An Earthquake Early Warning Method Based on Bayesian Inference

Jingsong Yang 1,2, Fulin He 1, Zhitao Li 3,* and Yinzhe Zhang 1

1 College of Electronic Science and Control Engineering, Institute of Disaster Prevention, Langfang 065201, China
2 Hebei Key Laboratory of Seismic Disaster Instrument and Monitoring Technology, Langfang 065201, China
3 National Institute of Natural Hazards, Beijing 100085, China
* Correspondence: zhitao.lee@163.com; Tel.: +86-137-1843-2722

Abstract: Earthquake early warning (EEW) systems can give people warnings before damaging ground motions reach their site and can reduce earthquake-related losses. It has been shown that Bayesian reference can be used to estimate earthquake magnitude, location, and the distribution of peak ground motion using observed ground-motion amplitudes, predefined prior information, and appropriate attenuation relationships. In this article, we describe a Bayesian approach to earthquake early warning (EEW) systems for the estimation of magnitude (M) and peak ground-motion velocity (PGV) using the Gutenberg–Richter relation as a priori probability derived from statistical historical earthquake information from the Japanese KiK-net. Moreover, using the magnitude $M_j = 4.5$ and $PGV = 0.4 \, \text{cm/s}$ as the warning threshold, an earthquake hazard discrimination model was established. Following this, the proposed approach was compared with the traditional fitting method to analyze the earthquake hazard discrimination using an $M_j = 4.3$ earthquake in the Mt. Fuji area and an $M_j = 5.5$ earthquake in the Ibaraki area of Japan as examples. The results are as follows: (1) The probabilistic algorithm founded on the predictive model of the magnitude from the average period ($\tau_c$) and PGV from the displacement amplitude ($P_d$) from the initial 3 s of the P wave is able to provide a fast and accurate estimation of the final magnitude and location. (2) The Bayesian inference approach for the earthquake hazard discrimination model has a 2.94% miss rate and a 5.88% false alarm rate, which is lower than that of the traditional fitting method, thus increasing the accuracy and reliability of earthquake hazard estimation.

Keywords: earthquake early warning; Bayesian inference; magnitude prediction; PGV prediction

1. Introduction

The basic principle of earthquake early warnings is the application of the P-wave information received by the corresponding station to make basic judgments and predictions about an earthquake that has just occurred or is currently occurring prior to the arrival of destructive seismic waves [1,2]. Target areas that could potentially be damaged are consequently alerted to take emergency measures [3]. The accuracy and timeliness of real-time earthquake intensity estimations have become the focus of key research on earthquake early warning systems [4]. Kanamori analyzed the relationship between the average P-wave period ($\tau_c$) and the magnitude M [5]. Zhao employed characteristic parameters (e.g., displacement amplitude $P_d$ and $\tau_c$) related to ground motion intensity and statistical regression methods for real-time predictions of peak ground acceleration (PGA) and peak ground velocity (PGV) [6]. Iervolino et al. proposed an earthquake magnitude prediction method based on Bayesian conditional probability distribution theory to calculate earthquake magnitude using historical seismicity information in Campania [7], southern Italy and real-time waveform data. Wu applied Bayesian inference to analyze the relationship between ground motion parameters [8], including $P_d$, amplitude of acceleration...
(P_d), initial P-wave velocity amplitude (P_v) and PGV. However, accurately evaluating earthquake hazards on the basis of a single parameter proves to be a complicated task [9]. Methods used for earthquake hazard assessments should be based on multiple parameters. Thus, this paper applies Bayesian inference to obtain prior probability distributions using the Gutenberg–Richter (G–R) relationship and historical earthquake information from Japan [10]. Combined with the average P-wave period \( \tau_c \) and the displacement amplitude \( P_d \), the magnitude and subsequent PGV are predicted. The four-level division method is then applied to complete the comprehensive judgment of earthquake hazards. The earthquake warning given according to the earthquake hazard level can reduce the rate of missed and false alarms of earthquake warnings.

2. Earthquake Data Selection and Preprocessing

We used KiK-net data (1 October 2007 to 1 October 2017) obtained from the Japan Institute of Disaster Prevention Technology and Science. The earthquake magnitudes in this dataset range between 3.0 and 8.0, and the majority exceed 4. Earthquake data were processed and screened according to the following standards:

1. We selected data from earthquakes within 10 km of the focal depth and 200 km from the epicenter.
2. The short–term average to long–term average (STA/LTA) and Akaike information criteria (AIC) were combined to automatically detect the arrival of the P-wave and manually correct it.
3. The acceleration record was corrected at the baseline, and the acceleration records were integrated to obtain the velocity and displacement records. Following this, 0.075 Hz Butterworth high-pass filtering was performed to eliminate the low-frequency drift caused by integration.
4. Earthquake records should meet certain Signal to Noise Ratio(SNR) requirements during the calculations. In this paper, velocity amplitudes greater than 0.05 cm/s within 3 s after P-wave triggering were selected as the SNR criteria. Noise reduction of the selected records was also performed to reduce the high-frequency interference of the acceleration records. A total of 145 earthquakes were screened out. Figure 1 presents the distribution of the magnitude, epicenter distance, and earthquake frequency of the dataset.

![Figure 1. Distribution of the number of earthquake records with selected epicentral distances and magnitudes.](image)

3. Traditional Calculation Method for Earthquake Magnitude and PGV

The magnitude and PGV of earthquakes are traditionally calculated using fitting methods based on the average period \( \tau_c \) and displacement amplitude \( P_d \) of the P-wave.
3.1. Magnitude Calculation using Average Period $\tau_c$

The average period $\tau_c$ is often used in earthquake early warning systems to calculate the magnitude. In particular, the magnitude is estimated by establishing a linear statistical relationship between $\tau_c$ and the magnitude within a few seconds prior to the P-wave [11,12]. $\tau_c$ can be calculated as follows:

$$\tau_c = \frac{2\pi}{\sqrt{r}}$$

for

$$r = \frac{\int_0^{\tau_0} u(t)^2 dt}{\int_0^{\tau_0} u^2(t) dt}$$

where $u(t)$ is the displacement of the ground motion at time $t$; $u(t)$ is the velocity of the ground motion at time $t$, which is counted from the trigger of the station and recorded as $t = 0$; and $\tau_0$ is the upper integration time window length.

In order to process the earthquake data, $\tau_c$ of the first-three-second P-wave was calculated and the fitting relationship with magnitude $M_j$ was established. Note that the magnitude difference between Japan and China is negligible between 4 and 8 [13], and this paper ignores the difference.

$$\lg \tau_c = aM_j + b \pm \delta,$$

In the above equation, $M_j$ is the magnitude, $a$ and $b$ are the fitting parameters, and $\delta$ is the standard deviation of Equation (3). Using the least-squares method, these parameters are determined as $a = 0.1353$, $b = -0.0495$, and $\delta = 0.1149$.

3.2. Calculation of PGV using Displacement Amplitude $P_d$

Previous studies have reported a correlation between PGV and earthquake intensity [14,15]. For the analysis of earthquake hazards, PGV is typically used to estimate the potential earthquake hazard in the local area [16]. Moreover, there is a strong log-linear relationship between $P_d$ and PGV:

$$\text{PGV} = \max[v(t_i)],$$

for

$$a(t_i) = \sqrt{a_{EW}(t_i)^2 + a_{NS}(t_i)^2 + a_{UD}(t_i)^2},$$

$$v(t_i) = \int_0^{\tau_0} a(t_i) dt,$$

where $t_i$ is the ith second after the arrival of the P wave; $a(t_i)$ is the combined acceleration in all three directions at $t_i$; $a_{EW}(t_i)$ is the integral of the resultant acceleration at $t_i$; and $a_{EW}(t_i)$, $a_{NS}(t_i)$, and $a_{UD}(t_i)$ are the acceleration records in the east–west, north–south, and vertical directions at $t_i$, respectively.

$P_d$ can be calculated as follows:

$$P_d = \max[d(t_i)],$$

where $d(t_i)$ is the displacement of the vertical earthquake record representing the displacement sequence from the initial arrival point of the P-wave to $t_i$. For the processed earthquake data, PGV and $P_d$ at the first-three-second P-wave are used, and the log-linear relationship between $P_d$ and PGV is established:

$$\lg \text{PGV} = a \lg P_d + b \pm \delta,$$
When the initial phase of the P-wave is picked up, various characteristic parameters are 
where $\beta$ with a certain magnitude and PGV. The likelihood function requires ground motion models 
y where $a$ and $b$ are the fitting parameters, and $\delta$ is the standard deviation of Equation (8). Using the least-squares method, the parameters are determined as $a = 1.049$, $b = -0.2691$, and $\delta = 0.16$.

4. A Dual-Parameter Earthquake Warning Model Based on Bayesian Inference

4.1. Prediction of Earthquake Magnitude and PGV Based on Bayesian Estimation

We want to estimate the magnitude and PGV of an earthquake given an available set of observed ground motions. According to Bayesian method, the state of belief regarding the magnitude and PGV (M, PGV) given a set of available observations $y_{obs}$ is given by:

$$f(x|y_{obs}) = \frac{f(y_{obs}|x) \cdot f(x)}{\int_{x_{min}}^{x_{max}} f(y_{obs}|x) \cdot f(x)dx},$$

(9)

where $y_{obs}$ is the ground motion observation data, namely, the average period $\tau_c$ and displacement amplitude $P_d$; $x$ denotes the parameters to be established, namely, the earthquake magnitude $M_i$ and PGV; and $f(x)$ is the a priori probability equation, representing all the information available for the real-time prediction of earthquake parameters prior to the earthquake. As a likelihood function, $f(x|y_{obs})$ is the posterior probability density function (PDF); it is the conditional probability that an earthquake of a certain magnitude and PGV generated the set of observations $y_{obs}$. The Bayesian estimates (M, PGV) are the most probable source estimates given the available observations; they maximize $f(x|y_{obs})A_i$; and $x_{max}$ and $x_{min}$ are the maximum and minimum values of the magnitude or PGV that can occur in the study area. On the right-hand side, $f(y_{obs}|x)$ is the likelihood function; it is the conditional probability of observing a set of ground motions $y_{obs}$ given an earthquake with a certain magnitude and PGV. The likelihood function requires ground motion models relating source descriptions (M, PGV) to observed ground motion amplitudes.

The prior probability function can be obtained from the G–R relation as follows:

$$f(x) = \begin{cases} \frac{\beta e^{-\frac{\beta x}{\tau_{min}}} - e^{-\frac{\beta x_{max}}{\tau_{max}}}}{e^{-\frac{\beta x_{min}}{\tau_{min}}} - e^{-\frac{\beta x_{max}}{\tau_{max}}}}, & x_{min} \leq x \leq x_{max} \\ 0, & x \notin [x_{min}, x_{max}] \end{cases},$$

(10)

where $\beta$ is the ratio of the size to the number of earthquakes and is determined as $\beta = 1.3543$ using the least-squares fitting method and earthquake records from 1 October 2007 to 1 October 2017 in Japan ($M_i - M_i$ 8). The joint probability density distribution $f(y_{obs}|x)$ is the likelihood function, that is, the conditional probability density of monitoring data $y_{obs}$ at a given $x$. The information collected in real time and the characteristic parameter data are integrated through the likelihood function and substituted into the Bayesian formula. Thus, the likelihood function can update the probability in real time according to the collected information. If the earthquake early warning information $y_{obs}$ received by each seismic monitoring station is independent and has an identical distribution, then $i$ is taken as the arrival time of the P-wave, and the early warning information from different stations is multiplied to obtain:

$$f(y_{obs}|x) = \prod_{k=1}^{n} f(y_{k,i}|x) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma y_{k,i}}} \exp\left[-\frac{1}{2}\left(\frac{\log y_{k,i} - \mu}{\delta}\right)^2\right].$$

(11)

In this paper, we adopt a single monitoring in-situ warning information calculation. When the initial phase of the P-wave is picked up, various characteristic parameters are extracted and are uniformly represented with $y_{obs}$, which can be obtained from Equation (11):

$$f(y_{1,i}|x) = \frac{1}{\sqrt{2\pi\sigma y_{1,i}}} \exp\left[-\frac{1}{2}\left(\frac{\log y_{1,i} - \mu}{\delta}\right)^2\right].$$

(12)
where \( f(x \mid y_{\text{obs}}) \) is the likelihood function; \( \mu \) is the mathematical expectation that is the average value of \( \lg y_{\text{obs}} \); and \( \sigma \) is the standard deviation of \( \lg y_{\text{obs}} \).

When an earthquake is detected by a seismograph, the prior probability function is obtained using the G–R relationship to calculate the historical earthquake information. As long as the characteristic parameters for magnitude prediction can be extracted, the prior probability function can be updated according to the newly extracted information, that is, it is determined by the G–R relationship before the arrival of the P-wave monitored by the station. When \( t \) is greater than the arrival time \( t_{\text{first}} \) of the P-wave, that is, when the station records new information, the size must be altered based on the new information:

\[
f_{t>t_{\text{first}}}(x) = f_{t-1}(x|y_{1,t-1})
\]  

(13)

By substituting Equations (10) and (11) into Equation (9), the posterior probability density function of the characteristic parameter \( y_{1,t} \), known, that is, the probability density value of \( x \):

\[
f(x|y_{1,t}) = \frac{\exp \left( \frac{2(\mu_{\lg y_{1,t}} - \mu)^2}{2\sigma^2} + \beta x \right)}{\int_{-\infty}^{\infty} \exp \left( \frac{2(\mu_{\lg y_{1,t}} - \mu)^2}{2\sigma^2} + \beta x \right) dx}
\]  

(14)

4.2. Prediction and Simulation of Magnitude Based on Bayesian Inference

From the fitting results of Equation (3) and \( \lg \tau_c \), the mathematical expectation \( \mu \) and standard deviation \( \sigma \) are described as:

\[
\mu_{\lg \tau_c} = 0.2455M - 0.9006, \\
\sigma_{\lg \tau_c} = 0.2220.
\]  

(15)

By substituting the results of Equation (15) into Equation (14), the magnitude probability density function of the average period parameter \( \tau_c \) at time \( t \) can be obtained. Assuming that the earthquake magnitudes are \( M_j = 4, M_j = 5 \), and \( M_j = 6 \), \( \tau_c \) takes the values of 1.206, 2.1228, and 3.7359, respectively. Figure 2 presents the probability density distribution obtained by substituting the three magnitudes into Equation (13). The estimated magnitudes corresponding to the maximum probability density are 3.8, 5.6, and 5.8. The verification of the model reveals an error of 0.4836, resulting from the model and calculation errors. Thus, the earthquake magnitude can be predicted using Bayesian inference.

![Figure 2](image-url)  

*Figure 2.* Probability density distribution of different \( M_j \) earthquake early warning method based on Bayesian inference.

4.3. Predictive Simulation of PGV Based on Bayesian Inference

From the fitting results of Equation (8) and \( \lg P_d \), the mathematical expectation \( \mu_{\lg P_d} \) and standard deviation \( \sigma_{\lg P_d} \) can be expressed as:

\[
\mu_{\lg P_d} = 0.9653\lg PGV + 0.0366, \\
\sigma_{\lg P_d} = 0.1045.
\]  

(16)
By substituting the parameters into Equation (14), the predicted PGV probability density function of the displacement amplitude parameter can be obtained. For PGV values of 0.6858, 0.3864, and 0.2886 cm/s, the $\lg P_G$ values are determined to be 0.2631, $-0.0549$ and $-0.2139$, respectively. The probability density distribution of different $\lg P_G$ can be obtained by substituting these three displacement amplitudes into Equation (13) (Figure 3). The maximum probability densities of $\lg P_G$ are determined as $-0.051$, $-0.3710$, and $-0.5310$, with corresponding PGV values of 0.8892, 0.4256, and 0.2944 cm/s. Due to the large dispersion of the PGV data, the calculation and model errors are also relatively large.

![Figure 3. Probability density distribution of different $\lg$PGVs.](image)

4.4. Earthquake Judgment Model Based on Dual-Parameter Prediction

A judgment model was subsequently established to identify earthquake size and distance based on Zollo et al. [8]. According to the relationship between the early warning level and magnitude and intensity [17], the early warning values of magnitude and PGV were set. The seismic events can be classified into four types (Figure 4). In this paper, the thresholds were set as $M_t = 4.5$ and $PGV = 0.4$ cm/s. Figure 4 depicts the human sensation, damage degree, and corresponding earthquake warning measures of different earthquake types.

![Figure 4. Seismic judgment model.](image)

5. Comparative Testing and Analysis

Using the processed seismic data, the $\tau_c$ of the first-three-second P-wave is selected to obtain the predicted magnitude $M_t$ using the traditional fitting approach. A dual-parameter prediction model based on Bayesian inference is then proposed to predict the earthquake
magnitude $M_b$. Finally, the predicted magnitudes $M_f$ and $M_b$ are compared with the actual magnitude $M_j$. Figure 5 compares the corresponding dispersion degree and error distribution.

The standard deviation of the magnitude predicted using Bayesian inference is 0.3202 lower than that obtained by the traditional fitting method (Figure 5a). This reveals the results of the Bayesian prediction method to be more stable and reliable than those of the traditional fitting approach. The prediction method based on Bayesian inference exhibits a smaller distribution range of errors, a higher fitting accuracy, and a lower deviation degree, which can provide more accurate seismic parameter information for the four-level discriminant model (Figure 5b).

Figure 5. Comparison of the actual magnitude and the predicted magnitude using the conventional method and the Bayesian inference method (a) and its error distribution (b).

Similarly, by determining the $P_d$ of the after-three-second $P$-wave, the PGV is predicted using the traditional fitting method (PGV_f). PGV is then predicted according to Bayesian inference (PGV_b). The predicted values obtained are then respectively compared with the real PGV. Figure 6 depicts the dispersion degree and error distribution. The slope of the red line in Figure 6a,b is 1, indicating that the Bayesian-inference-based PGV is more standard than that obtained with the traditional fitting method.
Figure 6. Comparison of true PGV with conventional fitted predictions (a) and Bayesian-inference-based predictions (b).

6. Discussion and Conclusions

In this paper, the 2014 $M_j = 4.3$ earthquake in Mt. Fuji, Japan (14 data points), and the 2017 $M_j = 5.5$ earthquake in Ibaraki, Japan (54 data points), were selected for the experimental analysis. Figure 7 displays the earthquake epicenters and the distribution of stations. The data of the first-three-second P-wave collected via 68 stations were selected to estimate the magnitude and PGV, and the earthquake hazard was predicted.

Figure 7. Distribution of epicenter stations for magnitudes $M_j 4.3$ (a) and $M_j 5.5$ (b). Available online: www.kyoshin.bosai.go.jp (accessed on 6 July 2022).

The values obtained with the Bayesian-inference-based prediction method and the traditional fitting method are substituted into the discriminative model in this paper. Figure 8 presents the distribution of all seismic hazard discrimination results. Table 1 shows the missed and false alarms of the Bayesian inference and traditional approaches.
while actually the magnitude or PGV is smaller than the threshold and an alarm should not be issued. In the current 68 experimental results, the prediction method based on Bayesian inference has a miss rate of 2.94% and a false alarm rate of 5.88%; the conventional fitting method has a miss rate of 4.41% and a false alarm rate of 16.18%. The results are as follows:

1. The probabilistic algorithm founded on the predictive model of the magnitude from the average period ($\tau_c$) and PGV from the displacement amplitude ($P_d$) from the initial 3 s of the P wave is able to provide a fast and accurate estimation of the final magnitude and location.

Table 1. Comparison of missed and false alarms of the experimental results of the two methods.

<table>
<thead>
<tr>
<th>Missed Alarm</th>
<th>False Alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian prediction method</td>
<td>Traditional fitting method</td>
</tr>
<tr>
<td>Number of Missed Alarms</td>
<td>Number of False Alarms</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2.94%</td>
<td>5.88%</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>4.41%</td>
<td>16.18%</td>
</tr>
</tbody>
</table>

Figure 8 and Table 1 reveal that both the Bayesian-inference-based prediction method and the traditional fitting method suffer from underreporting and misreporting. An under-prediction indicates that the estimated magnitude or PGV is smaller than the threshold and no earthquake alarm is issued, while actually the magnitude and PGV are greater than the threshold and an alarm should be issued. A false-prediction denotes that the estimated magnitude and PGV exceed the threshold and an earthquake alarm is issued, while actually the magnitude or PGV is smaller than the threshold and an alarm should not be issued. In the current 68 experimental results, the prediction method based on Bayesian inference has a miss rate of 2.94% and a false alarm rate of 5.88%; the conventional fitting method has a miss rate of 4.41% and a false alarm rate of 16.18%. The results are as follows:
(2) The Bayesian inference approach for the earthquake hazard discrimination model has a 2.94% miss rate and a 5.88% false alarm rate, which is lower than that of the traditional fitting method, thus increasing the accuracy and reliability of earthquake hazard estimation.

In this paper, a dual-parameter warning method of earthquake magnitude and PGV based on Bayesian inference is proposed to classify earthquake hazards into four levels. The results effectively improve the accuracy and reliability of early warning information. According to the real-time magnitude and PGV conditions, the $\tau_c$ and $P_d$ of the first-three-second P-wave are fitted to the magnitude and PGV, respectively, and the empirical prediction formula is obtained. Furthermore, combined with historical earthquake information, the Gutenberg–Richter relationship is employed as the initial prior probability, and the $\tau_c$ and $P_d$ of the first-three-second P-wave recorded in real time by the seismic station are used as the likelihood functions to predict the magnitude and PGV. The proposed method was compared with the traditional fitting approach [18], revealing the higher accuracy of the former. Finally, the $M_{J}4.3$ earthquake in the Mt. Fuji area and the $M_{J}5.5$ earthquake in the Ibaraki area of Japan were selected as examples to evaluate and discriminate the magnitude and distance of the earthquake and to analyze the under-prediction and false-prediction rates.

Despite the substantial progress made using the classification model of earthquake types in this paper, it has some limitations. For example, the early warning threshold needs to account for the characteristics of the local geographical environment, which vary across regions. This is particularly important, as the threshold settings play a large role in the model’s accuracy. Future research will focus on refining the model to further explore the available early warning indicators.

**Author Contributions:** Conceptualization and methodology, Z.L. and J.Y.; software, F.H. and Y.Z.; validation, F.H. and J.Y.; resources and data curation, Z.L., J.Y., F.H. and Y.Z.; writing—original draft preparation, Z.L. and J.Y.; writing—review and editing, Z.L., J.Y., F.H. and Y.Z.; supervision, Z.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** Funding: This research was funded by National Natural Science Foundation of China, grant number 41874019.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Earthquake data used in the study are available from Japan National Research Institute for Earth Science and Disaster Prevention: Available online: [https://www.kyoshin.bosai.go.jp/kyoshin/quake/index_en.html](https://www.kyoshin.bosai.go.jp/kyoshin/quake/index_en.html) (accessed on 6 July 2022).

**Acknowledgments:** We would like to thank the editor and reviewers for their valuable comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


