

Spectral decomposition of ground motions in New Zealand using the generalized inversion technique

Chuanbin Zhu¹, Sanjay Bora,² Brendon A. Bradley¹ and Dino Bindi³

¹Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch 8140, New Zealand. E-mail: chuanbin.zhu@canterbury.ac.nz

²Department of Earthquake Physics and Statistics, GNS Science, Lower Hutt 5011, New Zealand

³Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences, Potsdam 14473, Germany

Accepted 2024 May 6. Received 2024 April 15; in original form 2023 November 6

SUMMARY

To gain new insights into ground-motion phenomena in New Zealand (NZ), we apply the non-parametric generalized inversion technique (GIT) in the Fourier domain to isolate the systematic source, path, and site effects from 20 813 seismograms, recorded by 693 sensors at 439 unique locations, from 1200 shallow crustal events ($M_w > 3$) during the period 2000–2021. From the inverted source spectra, we derive Brune's stress parameter, $\Delta\sigma$, which is found to follow a lognormal distribution with a \log_{10} standard deviation of 0.36 or equivalently 0.83 in natural log unit. $\Delta\sigma$ slightly increases with focal depth and is practically independent of earthquake size (i.e. self-similar), but displays a statistically significant spatial clustering. Based on the inverted attenuation, a trilinear geometric-spreading function, and a distance-dependent quality-factor $Q(f)$ model are found to well describe the attenuation in NZ; though a single $Q(f)$ model is also obtained for the whole distance range: $Q(f) = 149.1 f^{0.62}$. Using the site response decomposed from GIT, we find that the soil classification scheme specified in NZ seismic code, NZS1170.5, has a limited capability in discerning the site-specific frequency-dependent amplification functions in comparison to a non-parametric clustering with the same number of discrete classes. The potential use of the spatial variation in source parameters from this GIT analysis in region-specific physics-based simulations is discussed.

Key words: Earthquake ground motions; Earthquake source observations; Seismic attenuation; Site effects.

1 INTRODUCTION

New Zealand (NZ) is one of the most seismically active and seismotectonically diverse countries in the world (Reyners 1989; Ristau 2008). Opposing subduction zones between the Pacific plate and the Australian plate exist at the margins of the North and South Islands, with an intense band of shallow seismicity diffuse over much of the country (e.g. Ristau 2008). During the NZ National Seismic Hazard Model 2022 update (NSHM22, Gerstenberger *et al.* 2024), a plethora of seismograms has been collected from a national seismograph network and processed (GNS Science 2022; Hutchinson *et al.* 2024). This new data set creates an opportunity to better understand ground motions and to improve their predictions.

One way to advance our understanding of ground motions is via gaining insights into the systematic effects governing them, that is, the effects of fault rupture, propagation path and near-surface geology on ground shaking, respectively. A commonly employed method to decompose ground motions into their constituent components is the generalized inversion technique (GIT, Andrews 1986).

GIT has been applied to many seismically active regions with a dense recording network, for example, Japan, California, Europe and China (e.g. Castro *et al.* 1990; Kawase & Matsuo 2004; Shearer *et al.* 2006; Drouet *et al.* 2008; Edwards *et al.* 2008; Fu *et al.* 2023).

In NZ, GIT has only been applied to specific earthquake sequences. Oth & Kaiser (2014) decomposed 2415 accelerograms from 205 events of the 2010–2011 Canterbury earthquake sequence (Bradley & Cubrinovski 2011; Kaiser *et al.* 2012) and then focused on the stress release and source scaling. They reported a median value of 5 MPa for stress parameter which was nearly independent of earthquake size but varied laterally throughout Canterbury. Ren *et al.* (2018) systematically analysed the source, path, and site effects of the 2016 M_w 7.8 Kaikōura earthquake sequence (e.g. Bradley *et al.* 2017) via spectral decomposition of 2445 recordings from 148 events recorded by 126 strong-motion stations. Brune stress parameters (Brune 1970) were found to have a geometric mean value of 1.25 MPa, which is lower than that of the Canterbury sequence (Oth & Kaiser 2014), and exhibited no dependence on earthquake magnitude, which is consistent with Oth & Kaiser

(2014). Though carefully analysed, results of the two prior investigations may be unique to the studied earthquake sequences and regions. To reveal, for instance, nationally consistent spatial patterns in parameters controlling ground shaking, it entails a comprehensive study on the national level, which is lacking in NZ.

In this research, we aim to gain a comprehensive understanding of the systematic source, path, and site effects of ground motions observed throughout NZ. To this end, we decompose the Fourier amplitude spectra (FAS) of ground-motion recordings compiled during NSHM22 using a nonparametric GIT. In the following sections, we first describe our data and method, followed by results on source, path and site components. Then we discuss potential directions for more in-depth analysis in the future and the use of results in region-specific ground-motion simulations.

2 DATA

We start with the raw waveforms recently compiled by Hutchinson *et al.* (2024), which contains 247 501 acceleration time-series, and then compute FAS after applying additional screening criteria and undertaking proper data processing. We focus on crustal events with moment magnitude $M_w \geq 3$, hypocentral distance $R \leq 300$ km, and focal depth $D \leq 30$ km. We then exclude recordings labelled as clipped and with multiple wave trains. We only keep records from either strong-motion (HN and BN: high and low sample-rate accelerometers, respectively), and seismological (HH and EH: high sample-rate broad-band and extremely short-period seismometers, respectively) channels (GNS Science 2022). Furthermore, we only retain records with peak ground acceleration, $PGA \geq 0.0002$ g, to exclude excessively weak signals and $PGA \leq 0.2$ g to avoid any significant impacts of soil non-linearity.

P arrivals are determined in this study using PPHASEPICKER (Kalkan 2016). We extract the signal section with a time-window length corresponding to 20–80 per cent of the cumulative integral of the squared acceleration (D_{S20-80}). Though there are other commonly used duration metrics (e.g. D_{S05-75} and D_{S05-95} , Trifunac & Brady 1975), we choose D_{S20-80} since it is better correlated with the duration of significant S waves than D_{S05-75} and D_{S05-95} (Boore & Thompson 2014), as illustrated in Fig. 1. In addition, we impose a minimum and maximum window length of 10 and 40 s, respectively. The minimum length is to ensure the low-frequency stability of FAS down to 0.3 Hz (at least three wavelengths for each oscillator frequency); and the maximum window length is chosen considering the specialty of our dataset which includes ground motions from large events, for example, the 2016 M_w 7.8 Kaikōura earthquake.

A noise window with up to the same duration as the signal is also extracted from the pre-event section of each record when possible. Each extracted acceleration waveform (both signal and noise) is then transformed to FAS after mean removal, tapering (5 per cent cosine taper to both ends) and zero-padding. FASs are then smoothed using the Konno & Ohmachi (1998) function with a smoothing coefficient $b = 40$. For each FAS, we only utilize its values at (a) $f \geq 3/D_{S20-80}$ (i.e. at least three wavelengths for each frequency), (b) $f \leq 0.8 \times \text{Nyquist frequency}$ (accounting for different sampling rates) and (c) with a spectral signal-to-noise ratio (SNR) ≥ 3 . We note that, in the computation of SNR, we use duration-scaled FAS to consider cases where the noise and signal durations differ (Perron *et al.* 2018). The two horizontal components (NS and EW) of each recording are processed independently and are then combined by

their root mean square:

$$H(f) = \sqrt{\frac{1}{2} [\text{FAS}_{\text{NS}}(f)^2 + \text{FAS}_{\text{EW}}(f)^2]}. \quad (1)$$

We impose further screening to enhance data quality. We exclude FAS of which usable frequency range is narrower than 0.5–19 Hz. This criterion ensures identical path coverage at each frequency in this range. Further, we only keep the events and stations with at least three records passing all the aforementioned screening criteria.

Our final data set contains 20 813 ground motions from 1200 crustal events, recorded by 693 sensors at 439 unique locations (Figs 2 and 3). Among the 439 sites, 184 sites are equipped with two sensors, and 35 sites with three sensors. We treat each sensor as an independent ‘site’ in the subsequent inversion which allows for examination of any dependence on instrument type, as discussed further subsequently. The M_w versus D , and M_w versus R distributions of the ground-motion data set are depicted in Fig. 2. We note that many events have integer depths provided. The maximum number of records per site and per event are 298 and 150, respectively, and 90 per cent of sites and 85 per cent of events have at least five recordings (Fig. S1, Supporting Information). Our data set has a reasonably good geographic path coverage (Fig. 3), which is critical for non-parametric GIT used in this study (e.g. Shible *et al.* 2022).

3 METHOD

3.1. Non-parametric generalized inversion

For a surface ground-motion recording during earthquake i at site j , the combined FAS of the two horizontal components, $H_{i,j}(f)$, can be represented as the combination of terms for the source $E_i(f)$, path $P_{i,j}(f)$ and site $S_j(f)$ in the natural logarithmic unit (ln):

$$\ln H_{i,j}(f) = \ln E_i(f) + \ln P_{i,j}(f) + \ln S_j(f). \quad (2)$$

For a data set with multiple recordings per event and station, and with interconnected ray paths (i.e. no isolated region), the above system of linear equations over i and j can be solved as a general inverse problem in the least-squares sense (Andrews 1986; Castro *et al.* 1990). We adopt a non-parametric one-step inversion approach in which source, path and site terms are completely described by data (Bindi *et al.* 2009; Oth *et al.* 2011), rather than rely on pre-defined functional forms as in parametric methods (e.g. Kawase & Matsuo 2004; Drouot *et al.* 2008; Edwards *et al.* 2008; Fu *et al.* 2023). Given sufficient data, non-parametric schemes are advantageous over parametric ones (Shible *et al.* 2022). However, in either approach, the path attenuation ($P_{i,j}$ in eq. 2) is often assumed to be homogeneous and isotropic, only depending on the scalar measure of distance (in 2 km bins) without explicitly considering the 3-D anelastic attenuation, and M_w -dependent near-source geometric spreading. We discuss the potential impact of some of these assumptions on our results subsequently.

Additional constraints are needed to minimize significant trade-offs among the three terms (eq. 2) from the non-parametric GIT. For instance, it would still be self-consistent even if one subtracts (in ln unit) an arbitrary function of frequency from site responses (S_i) and adds it to the source spectra (E_i). Therefore, various inversions schemes impose different constraints/assumptions to minimize the trade-offs. For example, the global EGF (empirical Greens’ function) fitting technique, proposed by Shearer *et al.* (2006), makes assumptions on the source spectral shape.

In this research, we apply two constraints. One is on attenuation, assuming $\ln P(f, R_{\text{ref}}) = 0$, where R_{ref} is the reference hypocentral

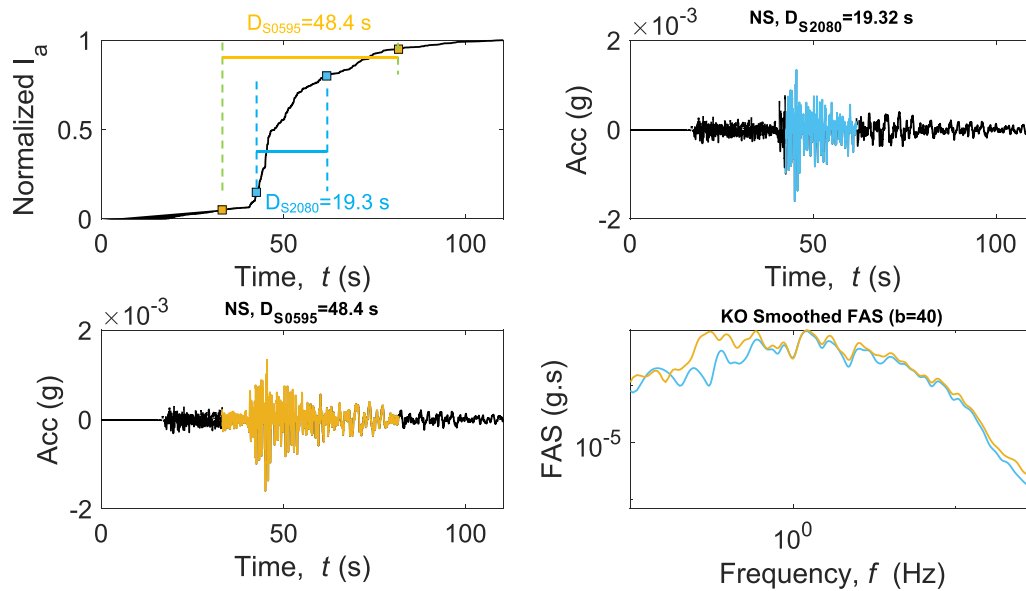


Figure 1. Comparison of window length D_{S2080} used in this study with D_{S0595} (event: 1 825 324, station: WHZ, channel: HH). I_a denotes Arias intensity, which is related to the integral squared acceleration.

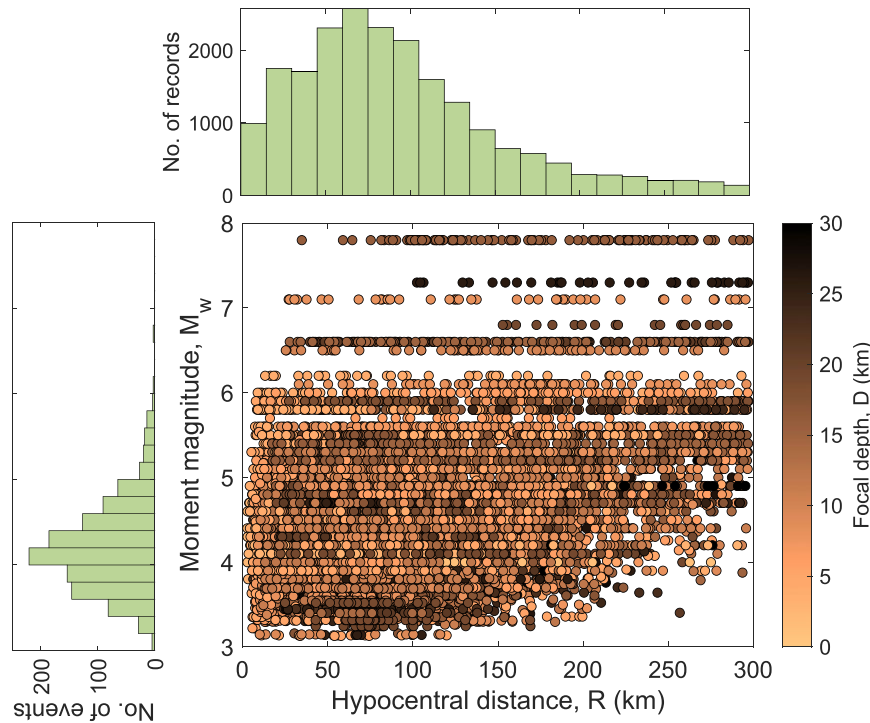


Figure 2. Ground motions selected in this study.

distance. We use $R_{\text{ref}} = 2$ km, which is approximately the minimum distance of our data set. This constraint leads to ‘shifted’ source spectra at $R_{\text{ref}} = 2$ km, which will be accounted for in subsequent interpretation of inversion results. The other constraint is on site response, assuming no site response at reference site(s), that is, $\ln S_{\text{ref}}(f) = 0$. The inverted site response at all other sites will then be relative to this reference condition, making the selection of reference sites non-trivial. The reference site(s) should be ideally free of site effects in the frequency range of interest, but it is challenging to identify good reference site(s) in reality (Steidl *et al.* 1996).

To identify proper reference site(s), as free of site effects as possible, we carried out a preliminary GIT calculation in which the mean site response over all sites in our data set was utilized as the reference. We then selected sites with (i) more than 20 records, (ii) site response varying smoothly with frequency (i.e. no significant peaks or troughs) and (iii) $V_{S30} > 600$ m s⁻¹, where V_{S30} denotes the time-averaged shear wave velocity in the topmost 30 m. After a few iterations, we select HUNS, RPZ and MRZ as reference sites (Fig. 3) which have V_{S30} values of 680, 1000 and 800 m s⁻¹, respectively (Wotherspoon *et al.* 2024). While $V_{S30} = 680$ m s⁻¹ at HUNS is

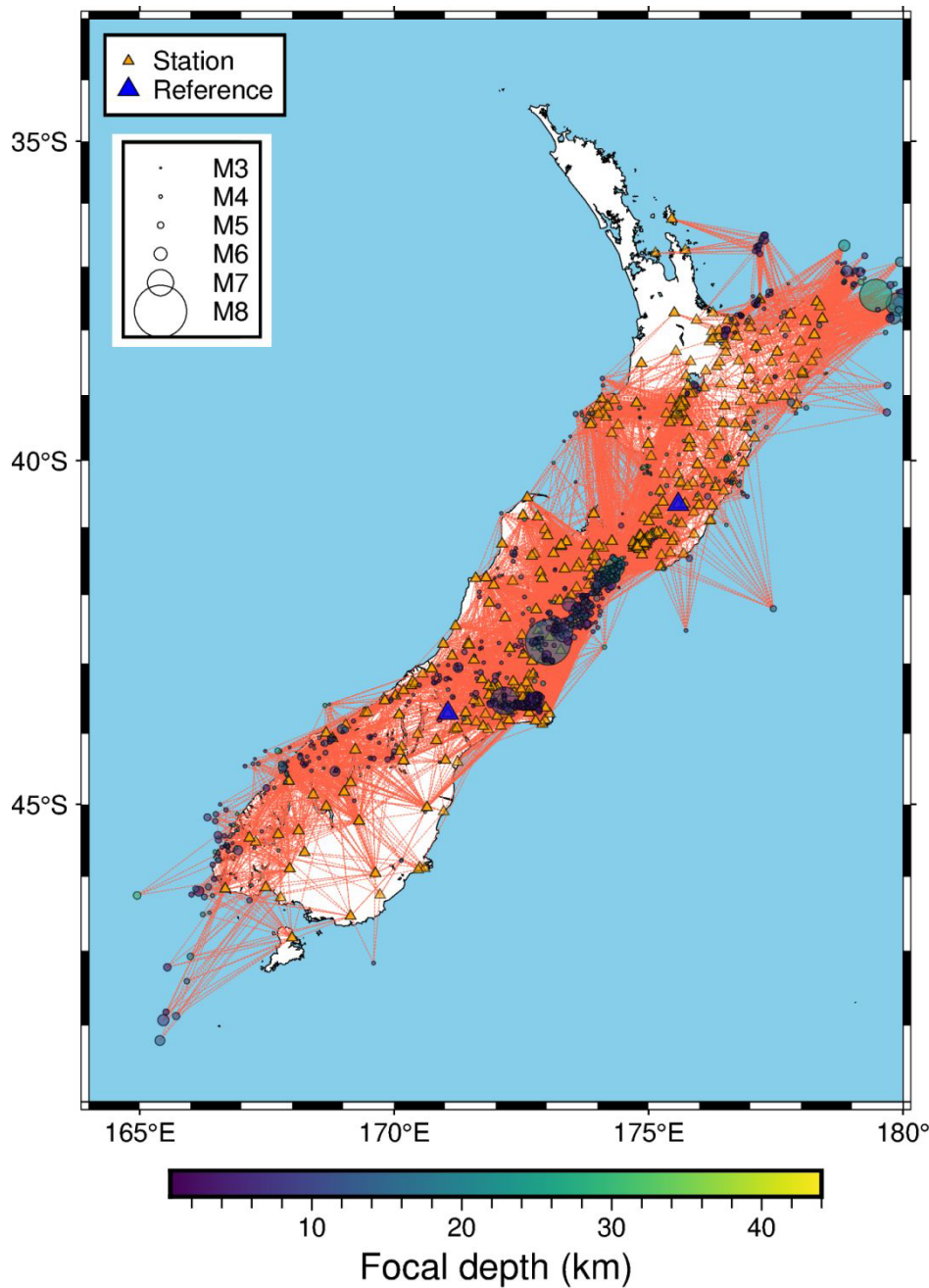


Figure 3. Path coverage of the 20 813 selected ground motions from 1200 crustal events (filled circles with colours indicating focal depth) recorded by 693 sensors at 439 unique locations (triangles). Lines represent simplified seismic ray paths, and blue triangles denote the three reference stations (from left to right: RPZ, HUNS and MRZ).

seemingly low for a reference site, it is derived from a regional V_{S30} map (Foster *et al.* 2019) and therefore has a relatively large uncertainty compared with that possible by direct measurement.

3.2. Parametrization of GIT results

3.2.1. Source parametrization

In our non-parametric inversion, each term is completely constrained by data. To obtain relevant source parameters, we parametrize the inverted acceleration source spectra in post-processing by fitting the ω^{-2} source model (Brune 1970, 1971)

as illustrated in Fig. S2 (Supporting Information). The Brune model is generally considered to provide an appropriate representation of the source spectrum of small and moderate earthquakes (e.g. Izutani & Kanamori 2001) which dominate our data set. Due to the trade-off between source and site terms, improper reference sites could systematically bias source spectra, especially at high frequencies. To investigate this potential, we apply a κ_{source} filter to quantify the deviation of the inverted source spectra $E(f)$ from the ω^{-2} model at $f > f_k$, where f_k indicates the frequency from which the filter is applied. We set $f_k = 10$ Hz to be well beyond the corner frequency range of the events (Oth & Kaiser 2014). If there is any systematic amplification or de-amplification at the reference sites, it will be

mapped into source spectra, and results in a large mean value of κ_{source} .

The theoretical source model used herein, $E(f)$, consists of both the standard ω^{-2} model and a κ_{source} filter:

$$E(f) = (2\pi f)^t \frac{R_{\theta\theta} V F}{4\pi\rho\beta^3 R_{\text{ref}}} \frac{M_0}{1 + \left(\frac{f}{f_c}\right)^2} e^{-\pi\kappa_{\text{source}}(f-f_c)}, \quad (3)$$

where $t = 0, 1, \text{ or } 2$ for displacement, velocity, or acceleration spectra, respectively; $R_{\theta\theta} = 0.55$ is the azimuthally averaged radiation pattern of S waves (Boore & Boatwright 1984); $V = 1/\sqrt{2}$ is the partition of S -wave energy into two horizontal components; $F = 2$ is the free-surface effect for vertically propagating SH waves; $\rho = 2700 \text{ kg m}^{-3}$ and $\beta = 3700 \text{ m s}^{-1}$ are the density and S -wave velocity at the source, respectively, derived from a 1-D velocity model used by Lee *et al.* (2022); $R_{\text{ref}} = 2 \text{ km}$ is the reference distance; M_0 (N-m) is the seismic moment; and f_c (Hz) is the corner frequency. For events with $M_w > 5.0$, f_c and M_0 may not be reliably determined simultaneously due to bandwidth limits (i.e. f_c may be lower than the lowest usable frequency of 0.5 Hz). Thus, M_0 is anchored to the associated moment magnitude M_w from the reference catalogue (Hutchinson *et al.* 2024), rather than freely inverted.

Assuming a circular fault rupture, the Brune stress parameter, $\Delta\sigma$ (unit: Pa), can be computed from M_0 and f_c (Eshelby 1957; Keilis-Borok 1959; Brune 1970):

$$\Delta\sigma = \frac{7}{16} M_0 \left(\frac{f_c}{k\beta} \right)^3, \quad (4)$$

where k is a constant, depending on the assumed rupture model, and is set to 0.37 for S wave (Brune 1970, 1971). Inverted and theoretical source spectra, as well as source parameters for each event are provided in Supporting Information ‘Source_parameters.xlsx’ and ‘DataSet_Source’ (Data and Resources).

3.2.2. Path parametrization

We parametrize the non-parametric attenuation from GIT via regression analyses using the following equation:

$$\ln P(f, R) = \ln G(R) + \ln A(f, R), \quad (5)$$

where $P(f, R)$ represents the total path attenuation, and $G(R)$ and $A(f, R)$ denote the parametric geometric spreading and anelastic attenuation functions, respectively. $A(f, R)$ is described as follows:

$$\ln A(f, R) = \frac{-\pi f(R - R_{\text{ref}})}{Q(f)\beta}, \quad (6)$$

where $Q(f)$ is the frequency-dependent quality factor modelled as $Q(f) = Q_0 f^\alpha$, Q_0 is the reference value of Q at $f = 1.0 \text{ Hz}$, and α is an empirical coefficient. An iterative procedure (Bindi & Kotha 2020) is adopted to parametrize $G(R)$ and then $A(f, R)$.

4 RESULTS

Using the non-parametric one-step GIT, we obtain the non-parametric acceleration source spectrum for each event, attenuation against distance, and site response at each site (Fig. 4). In Fig. 4(a), the inverted source terms are highlighted for three example events with $M_w = 4.0, 4.7$ and 5.0 , respectively. The theoretical source model in eq. (3) captures the non-parametric source terms of the three events reasonably well. Fig. 4(b) illustrates the distance and frequency dependency of path attenuation. High-frequency components tend to attenuate faster with distance than low-frequency ones.

At relatively low frequencies, a flattening of attenuation rate appears in the distance range from ~ 60 to 100 km , commonly attributed to post-critical reflections from the Moho discontinuity (e.g. Burger *et al.* 1987; Atkinson & Mereu 1992). In Fig. 4(c), each curve is the average site response over all events available at a given site and is relative to the selected reference sites (black lines). The coloured lines in Fig. 4(c) correspond to sites CCCC ($V_{S30} = 234 \text{ m s}^{-1}$, Wotherspoon *et al.* 2024) and WTYS ($V_{S30} = 230 \text{ m s}^{-1}$, Foster *et al.* 2019) in Christchurch and Wellington city centres, respectively. The azimuth coverage at each site is illustrated in Fig. S3 (Supporting Information).

To examine the stability of the inversion, we carried out residual analyses (observation minus prediction from eq. 2 using the three terms from GIT, as exemplified in Fig. S4, Supporting Information). But we found no systematic bias against earthquake magnitude and hypocentral distance. This indicates that the systematic source, path and site effects are well captured. In the following, we elaborate on each constituent component.

4.1. Source effects

4.1.1. Source parameters

Fig. 5(a) illustrates the distribution of κ_{source} (eq. 3) for the 1152 events for which source parameters are successfully derived. κ_{source} is normally distributed with a mean of -0.002 s and a standard deviation of 0.013 s . The small mean value of κ_{source} demonstrates that our reference sites do not significantly bias the source spectra at high frequencies ($f > 10 \text{ Hz}$), and that the inverted source spectra, on average, follow the Brune circular model at high frequencies. The variability of κ_{source} among individual events are attributable to factors, for instance, near-source attenuation and radiation energy effects (e.g. Bindi & Kotha 2020).

In addition to high frequencies, our reference site selection also does not appear to significantly bias the computed source spectra at relatively low frequencies which affect the inverted M_w . Fig. 5(b) compares the M_w from GIT with the centroid moment tensor solutions from the reference catalogue. For larger events ($M_w > 5.0$), M_w (or M_0) is anchored to the catalogue values in the spectral fitting due to bandwidth limits. However, for smaller events ($M_w < 5.0$), the inverted M_w is also consistent with the catalogue M_w with a negligible offset (e.g. the zero axis lies within the uncertainty bound for most magnitude bins), and specifically the difference has a mean of 0.016 and standard deviation of 0.187 . This confirms our source and site terms are not significantly biased at low frequencies either. Besides, using the catalogue M_w for large events will not introduce inconsistency in the M_0 - f_c scaling between large and small events, as investigated in Section 4.1.2.

Fig. 6(a) depicts the histogram of the stress parameter, $\Delta\sigma$, for the 1152 crustal events. $\Delta\sigma$ follows a lognormal distribution (quantile–quantile plot in Fig. S5, Supporting Information) with a \log_{10} standard deviation of $\sigma_{\log \Delta\sigma} = 0.36$, or $\sigma_{\ln \Delta\sigma} = 0.83$ in \ln unit. This $\Delta\sigma$ variability is consistent with those reported in other regions, for instance, Japan (e.g. $\sigma_{\log \Delta\sigma} = 0.42$ in Baltay *et al.* 2013; 0.34 – 0.36 in Nakano *et al.* 2015; 0.48 in Oth *et al.* 2017), California (e.g. 0.31 – 0.45 in Trugman & Shearer 2017) and Southern Kansas (0.35 by Trugman *et al.* 2017). Besides, we examine the distribution of $\Delta\sigma$ values for subsets of the 2010–2011 Canterbury sequence and the 2016 Kaikōura sequence included by Oth & Kaiser (2014) and Ren *et al.* (2018), respectively. Our results (Fig. 6a) indicates that the 2010–2011 Canterbury sequence has a mean $\Delta\sigma$ higher than the

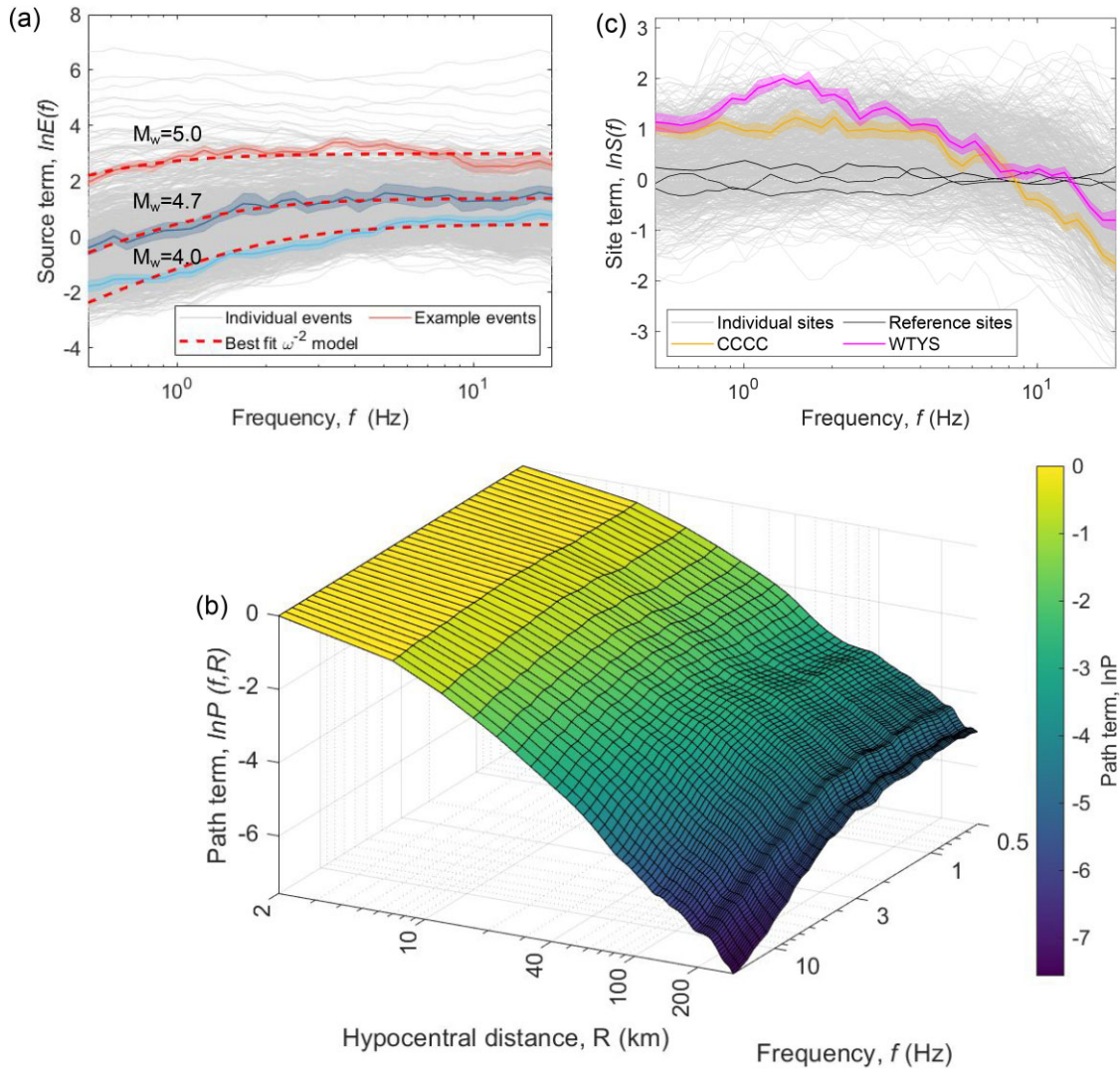


Figure 4. Inverted Fourier source, path and site terms. (a) Acceleration source spectra for 1200 crustal events. Colour lines are the non-parametric (solid lines) and parametric (dashed lines) source spectra for three example events: M_w 4.0 ‘2 825 685’, M_w 4.7 ‘1 797 520’ and M_w 5.0 ‘2 626 467’; and shaded areas represent the uncertainty via bootstrapping. (b) Path attenuation as a function of frequency and distance. (c) Site responses for 693 sites, including three reference sites (black lines), and two example non-reference sites CCCC and WTYS (colour lines) in Christchurch and Wellington city centres, respectively.

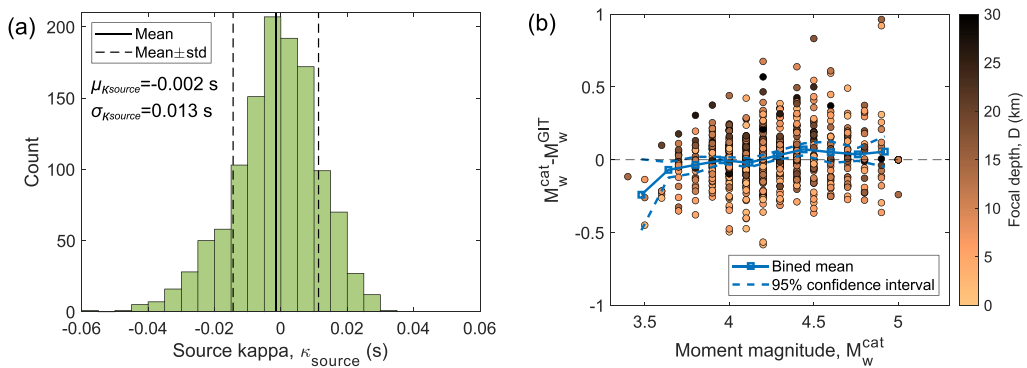


Figure 5. (a) Histogram of source kappa, κ_{source} , and (b) comparison between moment magnitudes from inversion, M_w^{GIT} , and centroid moment tensor catalogue, M_w^{cat} . Since M_w for larger events ($M_w > 5.0$) is anchored to the catalogue value during the inversion, thus the comparison is limited to $M_w \leq 5.0$.

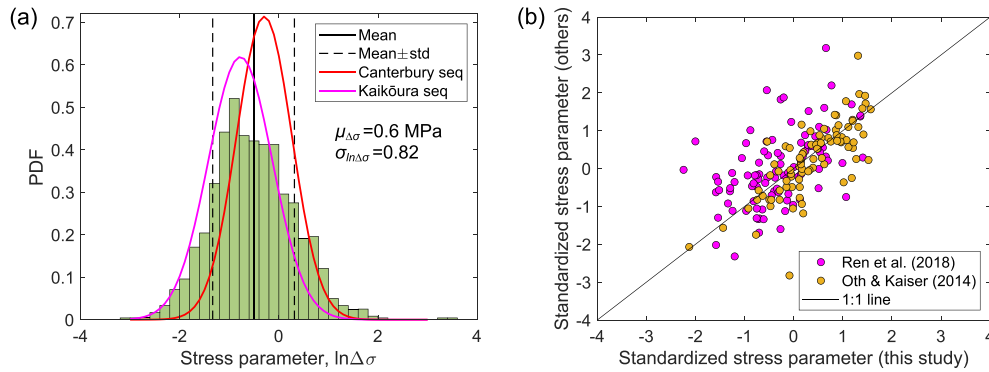


Figure 6. (a) Histogram of Brune's stress parameter, $\Delta\sigma$, for crustal events examined in this study as well as subsets of the 2010–2011 Canterbury and the 2016 Kaikōura sequences, and (b) comparison of the standardized stress parameter, $\Delta\sigma$, with those from Ren *et al.* (2018) and Oth & Kaiser (2014) for common events.

national average whereas the 2016 Kaikōura sequence has a lower mean $\Delta\sigma$. The $\sigma_{\log\Delta\sigma}$ values for the Canterbury and Kaikōura sequences are 0.25 and 0.28, respectively, both of which are smaller than the variability on the national level.

Because of the widely recognized variability in computing $\Delta\sigma$ values due to, for instance, model-specific assumptions (e.g. Pennington *et al.* 2021; Shible *et al.* 2022), we focus on the relative pattern in $\Delta\sigma$, rather than its absolute value, to compare trends between different studies. Ren *et al.* (2018) and Oth & Kaiser (2014) previously analysed the 2010–2011 Canterbury and the 2016 Kaikōura Earthquake Sequences, respectively. We thus compare $\Delta\sigma$ values for common events within these different studies after standardization within each data set:

$$\overline{\Delta\sigma}_i = \frac{\ln\Delta\sigma_i - \mu_{\ln\Delta\sigma}}{\sigma_{\ln\Delta\sigma}}, \quad (7)$$

where $\Delta\sigma_i$ denotes the stress parameter for event i , and $\mu_{\ln\Delta\sigma}$ and $\sigma_{\ln\Delta\sigma}$ are the logarithmic mean and standard deviation of $\Delta\sigma$, respectively, for a given data set. The intention of standardization via eq. (7) is that systematic deviations in $\Delta\sigma$ values across studies, as high as 1–2 orders of magnitude (Shible *et al.* 2022), would be significantly reduced.

There is a reasonably good agreement in $\overline{\Delta\sigma}$ values between our study and Oth & Kaiser (2014) (Fig. 6b), as indicated by a Pearson's correlation of $r = 0.76$ (with an associated $p = 8e-18$). The correlation with Ren *et al.*'s (2018) results is lower with $r = 0.49$ ($p = 5e-7$). While it is known that $\Delta\sigma$ depends on the choice of rupture model (e.g. circular, or elliptical crack model) and the value assumed for the high-frequency fall-off n , -for example, $n = 2$ in $(f/f_c)^n$ (Trugman 2020), all three studies (this work, Oth & Kaiser 2014; Ren *et al.* 2018) use a Brune-type circular rupture model. The lower correlation with Ren *et al.*'s (2018) results (Fig. 6b) may be attributed to the different modelling of the high-frequency decay in source spectra. Specifically, Ren *et al.* (2018) utilized the f_{\max} model, expressed as $[1 + f/f_{\max}^8]^{-1/2}$, whereas this study and Oth & Kaiser (2014) adopted the κ_{source} model (eq. 3). Other potential reasons for the scatter include different S -wave window length (e.g. Bindi *et al.* 2023). There are ongoing efforts, for example, the Community Stress Drop Validation Study (Baltay *et al.* 2021) to benchmark absolute $\Delta\sigma$ estimates. However, as illustrated in Fig. 6(b), $\overline{\Delta\sigma}$ values across studies are still highly correlated. Therefore, we emphasize on the relative pattern in $\Delta\sigma$ values in the following analyses and interpretation.

4.1.2. Source scaling

One topic of general interest to earthquake source physics is whether and how $\Delta\sigma$ scales with earthquake size (e.g. Oth 2013). For the shallow crustal events examined here, $\Delta\sigma$ displays a very slight increasing trend with M_w with a Pearson's $r = 0.06$ ($p = 0.04$) for the whole dataset (Fig. 7a). Though statistically significant (at the common 5 per cent threshold), such a low r value only suggests a negligible practical correlation (Schober *et al.* 2018). Thus, the practical independence of $\Delta\sigma$ on earthquake size supports self-similarity (Aki 1967), which is also the case using the 2010–2011 Canterbury and 2016 Kaikōura earthquake sequence subsets of events.

The near self-similarity can also be confirmed in terms of the M_0 - f_c scaling (Fig. 8). Kanamori & Rivera (2004) proposed a parameter ε in $M_0 \propto f_c^{-(3+\varepsilon)}$, where $\varepsilon = 0$ indicates self-similarity (Aki 1967). In this study, we find $\varepsilon = 0.09 \pm 0.12$, suggesting we cannot reject $\varepsilon = 0$ (i.e. self-similarity) at a 95 per cent confidence interval. This is consistent with other results in NZ using the ω^{-2} model, that is, Oth & Kaiser (2014) for the 2010–2011 Canterbury earthquake sequence ($\varepsilon = 0.16 \pm 0.17$), and Ren *et al.* (2018) for the 2016 Kaikōura earthquake sequence ($\varepsilon = 0.02 \pm 0.08$). For regions other than NZ, Oth *et al.* (2010) reported $\varepsilon = 0.12 \pm 0.12$ for crustal events in Japan. In contrast, deviations from self-similarity have been reported in Italy (e.g. Wang *et al.* 2019) and California (e.g. Trugman 2020; Bindi *et al.* 2023).

Stress parameter, $\Delta\sigma$, increases with focal depth, D , when a constant velocity is used in eq. (4) (Fig. 7b). The Pearson's correlation between $\Delta\sigma$ and D is $r = 0.22$ ($p = 2e-14$), indicating a weak correlation. The increasing trend of $\Delta\sigma$ with D could be caused by the increase in M_0 (eq. 4), but we rule this out since M_0 decreases slightly with D , with a Pearson's $r = -0.13$ ($p = 2e-05$, Fig. S6, Supporting Information). However, Abercrombie *et al.* (2017a) and Ren *et al.* (2018) found no depth-dependence of $\Delta\sigma$ in 176 subducting earthquakes and the 2016 Kaikōura earthquake sequence, respectively, in NZ. Whether $\Delta\sigma$ increases with source depth is still a subject of debate (e.g. Abercrombie *et al.* 2021). We thus further discuss the potential impact of our assumptions on the results in Section 5.

4.2. Path attenuation

Path attenuation is well constrained by data in the distance range considered herein. This is illustrated in Fig. 9(a) which depicts the FAS (at $f = 5$ Hz) corrected for source and site effects,

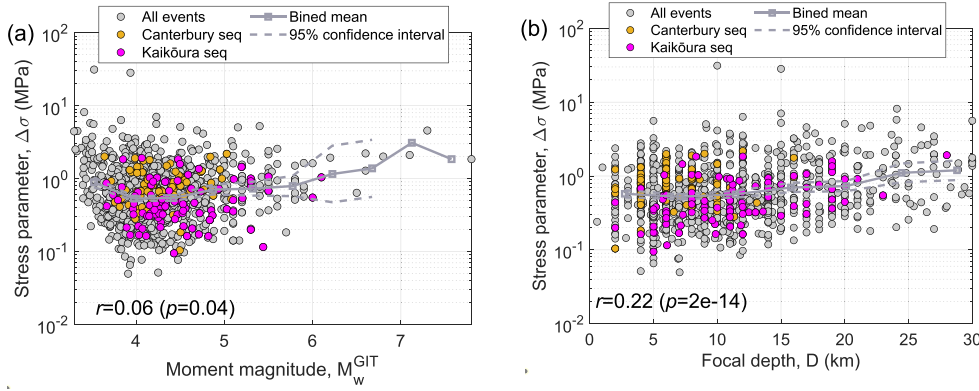


Figure 7. Dependence of Brune's stress parameter, $\Delta\sigma$, using a constant velocity model on (a) inverted earthquake magnitude, M_w^{GIT} , and (b) focal depth, D . r and p are the Pearson's correlation coefficient and corresponding p -value, respectively.

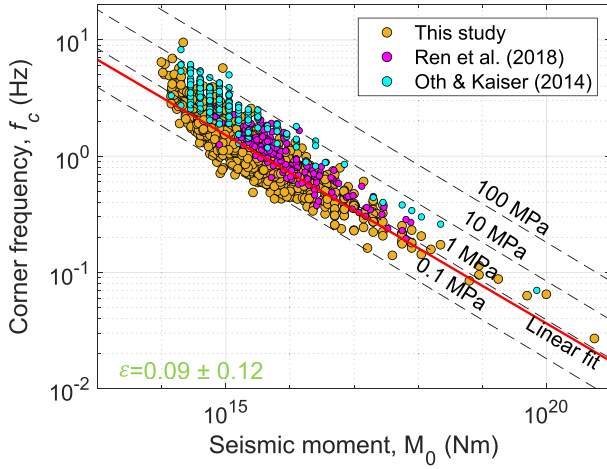


Figure 8. M_0 - f_c scaling. Dashed lines represent the theoretical relationship between M_0 and f_c for various levels of $\Delta\sigma$. Solid line is the linear fit to data from this study. ε is the parameter proposed by Kanamori & Rivera (2004) in $M_0 \propto f_c^{-(3+\varepsilon)}$, and its value is derived based on results in this study.

In $H_{i,j}(f) - \ln E_i(f) - \ln S_j(f)$, against hypocentral distance. In the following, we parametrize the data-driven attenuation functions.

4.2.1. Geometric spreading

To parametrize, in a simple manner, the non-parametric attenuation from GIT (Fig. 9a), geometric spreading, $G(R)$, is assumed to be frequency independent in this study. We first consider $f = 1$ Hz in eq. (6) to determine the geometric spreading. To boost the robustness of the results, the attenuation curves over frequencies $0.5 < f \leq 1.23$ Hz are fitted to represent that at $f = 1$ Hz. A simple nonlinear least-squares regression is performed by adopting the following hinged trilinear function:

$$\ln G(R) = \begin{cases} s_1 \ln \frac{R}{R_{\text{ref}}} & R \leq R_1 \\ s_1 \ln \frac{R_1}{R_{\text{ref}}} + s_2 \ln \frac{R}{R_1} & R_1 < R \leq R_2 \\ s_1 \ln \frac{R_1}{R_{\text{ref}}} + s_2 \ln \frac{R_2}{R_1} + s_3 \ln \frac{R}{R_2} & R > R_2, \end{cases} \quad (8)$$

where slope values and hinge distances are obtained in the regression as $[s_1, s_2, s_3] = [-0.31, -1.09, -0.33]$ and $[R_1, R_2] = [7.57, 59.62]$ km. Fig. 9(b) illustrates that eq. (8) well captures the distance-dependent geometric spreading.

The first hinge distance R_1 is devised to capture the near-source saturation effects, which can be observed in Fig. 9(a). The saturation

effects are known to be magnitude dependent (e.g. Yenier & Atkinson 2014). However, the sparsity of our data set at close distance from large events prevents us reliably modelling its magnitude-dependence in a fully data-driven manner. Thus, we use the first hinge R_1 to capture the distance saturation effects on average (across different M_w bins).

The second hinge distance R_2 is to model the attenuation rate change in the distance range of ~ 60 to 100 km (Fig. 9b). The rate change is potentially caused by the significant arrivals of S waves postcritically reflected at the Moho discontinuity. Such a variation in attenuation has long been observed in other regions, for example, North America (e.g. Burger *et al.* 1987; Atkinson & Mereu 1992), but its distance range depends on the focal depth, crustal thickness, and the crustal velocity gradient, and thus varies across regions (Burger *et al.* 1987).

4.2.2. Anelastic attenuation

After parametrizing the apparent geometric spreading, we then parametrize the anelastic attenuation. First, we correct the inverted path attenuation for the apparent geometric spreading effects using eq. (8). The non-parametric spreading-corrected attenuation, that is, $\ln P(f, R) - \ln G(R)$, are illustrated in Fig. 9(c). The slopes of the non-parametric anelastic curves vary with distance intervals, and thus we model the trend using the following piecewise function for different distance ranges with a break point at R_h :

$$\ln A(f, R) = \begin{cases} \frac{-\pi f(R-R_{\text{ref}})}{Q_1(f)\beta} & R \leq R_h \\ \frac{-\pi f(R-R_{\text{ref}})}{Q_1(f)\beta} + \frac{-\pi f(R-R_h)}{Q_2(f)\beta} & R > R_h, \end{cases} \quad (9)$$

where $Q_1(f) = Q_{0,1}f^{\alpha_1}$ and $Q_2(f) = Q_{0,2}f^{\alpha_2}$ describe the quality factors for distances smaller and larger than R_h , respectively. The fitted values of the parameters are obtained as $[Q_{0,1}, Q_{0,2}] = [369.44, 98.27]$, $[\alpha_1, \alpha_2] = [0.11, 0.95]$ and $R_h = 94.49$ km. These estimated values of quality factor suggest that at distances $R > R_h$ the entire attenuation function becomes independent of frequency with $\alpha_2 = 0.95$. This does not mean that the attenuation is indeed frequency independent because most of the frequency dependence of the attenuation (over the entire distance range) comes from distances $R \leq R_h$, and the slope of distance decay for $R > R_h$ is conditioned on the intercept at $R = R_h$ (eq. 9). Moreover, it is worth mentioning that at larger distances the trade-off between apparent geometric spreading and quality factor becomes difficult to resolve (e.g. Edwards *et al.* 2008). Fig. 9(d) demonstrates that our

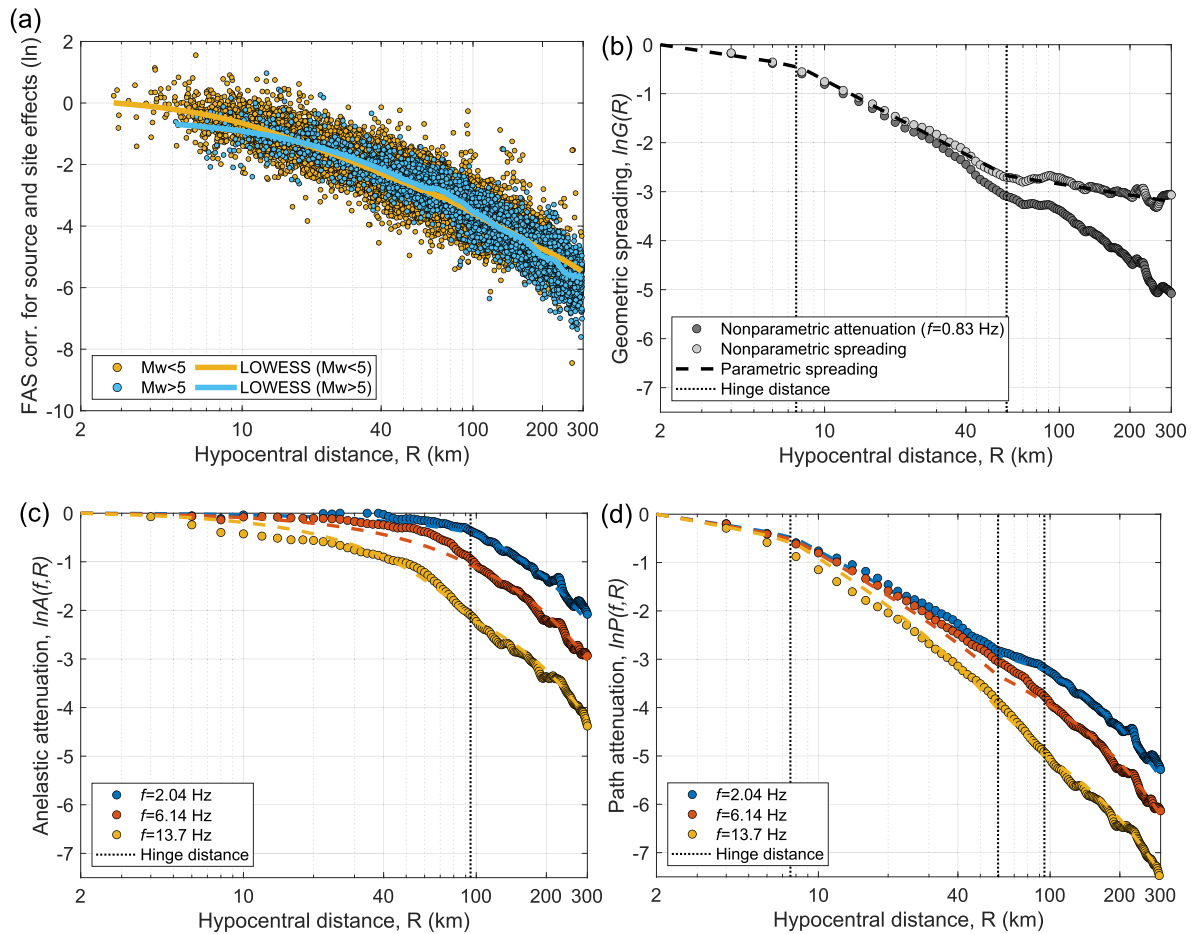


Figure 9. (a) Path attenuation with data at $f = 5$ Hz in which LOWESS denotes ‘locally weighted scatterplot smoothing’, and parametrization of (b) geometric spreading, $G(R)$ (eq. 8), (c) anelastic attenuation, $A(f, R)$ (eq. 9) and (d) combined path attenuation, $P(f, R)$ (eq. 5).

parametric equations (eqs 8 and 9) can well model the distance- and frequency-dependence of attenuation from the inversion.

In addition to the distance-dependent Q model (eq. 9), we also fit a single Q model (eq. 6) over the entire distance range, as shown in eq. (10), to facilitate comparisons with Ren *et al.* (2018) who presented a single Q model. While the single $Q(f)$ model appears simple, the distance-dependent $Q(f)$ model fits better the spreading-corrected attenuation curves, as shown in Fig. S7 in the Supporting Information:

$$Q(f) = 149.1 f^{0.62}. \quad (10)$$

We compare our single $Q(f)$ model (eq. 10) with that of Ren *et al.* (2018) who obtained $Q(f) = 130.87 f^{1.07}$ for the southern North Island and northeastern South Island of NZ. Our Q_0 values are comparable, but difference is evident in the frequency dependence of Q (i.e. α in eq. 6). The difference in α values is due to different modelling assumptions and target regions. First, Ren *et al.* (2018) directly adopted the geometric spreading model of Atkinson & Mereu (1992) for southeastern Canada, which follows R^{-1} for $R < 37.5$ km, while we use eq. (8) with coefficients obtained specifically for NZ data. Due to trade-offs between geometric spreading and anelastic attenuation (e.g. Edwards *et al.* 2008), differences in the former will result in deviation in the latter. In addition, trade-offs between attenuation and other effects may also contribute to the difference in attenuation across studies, as reported in the GIT benchmark study

by Shible *et al.* (2022). Furthermore, Ren *et al.*’s (2018) model was derived for the southern North Island and northeastern South Island while our eq. (10) is for the entire NZ. Thus, the difference in $Q(f)$ is also attributable to the lateral variations in attenuation within the crust in NZ, as revealed by Eberhart-Phillips *et al.* (2015). Nonetheless, herein we only provide simple parametrized attenuation functions, and examination of depth-, magnitude-, and region-dependence of attenuation will be left for future studies.

4.3. Site response

The observational dataset of FAS-based site responses from GIT (i.e. Fig. 4c) enables us to examine important questions regarding site effects in an NZ context. Since there are colocated sensors at some sites, we only use one site-response curve at each unique location (total no.: 439), adopting the preferential system: HN > BN > HH > EH. In addition, to avoid any potential impact of instrument response removal on data from seismometers, we only utilize the 319 available strong-motion channels in the subsequent analyses and interpretations. Site-response data (mean, and uncertainty quantified via bootstrapping) at each site, and its corresponding azimuth coverage are provided in Supporting Information ‘*Amp_mean_uncertainty.xlsx*’ and ‘*DataSet_Amp*’ (Data and Resources).

4.3.1. Data-driven clustering

Current seismic codes account for local site effects by differentiating elastic design spectra for different site categories in many countries (NZS1170.5 in NZ; Eurocode 8 in Europe). These discrete categories are known to be a simplistic representation of the complex and location-specific nature of site effects. Nevertheless, they are still a step forward relative to completely ignoring the impacts of surface geology, or the crude binary distinction of soil versus rock used in the past (Fig. S8, Supporting Information). None the less, site categories in current seismic codes are based on proxies of site response (e.g. V_{S30}), rather than site response itself. One important question is how effective the current site classification code in NZ (i.e. NZS1170.5) is in discerning differences in observed amplification.

We assess the performance of the current NZ site classification code against a data-driven approach. In the data-driven method, mutually exclusive clusters in the high-dimensional site-response data are recognized via an unsupervised technique, k -means clustering (MacQueen 1967). k is set to five primarily considering the number of site categories (i.e. five) in NZS1170.5. It is also appropriate to set k at 5 based on the metric of within-cluster sum of distances and silhouette criterion values (Fig. S9, Supporting Information).

Each k -means cluster is distinctly separated as can be observed in their distributions in the first two dimensions from principal component analysis (PCA, Fig. 10a). Individual curves in each cluster are presented in Fig. S10 (Supporting Information). The spatial distribution of each cluster is depicted in Fig. S11 (Supporting Information) for all sites in NZ, while Fig. 11 specifically illustrates sites in the Canterbury region for a better visualization.

Site clusters recognized by k -means clustering are also broadly consistent with independent geological information. For instance, cluster 2 exhibits the largest amplification among the five clusters at $f < \sim 2$ Hz, followed by a significant decay in amplitude with the increase in frequency (Fig. 10b), a typical feature of sites underlain by deep soft sediments. Sites from the Christchurch CBD, filled by Holocene river deposits (GNS Science 2012), are in this cluster (Fig. 11). Cluster 4 features a high-frequency resonance, which suggests a formation of thin soft layers on top of stiffer soils or rocks, for example, some sites in the Banks Peninsula (volcanic rocks), Southeast of the Christchurch CBD.

Besides the significant differences in mean site response, dispersion in the site-response data is also reduced by more than 35 per cent by k -means clustering. Specifically, within-cluster standard deviations are 36, 41, 47, 35 and 42 per cent (average over the examined frequency range) smaller for cluster 1–5, respectively, relative to the entire data set (Fig. 10c and Table 1). The reduction is more pronounced at lower frequencies ($f < \sim 4$ Hz) at which the standard deviation is approximately $\sigma_{\ln S} = 0.4$. However, for high frequencies, especially $f > 10$ Hz, the standard deviation still increases. We attribute this increase to the between-site variability in κ_0 (Anderson & Hough 1984; Hough & Anderson 1988), and it is a global challenge to effectively capture the high-frequency variability in forward predictions with existing approaches (e.g. Zhu *et al.* 2022, 2023).

In contrast to the k -means clustering results, the NZS1170.5 soil classification scheme is far less effective in differentiating site responses. Figs 10(d)–(f) illustrate the result for the same data set grouped by NZS1170.5 subsoil class which is obtained from

Wotherspoon *et al.* (2024). Among these 319 sites, 49 per cent (or 157) of sites are in class D, 29 per cent (or 94) in class B, and 18 per cent (or 57) in class C (Table 1). Classes A and E have less than 10 samples, and consequently within-class statistics are thus not presented. As illustrated in Fig. 10(d), site-response curves in each NZS1170.5 class are not distinctly separated. The reductions in standard deviation are 4, 4 and 15 per cent for classes B, C and D, respectively (Fig. 10f and Table 1), which are significantly smaller than those achieved by k -means clustering (i.e. 35–47 per cent). The within-class standard deviation for the class B is even higher than the whole data set at $f > \sim 4$ Hz, suggesting the ineffectiveness of the NZS1170.5 classification for this category. Overall, our results indicate a rather limited discerning power of the NZS1170.5 soil classification scheme.

4.3.2. Towards end-to-end classification

Though k -means clustering is very effective in differentiating site responses, the approach is not directly useful in forward classification in which observed site responses at target sites are generally unavailable. Thus, we still need to rely on parametrized classification. However, these k -means clusters can be utilized to develop new parametrized classification schemes (e.g. Kotha *et al.* 2018) which we refer to as ‘end-to-end’ classification herein. To this end, we examine the metadata distribution of the k -means clusters, specifically V_{S30} and the site period T_0 based on values in Wotherspoon *et al.* (2024), as illustrated in Fig. 12. In examining trends, it is important to note that 81 per cent of V_{S30} and 18 per cent of T_0 data are inferred values.

Differences in V_{S30} and T_0 values among k -means clusters can on average be clearly observed (Fig. 12 and Table 2). There is an increasing trend in the median values of V_{S30} from cluster 2, 1, 3, 4 to 5, and a decreasing trend in median T_0 accordingly (due to their correlation seen in Fig. 12a). Cluster 2 has the lowest median V_{S30} of 264 m s^{-1} and the highest median T_0 of 1 s, thus corresponding to soft soil sites. In contrast, cluster 5 has the largest median V_{S30} of 800 m s^{-1} and the lowest median T_0 of 0.05 s, corresponding to stiff soil or rock sites. This agrees with the shape of average amplification curves for these clusters shown in Fig. 10(b). Since V_{S30} and T_0 are an overly simplified characterization of the high-dimensional site responses, and have uncertainties of their own, there are overlaps in the 25th–75th percentile V_{S30} or T_0 values between neighbouring clusters (Table 2).

Including more site parameters, for example, depth to a 1.0 km s^{-1} velocity horizon $Z_{1.0}$, can more exclusively define each cluster. However, this requires a larger sample size, that is, the curse of dimensionality (e.g. Zheng & Casari 2018). In addition, there are studies questioning the use of simple site parameters in site classification. Yaghmaei-Sabegh & Rupakhety (2020) and Ji *et al.* (2022) utilized more informative site data, that is, horizontal-to-vertical spectral ratio (HVSr) curves to separate different site classes. However, in future works, the inherent clusters in site-response data (e.g. the k -means clusters in Fig. 10a) can be utilized as the ground truth (or labels in a machine learning context), rather than those specified by current seismic provisions. From these inherent clusters, one can ‘reverse engineer’ new parameter-based site classification schemes (e.g. Kotha *et al.* 2018), or train data-driven classification models using a variety of site information as features, for example, microtremor HVSr, geological age and terrain attributes.

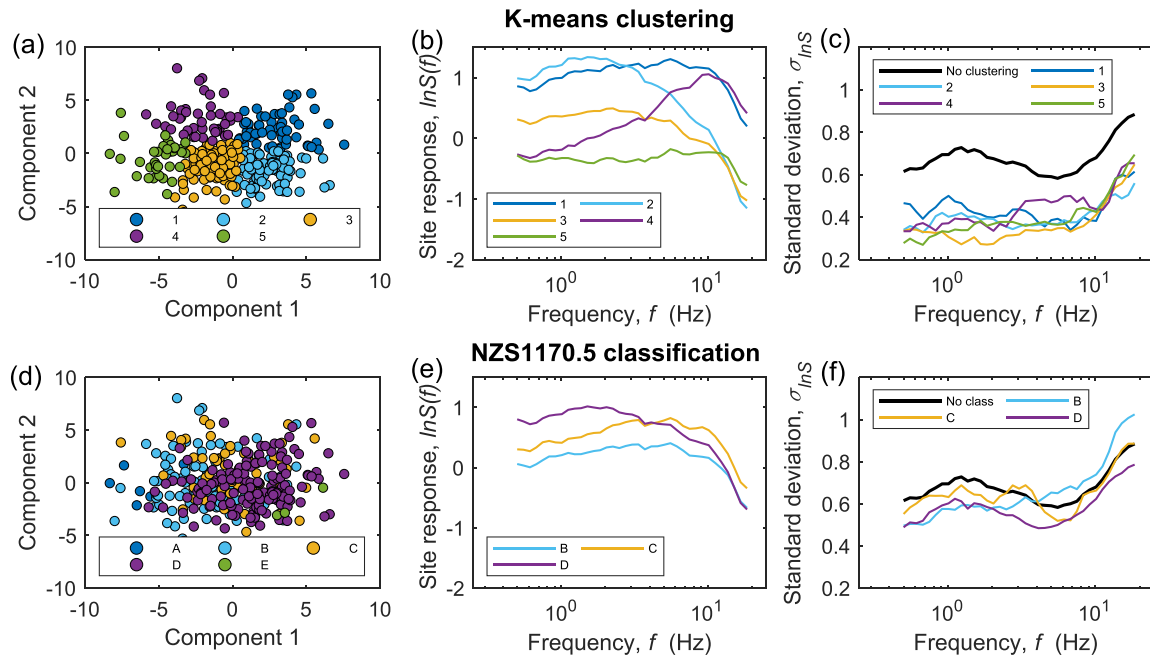


Figure 10. (a)–(c) *K*-means clustering versus (d)–(f) NZS1170.5 subsoil classification of site-response curves with frequency (Fig. 4c). (a) Five clusters by *k*-means clustering (an unsupervised machine learning) in the first two PCA dimensions, (b) within-cluster means and (c) standard deviations of site responses for each cluster. Panels (d)–(f) are similar to (a)–(c) but by NZS1170.5 subsoil classification.

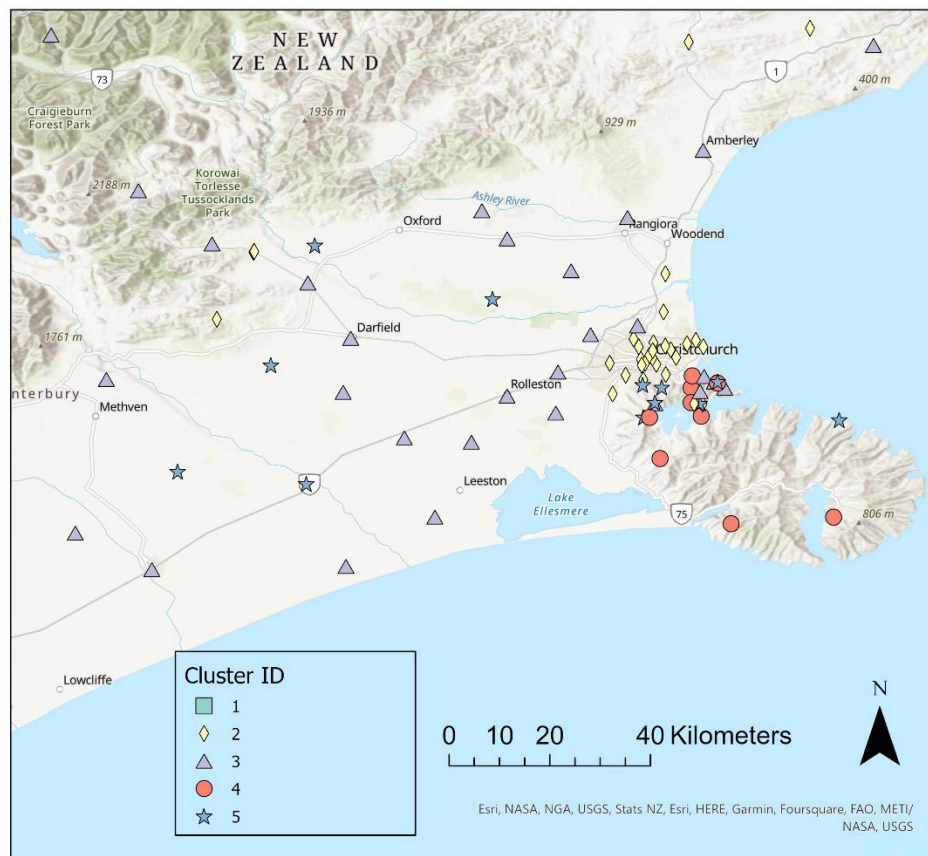


Figure 11. Spatial distribution of sites in each *k*-means cluster in the Canterbury region as an example. The spatial distribution for the whole of NZ is depicted in Fig. S11 (Supporting Information).

Table 1. Statistics of site clustering/classification.

	Cluster/Class	Count	Per cent	Reduction in std
K-means cluster	1	59	19	36 per cent
	2	95	30	41 per cent
	3	87	27	47 per cent
	4	42	13	35 per cent
	5	36	11	42 per cent
NZS1170.5 class	A	8	3	—*
	B	94	29	4 per cent
	C	57	18	4 per cent
	D	157	49	15 per cent
	E	3	1	—

*Mean and standard deviation for NZS1170.5 classes A and E are not presented due to small sample size (<10).

5 DISCUSSION

5.1. Impacts of depth-dependent V_s and Q

As shown previously in Fig. 7(b), stress parameter $\Delta\sigma$ slightly increases with focal depth D when a constant velocity, β , is assumed in eq. (4). However, when a 1-D velocity model (Lee *et al.* 2022) is considered, the trend in $\Delta\sigma$ with D weakens (Fig. 13a). For $D < \sim 7$ km, $\Delta\sigma$ decreases with D and then remains constant for larger D . The correlation between $\Delta\sigma$ and D becomes much weaker with $r = -0.07$ ($p = 0.02$). Therefore, if a depth-dependent velocity model is used, deeper events, which tend to have larger β , would have a lower $\Delta\sigma$, thus ‘absorbing’, to a certain degree, the increasing trend of $\Delta\sigma$ with D resulting from a constant velocity model.

In addition to a depth-dependent velocity, S -wave quality factor Q also varies with depth (Eberhart-Phillips *et al.* 2015), which is not explicitly considered in this study (only implicitly accounted for via the use of hypocentral distance). This could potentially bias the derived source spectra and associated parameters. To evaluate the impacts of depth-dependent Q , we re-ran GIT on two data subsets, one for 705 events with focal depth $D \geq 8$ km, and the other for 716 events with $D \leq 10$ km. 232 events with $8 \leq D \leq 10$ km are shared by both subsets. Other conditions (e.g. reference sites) remain the same.

Fig. 13(b) illustrates that, at a given hypocentral distance, deeper events tend to attenuate faster than shallower ones, particularly at high frequencies. This variation in attenuation will lead to different source spectra for a given event, as shown in Fig. 13(c) (an event shared by both subsets). Deeper events tend to have higher source spectral amplitudes at high frequencies, and accordingly higher f_c (dashed vertical line in Fig. 13c) and $\Delta\sigma$. The comparison of f_c is provided for all events common to both data sets in Fig. S12 in the Supporting Information. Our assessment on the impact of depth-dependent attenuation on source parameters is consistent with that of Bindi *et al.* (2021). They reported that $\Delta\sigma$ for events deeper than 7 km can double, on average, when depth is explicitly introduced in attenuation model.

Introducing depth-dependent velocity and attenuation exerts opposite effects on the depth-dependency of $\Delta\sigma$. The depth-dependence of $\Delta\sigma$ (Fig. 7b) initially absorbed by considering depth-dependent velocity (Fig. 13a) could be revived by explicitly including depth-dependent attenuation (Figs 13b and c). It is worth noting that, even though consistent with Bindi *et al.* (2021), the finding that deeper events tend to attenuate faster than shallower events (with the same hypocentral distance) appears to be counter intuitive, which we are yet to reconcile as of writing. Whether $\Delta\sigma$ increases with

source depth is still debatable (e.g. Abercrombie *et al.* 2021). Ultimately addressing this question may require simultaneously considering the 3-D variations of velocity and Q , which is out of scope of this study. However, we consider the overall pattern of $\Delta\sigma$ is not significantly biased by using a constant velocity model.

5.2. Implications on broad-band ground-motion simulations

It has long been recognized that dynamic properties of earthquake source, such as $\Delta\sigma$, control the between-event variability in the median ground motions (e.g. Cotton *et al.* 2013). Here, we investigate the connection between $\Delta\sigma$ and δB_e of hybrid broad-band ground-motion simulations, building on prior studies of $\Delta\sigma$ and δB_e obtained from parametric empirical ground-motion models (e.g. Trugman & Shearer 2018).

Recently, Lee *et al.* (2022) validated the predictions from hybrid broadband simulations (Graves & Pitarka 2010, 2015) against observations for 479 small magnitude (M_w 3.5–5.0) active shallow crustal earthquakes in NZ. The hybrid approach combines a comprehensive approach to 3-D modelling at low frequencies ($f < 1.0$ Hz) with a simplified-physics-based simulation at high frequencies ($f > 1.0$ Hz). Lee *et al.* (2022) then partitioned the total residual between observation (y_{es}) and prediction (f_{es}) using a partially crossed linear mixed-effects regression algorithm (Stafford 2014; Bates *et al.* 2015):

$$\ln y_{es} - \ln f_{es} = a + \delta B_e + \delta S2S_s + \delta W S_{es} \quad (11)$$

where y_{es} and f_{es} are the ground-motion observation and prediction, respectively, in the form of FAS for event e at site s ; a is the overall prediction bias (fixed effect); δB_e is the between-event residual; $\delta S2S_s$ is the event-corrected systematic residual at site s and $\delta W S_{es}$ is the remaining residual. δB_e , $\delta S2S_s$ and $\delta W S_{es}$ are zero-mean, random variables with orthogonal normal distributions (e.g. Stafford 2014). Subscripts e and s correspond to i and j , respectively, in eq. (2), and are adopted here to comply with commonly used notation for residual analysis (Al Atik *et al.* 2010). In the following, we investigate the correlation patterns between δB_e from Lee *et al.* (2022) and source parameters derived in this study.

Fig. 14(a) depicts the high-frequency decay in the source spectrum, κ_{source} (eq. 3), against δB_e for PGA and FAS ($f = 18.5$ Hz). κ_{source} is negatively correlated with δB_e at high frequencies. Their correlation has also been reported for δB_e of parametric ground-motion prediction models (e.g. Bindi *et al.* 2017; Morasca *et al.* 2023). For $f > 10$ Hz, the correlation between κ_{source} and δB_e becomes stronger with the increase in frequency, though still weak in absolute terms (Fig. 14b), which resembles the correlation pattern between κ_0 (Anderson & Hough 1984) and site response. For forward ground-motion prediction, we examine the dependence of κ_{source} on earthquake magnitude and source depth but find no association (Fig. S13, Supporting Information).

Stress parameter $\Delta\sigma$ is positively correlated with δB_e at relatively high frequencies (Fig. 15a), which is consistent across studies (Oth & Kaiser 2014; Ren *et al.* 2018). This correlation is not surprising since a constant $\Delta\sigma$ was assumed for all events in the simulations (Lee *et al.* 2022) due to the unavailability of event-specific values for forward prediction. Readers should interpret the results of Ren *et al.* (2018) and Oth & Kaiser (2014) with caution due to their much smaller sample size and sequence-specific nature. This study depicts a more complete picture for crustal events across NZ.

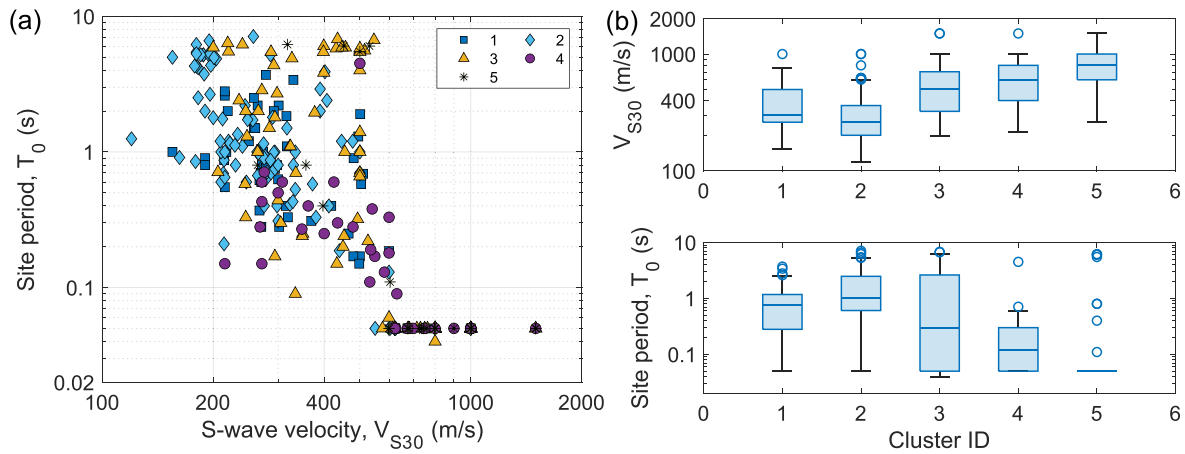


Figure 12. Distribution of V_{S30} and T_0 for the five k -means clusters. Box plots indicate the median, the 25th and 75th percentiles with the whisker extending to the minimum and maximum values that are not outliers.

Table 2. Statistics of V_{S30} and T_0 in each k -means cluster.

Cluster ID	Count	Per cent	V_{S30} (m s^{-1})		T_0 (s)	
			Median	25th–75th	Median	25th–75th
1	59	19	298	[261 497]	0.75	[0.28 1.18]
2	95	30	264	[202 363]	1	[0.61 2.48]
3	87	27	500	[323 706]	0.3	[0.05 2.63]
4	42	13	600	[400 800]	0.12	[0.05 0.3]
5	36	11	800	[602 1000]	0.05	[0.05 0.05]

The positive correlation between δB_e and $\Delta\sigma$ increases with frequency till $f = \sim 6$ Hz and then flattens out at approximately $r = 0.56$ (Fig. 15b). The small r values at relatively low frequencies are due to the weak dependence of $\Delta\sigma$ on Fourier amplitudes for $f < f_c$. However, the relatively large r values for $f > \sim 6$ Hz confirms the potential to reduce between-event variability at relatively high frequencies in ground-motion predictions by accounting for variations in $\Delta\sigma$ (e.g. Oth et al. 2017; Bindi et al. 2018; Trugman & Shearer 2018; Lee et al. 2022; Morasca et al. 2023).

Since $\Delta\sigma$ is generally unknown for a future scenario, one can only utilize its dependence on other source properties, and/or spatial-temporal variation patterns, if any, in forward predictions. As discussed in the prior Section 4.1.2, we only found a negligible to weak association of $\Delta\sigma$ with earthquake magnitude and depth. Thus, we further investigate its spatial variation. Given the difficulty in obtaining the absolute values of $\Delta\sigma$, we only identify regions with stress parameter lower or higher than the national average.

Stress parameter exhibits a statistically significant spatial dependence (Fig. 16a). Global Moran's I test (Moran 1950) gives a Moran's I Index of 0.31. The I Index ranges between -1.0 and 1.0 with a positive I value indicating a tendency toward clustering, while a negative I value for a tendency toward dispersion. Statistical significance measures, z -score = 2.46 and $p = 0.01$, reject the null hypothesis (the values associated with features are randomly distributed). Many prior studies (e.g. Allmann & Shearer 2007; Zhang et al. 2022) have also reported the spatial dependence of $\Delta\sigma$, which cannot be explained as an artefact of varying rupture velocity (e.g. Abercrombie et al. 2017b).

It is intriguing as to how/whether the spatial pattern extracted from historical data can improve forward predictions of even small-to-moderate events. Given the distinct pattern of $\Delta\sigma$ in space, we then utilize a spatially constrained clustering to detect spatially contiguous groups in $\Delta\sigma$. Based on Fig. 16(a), we set the number

of clusters at eight. Clustered data are shown in Fig. 16(b). The lateral variation pattern of $\Delta\sigma$ is consistent with that of δB_e for PGA (fig. 13 in Lee et al. 2022). The central region of the South Island, which has an overprediction of PGA (negative δB_e) as reported by Lee et al. (2022), is detected as Cluster 2 which has a lower-than-average $\Delta\sigma$ (Fig. 16c). This is consistent with the overprediction of PGA in this area (Lee et al. 2022). Likewise, the higher-than-average $\Delta\sigma$ in the Canterbury region (i.e. Cluster 8) is consistent with the underprediction of PGA (positive δB_e) for events in this area.

Spatial clustering can capture a sizable portion of the variability in $\Delta\sigma$. Within-cluster standard deviation $\sigma_{\ln\Delta\sigma}$ for Cluster 1–8 is 0.69, 0.55, 0.77, 0.62, 0.52, 0.67, 0.42 and 0.56, respectively (Fig. S14, Supporting Information), corresponding to a reduction of 17, 33, 8, 25, 37, 19, 49 and 32 per cent relative to Cluster 0 (the entire data set, i.e. 0.83). Noting that Cluster 7 has relatively smaller sample size (no.: 19, Fig. S14, Supporting Information) and thus has a higher uncertainty in the estimate of its standard deviation and corresponding reduction (i.e. 49 per cent). However, it is unequivocal that variance in $\Delta\sigma$ can be partially explained by its lateral variation for crustal events in NZ. A future line of study is to quantify to what extent the portion of $\Delta\sigma$ variability explained via spatial clustering can be translated to improvement in ground-motion prediction. While we have briefly illustrated here a spatial clustering analysis, application-orientated models could also be framed using conventional geostatistical approaches.

6 CONCLUSIONS

To gain new insights into the ground-motion phenomenon, we decompose two decades of seismicity data from shallow crustal events from 2000 to 2021 in NZ via a non-parametric GIT. From the inverted source spectra, we derive Brune's stress parameter, $\Delta\sigma$,

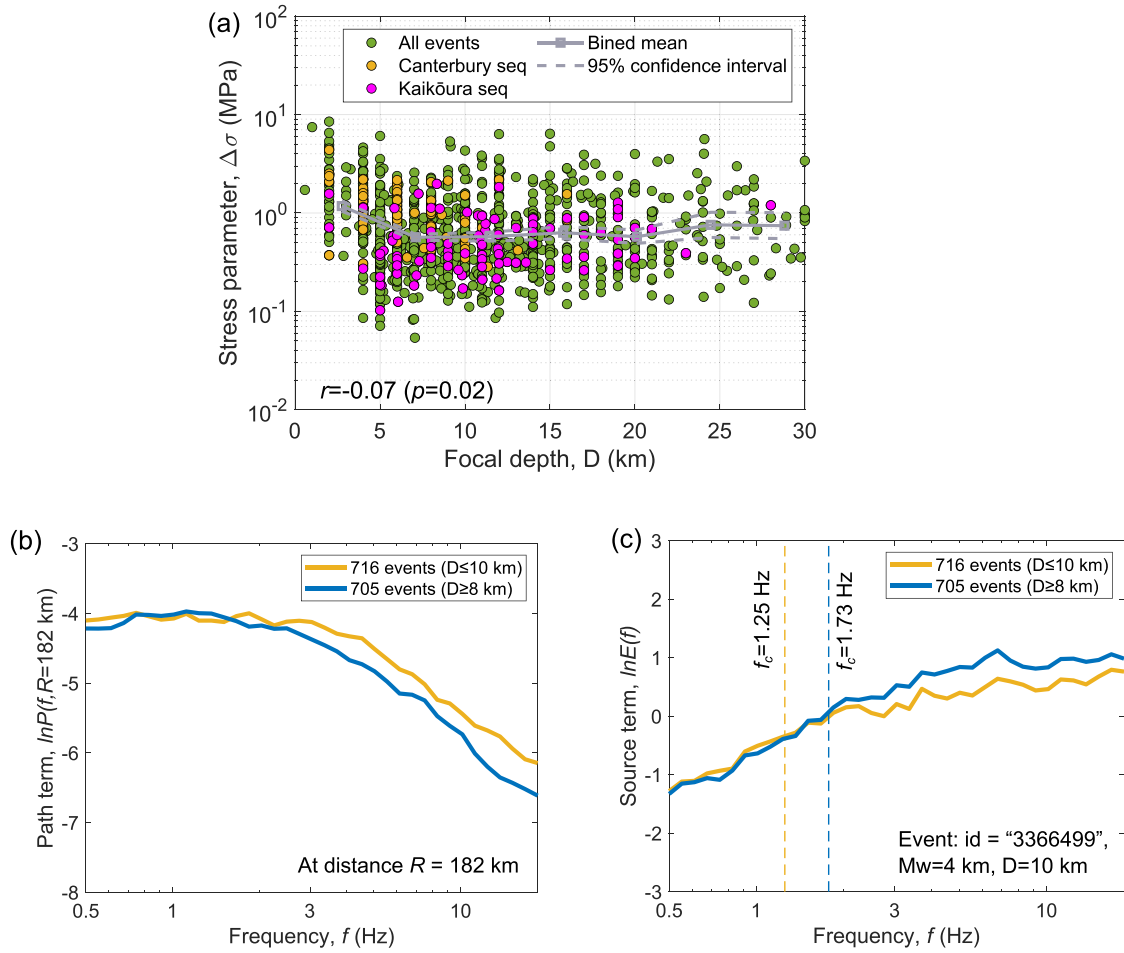


Figure 13. (a) Focal depth dependence of $\Delta\sigma$ considering depth-dependent velocity via a 1-D velocity model. (b) Path attenuation at a distance of $R = 182 \text{ km}$ inverted from two subsets, one containing 705 events with focal depth $D \geq 8 \text{ km}$, and the other consisting of 716 events with $D \leq 10 \text{ km}$, and 232 events with $8 \leq D \leq 10 \text{ km}$ are shared by both subsets. (c) Corresponding source spectra for one example event (id = '3366499') considering depth-dependent attenuation in panel (b).

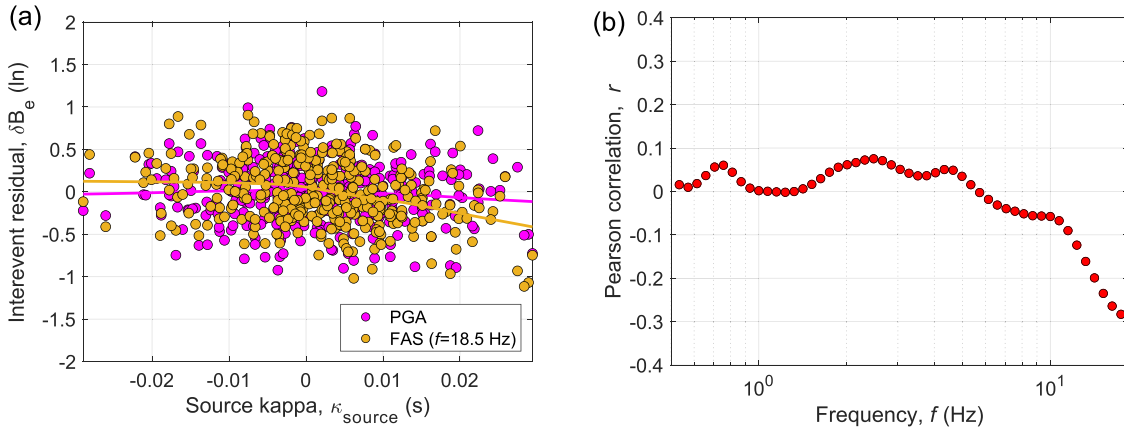


Figure 14. (a) Between-event term, δB_e , of 3-D simulations (Lee *et al.* 2022) for PGA and FAS ($f = 18.5 \text{ Hz}$) versus κ_{source} inverted herein, and (b) Pearson correlation between κ_{source} and δB_e at different frequencies.

which is found to follow a lognormal distribution with a \log_{10} standard deviation of 0.36 or equivalently 0.83 in natural log unit. $\Delta\sigma$ slightly increases with focal depth and is practically independent of earthquake size (i.e. self-similar), but exhibits a statistically significant spatial clustering, which can explain a sizable portion of its variability. Based on the inverted attenuation curves, a multisegment

apparent geometric-spreading function and a distance-dependent quality-factor $Q(f)$ model are found to best describe the attenuation on average in NZ, though a single $Q(f)$ is also obtained for the whole distance range: $Q(f) = 149.1 f^{0.62}$. Further, using the site response from GIT, we find that the soil classification scheme specified in NZ seismic code, NZS1170.5, has a limited capability

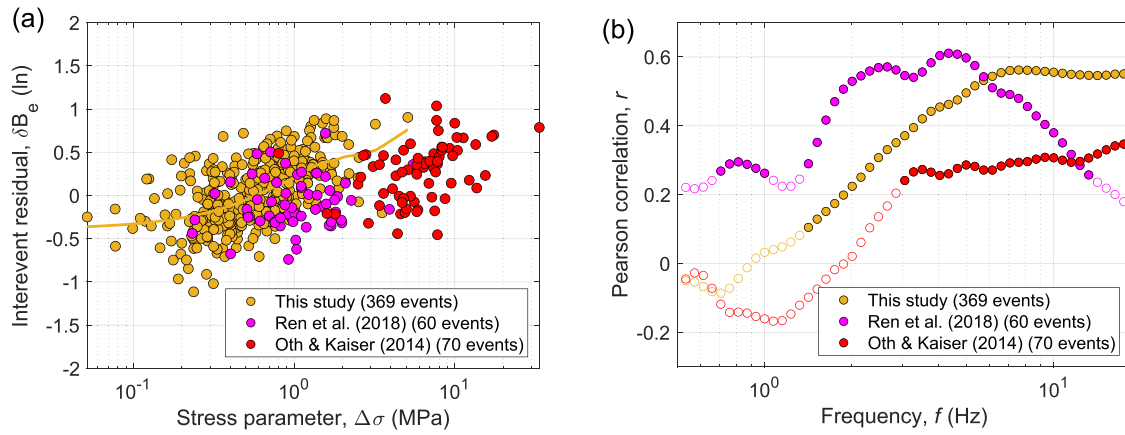


Figure 15. (a) Between-event term, δB_e , of 3-D simulations (Lee *et al.* 2022) for FAS ($f = 18.5$ Hz) versus Brune’s stress parameter, $\Delta\sigma$, inverted in this study, and (b) Pearson correlation between $\Delta\sigma$ and δB_e at different frequencies (open circles represent those with $p > 0.05$).

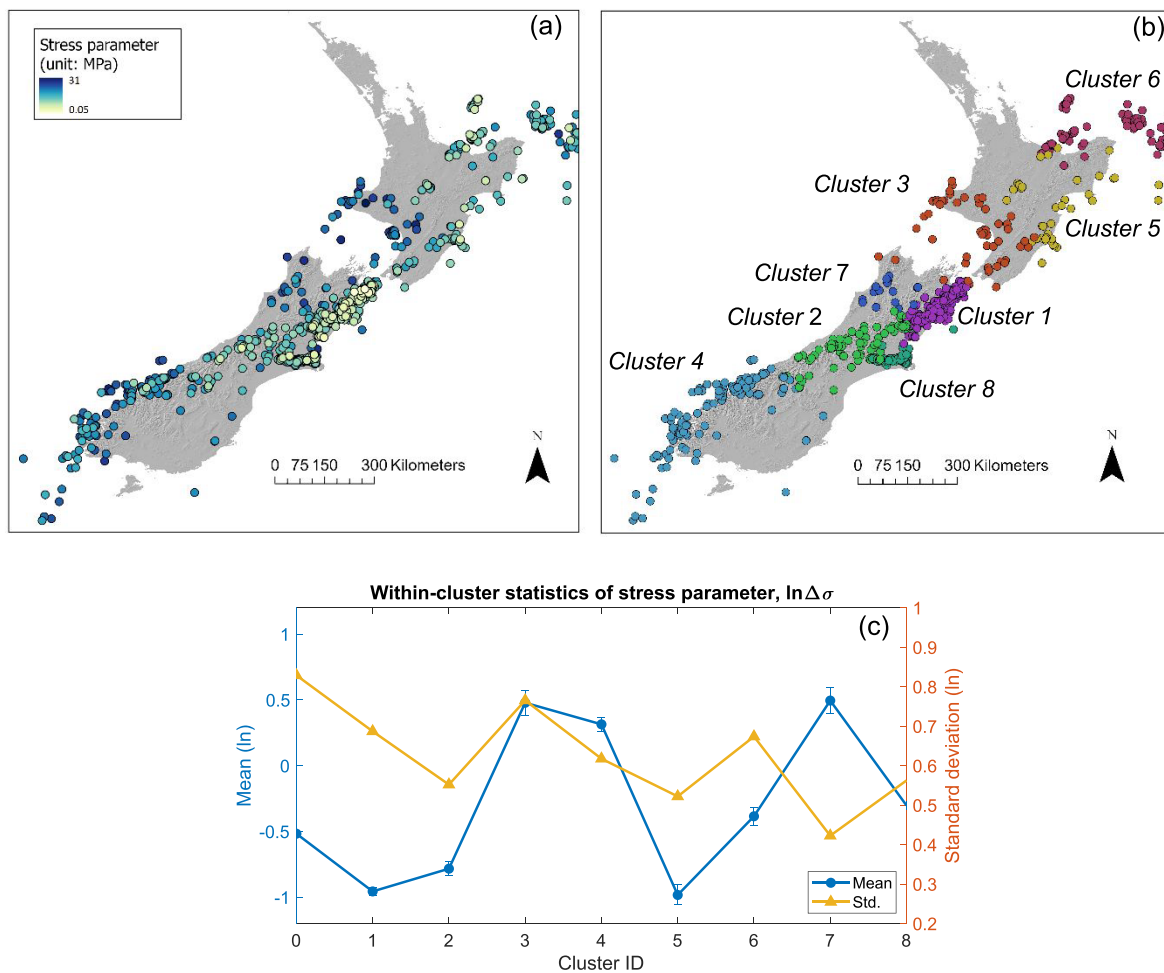


Figure 16. Spatial distribution of stress parameter, $\Delta\sigma_i$ (MPa), for crustal events, (a) continuous data, and (b) clustered data (no colour scale). (c) Statistics of stress parameters, $\ln\Delta\sigma_i$, within each cluster (b) where Cluster ID = 0 corresponds to the whole data set without clustering.

in discerning the site-specific frequency-dependent amplification functions.

This paper reveals the general trends and patterns in observations of past shallow crustal events in NZ. Despite this, it is important to keep in mind the necessity of probing some more subtle physical processes, for example, 3-D anelastic attenuation, M_w -dependent

near-source geometric spreading, the temporal variation in $\Delta\sigma$, and nonlinear soil effects during strong shaking, and other event-specific effects in site response at sites with 3-D features. Specifically, a refined attenuation model affects the computed site response and source parameters from the inversion. Nevertheless, the insights gained here from historical data on the properties of fault regions

and the propagation media can capture the first-order phenomena and potentially lead to better predictions of ground shakings during future events.

ACKNOWLEDGMENTS

This research is partially supported through funding from QuakeCoRE: The New Zealand Centre for Earthquake Resilience, the New Zealand Earthquake Commission, and The Royal Society of New Zealand Marsden Fund. The authors thank Hiroshi Kawase for the discussion, and Gail Atkinson for providing very constructive feedback on an early version of the manuscript. We also would like to thank John Thornley and one anonymous reviewer for their input on this article. This is QuakeCoRE Publication 903.

AUTHOR CONTRIBUTIONS

Chuanbin Zhu (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Methodology [lead], Validation [lead], Visualization [lead], Writing – original draft [lead], Writing – review & editing [lead]), Sanjay Bora (Formal analysis [supporting], Investigation [supporting], Methodology [supporting], Writing – original draft [supporting], Writing – review & editing [supporting]), Brendon A. Bradley (Formal analysis [supporting], Funding acquisition [lead], Investigation [supporting], Methodology [supporting], Project administration [lead], Resources [lead], Writing – review & editing [supporting]), and Dino Bindi (Formal analysis [supporting], Investigation [supporting], Methodology [supporting], Writing – review & editing [supporting])

SUPPORTING INFORMATION

Supplementary data are available at [GJIRAS](https://doi.org/10.1002/gjir.1403) online.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

DATA AND RESOURCES

The ground-motion database (GMDB, v3.0, Hutchinson *et al.* 2024) for New Zealand was downloaded at: https://osf.io/q9yrg/?view_only=05337ba1ebc744fc96b9924de633ca0e (last accessed on 2022-11-14). Data processing, *k*-means clustering and correlation analyses were carried out in MATLAB. Esri ArcGIS Pro (www.esri.com/software/arcgis) was utilized for spatial clustering of source stress parameter, PyGMT (<https://doi.org/10.5281/zenodo.4592991>) was used to produce some geospatial plots. Supplemental materials include one word file ‘Supplement_S1’. We adhere to the FAIR data principles. Other resources can be accessed online via <https://figshare.com/s/1a007b399856d743c407>. These online resources include inverted source spectra and parameters for each event, as well as site-response function for each site: ‘Source_parameters.xlsx’, ‘DataSet_Source’, ‘Amp_mean_uncertainty.xlsx’ and ‘DataSet_Amp’. Among them, ‘Source_parameters.xlsx’ contains source parameters for each event while plots of inverted and theoretical source spectra are provided in ‘DataSet_Source’. ‘Amp_mean_uncertainty.xlsx’ contains inverted site-response functions (mean with uncertainty estimate and site metadata) at each of the 319 strong-motion stations (HN and BN) presented in the ‘Site Response’ section. ‘DataSet_Amp’ provides

plots of site-response curves at each site, and its corresponding azimuth coverage.

REFERENCES

- EN 1998-1:2004 Eurocode 8: Design of structures for earthquake resistance - Part 1: General rules, seismic actions and rules for buildings. European Committee for Standardization (CEN), Brussels.
- Abercrombie, R.E., Bannister, S., Ristau, J. & Doser, D. (2017a). Variability of earthquake stress drop in a subduction setting, the Hikurangi Margin, New Zealand. *Geophys. J. Int.*, **208**, 306–320.
- Abercrombie, R.E., Poli, P. & Bannister, S. (2017b). Earthquake directivity, orientation, and stress drop within the subducting plate at the Hikurangi margin, New Zealand. *J. geophys. Res. Solid Earth*, **122**, 10 176–110 188.
- Abercrombie, R.E., *et al.* (2021). Does earthquake stress drop increase with depth in the crust? *J. geophys. Res. Solid Earth*, **126**, e2021JB022314, doi:10.1029/2021JB022314.
- Aki, K. (1967). Scaling law of seismic spectrum. *J. geophys. Res.*, **72**, 1217–1231.
- Al Atik, L., Abrahamson, N., Bommer, J.J., Scherbaum, F., Cotton, F. & Kuehn, N. (2010). The variability of ground-motion prediction models and its components. *Seismol. Res. Lett.*, **81**, 794–801.
- Allmann, B.P. & Shearer, P.M. (2007). Spatial and temporal stress drop variations in small earthquakes near Parkfield, California. *J. geophys. Res. Solid Earth*, **112**, B04305, doi:10.1029/2006JB004395.
- Anderson, J.G. & Hough, S.E. (1984). A model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies. *Bull. seism. Soc. Am.*, **74**, 1969–1993.
- Andrews, D.J. (1986). Objective determination of source parameters and similarity of earthquakes of different size. in *Earthquake Source Mechanics*, Das, S., Boatwright, J. & Scholz, C. H. (Eds), Geophysical Monograph Series, Vol., **37**, American Geophysical Union, Washington, DC, pp. 259–267.
- Atkinson, G.M. & Mereu, R.F. (1992). The shape of ground motion attenuation curves in southeastern Canada. *Bull. seism. Soc. Am.*, **82**, 2014–2031.
- Baltay, A.S., Abercrombie, R.E. & Taira, T. (2021). A community stress drop validation study using the 2019 Ridgecrest Earthquake dataset, in *Seismological Society of America Annual Meeting*, **2021**.
- Baltay, A.S., Hanks, T.C. & Beroza, G.C. (2013). Stable stress-drop measurements and their variability: implications for ground-motion prediction. *Bull. seism. Soc. Am.*, **103**, 211–222.
- Bates, D.M., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.*, **67**, 1–48.
- Bindi, D., Cotton, F., Spallarossa, D., Picozzi, M. & Rivalta, E. (2018). Temporal variability of ground shaking and stress drop in Central Italy: a hint for fault healing? *Bull. seism. Soc. Am.*, **108**, 1853–1863.
- Bindi, D. & Kotha, S.R. (2020). Spectral decomposition of the Engineering Strong Motion (ESM) flat file: regional attenuation, source scaling and Arias stress drop. *Bull. Earthq. Eng.*, **18**, 2581–2606.
- Bindi, D., Pacor, F., Luzi, L., Massa, M. & Ameri, G. (2009). The Mw 6.3, 2009 L’Aquila earthquake: source, path and site effects from spectral analysis of strong motion data. *Geophys. J. Int.*, **179**, 1573–1579.
- Bindi, D., Razafindrakoto, H.N.T., Picozzi, M. & Oth, A. (2021). Stress drop derived from spectral analysis considering the hypocentral depth in the attenuation model: application to the Ridgecrest Region, California. *Bull. seism. Soc. Am.*, **111**, 3175–3188.
- Bindi, D., Spallarossa, D. & Pacor, F. (2017). Between-event and between-station variability observed in the Fourier and response spectra domains: comparison with seismological models. *Geophys. J. Int.*, **210**, 1092–1104.
- Bindi, D., Spallarossa, D., Picozzi, M., Oth, A., Morasca, P. & Mayeda, K. (2023). The community stress-drop validation study—Part II: uncertainties of the source parameters and stress drop analysis. *Seismol. Res. Lett.*, **94**, 1992–2002.
- Boore, D.M. & Boatwright, J. (1984). Average body-wave radiation coefficients. *Bull. seism. Soc. Am.*, **74**, 1615–1621.

- Boore, D.M. & Thompson, E.M.(2014). Path durations for use in the stochastic-method simulation of ground motions. *Bull. seism. Soc. Am.*, **104**, 2541–2552.
- Bradley, B.A. & Cubrinovski, M.(2011). Near-source strong ground motions observed in the 22 February 2011 Christchurch earthquake. *Seismol. Res. Lett.*, **82**, 853–865.
- Bradley, B.A., Wotherspoon, L.M. & Kaiser, A.E.(2017). Ground motion and site effect observations in the Wellington region from the 2016 Mw 7.8 Kaikōura, New Zealand, earthquake. *Bull. New Zeal. Soc. Earthq. Eng.*, **50**, 94–105.
- Brune, J.N.(1970). Tectonic stress and the spectra of seismic shear waves from earthquakes. *J. geophys. Res.*, **75**, 4997–5009.
- Brune, J.N.(1971). Seismic sources, fault plane studies and tectonics. *EOS, Trans. Am. geophys. Un.*, **52**(5), IUGG 178–IUGG 187.
- Burger, R.W., Somerville, P.G., Barker, J.S., Herrmann, R.B. & HelMBERGER, D.V.(1987). The effect of crustal structure on strong ground motion attenuation relations in eastern North America. *Bull. seism. Soc. Am.*, **77**, 420–439.
- Castro, R.R., Anderson, J.G. & Singh, S.K.(1990). Site response, attenuation and source spectra of S waves along the Guerrero, Mexico, subduction zone. *Bull. seism. Soc. Am.*, **80**, 1481–1503.
- Cotton, F., Archuleta, R. & Causse, M.(2013). What is sigma of the stress drop? *Seismol. Res. Lett.*, **84**, 42–48.
- Drouet, S., Chevrot, S., Cotton, F. & Souriau, A.(2008). Simultaneous inversion of source spectra, attenuation parameters, and site responses: application to the data of the French accelerometric network. *Bull. seism. Soc. Am.*, **98**, 198–219.
- Eberhart-Phillips, D., Reyners, M. & Bannister, S. (2015). A 3D QP attenuation model for all of New Zealand. *Seismol. Res. Lett.*, **86**, 1655–1663.
- Edwards, B., Rietbrock, A., Bommer, J.J. & Baptie, B.(2008). The acquisition of source, path, and site effects from microearthquake recordings using Q tomography: application to the United Kingdom. *Bull. seism. Soc. Am.*, **98**, 1915–1935.
- Eshelby, J.D.(1957). The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc. R. Soc. Lond.*, **241**, 376–396.
- Foster, K.M., Bradley, B.A., McGann, C.R. & Wotherspoon, L.M.(2019). A VS30 map for New Zealand based on geologic and terrain proxy variables and field measurements. *Earthq. Spectra*, **35**, 1865–1897.
- Fu, L., Chen, S., Li, J., Zhang, L., Xie, J. & Li, X.(2023). Regional spectral characteristics derived using the generalized inversion technique and applications to stochastic simulation of the 2021 Mw 6.1 Yangbi Earthquake. *Bull. seism. Soc. Am.*, **113**, 378–400.
- Gerstenberger, M.C., et al.(2024). The 2022 Aotearoa New Zealand National seismic hazard model: process, overview, and results. *Bull. seism. Soc. Am.*, **114**, 7–36.
- GNS Science. (2012). Geological Map of New Zealand [Data set]. *GNS Science*, doi:10.21420/QF82-7D42.
- GNS Science. (2022). GeoNet Aotearoa New Zealand seismic digital waveform dataset [Data set]. *GeoNet*, doi:10.21420/G19Y-9D40.
- Graves, R.W. & Pitarka, A.(2010). Broadband ground-motion simulation using a hybrid approach. *Bull. seism. Soc. Am.*, **100**(5A), 2095–2123.
- Graves, R.W. & Pitarka, A.(2015). Refinements to the Graves and Pitarka (2010) broadband ground motion simulation method. *Seismol. Res. Lett.*, **86**, 75–80.
- Hough, S.E. & Anderson, J.G.(1988). High-frequency spectra observed at Anza, California: implications of Q structure. *Bull. seism. Soc. Am.*, **78**, 692–707.
- Hutchinson, J.A. et al.(2024). 2023 New Zealand ground motion database. *Bull. seism. Soc. Am.*, **114**, 291–310.
- Izutani, Y. & Kanamori, H.(2001). Scale-dependence of seismic energy-to-moment ratio for strike-slip earthquakes in Japan. *Geophys. Res. Lett.*, **28**, 4007–4010.
- Ji, K., Zhu, C., Yaghmaei-Sabegh, S., Lu, J., Ren, Y. & Wen, R.(2022). Site classification using deep-learning-based image recognition techniques. *Earthq. Eng. Struct. Dyn.*, **52**, 2323–2338.
- Kaiser, A.E. et al.(2012). The Mw 6.2 Christchurch earthquake of February 2011: preliminary report. *New Zeal. J. Geol. Geophys.*, **55**, 1, 67–90.
- Kalkan, E.(2016). An automatic P-phase arrival time picker. *Bull. seism. Soc. Am.*, **106**, 971–986.
- Kanamori, H. & Rivera, L.(2004). Static and dynamic scaling relations for earthquakes and their implications for rupture speed and stress drop. *Bull. seism. Soc. Am.*, **94**, 314–319.
- Kawase, H. & Matsuo, H. (2004). Amplification characteristics of K-NET, KiK-net, and JMA Shindokei network sites based on the spectral inversion technique, in *13th World Conference on Earthquake Engineering*, Vancouver, Canada, 1–6 August 2004, Paper No. 454.
- Keilis-Borok, V.(1959). On the estimation of the displacement in an earthquake source and of source dimension. *Ann. Geofisc.*, **12**, 205–214.
- Konno, K. & Ohmachi, T.(1998). Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. *Bull. seism. Soc. Am.*, **88**, 228–241.
- Kotha, S.R., Cotton, F. & Bindi, D.(2018). A new approach to site classification: mixed-effects ground motion prediction equation with spectral clustering of site amplification functions. *Soil Dyn. Earthq. Eng.*, **110**, 318–329.
- Lee, R.L., Bradley, B., A., S., P. J., Graves, R.W. & Rodriguez-Marek, A. (2022). Hybrid broadband ground-motion simulation validation of small magnitude active shallow crustal earthquakes in New Zealand. *Earthq. Spectra*, **38**, 2548–2579.
- MacQueen, J.B.(1967). Some methods for classification and analysis of multivariate observations. In Le Cam, L. M. & Neyman, J.(Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*(Vol. 1, pp. 281–297). California: University of California Press.
- Moran, P.A.P.(1950). Notes on continuous stochastic phenomena. *Biometrika*, **37**, 17–23.
- Morasca, P., D’Amico, M., Sgobba, S., Lanzano, G., Colavitti, L., Pacor, F. & Spallarossa, D.(2023). Empirical correlations between an FAS non-ergodic ground motion model and a GIT derived model for Central Italy. *Geophys. J. Int.*, **233**, 51–68.
- Nakano, K., Matsushima, S. & Kawase, H.(2015). Statistical properties of strong ground motions from the generalized spectral inversion of data observed by K-NET, KiK-net, and the JMA Shindokei Network in Japan. *Bull. seism. Soc. Am.*, **105**, 2662–2680.
- NZS1170.5(2004). *Structural Design Actions—Earthquake Actions. Section 3 - Site Hazard Spectra*, Standards New Zealand. p. 81.
- Oth, A.(2013). On the characteristics of earthquake stress release variations in Japan. *Earth planet. Sci. Lett.*, **377–378**, 132–141.
- Oth, A., Bindi, D., Parolai, S. & Di Giacomo, D.(2010). Earthquake scaling characteristics and the scale-(in)dependence of seismic energy-to-moment ratio: insights from KiK-net data in Japan. *Geophys. Res. Lett.*, **37**, L19304, doi:10.1029/2010GL044572.
- Oth, A., Bindi, D., Parolai, S. & Di Giacomo, D.(2011). Spectral analysis of K-NET and KiK-net data in Japan, part II: on attenuation characteristics, source parameters, and site response of borehole and surface stations. *Bull. seism. Soc. Am.*, **101**, 667–687.
- Oth, A. & Kaiser, A.E.(2014). Stress release and source scaling of the 2010–2011 Canterbury, New Zealand, earthquake sequence from spectral inversion of ground motion data. *Pure appl. Geophys.*, **171**, 2767–2782.
- Oth, A., Miyake, H. & Bindi, D.(2017). On the relation of earthquake stress drop and ground motion variability. *J. geophys. Res. Solid Earth*, **122**, 5474–5492.
- Pennington, C.N., Chen, X., Abercrombie, R.E. & Wu, Q.(2021). Cross validation of stress drop estimates and interpretations for the 2011 Prague, OK, earthquake sequence using multiple methods. *J. geophys. Res. Solid Earth*, **126**, e2020JB020888, doi:10.1029/2020JB020888.
- Perron, V., et al.(2018). Selecting time windows of seismic phases and noise for engineering seismology applications: a versatile methodology and algorithm. *Bull. Earthq. Eng.*, **16**, 2211–2225.
- Ren, Y., Zhou, Y., Wang, H. & Wen, R.(2018). Source characteristics, site effects, and path attenuation from spectral analysis of strong-motion recordings in the 2016 Kaikōura Earthquake sequence. *Bull. seism. Soc. Am.*, **108**, 1757–1773.
- Reyners, M.(1989). New Zealand seismicity 1964–87: an interpretation. *New Zeal. J. geol. Geophys.*, **32**, 307–315.

- Ristau, J.(2008). Implementation of routine regional moment tensor analysis in New Zealand. *Seismol. Res. Lett.*, **79**, 400–415.
- Schober, P., Boer, C. & Schwarte, L.A.(2018). Correlation coefficients: appropriate use and interpretation. *Anesth. Analg.*, **126**, 1763–1768.
- Shearer, P.M., Prieto, G.A. & Hauksson, E.(2006). Comprehensive analysis of earthquake source spectra in Southern California. *J. geophys. Res. Solid Earth*, **111**, B06303.
- Shible, H. et al.(2022). GITEC: a generalized inversion technique benchmark. *Bull. seism. Soc. Am.*, **112**, 850–877.
- Stafford, P.J.(2014). Crossed and nested mixed-effects approaches for enhanced model development and removal of the ergodic assumption in empirical ground-motion models. *Bull. seism. Soc. Am.*, **104**, 702–719.
- Steidl, J.H., Tumarkin, A.G. & Archuleta, R.J.(1996). What is a reference site? *Bull. seism. Soc. Am.*, **86**, 1733–1748.
- Trifunac, M.D. & Brady, A.G.(1975). A study on the duration of strong earthquake ground motion. *Bull. seism. Soc. Am.*, **65**, 581–626.
- Trugman, D.T.(2020). Stress-drop and source scaling of the 2019 Ridgecrest, California, earthquake sequence. *Bull. seism. Soc. Am.*, **110**, 1859–1871.
- Trugman, D.T., Dougherty, S.L., Cochran, E.S. & Shearer, P.M.(2017). Source spectral properties of small to moderate earthquakes in southern Kansas. *J. geophys. Res. Solid Earth*, **122**, 8021–8034.
- Trugman, D.T. & Shearer, P.M.(2017). Application of an improved spectral decomposition method to examine earthquake source scaling in Southern California. *J. geophys. Res. Solid Earth*, **122**, 2890–2910.
- Trugman, D.T. & Shearer, P.M.(2018). Strong correlation between stress drop and peak ground acceleration for recent M 1–4 earthquakes in the San Francisco Bay Area. *Bull. seism. Soc. Am.*, **108**, 929–945.
- Wang, H., Ren, Y., Wen, R. & Xu, P.(2019). Breakdown of earthquake self-similar scaling and source rupture directivity in the 2016–2017 central Italy seismic sequence. *J. geophys. Res. Solid Earth*, **124**, 3898–3917.
- Wotherspoon, L.M., Kaiser, A.E., Stolte, A.C. & Manea, E.F.(2024). Development of the site characterization Database for the 2022 New Zealand national seismic hazard model. *Seismol. Res. Lett.*, **95**, 214–225.
- Yaghmaei-Sabegh, S. & Rupakhety, R.(2020). A new method of seismic site classification using HVSR curves: a case study of the 12 November 2017 Mw 7.3 Ezgeleh earthquake in Iran. *Eng. Geol.*, **270**, 105574, doi:10.1016/j.enggeo.2020.105574.
- Yenier, E. & Atkinson, G.(2014). Equivalent point-source modeling of moderate-to-large magnitude earthquakes and associated ground-motion saturation effects. *Bull. seism. Soc. Am.*, **104**, 1458–1478.
- Zhang, J., Chen, X. & Abercrombie, R.E.(2022). Spatiotemporal variability of earthquake source parameters at Parkfield, California, and their relationship with the 2004 M6 earthquake. *J. geophys. Res. Solid Earth*, **127**, e2021JB022851, doi:10.1029/2021JB022851.
- Zheng, A. & Casari, A., 2018. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*, O'Reilly Media.
- Zhu, C., Cotton, F., Kawase, H., Haendel, A., Pilz, M. & Nakano, K.(2022). How well can we predict earthquake site response so far? Site-specific approaches. *Earthq. Spectra*, **38**, 1047–1075.
- Zhu, C., Cotton, F., Kawase, H. & Nakano, K.(2023). How well can we predict earthquake site response so far? Machine learning vs. physics-based modeling. *Earthq. Spectra*, **39**, 478–504.