hist/index.html`, [Accessed 25 May, 2014].

Kamigaichi, O., 2004. JMA Earthquake Early Warning, *Journal of Japan Association for Earthquake Engineering*, **4**, 134–137.

Kamigaichi, O., Saito, M., Doi, K., Matsumori, T., Tsukada, S., Takeda, K., Shimoyama, T., Nakamura, K., Kiyomoto, M., & Watanabe, Y., 2009. Earthquake early warning in japan: Warning the general public and future prospects, *Seismological Research Letters*, **80**(5), 717–726.

Katsumata, A., 2008. Recursive digital filter with frequency response of a mechanical seismograph, *Quarterly Journal of Seismology*, **71**, 89–91.

Liu, A. & Yamada, M., 2014. Bayesian approach for identification of multiple events in an early warning system, *Bulletin of the Seismological Society of America*, **104**(3), [Accepted].

Liu, J., 2002. *Monte Carlo Strategies in Scientific Computing*, Berlin: Springer-Verlag.

National Research Institute for Earth Science and Disaster Prevention (NIED), 2014. JMA earthquake catalogue, `http://www.hinet.bosai.go.jp/`, [Accessed 10 April, 2014].

Newton, N., 1994. Variance reduction for simulated diffusions, *SIAM Journal on Applied Mathematics*, **54**(6), 1780–1805.

Satriano, C., Lomax, A., & Zollo, A., 2008. Real-time evolutionary earthquake location for seismic early warning, *Bulletin of the Seismological Society of America*, **98**(3), 1482–1494.

Satriano, C., Elia, L., Martino, C., Lancieri, M., Zollo, A., & Iannaccone, G., 2011. PRESTo, the earthquake early warning system for Southern Italy: Concepts, capabilities and future perspectives, *Soil Dynamics and Earthquake Engineering*, **31**(2), 137–153.

Shiomi, K., Obara, K., & Kasahara, K., 2005. Amplitude saturation of the NIED Hi-net waveforms and simple criteria for recognition, *Zisin: Journal of the Seismological Society of Japan, 2nd Series*, **57**(4), 451–461.

Si, H. & Midorikawa, S., 2000. New attenuation relations for peak ground acceleration and velocity considering effects of fault type and site condition, in *Proceedings of 12th World Conference on Earthquake Engineering*, no. 0532, Auckland, New Zeland.

Tamaribuchi, K., Yamada, M., & Wu, S., 2014. A new approach to identify multiple concurrent events for improvement of Earthquake Early Warning, *Zisin2*, [Accepted].

Ueno, H., Hatakeyama, S., Aketagawa, T., Funasaki, J., & Hamada, N., 2002. Improvement of hypocenter determination procedures in the Japan Meteorological Agency, *Quarterly Journal of Seismology*, **65**(1-4), 123–134.

Yamada, M., Liu, A., & Mori, J., 2012. Classification of simultaneous multiple earthquakes for the Earthquake Early Warning system, in *JpGU Annual Meeting*, [Abstract].

Yamada, M., Tamaribuchi, K., & Wu, S., 2014. Faster and more accurate Earthquake Early Warning system - combination of velocity and acceleration-type seismometers, *Journal of Japan Association for Earthquake Engineering*, [Accepted].

Zollo, A., Iannaccone, G., Lancieri, M., Cantore, L., Convertito, V., Emolo, A., Festa, G., Gallovic, F., Vassallo, M., Martino, C., Satriano, C., & Gasparini, P., 2009. The earthquake early warning system in Southern Italy: Methodologies and performance evaluation, *Geophysical Research Letters*, **36**(L00B07).

APPENDIX A: EARTHQUAKE LISTS FOR JMA EEW

Table of the details of 71 earthquakes for which the JMA EEW system issued a warning between March 11 and April 30, 2011. Results of the maximum seismic intensity estimates from the catalogue, JMA EEW system (JMA), an algorithm proposed by Tamaribuchi et al. (2014) (TYW2014) and the proposed algorithm in this paper (RBIS) are listed. If the error of the seismic intensity is within ± 1, or the observed and estimated intensities are both less than 5– (no warning), we define it is an accurate warning. The inaccurate warnings produced by each of the three algorithms are shown by an underline on the seismic intensity values.

No.	Date	Time	Location	Magnitude	Catalog	JMA	TYW2014	RBIS
1	2011-03-11	14:46	Off Miyagi Pref.	9	7	6–	7	6–
2		17:40	Fukushima Pref.	6	5+	5+	5–	5–[b]
3		19:35	Off Fukushima Pref.	5.1	4	5+	4	3
4	2011-03-12	3:11	Off Fukushima Pref.	6	3	5–	5–	4
5		3:59	Niigata Pref.	6.7	6+	6–	6+	6+
6		4:08	Off Ibaraki Pref.	5.2	4	5–	5–	4
7		4:16	Off Fukushima Pref.	4	3	5+	4	0[a]
8		4:31	Niigata Pref.	5.9	6–	5+	6–	5+
9		5:11	Off Miyagi Pref.	6.4	3	5+	4	3
10		5:42	Niigata Pref.	5.3	6–	5–	4	5–
11		6:19	Nagano Pref.	3.7	3	6–	4	3
12		6:34	Off Fukushima Pref.	4.8	3	6+	4	0[a]
13		6:48	E off Chiba Pref.	4.6	3	5–	4	0[a]
14		22:15	Off Fukushima Pref.	6.2	5–	5–	5+	4
15		22:24	Off Iwate Pref.	5	3	5–	0[a]	3
16		22:26	Off Iwate Pref.	5.4	2	5–	4	0[a]
17		23:34	Niigata Pref.	3.7	5–	6+	4	4
18		23:43	Off Iwate Pref.	5.9	4	5–	5–	4
19	2011-03-13	8:25	Off Miyagi Pref.	6.2	5–	5–	5–	4
20		10:26	Off Ibaraki Pref.	6.6	4	5–	5–	5–
21	2011-03-14	10:02	Off Ibaraki Pref.	6.2	5–	5–	5+	5–
22		15:12	Off Fukushima Pref.	5.2	4	6–	4	4
23		16:25	Off Ibaraki Pref.	5	3	6–	3	3
24	2011-03-15	1:36	Tokyo Bay	3.3	2	5–	3	0[a]
25		5:33	E off Chiba Pref.	3.6	1	5+	3	0[a]
26	2011-03-15	7:29	Fukushima Pref.	4.3	3	6+	4	3
27		22:31	Yamanashi Pref.	6.4	6+	5–	6–	5+
28	2011-03-16	2:40	Chiba Pref.	4	2	5+	4	0[a]
29		12:23	Off Fukushima Pref.	4.7	2	5–	3	0[a]
30		12:52	E off Chiba Pref.	6.1	5–	6–	6–	5+
31	2011-03-17	21:32	E off Chiba Pref.	5.7	4	5–	5–	5–
32	2011-03-19	6:18	Off Ibaraki Pref.	4.8	2	5–	3	0[a]
33		8:32	Off Iwate Pref.	5.7	4	5–	4	3
34		18:56	Ibaraki Pref.	4	4	5+	6–	5+[c]
35		18:57	Ibaraki Pref.	6.1	5+	5–	6–	5+
36		18:57	Ibaraki Pref.	6.1	5+[e]	5+[e]	6–[e]	5+[e]
37	2011-03-20	14:19	Fukushima Pref.	4.7	3	6–	4	4
38	2011-03-22	12:38	E off Chiba Pref.	5.9	4	6+	4	4
39	2011-03-23	1:12	Off Ibaraki Pref.	5.4	3	6+	3	3
40		7:12	Fukushima Pref.	6	5+	5+	6–	5+

No.	Date	Time	Location	Magnitude	Catalog	JMA	TYW2014	RBIS
41	2011-03-23	7:36	Fukushima Pref.	5.8	5+	5+	5+	5+
42		8:46	E off Chiba Pref.	5	2	5–	3	2
43	2011-03-25	20:36	Off Miyagi Pref.	6.3	4	6–	5–	4
44	2011-03-27	19:23	E off Chiba Pref.	5	2	5+	3	3
45	2011-03-28	7:23	Off Miyagi Pref.	6.5	5–	6–	5–	4
46	2011-04-01	19:49	Akita Pref.	5	5+	5–	5–	4[d]
47	2011-04-03	16:38	Off Fukushima Pref.	5.4	4	5+	4	4
48	2011-04-04	18:29	Off Fukushima Pref.	4	2	5+	3	0[a]
49	2011-04-07	23:33	Off Miyagi Pref.	7.2	6+	6–	6–	6–
50	2011-04-11	17:16	Fukushima Pref.	7	6–	6+	7	6–
51		17:26	Fukushima Pref.	5.4	5–	5–	5+	4
52		18:05	Fukushima Pref.	5.1	4	5–	5–	4
53		20:42	Off Fukushima Pref.	5.9	5–	6–	4	5–
54	2011-04-12	8:08	E off Chiba Pref.	6.4	5–	7	5–	5–
55		8:08	E off Chiba Pref.	6.4	5–	5–	5–	5–
56		10:23	Chiba Pref.	4.2	2	5+	2	0[a]
57		12:20	E off Chiba Pref.	3.8	2	5–	0[a]	0[a]
58		14:07	Fukushima Pref.	6.4	6–	6–	6+	5+
59		16:14	Nagano Pref.	3.1	2	6–	4	3
60	2011-04-13	10:07	Fukushima Pref.	5.7	5–	6+	5–	4
61	2011-04-14	6:43	Fukushima Pref.	4.1	3	6–	3	2
62		12:08	Fukushima Pref.	5.4	4	5–	4	5–
63		20:24	Fukushima Pref.	4.4	3	6–	5–	3
64		21:24	Fukushima Pref.	3.9	3	5–	4	3
65	2011-04-15	23:34	Off Iwate Pref.	5	3	6+	3	0[a]
66	2011-04-16	11:19	Ibaraki Pref.	5.9	5+	5–	6+	5–
67	2011-04-19	4:14	Akita Pref.	4.9	5–	5–	6–	5–
68		6:33	Ibaraki Pref.	4.8	3	5–	3	0[a]
69	2011-04-21	22:37	Chiba Pref.	6	5–	5–	5–	4
70	2011-04-24	20:50	Fukushima Pref.	3.1	3	6–	4	2
71	2011-04-30	2:04	E off Chiba Pref.	4.7	3	6+	4	3

a: Events with small seismic intensity that did not trigger a new event

b: Correctly identified earthquake, but alarm later cancelled

c: Two overlapping concurrent events

d: Accurate hypocenter location, origin time and magnitude estimate with bad seismic intensity estimate

e: JMA EEW system created an extra fake event (other two systems did not)