
Site-Specific Effects on Rapid Shaking Intensity Estimation for Earthquake Early Warning in Aotearoa New Zealand

*Dr. Chandrakumar Chanthujan**

Data Scientist, Global Seismic Data, Wellington, New Zealand.

Dr. A Mugesh

Project Scientist, Central Building Research Institute (CBRI-CSIR), Roorkee, India

ABSTRACT

Rapid estimation of expected shaking intensity from the first few seconds of an earthquake's P-wave is crucial for effective Earthquake Early Warning (EEW). This study investigates the feasibility of predicting the Modified Mercalli Intensity (MMI) using only the initial three seconds of P-wave data, considering both a combined and site-specific classified dataset across New Zealand. A dataset of 2,155 GeoNet-recorded earthquakes with magnitudes ranging from 3.0 to 6.6 recorded during the period 2012–2025 was analysed to establish empirical relationships between early P-wave amplitude parameters, peak ground acceleration (P_a), peak ground velocity (P_v), and peak ground displacement (P_d) and shaking intensity expressed as MMI. The analysis was performed for the combined dataset as well as individually for site classes B, C, and D. Univariate analysis shows that the P-wave's P_v is the most reliable single predictor of MMI across all site classes. Site-class-specific relationships show distinct MMI = 4 thresholds, demonstrating that a single nationwide threshold is not suitable for consistent and reliable EEW performance across different site conditions. Further multivariate analysis improves intensity estimation. Across all site classes, the three-parameter model combining P_a , P_v , and P_d provides the best overall performance, with R^2 values of approximately 0.63–0.66 and RMSE values of 0.55–0.57 MMI units. The multivariate threshold curves indicate that each site class requires a distinct combination of P-wave amplitudes to reach the same intensity level. The findings confirm that site-class-specific empirical relationships and alert thresholds are essential for reliable and timely EEW.

1 INTRODUCTION

Earthquake Early Warning (EEW) systems are designed to provide advance notice of impending strong ground shaking by detecting an earthquake in its earliest stages and issuing alerts before damaging seismic waves arrive at populated or critical infrastructure locations. Although earthquakes cannot be predicted in advance, EEW systems use the physical separation between the fast-travelling but weak P-waves and the slower, more damaging S-waves to gain seconds to tens of seconds of warning.

Operational EEW systems detect the initial P-wave arrivals at seismic stations and rapidly extract amplitude and frequency features from the first few seconds of motion. These early measurements are then used to estimate key earthquake characteristics, such as magnitude, source location, or expected shaking intensity at target sites, and to generate alerts before the arrival of strong S-wave shaking (Allen et al., 2009; Jin et al., 2013; Kanamori, 2005; Peng et al., 2011; Wang & Zhao, 2018; Wu & Hiroo Kanamori, 2008; Wu & Zhao, 2006a; Zhang et al., 2016; Zollo et al., 2009).

Central to this process are empirical relationships that link early P-wave parameters—such as peak ground acceleration (P_a), peak ground velocity (P_v), and peak ground displacement (P_d) to later ground-motion measures or intensity metrics (Abdalzaher et al., 2023; Wu & Kanamori, 2005a, 2005b; Wu & Mittal, 2021). Since only a limited portion of the waveform is available in real time, these empirical models must be both physically meaningful and statistically robust. Their accuracy directly controls the reliability of alert timing and intensity estimation, making them a critical component of an EEW framework.

Another important factor when it comes to ground shaking is local geological conditions at the earth's surface which influences the wave amplification, frequency content and duration of the shaking (Shoji et al., 2005). To capture these effects, stations are commonly grouped into site classes based on near-surface properties such as shear-wave velocity and soil stiffness. In New Zealand, GeoNet classifies sites into categories ranging from stiff rock (e.g., class A) to softer soil conditions (e.g., classes C, D and E) (Kaiser et al., 2017). These site effects systematically modify the relationship between early P-wave amplitudes and subsequent shaking intensity. However, in EEW systems used globally, the empirical relationships are derived from pooled datasets that combine observations across multiple site classes, implicitly assuming uniform behaviour (Wu & H Kanamori, 2008). This approach masks site-specific differences and can lead to systematic over- or underestimation of shaking, particularly when fixed nationwide thresholds are applied.

Recent research conducted by Chandrakumar et al. (2024) explored the potential to estimate the damaging S-wave amplitude before it arrives only focusing on a single site class D using the initial P-wave signals. Similarly, Mugesh et al. (2022) presented an approach for predicting earthquake magnitudes using the parameter P_d , while accounting for variations in local site conditions based on site class, in the context of EEW and showed that site class influences magnitude estimation using Japan's ground motion data.

In this context, this paper investigates the feasibility of estimating MMI using only the first three seconds of P-wave data, with explicit consideration of site-class effects across New Zealand. Using an extensive set of GeoNet strong-motion waveforms, we develop univariate and multivariate empirical relationships linking early P-wave amplitude measures (P_a , P_v , and P_d) to subsequent shaking intensity, both for a pooled dataset and for individual Site Classes B, C, and D. We further assess how these relationships translate into EEW alert thresholds for a target intensity of $\text{MMI} = 4$, showing that the combinations of P_a , P_v and P_d required to reach this intensity vary systematically between site classes. The results demonstrate that a single universal relationship or threshold is insufficient for reliable EEW performance and that site-class-specific formulations are required for accurate calibration.

2 DATASET SELECTION AND ANALYSIS

The dataset used in this study was compiled from earthquakes that occurred across New Zealand. Events were selected based on epicentral locations constrained within the geographic bounds of latitudes -47.5617 to -34.2192 and longitudes 165.8271 to 179.6050 . Earthquake waveforms were retrieved using the GeoNet FDSN communication service, drawing on the national strong-motion sensor network (GNS Science, 2022). The study focuses on the period from January 2012 to December 2025. This timeframe was chosen because it corresponds to the interval during which GeoNet consistently catalogued and archived both P-wave and S-wave arrival picks for the majority of strong-motion ground recordings. The availability of these pre-existing phase picks was critical, as it substantially reduced the need for manual phase identification across the large volume of waveform data analysed. Waveforms that did not contain either a P-wave pick or an S-wave pick were removed from the dataset. The final dataset comprised 26,765 earthquake waveforms recorded by MEMS-based accelerometers, representing 4,446 earthquakes with magnitudes greater than 3 ($M > 3$).

As a first step, all waveforms were grouped using the site-classification metadata provided by GeoNet (Kaiser et al., 2017). To ensure data quality and relevance, only ground-motion records from stations located within 50 km of the earthquake epicentre were selected. This distance limit is consistent with previous studies and was chosen for two main reasons. First, waveforms closer to the epicentre reduce interference from later seismic phases and coda waves, allowing clearer identification of P- and S-wave arrivals. Second, near-source recordings generally exhibit higher signal-to-noise ratios, which improves the reliability of the analysis.

Further selection criteria were applied to strengthen the dataset. Only earthquakes recorded by at least three stations were retained, and stations that recorded fewer than three earthquake events were excluded. After applying these filters, the final dataset consists of 2,155 earthquake events with magnitudes ranging from 3.0 to 6.6, producing a total of 10,350 ground-motion waveforms (Table 1). The final dataset includes 96 waveforms from Site Class A, 3,261 from Site Class B, 1,524 from Site Class C, 4,146 from Site Class D, and 233 from Site Class E. The study therefore focuses on Site Classes B, C, and D. Due to the limited number of waveforms in Site Classes A and E, these classes were excluded from further analysis. Following this exclusion, the final dataset used in the study consists of 2,145 earthquake events from Site Classes B, C, and D.

Table 1. Distribution of Earthquake Events by Magnitude Range.

Magnitude (M) range	Number of events
3 to 4	1698
4 to 5	394
5 and above	53
Total	2145

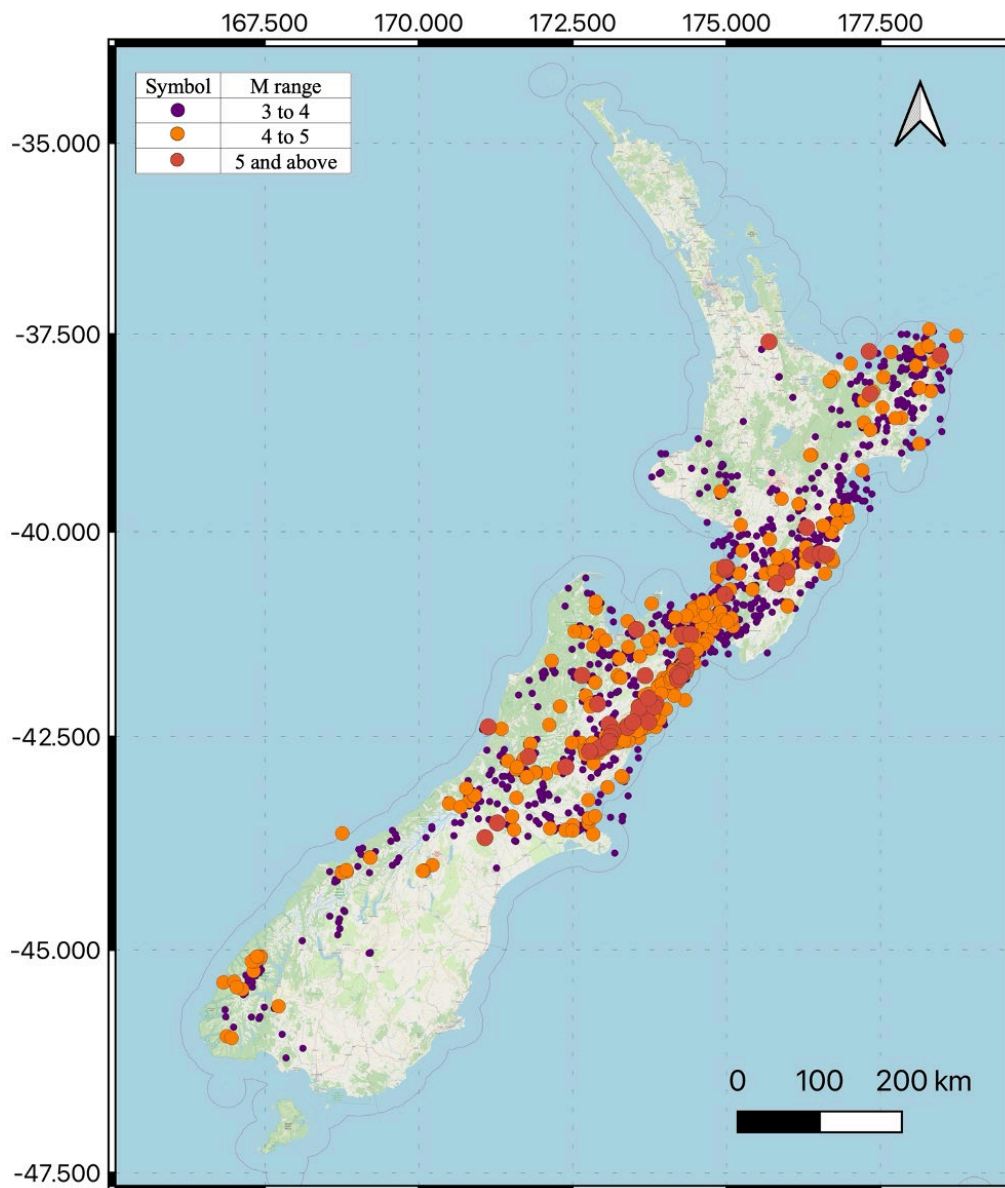


Figure 1. Epicentre locations of the earthquakes and their spatial distribution across New Zealand.

3 METHOD

3.1 Parameters used for analysis

The analysis focuses on two distinct time windows: (1) the first three seconds following the P-wave arrival, and (2) the subsequent window corresponding to the arrival of the S-wave. The selection of a 3-s P-wave window is well supported in the literature, as it represents an effective compromise between reliable amplitude estimation and maximising the available warning time while reducing the EEW blind zone (Böse et al., 2009; Caruso et al., 2017; Wang et al., 2020; Wu & Mittal, 2021).

Standard ground-motion parameters Peak Ground Acceleration, Peak Ground Velocity, and Peak Ground Displacement can be used to represent the amplitude of seismic waves for an earthquake (Wu and Zhao, 2006; Wu and Kanamori, 2008). For the selected P-wave window, the amplitude of the initial motion is quantified using the peak ground acceleration (P_a), peak ground velocity (P_v), and peak ground displacement

(P_d) derived from the vertical components of acceleration, velocity, and displacement. This choice reflects the predominantly vertical particle motion of P-waves (Wu & H Kanamori, 2008; Wu & Zhao, 2006b).

For the S-wave window, instead of directly relating S-wave amplitude parameters to the P-wave metrics as in earlier studies (e.g., Chandrakumar et al. (2024), Colombelli et al. (2015), Wu & H Kanamori (2008)), this study estimates shaking intensity in terms of MMI. Peak Ground Acceleration (PGA) and peak Ground Velocity (PGV) values from the S-wave phase are converted to MMI using the Ground Motion to Intensity Conversion Equations (GMICEs) proposed by Moratalla et al. (2020). Horizontal components are used to capture peak S-wave motion, consistent with the predominantly horizontal particle motion of S-waves (Shearer, 2009). PGA and PGV of the S-waves are computed using the RotD50 approach (Boore, 2010). MMI is then calculated separately from PGA and PGV, referred to as MMI (Acc) and MMI (Vel), respectively. A combined average intensity is subsequently derived from both estimates, denoted as MMI (Avg).

Prior to parameter extraction, all selected ground-motion waveforms are filtered using a Butterworth band-pass filter between 0.1 and 20 Hz. This processing step preserves the frequency content relevant to earthquake ground motion while suppressing low-frequency drift and high-frequency noise (Claerbout, 1964; Virtanen et al., 2020).

Table 2. Summary of amplitude and intensity parameters including their definitions and abbreviations.

Abbreviation	Description
P_a	Peak Ground Acceleration calculated for the P-wave window
P_v	Peak Ground Velocity calculated for the P-wave window
P_d	Peak Ground Displacement calculated for the P-wave window
MMI (Acc)	MMI calculated from S-wave Peak Ground Acceleration
MMI (Vel)	MMI calculated from S-wave Peak Ground Velocity
MMI (Avg)	Average Modified Mercalli Intensity derived from MMI (Acc) and MMI (Vel)

3.2 Linear regression analysis

This study adopts linear regression to model the relationship between early P-wave amplitudes and subsequent shaking intensity. Linear regression is widely used in seismology because of its transparency, robustness, and ease of implementation, making it well-suited for establishing reliable baseline relationships between seismic parameters.

The choice of a simple regression framework is driven by both practical and scientific considerations. Although machine-learning methods such as neural networks can capture complex, non-linear behaviour, they typically require very large and diverse datasets to generalise reliably. In this study, the available dataset consists of MEMS-based strong-motion recordings with limited coverage across some site classes. Under such conditions, complex models risk overfitting and producing unstable predictions, which is undesirable for operational EEW applications where reliability and interpretability are critical. Linear regression, therefore, provides a more appropriate and dependable modelling approach for the present dataset.

To improve model stability, all ground-motion amplitude parameters are transformed into $\log_{10}$ before regression. This reduces the influence of extreme values, stabilises variance and promotes a more linear

relationship between predictors and targets. Model performance is evaluated using the coefficient of determination (R^2) and Root Mean Square Error (RMSE), providing a quantitative basis for comparing predictive skill across site classes.

4 RESULTS AND DISCUSSION

Before establishing empirical relationships between P-wave amplitude parameters and MMI (Avg), all independent and dependent variables were subjected to a systematic outlier screening process to enhance the robustness and reliability of the models. Outliers, which can disproportionately influence regression results, were identified and removed using the Interquartile Range (IQR) method (Courtney Taylor, 2018; Gianluca Malato, 2021). This approach ensures that the derived relationships are representative of the underlying physical trends in the data rather than being driven by a small number of extreme observations.

4.1 Single amplitude parameters vs MMI (Avg)

As an initial step in the analysis, individual P-wave amplitude parameters were used to estimate earthquake intensity, following a similar approach to Chandrakumar et al. (2024). While that earlier study focused on the Canterbury region and examined relationships between P- and S-wave amplitudes on for Site Class D, the present study extends the analysis to the whole of New Zealand and places emphasis on directly estimating shaking intensity in terms of MMI.

Table 3. Univariate empirical relationships between P-wave amplitude parameters (P_a , P_v , and P_d) and MMI (Avg) for Site Classes B, C, and D, and for the pooled dataset (ALL).

Site Class	Predictor	Empirical Relationship	R^2	RMSE
B	P_a	$\text{MMI (Avg)} = 1.34 \log(P_a) + 2.20$	0.55	0.6
B	P_v	$\text{MMI (Avg)} = 1.45 \log(P_v) + 4.64$	0.61	0.56
B	P_d	$\text{MMI (Avg)} = 1.32 \log(P_d) + 5.61$	0.53	0.62
C	P_a	$\text{MMI (Avg)} = 1.41 \log(P_a) + 2.28$	0.58	0.61
C	P_v	$\text{MMI (Avg)} = 1.54 \log(P_v) + 4.95$	0.62	0.58
C	P_d	$\text{MMI (Avg)} = 1.43 \log(P_d) + 6.09$	0.52	0.65
D	P_a	$\text{MMI (Avg)} = 1.42 \log(P_a) + 2.23$	0.58	0.6
D	P_v	$\text{MMI (Avg)} = 1.56 \log(P_v) + 4.86$	0.65	0.55
D	P_d	$\text{MMI (Avg)} = 1.43 \log(P_d) + 5.92$	0.52	0.64
ALL	P_a	$\text{MMI (Avg)} = 1.39 \log(P_a) + 2.24$	0.58	0.6
ALL	P_v	$\text{MMI (Avg)} = 1.50 \log(P_v) + 4.77$	0.63	0.56
ALL	P_d	$\text{MMI (Avg)} = 1.36 \log(P_d) + 5.77$	0.52	0.64

Across all site classes as shown in Table 3, clear differences are observed in the performance of the three P-wave parameters. For Site Class B, P_v provides the strongest relationship with MMI, showing a higher R^2

and lower RMSE than both P_a and P_d . A similar pattern is observed for Site Classes C and D, where P_v consistently outperforms the other parameters and produces the most accurate intensity estimates. In contrast, P_d shows the weakest performance for all site classes, with lower R^2 values and higher prediction errors, indicating that P-wave displacement alone is less suitable for estimating shaking intensity. A summary of the best-performing predictor for each site class is presented in Table 4. P_v is identified as the optimal predictor for Site Classes B, C, and D, as well as for the combined dataset including all site classes. This indicates that the P-wave's P_v provides the most stable and reliable link to perceived shaking intensity across a range of site conditions.

Table 4. Best-performing univariate P-wave predictor for estimating MMI (Avg) for each site class and for the pooled dataset.

Site Class	Best Predictor	R^2	RMSE
B	P_v	0.61	0.56
C	P_v	0.62	0.58
D	P_v	0.65	0.55
ALL	P_v	0.63	0.56

These results are broadly consistent with Chandrakumar et al. (2024), who found P_v to be the most reliable P-wave parameter for estimating S-wave amplitudes at soft sites using the Canterbury dataset. The present study extends this finding by showing that P_v is also the most effective parameter for directly predicting MMI across New Zealand, even when site effects are considered explicitly. The use of different empirical relationships for each site class highlights the strong influence of site effects on EEW. The results show that using a single relationship to estimate intensity across all site classes is not optimal. Instead, the relationship used to estimate intensity should reflect local site conditions to improve accuracy and reliability.

To further investigate this, all site classes were combined into a single pooled dataset to derive a nationwide empirical relationship. For the pooled dataset, P_v again emerged as the most effective predictor of MMI, producing the highest R^2 and lowest RMSE among the three P-wave parameters. While this pooled relationship provides a useful first-order approximation for nationwide application, it conceals important site-dependent behaviour that becomes evident when the site classes are analysed separately. Figure 2 illustrates the $\text{MMI} = 4$ alert thresholds derived from the univariate P_v –MMI relationships for Site Classes B, C, and D, together with the pooled (ALL) dataset. The figure clearly shows that the P_v threshold required to reach the same intensity level differs across site classes, demonstrating that alert thresholds are sensitive to local site conditions.

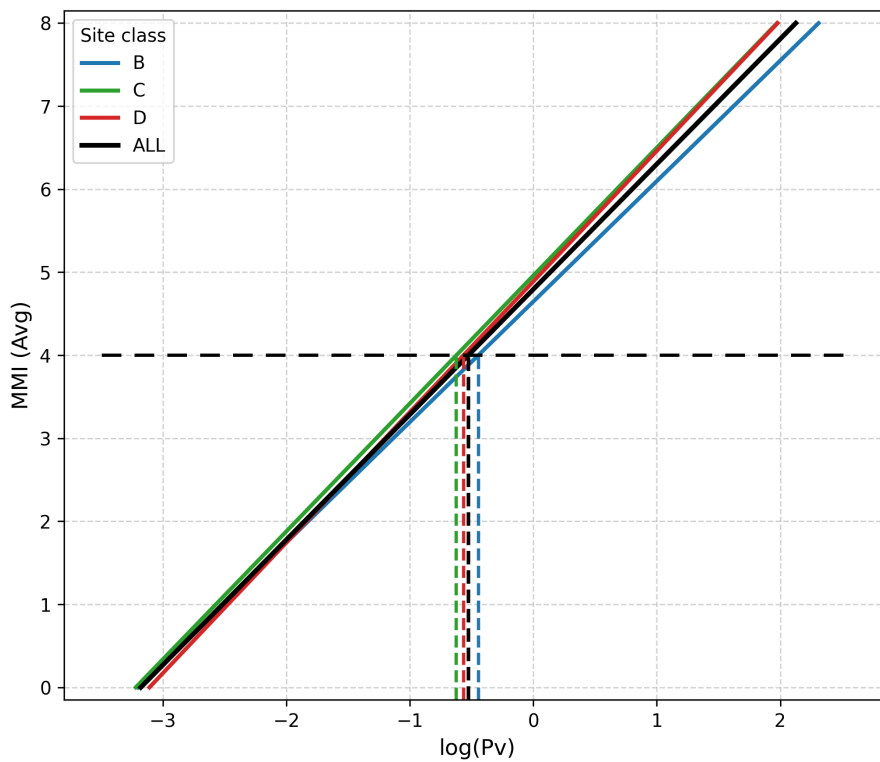


Figure 2. Univariate P_v –MMI relationships showing the $MMI = 4$ alert thresholds for Site Classes B, C, and D, together with the pooled (ALL) dataset.

A clear separation is observed among the site classes, highlighting the influence of local site conditions on the conversion of early P-wave motion into perceived shaking intensity. Site Class D (soft soil) reaches the same intensity level at a lower $\log_{10}(P_v)$ threshold compared to Site Classes C and B, indicating stronger amplification effects and higher sensitivity to incoming P-wave energy. Site Class C exhibits intermediate behaviour, while Site Class B (stiffer soil/rock sites) requires the highest P_v threshold to achieve $MMI = 4$, reflecting reduced amplification and greater attenuation of ground motion.

The pooled “ALL” relationship lies between the site-class-specific curves and represents an average response across all site conditions. However, this averaged threshold conceals systematic differences between individual site classes. As a result, applying a single pooled threshold would produce inconsistent alerting behaviour, with delayed warnings at softer sites and potentially premature or missed alerts at stiffer sites.

For a target intensity of $MMI = 4$, the derived P_v thresholds are 0.36 cm s^{-1} for Site Class B, 0.24 cm s^{-1} for Site Class C, and 0.27 cm s^{-1} for Site Class D, while the pooled dataset yields a threshold of 0.30 cm s^{-1} . Using this pooled value uniformly would delay warnings at softer sites, particularly for Site Class C, thereby reducing the available warning time.

Overall, these results demonstrate that a single universal alert threshold is not suitable for EEW applications. Instead, site-class-specific P_v thresholds are required to ensure consistent, timely, and reliable warning performance across different geological conditions. Incorporating site effects into threshold selection improves intensity prediction, reduces false or delayed alerts, and enhances the robustness of EEW systems operating across heterogeneous site conditions.

4.2 Multi-amplitude parameter vs MMI

Following the single-parameter exploration of shake intensity (MMI), the analysis was extended to examine the relationship between MMI and multiple P-wave amplitude measures. This approach is similar to the study by Colombelli et al. (2015) in which a multivariate framework was developed to estimate S-wave amplitudes using three P-wave parameters (P_a , P_v , and P_d). Building on this methodology, multivariate regression was performed separately for Site Classes B, C, and D, as well as for the pooled dataset including all site classes. Initially, two-parameter combinations (P_a – P_v , P_a – P_d , and P_v – P_d) were tested, followed by a three-parameter model incorporating P_a , P_v , and P_d . The resulting empirical relationships and their performance metrics are summarised in Table 5.

Table 5. Multivariate empirical relationships between early P-wave amplitude parameters (P_a , P_v , and P_d) and MMI (Avg), derived for Site Classes B, C, and D, and for the pooled dataset.

Site Class	Predictors	Empirical Relationship	R ²	RMSE
B	P_a , P_v	$\text{MMI (Avg)} = 0.16 \log(P_a) + 1.30 \log(P_v) + 4.40$	0.61	0.56
B	P_a , P_d	$\text{MMI (Avg)} = 0.80 \log(P_a) + 0.70 \log(P_d) + 4.12$	0.61	0.56
B	P_v , P_d	$\text{MMI (Avg)} = 1.16 \log(P_v) + 0.32 \log(P_d) + 5.02$	0.62	0.56
B	P_a , P_v , P_d	$\text{MMI (Avg)} = 0.34 \log(P_a) + 0.76 \log(P_v) + 0.41 \log(P_d) + 4.59$	0.63	0.55
C	P_a , P_v	$\text{MMI (Avg)} = 0.17 \log(P_a) + 1.36 \log(P_v) + 4.65$	0.62	0.58
C	P_a , P_d	$\text{MMI (Avg)} = 0.94 \log(P_a) + 0.65 \log(P_d) + 4.07$	0.62	0.58
C	P_v , P_d	$\text{MMI (Avg)} = 1.25 \log(P_v) + 0.34 \log(P_d) + 5.38$	0.63	0.57
C	P_a , P_v , P_d	$\text{MMI (Avg)} = 0.33 \log(P_a) + 0.87 \log(P_v) + 0.39 \log(P_d) + 4.88$	0.64	0.57
D	P_a , P_v	$\text{MMI (Avg)} = 0.06 \log(P_a) + 1.50 \log(P_v) + 4.76$	0.65	0.55
D	P_a , P_d	$\text{MMI (Avg)} = 0.96 \log(P_a) + 0.68 \log(P_d) + 4.06$	0.64	0.56
D	P_v , P_d	$\text{MMI (Avg)} = 1.35 \log(P_v) + 0.26 \log(P_d) + 5.19$	0.65	0.54
D	P_a , P_v , P_d	$\text{MMI (Avg)} = 0.23 \log(P_a) + 1.08 \log(P_v) + 0.31 \log(P_d) + 4.89$	0.66	0.54

ALL	P_a, P_v	$\text{MMI (Avg)} = 0.20 \log(P_a) + 1.31 \log(P_v) + 4.45$	0.63	0.56
ALL	P_a, P_d	$\text{MMI (Avg)} = 0.92 \log(P_a) + 0.64 \log(P_d) + 3.98$	0.63	0.56
ALL	P_v, P_d	$\text{MMI (Avg)} = 1.27 \log(P_v) + 0.27 \log(P_d) + 5.10$	0.64	0.56
ALL	P_a, P_v, P_d	$\text{MMI (Avg)} = 0.37 \log(P_a) + 0.84 \log(P_v) + 0.35 \log(P_d) + 4.61$	0.65	0.56

Across all site classes, the multivariate models outperform the corresponding univariate relationships. The inclusion of multiple P-wave amplitude parameters leads to higher R^2 values and lower RMSE, demonstrating that different amplitude measures capture complementary aspects of the early P-wave signal. For Site Class B, modest but improvements are observed when moving from single-parameter to multivariate formulations. The three-parameter model ($P_a + P_v + P_d$) provides the best overall performance ($R^2 = 0.63$, $\text{RMSE} = 0.55$), slightly improving upon the two-parameter combinations. This suggests that incorporating displacement information alongside acceleration and velocity contributes to better estimation, even at stiffer sites. Similarly, for Site Classes C and D, the benefits of multivariate modelling are more pronounced. In both cases, the three-parameter formulation produces the highest R^2 and lowest RMSE, with performance improving progressively as additional parameters are included. This indicates that softer sites are more sensitive to a broader range of motion characteristics, including both peak amplitudes and longer-period components captured by P_v and P_d .

A summary of the best-performing multivariate model for each site class is provided in Table 6. In all cases, the combination of P_a , P_v , and P_d yields the strongest relationship with MMI. The pooled dataset also shows improved performance with the three-parameter model, although this approach conceals important site-dependent behaviour observed in the individual site-class analyses.

Table 6. Best-performing multivariate P-wave amplitude models for estimating MMI (Avg) for each site class and for the pooled dataset.

Site Class	Best Multivariate Model	R^2	RMSE
B	$P_a + P_v + P_d$	0.63	0.55
C	$P_a + P_v + P_d$	0.64	0.57
D	$P_a + P_v + P_d$	0.66	0.54
ALL	$P_a + P_v + P_d$	0.65	0.56

Figure 3 further supports the key argument of this study by comparing site-class-specific multivariate relationships with the pooled all-data relationship. The figure highlights the differences in EEW alert thresholds for a common target intensity of $\text{MMI} = 4$.

Visualising the full three-parameter relationship (P_a , P_v , and P_d) in a single figure is not straightforward. To enable clear comparison, P_d was fixed at a constant value of 0.35 cm, and the resulting relationships between

P_a and P_v were plotted for each site class. This P_d value follows Wu & H Kanamori (2008), who adopted 0.35 cm as a representative P-wave displacement threshold for EEW applications.

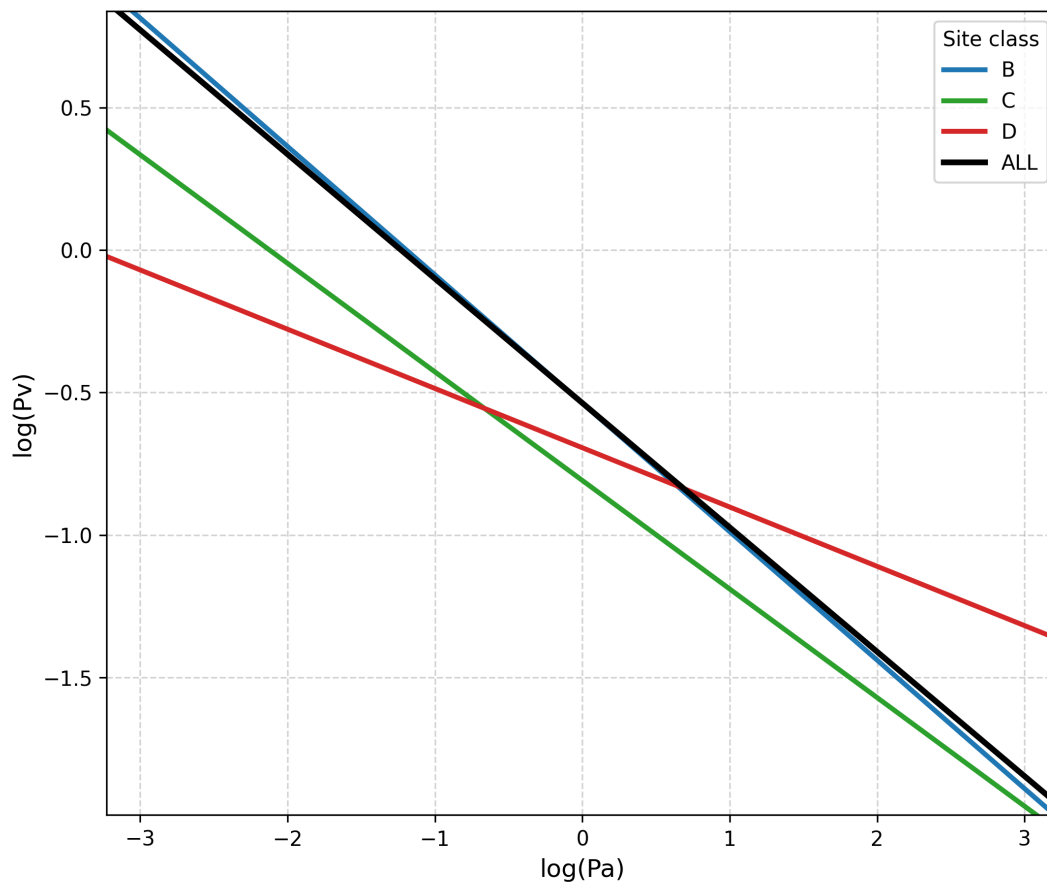


Figure 3. Multivariate MMI (Avg) = 4 alert threshold curves derived from the P_a – P_v – P_d relationships for Site Classes B, C, and D, and for the pooled (ALL) dataset. P_d is fixed at 0.35 cm, allowing comparison of site-class-specific threshold behaviour in the P_a , P_v space.

In Figure 3, while all relationships show a broadly consistent negative slope, clear site-dependent differences are evident in both offset and curvature. Site Class B (stiffer soil and rock sites) generally exhibits higher PGV values for a given PGA, reflecting a more efficient transfer of high-frequency acceleration into velocity components. Site Class C displays intermediate behaviour between the stiff and soft site classes. In contrast, Site Class D (soft soil) shows a distinct departure from the other classes at higher PGA levels ($\log_{10}(P_a) \geq 1$), where the relationship becomes noticeably flatter and produces higher PGV values than both the pooled regression and the stiffer-site relationships. This divergence at higher PGA levels suggests a nonlinear site response at soft-soil sites, likely associated with soil softening and strain-dependent damping effects that modify the coupling between acceleration and velocity under strong shaking. The pooled relationship closely follows the Site Class B trend at lower PGA levels but fails to capture this high-PGA behaviour observed for Site Class D, thereby obscuring important site-specific effects.

Overall, the systematic separation between site classes and the divergence observed at strong shaking levels highlight the critical role of local site conditions in controlling the PGA–PGV scaling. These findings reinforce the need for site-class-dependent relationships in ground-motion characterisation and EEW intensity estimation, as pooled models may not reliably represent the behaviour of softer sites during strong motion.

5 CONCLUSION

This study evaluated the feasibility of rapidly estimating shaking intensity from the first three seconds of P-wave recordings for EEW in New Zealand using strong-motion data from the GeoNet network for earthquakes with magnitudes between 3.0 and 6.6. Empirical relationships were developed between early P-wave amplitudes (P_a , P_v , and P_d) and MMI (Avg) for Site Classes B, C, and D, with explicit consideration of site-dependent effects. Univariate analysis identified peak P-wave's P_v as the most reliable single predictor of MMI, consistently yielding higher coefficients of determination and lower prediction errors than P_a or P_d across all site classes. While a pooled nationwide P_v –MMI relationship can be derived, the corresponding MMI = 4 thresholds differ systematically among site classes, indicating that a single national threshold would lead to inconsistent warning performance and delayed alerts at softer sites, thereby reducing available lead time.

Multivariate formulations further improved intensity estimation. Models combining P_a , P_v , and P_d achieved the best overall performance, with R^2 values of approximately 0.63–0.66 and RMSE values of about 0.55–0.57 MMI units. Although the inclusion of P_d provided only modest gains over P_a – P_v combinations, the three-parameter model produced the most stable and robust results. Site-class-specific MMI = 4 threshold curves demonstrate that each class requires a distinct combination of P-wave amplitudes to reach the same intensity level, while the pooled relationship represents an average response that does not correspond to any individual site condition. These results highlight the decisive role of local site conditions in controlling the relationship between early P-wave motion and perceived shaking intensity. Stiffer sites (Class B) generally require higher P-wave amplitude thresholds to reach a given MMI level, reflecting reduced amplification, whereas softer sites (Class D) reach equivalent intensity at lower thresholds due to enhanced amplification and strain-dependent soil response, with Class C exhibiting intermediate behaviour.

Overall, this study highlights the importance of site effects in improving both the physical consistency and statistical robustness of P-wave-based intensity estimation for EEW. The proposed site-class-specific univariate and multivariate relationships provide a practical framework for improving EEW performance in New Zealand. While pooled relationships may serve as a useful first-order approximation, they are inadequate for operational deployment, where accurate, location-specific alert thresholds are essential to ensure reliable and equitable warning performance across diverse geological conditions.

References

- Abdalzaher, M. S., Soliman, M. S., & El-Hady, S. M. (2023). Seismic Intensity Estimation for Earthquake Early Warning Using Optimized Machine Learning Model. *IEEE Transactions on Geoscience and Remote Sensing*, 61. <https://doi.org/10.1109/TGRS.2023.3296520>
- Allen, R. M., Gasparini, P., Kamigaichi, O., & Böse, M. (2009). The status of earthquake early warning around the World: An introductory overview. *Seismological Research Letters*, 80(5), 682–693. <https://doi.org/10.1785/gssrl.80.5.682>
- Boore, D. M. (2010). Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. In *Bulletin of the Seismological Society of America* (Vol. 100, Number 4). <https://doi.org/10.1785/0120090400>
- Böse, M., Hauksson, E., Solanki, K., Kanamori, H., & Heaton, T. H. (2009). Real-time testing of the on-site warning algorithm in southern California and its performance during the July 29 2008 Mw5.4 Chino Hills earthquake. *Geophysical Research Letters*, 36(5). <https://doi.org/10.1029/2008GL036366>
- Caruso, A., Colombelli, S., Elia, L., Picozzi, M., & Zollo, A. (2017). An on-site alert level early warning system for Italy. *Journal of Geophysical Research: Solid Earth*, 122(3). <https://doi.org/10.1002/2016JB013403>
- Chandrakumar, C., Tan, M. L., Holden, C., Stephens, M. T., Punchihewa, A., & Prasanna, R. (2024). Estimating S-wave amplitude for earthquake early warning in New Zealand: Leveraging the first 3 seconds of P-Wave. *Earth Science Informatics*. <https://doi.org/10.1007/s12145-024-01403-6>
- Claerbout, J. F. (1964). Detection of P-waves from weak sources at great distances. *Geophysics*, 29(2), 197–211. <https://doi.org/10.1190/1.1439350>
- Colombelli, S., Caruso, A., Zollo, A., Festa, G., & Kanamori, H. (2015). A P wave-based, on-site method for earthquake early warning. *Geophysical Research Letters*, 42(5). <https://doi.org/10.1002/2014GL063002>
- Courtney Taylor. (2018, April 26). *What Is the Interquartile Range Rule?* <https://www.thoughtco.com/what-is-the-interquartile-range-rule-3126244>
- Gianluca Malato. (2021, November 1). *Outlier identification using Interquartile Range*. Towards Data Science. <https://towardsdatascience.com/outlier-identification-using-interquartile-range-74f5de12932a>
- GNS Science. (2022). *GeoNet Aotearoa New Zealand Strong Motion Data Products*. <https://data.gns.cri.nz/metadata/srv/eng/catalog.search#/metadata/25c52e65-dbf9-4687-8343-3ca0b60961c1>
- Jin, X., Zhang, H., Li, J., Wei, Y., & Ma, Q. (2013). Earthquake magnitude estimation using the τ_c and P d method for earthquake early warning systems. *Earthquake Science*, 26(1), 23–31. <https://doi.org/10.1007/s11589-013-0005-4>
- Kaiser, A., Van Houtte, C., Perrin, N., Wotherspoon, L., & Mcverry, G. (2017). Site characterisation of geonet stations for the New Zealand strong motion database. *Bulletin of the New Zealand Society for Earthquake Engineering*, 50(1). <https://doi.org/10.5459/bnzsee.50.1.39-49>
- Kanamori, H. (2005). Real-time seismology and earthquake damage mitigation. *Annual Review of Earth and Planetary Sciences*, 33, 195–214. <https://doi.org/10.1146/annurev.earth.33.092203.122626>
- Moratalla, J. M., Goded, T., Rhoades, D. A., Canessa, S., & Gerstenberger, M. C. (2020). New ground motion to intensity conversion equations (GMICEs) for New Zealand. *Seismological Research Letters*, 92(1). <https://doi.org/10.1785/0220200156>

- Mugesh, A., Desai, A., Jakka, R. S., & Kamal. (2022). Local Site Effects Influence on Earthquake Early-Warning Parameters. *Lecture Notes in Civil Engineering*, 187. https://doi.org/10.1007/978-981-16-5669-9_10
- Peng, H., Wu, Z., Wu, Y. M., Yu, S., Zhang, D., & Huang, W. (2011). Developing a prototype earthquake early warning system in the Beijing capital region. *Seismological Research Letters*, 82(3), 394–403. <https://doi.org/10.1785/gssrl.82.3.394>
- Shearer, P. M. (2009). Introduction to seismology, second edition. *Cambridge University Press*, 4(3).
- Shoji, Y., Tanii, K., & Kamiyama, M. (2005). A study on the duration and amplitude characteristics of earthquake ground motions. *Soil Dynamics and Earthquake Engineering*, 25(7–10). <https://doi.org/10.1016/j.soildyn.2004.11.033>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3). <https://doi.org/10.1038/s41592-019-0686-2>
- Wang, Y., Li, S., & Song, J. (2020). Threshold-based evolutionary magnitude estimation for an earthquake early warning system in the Sichuan–Yunnan region, China. *Scientific Reports*, 10(1), 1–12. <https://doi.org/10.1038/s41598-020-78046-2>
- Wang, Z., & Zhao, B. (2018). A new Mw estimation parameter for use in earthquake early warning systems. *Journal of Seismology*, 22(1), 325–335. <https://doi.org/10.1007/s10950-017-9708-7>
- Wu, Y. M., & Kanamori, H. (2005a). Experiment on an onsite early warning method for the Taiwan early warning system. *Bulletin of the Seismological Society of America*, 95(1), 347–353. <https://doi.org/10.1785/0120040097>
- Wu, Y. M., & Kanamori, H. (2005b). Rapid assessment of damage potential of earthquakes in Taiwan from the Beginning of P waves. *Bulletin of the Seismological Society of America*, 95(3), 1181–1185. <https://doi.org/10.1785/0120040193>
- Wu, Y. M., & Kanamori, Hiroo. (2008). Development of an earthquake early warning system using real-time strong motion signals. *Sensors*, 8(1), 1–9. <https://doi.org/10.3390/s8010001>
- Wu, Y. M., & Kanamori, H. (2008). Development of an earthquake early warning system using real-time strong motion signals. *Sensors*, 8(1), 1–9. <https://doi.org/10.3390/s8010001>
- Wu, Y. M., & Mittal, H. (2021). A Review on the Development of Earthquake Warning System Using Low-Cost Sensors in Taiwan. *Sensors (Basel, Switzerland)*, 21(22). <https://doi.org/10.3390/s21227649>
- Wu, Y. M., & Zhao, L. (2006a). Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophysical Research Letters*, 33(16), 4–7. <https://doi.org/10.1029/2006GL026871>
- Wu, Y. M., & Zhao, L. (2006b). Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophysical Research Letters*, 33(16), 4–7. <https://doi.org/10.1029/2006GL026871>
- Zhang, H., Jin, X., Wei, Y., Li, J., Kang, L., Wang, S., Huang, L., & Yu, P. (2016). An earthquake early warning system in Fujian, China. *Bulletin of the Seismological Society of America*, 106(2), 755–765. <https://doi.org/10.1785/0120150143>

Zollo, A., Iannaccone, G., Lancieri, M., Cantore, L., Convertito, V., Emolo, A., Festa, G., Gallovič, F., Vassallo, M., Martino, C., Satriano, C., & Gasparini, P. (2009). Earthquake early warning system in southern Italy: Methodologies and performance evaluation. *Geophysical Research Letters*, 36(5). <https://doi.org/10.1029/2008GL036689>