Improved Epicentral Distance Estimations for Railway Earthquake Early Warning

Heather STEELE
Seismic Data Analysis Laboratory, Center for Railway Earthquake Engineering Research (Former)

Shunta NODA
Seismic Data Analysis Laboratory, Center for Railway Earthquake Engineering Research

Shunroku YAMAMOTO
Center for Railway Earthquake Engineering Research

Masahiro KORENAGA
Seismic Data Analysis Laboratory, Center for Railway Earthquake Engineering Research

Accurate, rapid estimation of epicentral distance, \( \Delta \), is essential for earthquake early warning systems. To improve \( \Delta \) estimation, new methods to calculate the empirical relationship between the amplitude growth rate parameter, \( C \), and \( \Delta \) are investigated. Using orthogonal regression is most appropriate for application to EEW systems for Japanese high-speed trains. Evaluation using a K-NET dataset, of earthquake epicenters up to 200 km from the recording station, showed that the proposed method reduces maximum error from 546.2 km to 209.8 km. The percentage of correct estimations, defined as estimates within \( \pm 30\% \) of the measured epicentral distance, is increased from 38.0\% to 55.3\%.

Keywords: earthquake early warning (EEW), epicentral distance estimation, \( C-\Delta \) algorithm, seismic parameter, P-wave growth rate

1. Introduction

To stop high-speed trains as soon as possible and minimize damage in the event of an earthquake, the Japan Railway earthquake early warning (JR EEW) systems, developed by the Railway Technical Research Institute, control trains based on P-wave information [1, 2]. Seismographs dedicated to railway operations are deployed nationally to detect P-wave onset, and within seconds estimate the seismic parameters of epicentral distance, back-azimuth and magnitude from single station information [3, 4, 5]. Since 2003, the JR EEW systems have estimated epicentral distance, \( \Delta \), using an empirical relationship, determined from recorded data, between the measured epicentral distance of earthquakes from single seismic stations and a parameter describing the amplitude growth rate of some period of the initial P-wave [3]. As epicentral distance increases, the amplitude growth rate decreases due to geometrical scattering and seismic wave attenuation [6]. Provided the calculation time period is short, the relationship appears to be independent of magnitude [7]. The two main approaches are commonly referred to as the B-\( \Delta \) [2, 3, 6, 8, 9] and C-\( \Delta \) methods [10, 11, 12] where \( B \) and \( C \) refer to amplitude growth rate parameters. Poor \( \Delta \) estimation may cause missed alarms that waste lead time and increase the risk of damage, or false alarms that unduly disrupt train operation.

The relationship of \( B \) or \( C \) and \( \Delta \) exhibits a large amount of scattering, which has led to efforts to improve the definition of the amplitude growth rate parameter and empirical relationship. To reduce estimation error, distinct relationships for earthquakes grouped by factors such as the correlation distance [6] or fractional fluctuation [12], which categorize local heterogeneity, have been considered in previous studies. This work instead considers how the estimation error can be reduced by changing the method used to calculate the empirical relationship between \( C \) and \( \Delta \).

2. The C-\( \Delta \) method

2.1 The C-\( \Delta \) relationship

\( C \) is found by fitting \( y(t) = C t \) to the first 0.5s of the envelope curve of the Up-Down component of the acceleration of the observed P-wave’s initial phase, using least squares regression [7, 10]. The envelope is created by filtering the initial waveform using a 10 – 20 Hz recursive band-pass filter and taking the maximum acceleration of the absolute filtered waveform (Fig. 1). The arrival of the P-wave is determined using the STA/LTA method applied to unfiltered traces [13], with a level-trigger based on acceleration used in cases where STA/LTA fails to detect P-wave arrival [10].

The calculated \( C \) values (amplitude growth rate of the initial P wave) are correlated to epicentral distance, \( \Delta \), with variance due to seismological factors, e.g. source effect, heterogeneity of the crustal structure, and technical factors, e.g. picking error and filtering parameters. The purpose of this study is to improve the estimation accuracy using statistical analyses of the relationship between \( \Delta \) and \( C \), not to evaluate the accuracy of the \( C \) calculation itself.

In the original linear method, henceforth referred to as C-\( \Delta _L \), the empirical relationship is described as:
where $\alpha_L$ and $\beta_L$ are linear regression coefficients determined using recorded earthquake data. For a given $C$ value, the epicentral distance can be estimated by rearranging the equation:

$$A_{\Delta L} = 10^\left(\frac{\log_{10} C - \beta_L}{\alpha_L}\right)$$  \hspace{1cm} (2)$$

A damping member was incorporated to account for attenuation at longer epicentral distances. In this ‘curved’ method, henceforth referred to as C-$\Delta_c$, the relationship is described as:

$$\log_{10} C = \alpha_c \log_{10} A + \beta_c + \gamma A$$  \hspace{1cm} (3)$$

where $\alpha_c$, $\beta_c$, and $\gamma$ are multiple linear regression coefficients that give rise to a curved relationship between $\log_{10} C$ and $A$. Equation (3) was rearranged for $A$ using Wolfram\textsuperscript{\textregistered}Alpha [14], to enable direct calculation of epicentral distance estimates. $W$ represents the Lambert $W$ function [15]:

$$A_{\Delta c} = \frac{10^{\frac{\beta_c}{\alpha_c}}}{\gamma \ln(10)} \left( \sqrt{\frac{C^\gamma \ln(10)}{\alpha_c W}} \right)$$  \hspace{1cm} (4)$$

Inclusion of the damping term was shown to reduce the root mean square logarithm error (RMSLE) by 5.7% compared to the linear method [10]. RMSLE is an index which quantifies the difference between measured values and those estimated using a specified relationship. The measured values are the JMA earthquake epicentral distances and the estimates are the $A$ values calculated using the C-$\Delta_L$ or C-$\Delta_c$ methods. Figure 2 shows both relationships. The coefficients of C-$\Delta_c$ are from the currently implemented method. The coefficients of C-$\Delta_L$ were calculated using a dataset comprised of measured epicentral distances and $C$ values computed using seismic waveforms observed at K-NET stations operated by the NIED in Japan [16].

### 2.2 Earthquake dataset

The dataset, shown in Fig. 3, contains 8965 recordings for seismic stations within 200 km of the epicenter of 195 Japanese earthquakes, of magnitudes between 4.5 and 9. Although damaging earthquakes can occur at longer distances [17], the relationship between the growth amplitude parameter and $\Delta$ does not exhibit significant correlation beyond 200 km [8]. The peak number of entries occur around 80 km and there are few entries in the range 0-40 km. This happens because of the distribution of seismographs and is therefore inherent to the K-NET system. Within a small radius from the epicenter there are few seismographs, but as the radius increases data from more seismic stations is included. Due to attenuation of the waveform, fewer traces register at further stations, thus creating the observed peak.

The dataset was split into a test set for calculating different C-$\Delta$ relationships, and a validation set to assess the difference between measured and estimated $\Delta$ values. To reduce the dataset bias for the regression, the data was binned into 50 km epicentral distance ranges containing an even number of entries. The first bin contains all entries where $\Delta$ is less than 50 km. The following bins contain a random selection of entries from their range, equal to the number of entries in the first bin. The first half of each bin was used to form a ‘non-biased’ test set for regression. The remaining data formed the validation set.

### 2.3 Estimation accuracy

Plotting the estimated epicentral distance values, $A_{\Delta M}$, against the measured values, $A_{\Delta m}$ on a log scale shows a linear relationship with scattering, quantified by RMSLE. Previous studies have largely evaluated the estimation
error based on this index. This study also considers the slope, $S_{\text{est}}$, and intercept, $I_{\text{est}}$, values of the regression, to describe how well the estimates are centered on the measured values. The percentage of correct estimations index $\%\text{Corr}$ quantifies the percentage of the $\Delta_{\text{est}}$ values that are within 30% of the $\Delta_{M}$ values. A percentage limit rather than a distance limit places stricter requirements on the estimation of shorter epicentral distances. The maximum error in km, $\text{Err}_{\text{max}}$, is also used. Figure 4 shows the relationship between $\Delta_{\text{est}}$ and $\Delta_{M}$ for the C-$\Delta_{L}$ and C-$\Delta_{C}$ methods on a log scale, from which the RMSLE, $S_{\text{est}}$, and $I_{\text{est}}$ are derived. Figure 5 shows the relationship on a km scale, with the estimates meeting $\Delta_{\text{est}} \pm 30\%$ highlighted. The values of each index are given in Table 1, alongside the desired values. Although the $\Delta_{\text{est}}$ values for both methods are well centered on the $\Delta_{M}$ values - as indicated by $S_{\text{est}}$ and $I_{\text{est}}$ being very close to the desired values of 1 and 0 respectively - the RMSLE is high at over 0.3. The addition of the damping term in the C-$\Delta_{C}$ method increases $\%\text{Corr}$ from $\sim 34\%$ to $38\%$, and reduces the maximum error considerably, from over 3000 km to approximately 550 km. However, despite this improvement, the majority of estimations do not fall within the 30% range when using the C-$\Delta_{C}$ method, and the maximum error remains high. In order to perform satisfactorily for the majority of earthquakes, the C-$\Delta_{C}$ method requires improvement.

3. The reverse C-$\Delta$ method

The C-$\Delta$ methods use linear least squares regression to find the relationship between $C$ and $\Delta$. Linear least squares minimizes the sum of the squares of residuals in the $y$ direction only, meaning that the error is reduced for the $y$ variable, $\log_{10} C$. As previously stated, there is uncertainty in the relationship between $C$ and $\Delta$ due to various factors. This suggests that the original equations need other terms to more exactly explain the relationship between $C$ and $\Delta$, that is the inclusion of errors. Since the exact relationship between $C$ and $\Delta$ is not clear, the reverse C-$\Delta$ method, C-$\Delta_{\text{rev}}$, proposes to switch the $x$ and $y$ variables in order to explore the uncertainty in this case. This approach minimizes the error for $\log_{10} \Delta$, using a relationship defined as:

$$\log_{10} \Delta = Z \log_{10} C + i$$

where $Z$ and $i$ are the linear regression coefficients. As before, the true relationship can be considered to include errors. Figure 6 shows the empirical C-$\Delta_{\text{rev}}$ relationship calculated using the regression dataset. Equivalent coefficients, $Z_{\text{eq}} = 1/\alpha_{L}$, $i_{\text{eq}} = \beta_{L}/\alpha_{L}$, are calculated for the C-$\Delta_{L}$ method and this relationship is also shown.

The estimation accuracy was assessed using the same evaluation indices as the original methods, and results can be found in Table 1. The RMSLE is reduced by 34%, from 0.314 to 0.207. Although this is a vast improvement, the estimated epicentral distances are not well centered on the measured values, as shown in Fig. 7. $S_{\text{est}}$ is 0.45, rather than 1 and there is a clear offset, indicated by the value of $I_{\text{est}}$ at 1 rather than 0. For $\log_{10} \Delta_{L}$ values below 1.5 ($\Delta ~ 30$ km), the epicentral distance is exclusively overestimated (all estimates are above the desired result line). However, $\%\text{Corr}$ is
increased to 51.13% due to improved estimations at longer distance, as is evident in Fig. 8. $\text{Err}_{\text{max}}$ is slightly reduced to 404.5 km.

The estimation accuracy of the reverse C-Δ method is improved over the current C-ΔC method. However, it cannot estimate correctly for distances less than $\sim 30$ km. Despite binning the data at 50 km to reduce bias in the regression, there are still comparatively few entries below 30 km. This is likely causing the bias and resultant overestimation. It is difficult to obtain the same quantity of data for short epicentral distances due to the nature of the seismograph network.

### 4. Orthogonal C-Δ methods

The original and reverse C-Δ methods use linear least squares regression to minimize error in the $y$ direction, for $\log_{10} C$ and $\log_{10} \Delta$ respectively. However, the relationship between $C$ and $\Delta$ is not straightforward and errors are present in each case. The proposed orthogonal methods use orthogonal least squares regression to account for errors in both the $x$ and $y$ variables. For linear relationships the error is reduced perpendicular to the regression line, rather than the $x$ or $y$ direction, so the orthogonal regression results for $\log_{10} C = a\log_{10} \Delta + b$ and $\log_{10} \Delta = Z\log_{10} C + i$ are equivalent.

Two orthogonal methods are considered, based on the previous equations defining linear and curve relationships. By defining the empirical relationship to be optimized, orthogonal regression parameters in the form of both equation (1) and (3) were determined and evaluated. These methods are henceforth referred to as Orth-C-ΔL and Orth-C-ΔC respectively.

Figure 9 shows the empirical relationships for all of the linear and curve methods. It can be seen that the Orth C-ΔL and C-Δrev methods have similar regression coefficients. The evaluation indices given in Table 1 confirm that they perform similarly, with the Orth-C-ΔL producing marginally worse results. For linear methods, higher $Z$ values lead to greater overestimation of $\Delta$ for low $C$ values and a worse overall performance.

The Orth-C-ΔC method performs better. Figure 10 shows that a number of estimates are correct within the range $\Delta_M < 30$ km and that $\text{Err}_{\text{max}}$ is reduced by $\sim 200$ km compared to the C-Δrev estimates in Fig. 8. A higher percentage of estimates are correct overall, 55.29% compared to 51.13%, and the bias of estimates is reduced, as indicated by the higher value of $S_{\text{Est}}$ and lower value of $I_{\text{Est}}$. Although there still is a bias, the overestimation is not as significant as for the C-Δrev method. The methods will now be discussed in the context of JR EEW systems, based on their potential for damage mitigation.
5. Discussion

The JR EEW systems aim to slow high-speed trains in advance of S-wave arrival using P-wave information. However, warning is also issued after S-wave arrival, meaning there is redundancy built in [2]. A previous study defined a damage index of 0 (low) to 1 (high), based on the probability of a train encountering a damaged structure in an earthquake, and investigated the effect of lead time and running speed on this index [18]. It was determined that when travelling at maximum line speed, 270 km/h, a 10 s lead time would reduce the index to 0.7. Below 10 s it remains high. To fully understand the system requirements, thought should be given to the distribution of trains’ running speeds on the network as determined by the timetable. However, for the preliminary discussion here a 10 s lead time is chosen to provide context, in the sense that this would reduce damage for the fastest travelling trains. Assuming a P-wave velocity of 6 km/h, an S-wave velocity of 3.5 km/h and 1 s for the seismic station to calculate the necessary seismic parameters, a 10 s lead time is only achievable if an earthquake occurs further than 100 km from the railway line.

Under these conditions, the Orth-C-Δc method appears most suitable for high-speed rail EEW systems in Japan, as it provides a higher percentage of correct estimations for Δ values > 100 km than both the current method and C-Δrev method (see Fig. 11). It is also more consistently correct for lower epicentral distances and Errmax is small. Although the C-Δrev method is better able to estimate Δ in the range 30-80 km, given that at this range the lead time is not sufficient to minimize damage, this is not considered as valuable as the improved accuracy afforded for longer distances by the Orth-C-Δc method. For systems travelling at slower speeds, or in areas where seismic events do not occur at such long range, the C-Δrev method may be better suited.

6. Summary

This paper evaluated different approaches to calculate the empirical relationship relating C, the initial P-wave amplitude growth rate parameter, and measured earthquake epicentral distances, Δ, by their ability to improve epicentral distance estimation accuracy. In order to do so a number of evaluation indices were introduced: RMSLE, SEst, IEst, %Corr, and Errmax. By switching the x and y variables, thereby changing the target of error minimization in linear regression, the C-Δrev method was able to improve estimation accuracy but led to biased overestimation of small epicentral distances. To overcome this, orthogonal regression was suggested as a way to minimize error in both directions thus affording a compromise between low estimation bias and low error. Although the Orth-C-Δc method performed marginally worse than the C-Δrev method, the Orth-C-Δc method was able to give comparable RMSLE and %Corr with reduced bias, while also reducing the maximum estimation error. The C-Δrev method gave a higher percentage of correct Δ estimations than the Orth-C-Δc method in the range 30-80 km. However, given the EEW systems considered in this paper deal with high-speed trains, the estimation accuracy for closer epicentral distances can be considered less important than accuracy for distances > 100 km. When travelling at high-speed, close earthquakes often do not afford enough lead time to allow meaningful damage mitigation. For these reasons, it is recommended that the Orth-C-Δc method be further investigated for implementation within JR EEW systems. Compared to the currently used method, the RMSLE was reduced from 0.314 to 0.207. The percentage of correct estimations for the analysis dataset, defined as Δrev values within Δerr ± 30%, was increased from 38.00% to 55.29%.

Acknowledgement

The authors would like to thank the Daiwa Anglo Japanese Foundation for supporting Dr. Heather Steele.

---

5.1 Limitations and further work

The measured C values have variance due to seismological and technical factors. It is suggested to assess the impact of using the Orth C-Δc method to calculate C-Δ relationships for earthquakes grouped by previously investigated seismological factors, over the current method. If the filtering parameters are improved or picking error reduced, it is also suggested to incorporate the Orth C-Δc method into the study assessing the impact of this. After epicentral distance has been estimated, the JR EEW system uses this value to estimate the magnitude of the earthquake and assess whether trains need to be stopped. Further studies should evaluate whether the improved epicentral distance values correspond to improved magnitude estimates and reduced false warnings etc. Despite the suggested methods giving improved estimations for longer epicentral distances, none of the methods perform well for short epicentral distances. The reasons for this should be investigated further.

---

Table 1 Evaluation indices for all methods

| Method   | RMSLE | SEst | IEst | %Corr | Errmax km |
|----------|-------|------|------|-------|-----------|
| Desired  | 0     | 1    | 0    | 100   | 0         |
| C-Δl     | 0.331 | 0.96 | 0.01 | 34.15 | 3059.6    |
| C-Δrev   | 0.207 | 0.45 | 1.02 | 51.13 | 404.5     |
| Orth C-Δl| 0.212 | 0.49 | 0.94 | 50.09 | 480.55    |
| C-Δc     | 0.314 | 1.00 | -0.06| 38.00 | 546.2     |
| Orth C-Δc| 0.207 | 0.57 | 0.80 | 55.29 | 209.76    |

Fig. 11 % of correct estimates for 10 km Δ ranges
and facilitating her internship at the Railway Technical Research Institute. Dr. Steele spent 18 months in Japan on the Daiwa Scholarship program, 6 months of which were spent conducting research with RTRI. The authors are also grateful to the NIED for making the K-NET data used in this study available via the K-NET site [16].

References

[1] Yamamoto, S., Sato, S., “History of earthquake early warning system for railway,” Railway Research Review, Vol. 67, No. 3, pp. 16-19, 2010 (in Japanese).

[2] Yamamoto, S., Tomori, M., “Earthquake early warning system for railways and its performance,” Journal of Japan Society of Civil Engineers, Vol. 1, No. 1, pp. 322-328, 2013.

[3] Odaka, T., Ashiya, K., et al., “A new method of quickly estimating epicentral distance and magnitude from a single seismic record,” Bulletin of the Seismological Society of America, Vol. 93, No. 1, pp. 526-532, 2003.

[4] Noda, S., Yamamoto, S., et al., “Improvement of back-azimuth estimation in real-time by using a single station record,” Earth Planets and Space, Vol. 64, No. 1, pp. 305-308, 2012.

[5] Noda, S., “Applicability of Magnitude Determination Method Considering Time Dependence of P-Wave Growth to Railway Earthquake Early Warning System,” QR of RTRI, Vol.61, No.3, pp. 210-214, 2020.

[6] Okamoto, K., Tsuno, S., “Investigation on relationship between epicentral distance and growth curve of initial P-wave propagating in local heterogeneous media for earthquake early warning system,” Earth, Planets and Space, Vol. 67, No. 167, pp. 1-8, 2015.

[7] Yamamoto, S., Noda, S., Korenaga, M., “An estimation method of epicentral distance based on characteristics of P-wave initial envelope,” RTRI Report, Vol. 26, No. 9, pp. 5-10, 2012 (in Japanese).

[8] Tsukuda S., Odaka, T., et al., “Analysis of the envelope waveform of the initial part of P-waves and its application to quickly estimating the epicentral distance and magnitude,” Zisin, Vol. 56, No. 4, pp. 351-361, 2004; (in Japanese).

[9] Noda, S., Yamamoto, S., Sato, S., “New method for estimating earthquake parameters for earthquake early warning,” QR of RTRI, Vol. 53, No. 2, pp. 102-106, 2012.

[10] Iwata, N., Yamamoto, S., et al., “Improved Algorithms of Seismic Parameters Estimation and Noise Discrimination in Earthquake Early Warning,” QR of RTRI, Vol. 56, No. 4, pp. 291-298, 2015.

[11] Eshaghi, A., Shahvar, M., et al., “Empirical relationships based on the P-wave envelope for distance and magnitude estimation for earthquake early warning in Iran,” JSEE, Vol. 19, No. 3, pp. 207-217, 2017.

[12] Okamoto, K., Tsuno, S., “Accuracy improvement in the estimation of epicentral distance inferred from initial P-waves by considering local heterogeneity,” Exploration Geophysics, Vol. 50, No. 2, pp. 159-171, 2019.

[13] Allen, R., “Automatic earthquake recognition and timing from single traces,” Bulletin of the Seismological Society of America, Vol. 68, No. 5, pp. 1521-1532, 1978.

[14] WolframAlpha, search query [solve for D \log C = A \log D + B + E^D], https://www.wolframalpha.com/ [Accessed 8 11 2019].

[15] Weisstein, E., Lambert W-Function, http://mathworld. wolfram.com/LambertW-Function.html [Accessed 1 3 2019].

[16] National Research Institute for Earth Science and Disaster Resilience, NIED K-NET, KiK-net, doi:10.17598/ NIED.0004, 2019.

[17] Noda, S., Ellsworth, W., “Determination of Earthquake Magnitude for Early Warning from the Time Dependence of P-Wave Amplitudes,” Bulletin of the Seismological Society of America, Vol. 107, No. 4, pp. 1860-1867, 2017.

[18] Iwata, N., Yamamoto, S., Ashiya, K., “Quantitative evaluation of damage mitigation on running trains by earthquake early warning,” presented at the 15th World Conference on Earthquake Engineering, Lisbon, Portugal, September 24-28, 2012.

Authors

Heather STEELE, Dr. Visiting Researcher, Seismic Data Analysis Laboratory, Center for Railway Earthquake Engineering Research (Former) Research Fellow, Birmingham Center for Railway Research and Education, University of Birmingham (Current) Research Areas: Simulation and Modelling, Data Analysis, Data Visualization

Shunta NODA Assistant Senior Researcher, Seismic Data Analysis Laboratory, Center for Railway Engineering Research Research Areas: Earthquake Engineering, Seismology

Shunroku YAMAMOTO, Dr. Eng. Director, Center for Railway Earthquake Engineering Research Research Areas: Earthquake Engineering

Masahiro KORENAGA Senior Researcher, Seismic Data Analysis Laboratory, Center for Railway Earthquake Engineering Research Research Areas: Earthquake Engineering