Hybrid broadband ground-motion simulation validation of small magnitude active shallow crustal earthquakes in New Zealand

**Robin L Lee, M. EERI[1], Brendon A Bradley, M. EERI[1] (iD),
Peter J Stafford, M. EERI[2] (iD), Robert W Graves, M.
EERI[3], and Adrian Rodriguez-Marek, M. EERI[4] (iD)**

Abstract

This article presents a comprehensive validation of the hybrid broadband ground-motion simulation approach (via the commonly used Graves and Pitarka method) in a New Zealand context with small magnitude point source ruptures using an extensive set of 5218 ground motions recorded at 212 sites from 479 active shallow crustal earthquakes across the country. Modifications to the simulation method inferred from a previous New Zealand validation are implemented, and the improvements are explicitly quantified. Empirical ground-motion models are also considered to provide a benchmark for simulation prediction accuracy and precision. Examination of intensity measure residuals identifies that the simulation method modifications lead to reduced model prediction bias and within-event variability and provides evidence toward the use of spatially varying coefficient models for simulation parameters, such as the high-frequency Brune stress parameter. Additional biases identified include, among others, underprediction of significant durations at soft soil sites and overprediction of short-period pseudo-spectral accelerations at stiff alluvial gravel and rock sites due to low-estimated 30 m time-averaged shear-wave velocity values.

Keywords

Ground-motion simulation, hybrid broadband ground-motion simulation, ground-motion prediction, ground-motion simulation validation, New Zealand, small magnitude earthquake

[1]Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand
[2]Department of Civil and Environmental Engineering, Imperial College London, London, UK
[3]US Geological Survey, Pasadena, CA, USA
[4]The Charles E. Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, USA

Corresponding author:
Robin L Lee, Civil and Natural Resources Engineering, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand.
Email: robin.lee@canterbury.ac.nz

Introduction

Validation of physics-based ground-motion models (GMMs) is important toward their utilization in seismic hazard analysis (Graves et al., 2011) and earthquake engineering applications (Bijelić et al., 2018; Bradley et al., 2017a; Galasso et al., 2013). In the recent decade, several significant ground-motion simulation validation efforts have been carried out to assess the predictive capability of ground-motion simulation methods (e.g. Dreger and Jordan, 2015; Maufroy et al., 2015) In particular, the Southern California Earthquake Centre (SCEC) Broadband Platform (BBP) exercise (Goulet et al., 2015) rigorously evaluated four widely used simulation methods implemented on the SCEC BBP (Atkinson and Assatourians, 2015; Crempien and Archuleta, 2015; Graves and Pitarka, 2015; Olsen and Takedatsu, 2015). In that study, 12 events of magnitude $M_w \geqslant 4.60$ and a total of 394 pairs of orthogonal horizontal component observed ground-motion records were considered (from a potential set of 918 records within 200 km of their causative fault). From comparisons between simulated and observed pseudo-spectral accelerations (pSAs), and empirically predicted pSA, the predictive capability of the methods was quantified. In addition, several improvements to each of the simulation methods were identified as a part of the validation. While a few other validation studies have used a comparable number of events to the SCEC BBP exercise (e.g. Maufroy et al., 2016; Taborda et al., 2016; Yenier and Atkinson, 2015), the majority of ground-motion simulation validation studies focus on one or few specific events, such as the 1994 M_w 6.7 Northridge and 2008 M_w 5.4 Chino Hills earthquakes, to demonstrate new ground-motion simulation methods or input developments. For example, several recent studies have focused on validation of simulations which are deterministically calculated to frequencies as high as 10.0 Hz and hence require additional fault and velocity model complexities (Graves and Pitarka, 2016; Savran and Olsen, 2019), compared to previous ground-motion simulation modeling studies, or explicitly consider plasticity (Withers et al., 2019). Other studies have also focused on the validation of earthquake ruptures that are currently not as well understood, such as great subduction zone earthquakes (Wirth et al., 2017).

As the literature cited in the previous paragraph illustrates, validation against observations in an earthquake engineering context has usually been limited in the number of earthquake events or number of ground-motion recording stations (and consequently regional extent) considered. This is principally due to the scope of the particular studies and the choice to consider large M_w earthquakes which are spatially sparse and temporally infrequent. Small magnitude earthquakes ($M_w \leqslant 5.0$) occur significantly more frequently, and although their direct relevance in earthquake engineering applications is limited, they provide valuable information toward validation as their underlying processes are fundamentally the same as for large M_w earthquakes. In addition, the relative simplicity in source modeling of small M_w earthquakes allows for greater focus on the ground-motion simulation method, crustal velocity model, and near-surface site response to weak motions. Therefore, arguably, the first step toward an exhaustive validation study of a spatially extensive region is to consider small M_w earthquakes.

Recently in a New Zealand (NZ) context, Lee et al. (2020) presented a comprehensive hybrid broadband (BB) ground-motion simulation (Graves and Pitarka, 2010) validation study in the Canterbury region considering 148 small $M_w(3.5 \leqslant M_w \leqslant 5.0)$ earthquakes, with 1896 observed ground-motion records. Intensity measures (IMs) calculated from observed and simulated ground motions were compared and an analysis of prediction residuals across the aggregated dataset was carried out to quantify the biases in the simulations and their causes. Several potential improvements to the ground-motion simulation method in

a NZ context were identified. Prior to the Lee et al. (2020) study, only large M_w scenarios had been considered in NZ for several prominent events, such as the 2011 M_w 6.2 Christchurch (Razafindrakoto et al., 2018) and 2016 M_w 7.8 Kaikōura earthquakes (Bradley et al., 2017b).

Building on Lee et al. (2020), this study extends from that regional Canterbury study to a nation-wide application in NZ. This article provides a comprehensive validation of the Graves and Pitarka (2010, 2015, 2016) hybrid BB ground-motion simulation method, with modifications informed by the previous validation, in NZ using small M_w earthquakes. New and updated input models for source, 3D crustal velocity, and near-surface 30 m time-averaged shear-wave velocity (V_{s30}) are also utilized. First, a description of the data and methods used for the simulations is provided, followed by an outline of the validation framework. Next, the results of the validation are presented, and finally the near-term directions of opportunity are discussed.

Earthquake sources and ground-motion recording stations considered

Earthquake sources

Earthquake source descriptions used in this study were obtained from the GeoNet centroid moment tensor (CMT) catalog (Ristau, 2008, 2013; https://github.com/GeoNet/data/tree/master/moment-tensor). While the catalog contains more than 2400 earthquakes, dating back to the beginning of 2003, attention is restricted to small M_w, between 3.5 and 5.0, shallow crustal events, based on having a centroid depth (CD) less than 20 km. Figure 1 presents the locations of earthquakes considered and ground-motion recording stations, highlighting their spatial distribution and surface projections of schematic ground-motion ray paths which provide a qualitative illustration of the spatial domain being robustly assessed in the simulation validation. Most earthquakes are clustered near the tectonic plate boundary, with a significant proportion of earthquakes having occurred due to the 2010–2011 Canterbury, 2013 Seddon, and 2016 Kaikōura earthquake sequences which were all located on the north-east region of the South Island. Earthquakes located far off-shore, such that they do not produce reliable ground motions at onshore ground-motion recording stations due to large distances, have been excluded. The minimum $M_w = 3.5$ was chosen to ensure there is adequate constraint on earthquake source parameters (i.e. related to the CMT solution) and sufficient signal-to-noise ratio (SNR) of the ground-motion records. The maximum $M_w = 5.0$ was chosen to ensure that the point-source modeling approximation used is generally valid for the majority of observed ground motions (which significantly simplifies the source modeling and consequent uncertainties) and that there is no appreciable off-fault near-surface nonlinear soil behavior to reduce uncertainties associated with modeling nonlinear site response. The range of CD considered (0–20 km) is expected to broadly cover earthquakes which are mostly located above the crustal seismogenic depth and therefore likely classified as active shallow crustal. A minimum requirement of three high-quality observed ground motions per event was also enforced (in an effort to limit the potential influence of statistical outliers) resulting in 479 earthquakes in the final dataset. This requirement is less strict than the five high-quality ground motions required in Lee et al. (2020) as the density of earthquakes surrounding many recording stations in NZ is much less than in the Canterbury region.
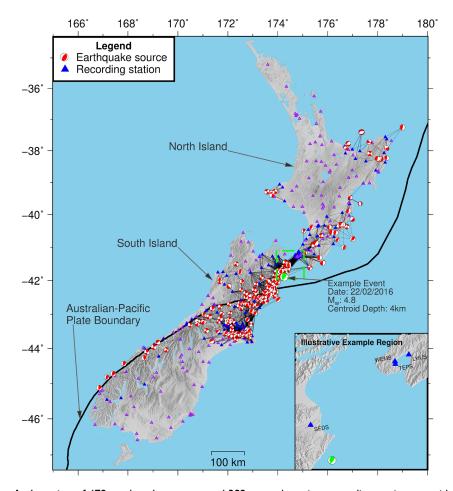
Figure 1. Location of 479 earthquake sources and 382 ground-motion recording stations considered across New Zealand. The 212 stations used in the validation are shown as blue-filled markers while the 170 stations not used due to insufficient records are shown as purple-unfilled markers. Schematic ray paths between the causal source and recording station of the final set of 5218 ground motions are also shown as black lines. The Australian-Pacific tectonic plate boundary is shown by the thick black line. The earthquake corresponding to the subsequent illustrative example simulation is highlighted, and the inset shows the station locations of the presented waveforms.

Figure 2 illustrates the M_w and source-to-site distance (R_{rup}) distributions of the earthquakes and corresponding recorded ground motions considered. Figure 2b shows that the M_w-R_{rup} distribution of the recordings has relatively widespread coverage in the M_w-R_{rup} space of typical interest. The histograms in Figure 2a and c illustrate the expected increase in the number of events with decreasing M_w down to $M_w = 4$ and the majority of ground-motion records have $R_{rup} \leqslant 80$km. The decrease in number of events below $M_w = 4$ occurs due to the inability to determine CMT solutions for many such events due to insufficient low-frequency energy (Ristau, 2018). No maximum R_{rup} is enforced; however, the largest R_{rup} considered in this study is within the range used for similar M_w in several empirical GMM studies, for example, over 200 km (Chiou et al., 2010). Thus, bias associated with instrument triggering (Boore et al., 1993) is not expected.
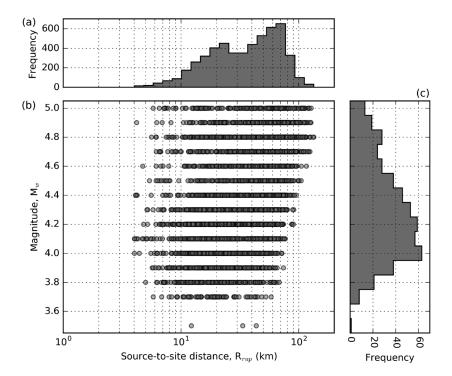
Figure 2. Earthquake source and ground-motion dataset distributions: (a) source-to-site distance histogram, (b) magnitude versus source-to-site distance scatter plot, and (c) magnitude histogram.

Figure 3a and b present histograms of the earthquake CD and the number of sites per earthquake (NS_e). The distribution of CD is relatively uniform with the exception of more shallow events at 4, 6, and 8 km (many of which come from the shallow 2010–2011 Canterbury and 2016 Kaikōura earthquake sequences). Figure 3c presents a plot of M_w against NS_e which illustrates that larger values of NS_e generally correspond to the larger M_w earthquakes. Supplemental Table A.1 in Electronic Supplement A provides further details of all the earthquakes events considered.

Ground-motion recording stations

Of the 382 ground-motion recording stations considered in this study (Figure 1), 212 stations recorded a sufficient quantity of high-quality ground motions from different earthquake events (at least three), while 170 stations did not (and hence were not used). Of these 212 stations, 23 were BB stations (commonly installed at rock sites) and 189 were strong-motion stations. Stations are relatively evenly distributed across most of the country, and are therefore located on highly variable site conditions, with the exception of larger densities in population centers, such as Canterbury and Wellington (whose locations are shown in Supplemental Figure B.1 in Electronic Supplement B). To quantify the site conditions, this study uses V_{s30}. Measured values of V_{s30} (through either invasive or non-invasive testing methods) are used, where available, at 32 stations. Where measured values are not available, the Foster et al. (2019) NZ-wide V_{s30} model is used, which predicts values based on surface geology and terrain categories conditioned on assumed prior
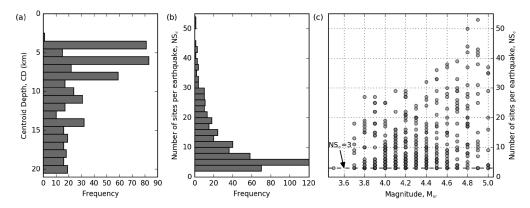
Figure 3. Earthquake source-related dataset distributions: (a) centroid depth histogram, (b) number of sites per earthquake histogram, and (c) magnitude versus number of sites per earthquake NS_e scatter plot. In panel (c), all markers are the same transparency so apparently darker symbols are multiple events with the same NS_e.
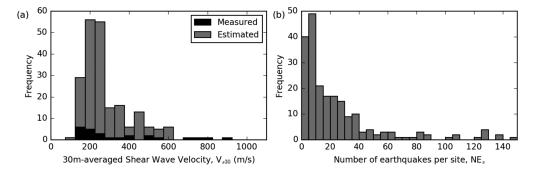

Figure 4. Ground-motion recording station distributions: (a) 30 m time-averaged shear-wave velocity (V_{s30}) directly measured or estimated from the Foster et al. (2019) NZ-wide V_{s30} model and (b) number of earthquake-induced ground motions per site (NE_s).

distributions and measured values. Figure 4a presents the distribution of V_{s30} values at the 212 sites, explicitly showing the distribution of measured and estimated V_{s30}, where there is notably few high values (i.e. $V_{s30} > 600$ m/s) despite many sites being nominally located on rock. Figure 4b presents the distribution of the number of earthquake-induced ground motions per site (NE_s), where most sites have recorded less than 40 earthquakes, although several have more than 100.

Observed ground-motion records

Volume 1 (unprocessed) ground-motion records were obtained from the GeoNet file transfer protocol (see "Data and Resources"). All ground motions were baseline-corrected, detrended, and processed with a fourth-order Butterworth filter with a low-pass frequency of 50 Hz (for sample rate of 200 Hz) or 20 Hz (for sample rate of 50 Hz), and a high-pass frequency of 0.08 Hz to reliably retain the Fourier amplitudes at $f \geqslant 0.1$ Hz (Ancheta et al.,

2014; Boore and Bommer, 2005). From the 479 earthquake events, 19,064 ground motions were recorded. However, the low amplitudes of many records (a consequence of small M_w earthquakes) lead to variable quality such that many are not usable for ground-motion simulation validation. Issues include low SNR, erroneous or late triggering, premature ending of records, and other instrument problems, which prevent the earthquake ground-motion signal from being adequately resolved. To quantify the quality of the ground motions and hence determine which ground motions are usable, the ground-motion quality classification neural network developed by Bellagamba et al. (2019) was used. The neural network provides a score on how well the ground motion resembles the good records which it was trained on through features, such as frequency-dependent SNR, Fourier amplitude ratios, and time-domain acceleration amplitude ratios. A quality score threshold of 0.5 was used such that a score below would be considered a low-quality record and a score above would be considered a high-quality record. This is the neural network's delineating threshold between high- and low-quality records based on the labeled training data. From the suite of 19,064 candidate ground motions, the final ground-motion dataset consists of 5218 ground-motion recordings (i.e. 27.4% of records).

GMM methods and inputs

Hybrid BB method

This study adopts the hybrid BB ground-motion simulation approach developed by Graves and Pitarka (2010, 2015, 2016). This approach computes the low-frequency (LF) and high-frequency (HF) ground-motion components separately using comprehensive and simplified physics, respectively. The comprehensive physics method uses a finite difference formulation to explicitly model 3D wave propagation and the simplified physics method (often referred to as a "stochastic" method) is based on a simpler theoretical representation of wave propagation. The two components are then merged in the time domain using a fourth-order Butterworth filter, with an LF–HF transition frequency of $f_t = 1$ Hz, to produce a single BB time series. A summary of the simulation method pertaining to this specific study is included in Electronic Supplement C, while comprehensive details can be found in Graves and Pitarka (2010, 2015, 2016). For HF simulation parameters, this study adopts values calibrated for the California region in lieu of specific evidence to the contrary but simulation validation studies can drive refinements of these choices in the future. Computational considerations and optimizations for this study are also included in Electronic Supplement C.

Method modifications

Two modifications are made to the simulation method based on prior validation by Lee et al. (2020). First, validation results illustrated that the significant durations of simulated ground motions were too short compared to observed records (which also influenced short-period pSA). Therefore, the HF path duration model, previously defined as $D_p = 0.07R_{rup}$ (GP10, Graves and Pitarka, 2010), was replaced with the Boore and Thompson (2014) (BT14) path duration model for active shallow crustal earthquakes. Boore and Thompson (2014) identified that conventional path duration models for simplified physics ground-motion simulation methods resulted in underprediction of durations and developed their model specifically to be used with these methods. Figure 5a presents a plot of the two path duration models, highlighting the increase across the range of R_{rup}
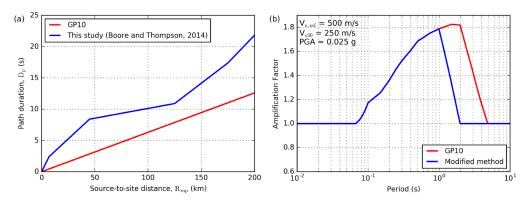
Figure 5. Simulation method modifications: (a) HF path duration models as a function of source-to-site distance and (b) empirical V_{s30}-based site amplification factor.

relevant to this study. The BT14 path duration model gives larger path durations at all R_{rup}, with the extent of this difference depending on the particular distance.

The second simulation method modification is motivated by Lee et al. (2020) identifying that the V_{s30}-based empirical site amplification factor model used (Campbell and Bozorgnia, 2014) was likely double-counting some long-period site effects, which were already explicitly modeled through the 3D velocity model (e.g. via sedimentary basin models) used in the LF finite difference simulation. Therefore, the empirical site amplification was modified to no longer be applied to the LF simulation component, but is still applied to the HF simulation component in the conventional manner. The impact of this modification depends on the quality of the crustal velocity model, which varies spatially, and the quality of V_{s30} values that would be used in the empirical site amplification factor. Figure 5b presents examples of the empirical site amplification functions, as applied in this study and also previous studies (e.g. Graves and Pitarka, 2010; Lee et al., 2020), for one set of input variables (V_{sref} = 500 m/s, V_{s30} = 250 m/s, and peak ground acceleration (PGA) = 0.025 g). Although the empirical site amplification is only applied to the HF simulation component (primarily corresponding to periods T ⩽ 1 s), the modified function used has reduced long-period amplification to reduce frequency roll-off effects when merging.

The subsequent validation presented in this article will highlight and quantify the improvements to prediction of ground-motion IMs made by these two modifications. For brevity herein, the Graves and Pitarka (2010, 2015, 2016) method as outlined in the respective papers, and used in Lee et al. (2020), will be referred to as the "Standard" simulation method and the method with the two modifications will be referred to as the "Modified" simulation method.

Crustal velocity and site response models

To simulate LF ground motions, the New Zealand Velocity Model (NZVM, Thomson et al., 2019, see "Data and Resources") is used to provide the P-wave and S-wave velocities, and density required (V_p, V_s, and ρ, respectively). This study uses a recent version of the NZVM (v2.02), with several improvements over the velocity model used in the related Canterbury validation study of Lee et al. (2020) (NZVM v1.66) and other previous
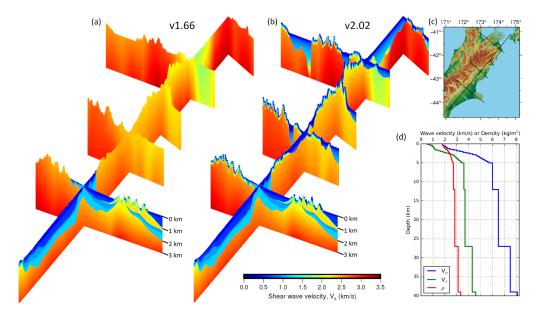
Figure 6. Crustal velocity models for ground-motion simulations. Fence diagram of six shear-wave velocity (V_s) cross-sections, extending to 3 km below mean sea level, through the 3D velocity model in the upper South Island region for LF simulations: (a) NZVM v1.66, (b) NZVM v2.02, (c) location of cross-section transects, and (d) 1D velocity model for HF simulations. The 1D velocity model has constant values below 39 km.

validation studies (Bradley et al., 2017b; Razafindrakoto et al., 2018). Figure 6a and b present fence diagrams of six cross-sections (whose transects are shown in Figure 6c) throughout the upper South Island of NZ for NZVM v1.66 and v2.02, respectively, highlighting their general features and recent modifications. The primary components of the velocity model are the embedded sedimentary basin models and a background travel-time tomography-based seismic velocity model (Eberhart-Phillips et al., 2010). NZVM v2.02 includes seven additional sedimentary basins. Specifically, from north to south: Wellington, Nelson, Marlborough, Kaikōura, Hanmer, North Canterbury, and Cheviot. The second major improvement is the inclusion of a geotechnical layer throughout the top 350 m of the model outside of the embedded sedimentary basins (to effectively represent weathered rock and shallow soil conditions). This follows the commonly used V_{s30}-dependent formulation of Ely et al. (2010) in combination with the Foster et al. (2019) V_{s30} model. In this study, the Standard simulation uses the NZVM v1.66 (consistent with previous validation by Lee et al. (2020)) while the Modified simulation uses the NZVM v2.02. For the HF component simulations, a generic 1D velocity model is used to account for wave propagation and amplification due to the crustal structure while the aforementioned empirical site amplification factor (Campbell and Bozorgnia, 2014) is used to account for shallow soil amplification. Figure 6d presents the V_p, V_s and ρ profiles which comprise the 1D velocity model. The reference site condition for the empirical site amplification factor is $V_s = 500$ m/s corresponding to the top layer of the 1D velocity model of 50 m thickness.

It is important to note that such nation-wide scale input models have large spatial variability in quality. For example, the crustal velocity model for seismic wave propagation may be better characterized in some regions (e.g. in the NZVM, the Canterbury

region has a high-resolution sedimentary basin model, whereas the rest of NZ has either simplistic basin models or the basin model is absent). Likewise for near-surface velocity models (e.g. V_{s30}), only a subset of ground-motion recording stations have "measured" values. However, it is usually not feasible to carry out field testing to obtain these measurements at every station, and therefore nation-wide models which use broader and more general sets of data, coupled with correlations, must be utilized. The effect of such aspects in the inputs will likely lead to variability in predictive performance of simulation methods. However, for forward prediction applications, with many locations of interest that are not explicitly characterized with measurements, the use of such models is a necessity.

Empirical GMMs

In addition to the comparison between observed and simulated ground motions, the performance of selected horizontal component empirical GMMs is also evaluated. The empirical prediction models considered in this study are the Bradley (2013) NZ-specific GMM for PGA, peak ground velocity (PGV), and pSA; Campbell and Bozorgnia (2012) for Arias intensity (AI), Campbell and Bozorgnia (2010) for cumulative absolute velocity (CAV); and Afshari and Stewart (2016) for 5%–75% and 5%–95% significant durations (D_{s575} and D_{s595}, respectively). The quantity of records that are both included in this validation study's dataset and also used in the development of the empirical models is small since relatively few NZ earthquakes were included in the relevant international databases and the Bradley (2013) model was developed before most earthquakes considered in this study. Hence, favorable predictive performance from the empirical models is unlikely to be due to overlapping data. Prediction using empirical models is intended to provide a relative performance benchmark for the simulations, and therefore exhaustive empirical predictions making use of a logic tree of alternate models were not considered.

Illustrative example simulation: 22 February 2016 M_w 4.8

As an example, simulation results from 22 February 2016 M_w 4.8 earthquake (GeoNet Public ID: 2016p140897) located at CD = 4 km in the Marlborough region (see Figure 1) are presented to provide insight on the salient attributes of the ground-motion simulations. The focal mechanism of the event was associated with a reverse fault (strike $\phi = 39$, dip $\delta = 62$, and rake $\lambda = 115$) and had $NS_e = 20$ high-quality recordings.

Figure 7 presents velocity waveforms at four stations of interest (locations shown in Figure 1) which highlight the simulation modifications. The black, red, and blue waveforms correspond to observed, Standard simulation, and Modified simulation ground motions, respectively. At the LHUS site ($R_{rup} = 90.8$km and $V_{s30} = 195$m/s), the Modified simulation waveform has longer duration of HF ground motion than the Standard simulation, and lower velocity amplitudes, as a result of the increased HF path duration model (Figure 5a), which are more comparable to the observed waveforms. For the WEMS ($R_{rup} = 81.5$km and $V_{s30} = 257$m/s) and TEPS ($R_{rup} = 80.0$km and $V_{s30} = 195$m/s) sites, the Standard simulation has LF velocity amplitudes which are too large as a result of the V_{s30}-based empirical site amplification (Figure 5b). The Modified simulation does not apply this empirical site amplification and therefore produces lower LF velocity amplitudes. The SEDS site is located in Marlborough, north-west of the causative fault ($R_{rup} = 28.8$km and $V_{s30} = 289$m/s). The Standard simulation does not have significant late-arriving LF amplitudes or coda waves, whereas the Modified simulation produces basin-generated waves,
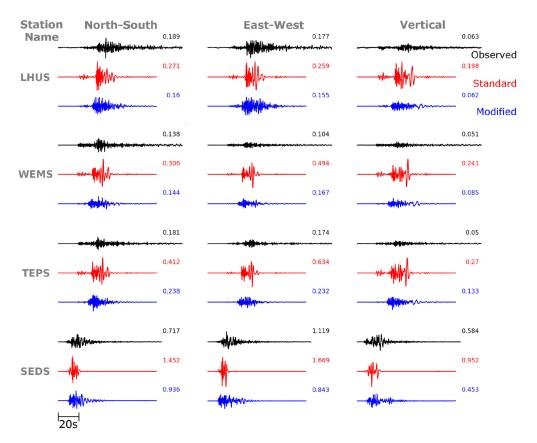
Figure 7. Comparison of observation (black), Standard simulation (red), and Modified simulation (blue) broadband velocity waveforms at four strong-motion stations of interest (LHUS, WEMS, TEPS in Wellington, and SEDS in Marlborough) for 22 February 2016 M_w 4.8 event. PGV values are provided to the right of each waveform in cm/s.

due to the inclusion of the Marlborough sedimentary basin in the NZVM, which are also present in the observed waveforms.

Figure 8 shows PGA, pSA(2.0 s), and D_{s595} as a function of R_{rup} for the 20 observed and simulated ground motions for this event. The median and $\pm$ 1 standard deviation of the relevant empirical prediction models are shown as the solid and dashed lines, respectively. To show the empirical model median as a single line against R_{rup}, a reference V_{s30} of 250 m/s is used, a representative value for Site Class D (New Zealand Standards, 2004), while subsequent site-specific prediction bias comparisons use the value of each individual station. All predictions provide a generally good comparison with observed PGA values, although the Modified simulation is marginally better, as measured using misfit residuals. For pSA(2.0 s), the Standard simulation significantly overpredicts for $R_{rup} = 70 - 100$km (in Wellington), likely due to the aforementioned double-counting of site amplification, while the Modified simulation and empirical model only slightly overpredict. The Standard simulation significantly underpredicts D_{s595} values at all distances while the Modified simulation provides better predictions due to the new HF path duration model.
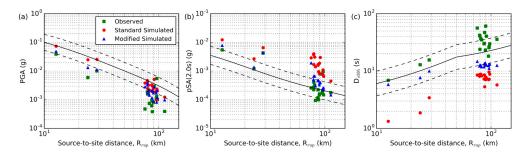
Figure 8. Observed, simulated, and empirically predicted horizontal geometric mean ground-motion intensity measures as a function of source-to-site distance, R_{rup}, for 22 February 2016 M_w 4.8 event: (a) PGA, (b) pSA(2.0 s), and (c) D_{s595}.
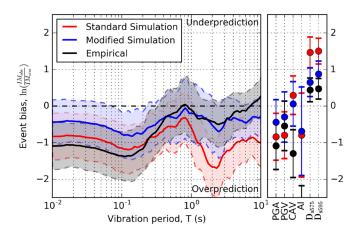

Figure 9. Systematic bias between observed and predicted IMs for all ground-motion recordings from 22 February 2016 M_w 4.8 earthquake. Solid lines and points indicate the systematic bias (mean residual), while the shaded regions and horizontal bars indicate $\pm$ 1 standard deviation.

Nonetheless, the empirical model provides a better prediction of significant duration than both simulations.

Figure 9 summarizes the bias and variability (as the mean and standard deviation, respectively) of the IM total residuals (more rigorously discussed in the next section) for this event. Overall, the biases for the Modified simulation are closer to zero than the Standard simulation for all IMs, where the latter is most overpredicted for 1.0–4.0 s pSA and significant durations. While this analysis is for only one event, and therefore may be subject to event-specific features, it illustrates the general simulation improvements.

Validation analysis method

Due to the large number of earthquake events considered, the primary focus of the analysis is the systematic effects across the entire dataset of ground motions. To achieve this, the total prediction residuals are partitioned into various components associated with ground-

motion variability using a partially crossed linear mixed-effects regression algorithm (Bates et al., 2015; Stafford, 2014). Following the notation of Al Atik et al. (2010), the general form of a GMM for an event, e, and site, s, pairing can be written as:

$$\ln \text{IM}_{es} = f_{es} + \Delta \tag{1}$$

where $\ln \text{IM}_{es}$ is the natural logarithm of the observed IM; f_{es} is the mean of the predicted logarithmic IM (given by either a single ground-motion simulation for each event or an empirical GMM) which is a function of the earthquake rupture, e, and site location, s; and Δ is the total residual. The total residual can be further decomposed into fixed and random effects:

$$\ln \text{IM}_{es} = f_{es} + a + \delta B_e + \delta W_{es} \tag{2}$$

where a is the (global) model prediction bias (fixed effect); δB_e is the between-event residual (random effect) with zero mean and variance τ^2; and δW_{es} is the within-event residual (random effect) with zero mean and variance ϕ^2. Comparison of Equations 1 and 2 illustrates that Δ has mean a and variance $\sigma^2 = \tau^2 + \phi^2$ assuming that δB_e and δW_{es} are independent random variables. δB_e represents the systematic misfit between observation and bias-corrected mean prediction for a given earthquake, e. Systematic location-to-location effects are not modeled, thus those effects will be included in the distribution of δB_e. δW_{es} represents the difference between observation and the bias-and-event-corrected mean prediction for a ground-motion record corresponding to earthquake e and site s. Finally, δW_{es} can be further partitioned into a systematic effect and a remaining residual:

$$\ln \text{IM}_{es} = f_{es} + a + \delta B_e + \delta S2S_s + \delta W_{es}^0 \tag{3}$$

where $\delta S2S_s$ is the systematic site-to-site residual, and δW_{es}^0 is the "remaining" within-event residual which represents factors which are not systematically accounted for by the δB_e or $\delta S2S_s$, or not accounted for in the models themselves. $\delta S2S_s$ is a zero-mean random effect with variance ϕ_{S2S}^2, and δW_{es}^0 has residual variance ϕ_{ss}^2 so that $\phi^2 = \phi_{S2S}^2 + \phi_{ss}^2$.

Results and interpretation

Model prediction bias (a)

Model prediction bias (a) and total standard deviation (σ) for various IMs predicted via the Standard and Modified simulations, and the relevant empirical GMMs are presented in Figure 10. The total standard deviations (and standard deviations of partitioned components shown in subsequent sections) correspond to the bias-corrected variances and thus reflect the variability around the global biases. Therefore, the total standard deviations presented should be considered in tandem with the global biases in the assessment of the precision of each GMM. The Standard simulation shows the same trends as previous validation carried out by Lee et al. (2020) for the Canterbury region, with overprediction of PGA, PGV, and pSA across all periods, and severe underprediction of significant durations. D_{s575} and D_{s595} have $a = 1.63$ (a factor of 5.1) and $a = 1.94$ (a factor of 7.0), respectively, which are beyond the axes of the plot but are placed at $a = 1.5$ for visual completeness. The Modified simulation has less overprediction than the Standard simulation for PGA, PGV, and pSA across all periods. At short periods, $T \leqslant 1.0$ s, this is a general result of the improved HF path duration model reducing HF acceleration amplitudes
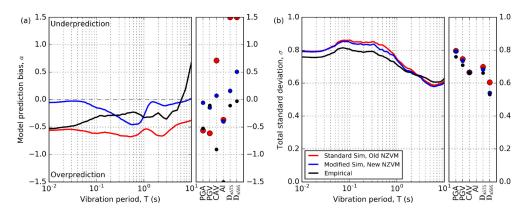
Figure 10. Simulated and empirical prediction of IMs for the entire dataset considered. (a) Systematic model prediction bias, *a*. (b) Total standard deviations, σ. For visual completeness, $a = 1.63$ and $a = 1.94$ for Standard simulation D_{s575} and D_{s595}, respectively, are plotted at $a = 1.5$ and $a = -2.24$ for empirical AI is plotted at $a = -1.5$.

(as discussed in Lee et al. (2020)), and at long periods, $T > 1.0$ s, due to the removal of the LF empirical V_{s30}-based site amplification. The decrease in HF acceleration amplitudes also contributes toward some bias reduction at long periods as HF acceleration amplitudes are generally larger than LF acceleration amplitudes for small M_w earthquakes and tend to also influence long-period pSA values slightly beyond the period corresponding to the LF–HF transition frequency, $T = 1/f_t = 1.0$s (Bora et al., 2016). The improvements to the NZVM result in minor increases to the moderate period pSA (not explicitly evident) which are secondary compared to the simulation method changes. The triangular-shaped feature in the Modified simulation pSA bias at short periods indicates increasing overprediction with period up to roughly $T = 0.8$ s. This overprediction has the same general shape as the empirical site amplification factor (for weak motions) which suggests they may be related. One possibility is that the empirical V_{s30}-based site amplification is too large due to estimated V_{s30} values that are too low. This is examined subsequently in the discussion of systematic site-to-site variability. The significant durations of the Modified simulation are drastically less biased than the Standard simulation, again due to the HF path duration model change, but still underpredicted. This underprediction may be due to the omission of features which cause increase ground-motion duration, such as small-scale heterogeneities in the 3D crustal velocity model which would cause wave scattering. In addition, the Boore and Thompson (2014) path duration model used in the HF simulations does not consider site conditions; hence, softer sites, which would normally have longer duration, are not properly accounted for. The Afshari and Stewart (2016) empirical model does explicitly account for site conditions and has practically no bias. The empirical prediction of PGA, PGV, and most short-period pSA has relatively small bias, while long-period pSA has more substantial underprediction, which is similar to the trends identified in other studies Bradley (2015) and Lee et al. (2020).

The σ of the Modified simulation compared to the σ of the Standard simulation shows a slight decrease across all IMs indicating that the Modified simulation has more precise predictions. Both simulation σ are slightly higher than empirical prediction σ at $T \leqslant 2.5$ s and slightly lower at $T > 2.5$ s. Several factors may cause these differences, a few of which are provided. At short periods, this may be partially due to variability from using only one
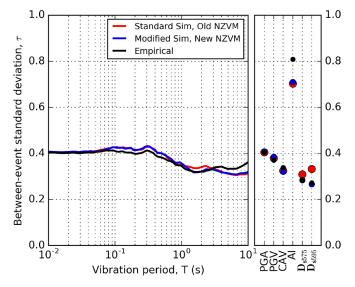
Figure 11. Between-event standard deviation, τ, for simulated and empirical predictions.

random white noise realization in the HF simulation for each record. At long periods, this may be due to better modeling of the physical processes with the comprehensive physics (e.g. kinematic rupture and wave propagation through a crustal velocity model) although this depends on the quality of models utilized. Compared to previous validation in the Canterbury region only (Lee et al., 2020), the σ for simulation and empirical predictions in this study are generally larger (previously pSA σ ranged 0.7–0.5 compared to 0.8–0.6 in this study) as a result of moving to an NZ-wide application where there is a broader range of source, path, and site conditions (emphasized by the change in the empirical σ), and more variable and lower average quality of inputs (particularly for site characterization and crustal velocity modeling).

Additional analysis of the bias associated with smoothed Fourier amplitude spectra (FAS) was carried out to identify the salient changes to the frequency content in the simulations and is included in Electronic Supplement D.

Between-event residual, δB_e

In the ground-motion simulations, the between-event residuals, δB_e, are generally associated with errors in source attributes or deviations from average source model scaling. The primary parameters in this regard are earthquake magnitude, location in the Earth's crust, faulting mechanism (Ristau, 2008), and HF stress parameter (Brune, 1970; Frankel, 2009; Graves and Pitarka, 2010) as point sources are used in this study. The empirical GMMs considered in this study represent the source effects through magnitude scaling, depth-to-rupture measures (e.g. Z_{TOR}), and faulting mechanism modifications.

Comparison for entire dataset. Figure 11 presents the between-event standard deviations, τ, for the simulated and empirical predictions. For PGA, PGV, and pSA at all periods, τ is similar for all prediction methods, with values between roughly 0.36 and 0.43, indicating similar between-event variability. The larger values of τ occur at shorter periods, while the
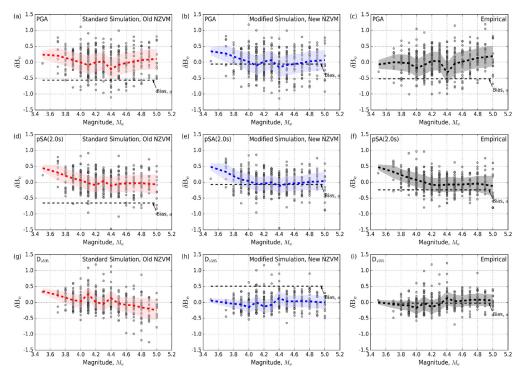
Figure 12. Comparison of between-event residual, δB_e against M_w for: (a) Standard simulation PGA, (b) Modified simulation PGA, (c) empirical prediction PGA, (d) Standard simulation pSA(2.0 s), (e) Modified simulation pSA(2.0 s), (f) empirical prediction pSA(2.0 s), (g) Standard simulation D_{s595}, (h) Modified simulation D_{s595}, and (i) empirical prediction D_{s595}. The locally weighted scatterplot smoothing regression trend lines are represented as the thick dashed lines with shaded regions corresponding to 16 -84th percentile ranges, while the associated model prediction biases, a, are represented as the thin dashed line. For Standard simulation D_{s595}, (g), the model prediction bias is a = 1.94 and is therefore outside of the axes shown.

smaller values occur at longer periods. This may be a result of the spectral ordinates at the longer periods being driven by the Fourier amplitudes below the source corner frequency, so variability in factors such as stress parameter is not evident. The Modified simulation and empirical model significant durations have the smallest τ, below 0.3 while the Standard simulation significant durations have slightly larger τ, at around 0.35.

Dependence on source parameters. The obtained δB_e values were compared against several source parameters, M_w, CD, and focal mechanism, to determine the causes of the variability and identify any parameter dependence within δB_e. However, interpretation of CD and focal mechanism dependence did not yield any significant trends, and hence are omitted here, but presented in Electronic Supplement E. Figure 12 provides comparisons between δB_e and M_w for PGA, pSA(2.0 s), and D_{s595} for the Standard simulation, Modified simulation, and empirical predictions. In these plots, the average trend is also indicated quantitatively via locally weighted linear regression. For PGA and pSA(2.0 s), predicted from both Standard and Modified simulation methods, there is no significant trend except some slight relative underprediction for $M_w < 4.0$. This is likely due to the selective CMT solution
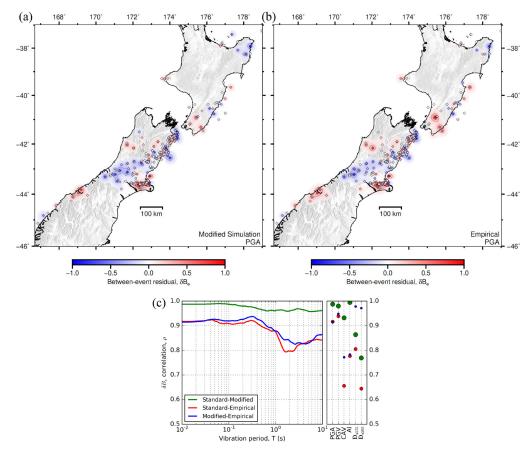
Figure 13. Spatial distribution of PGA between-event residual, δB_e, for (a) Modified simulation, (b) empirical prediction, for all 479 earthquake events across NZ and (c) a summary of the correlation, ρ, between prediction methods for all IMs.

generation for earthquakes $M_w < 4.0$, where only earthquakes with sufficient low-frequency energy are calculated, which would correspond to statistically stronger earthquake ground motions for a given M_w. The result on biases would be similar to instrument triggering biases, where statistically weaker ground motions with amplitudes near the instrument trigger threshold would be omitted. For empirical prediction, pSA(2.0 s) shows this same feature but it is not present in PGA. The D_{s595} δB_e for the Standard simulation has a slight negative linear trend which suggests better prediction at larger M_w when considering the model prediction bias ($a = 1.94$), which was also identified in Lee et al. (2020). This trend is expectedly absent in the Modified simulation, as a result of changing the HF path duration model, while the empirical prediction also has no trend.

Spatial dependence. Figure 13a and b present plots of the spatial variation of PGA δB_e for the Modified simulation and empirical prediction, respectively, showing the values at epicenter locations and a surface developed using geostatistical Kriging to illustrate spatial trends. Spatial plots of δB_e for other IMs are included in Electronic Supplement F. The spatial distributions are relatively similar between the Modified simulation and empirical

prediction, and differences are difficult to identify visually. A primary feature is the region of relative overprediction (blue) at the center of the South Island. For the simulations, it is speculated that this overprediction may be due to the constant stress parameter adopted, $\Delta\sigma = 5$ MPa, as the stress conditions may be weaker in this region. This trend would be mirrored in empirical prediction as there is no explicit consideration of stress parameter in the Bradley (2013), or other, empirical GMMs. Although depth-dependent stress parameter effects may be implicitly considered in Z_{TOR} (Boore et al., 2014), as all earthquakes considered are active shallow crustal, the variations due to stress parameter are more likely to be associated with regional differences rather than depth-dependence. Figure 13c summarizes the correlation of δB_e between the different prediction methods for all IMs. The correlation between the two simulation methods is expectedly higher than the correlations between each simulation and empirical prediction. The high correlations at short vibration periods further indicate source properties such as the stress parameter could be considered regionally variable and a discussion of this is presented subsequently.

Site-to-site and within-event residuals, $\delta S2S_s$ and δW_{es}

A broad range of site conditions and crustal properties exist in NZ at the ground-motion station locations. Within-event residuals, δW_{es}, are inherently associated with wave propagation path and near-surface site effects given the number of ground motions per event considered. For simulated ground motions, the path effect is dependent on the wave propagation through the specific velocity model (3D NZVM and 1D velocity model for LF and HF simulations, respectively). In this study, path effects are simply investigated with respect to R_{rup} and azimuth-specific path effects are not investigated. Site effects are also dependent on the velocity model; however, the spatial resolution of the velocity models used in this study is unable to accurately capture all near-surface site effects, which is the essential motivation for using empirical site amplification factors to adjust the simulations for reference site conditions. Empirical GMMs commonly treat path effects through various source-to-site distance parameters, and site effects through V_{s30} and depth-to-rock parameters.

Systematic site-to-site and within-event residuals for all stations. Figure 14a and b present the within-event standard deviations, ϕ_{S2S} and ϕ_{ss}, respectively, for the simulated and empirical predictions. For all IMs considered, ϕ_{S2S} are similar between simulation and empirical predictions, roughly 0.30–0.52, with the largest differences at long periods. Previously, in the Canterbury-specific study by Lee et al. (2020), there was a relative increase in ϕ_{S2S} for T = 1.0–4.0 s which is not present in any of the predictions in this study. This was found to be primarily due to a few sites which had poor estimates of V_{s30} (e.g. CSHS previously had a V_{s30} consistent with rock as it was located in a mountainous region but has since been identified to be on relatively soft soil). While the removal of empirical site amplification for the LF simulation component removes such inconsistencies, the revision of V_{s30} values still causes differences at short periods (as it is still applied to the HF simulation component). The empirical prediction has ϕ_{S2S} slightly lower than the simulations at short periods and slightly larger at long periods. The size of ϕ_{S2S}, relative to σ, indicates that significant improvements in ground-motion prediction are possible through being able to capture this systematic phenomena (Rodriguez-Marek et al., 2011) through further improvements in the simulation input models and method itself.
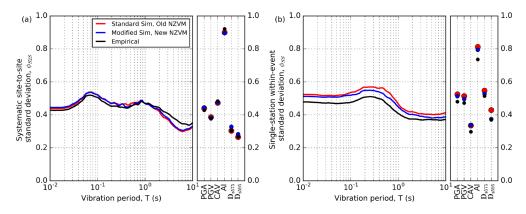
Figure 14. Within-event standard deviations. (a) Systematic site-to-site uncertainty, ϕ_{S2S}. (b) Single-station within-event variability, ϕ_{ss}.

The ϕ_{ss} from the Modified simulation is slightly smaller than the ϕ_{ss} for Standard simulation for all IMs considered suggesting that the adopted modifications capture slightly better the "remaining" physical phenomena, most likely path effects. This may be through changes resulting from the improved HF path duration model as there are record-specific influences due to the R_{rup} dependence. Empirical GMMs have lower ϕ_{ss} than both simulations, implying that they currently have less unexplained variability. However, as only a single ground-motion simulation is run for each event, inherent stochastic features such as the HF random phase spectrum can also contribute to ϕ_{ss}, particularly for short-period pSA. Running several simulations for each event and considering the mean of all realizations for each event could provide a more consistent comparison between empirical and simulated ϕ_{ss}. This would reduce the aleatory variability from considering only one random seed of white noise in the HF simulation for each record and thus reduce simulated ϕ_{ss} as the white noise would be averaged over many realizations. It is also important to note that τ (Figure 11), ϕ_{S2S}, and ϕ_{ss} are all of similar size implying that the variability in δB_e, $\delta S2S_s$, and δW_{es}^0 is similar. While the simulation method modifications have source-, site- and record-specific influences, their effects are still relatively general, which has led to only small reductions in standard deviations. It is expected that further nonergodic modifications (e.g. event-specific stress parameter, explicit physics-based site response analysis for modeling near-surface site effects, among others) have the potential to more significantly reduce simulation standard deviations.

Dependence on site and path parameters. To evaluate the variability and biases resulting from site effects, $\delta S2S_s$ are compared against V_{s30}. A comparison between δW_{es}^0 and R_{rup} did not identify any conclusive trends and is therefore omitted from the main text but included in Electronic Supplement E.

Figure 15 presents the comparisons of $\delta S2S_s$ with V_{s30} for PGA, pSA(2.0 s), and D_{s595} for the Standard and Modified simulations, and empirical prediction. The comparison for PGA shows no significant trend while the comparison for pSA(2.0 s) shows a slight negative trend at low V_{s30} for the Modified simulation which is likely a result of removing empirical LF site amplification (i.e. lower V_{s30} is relatively underpredicted). For both simulations and empirical prediction, the comparison for D_{s595} shows a strong negative
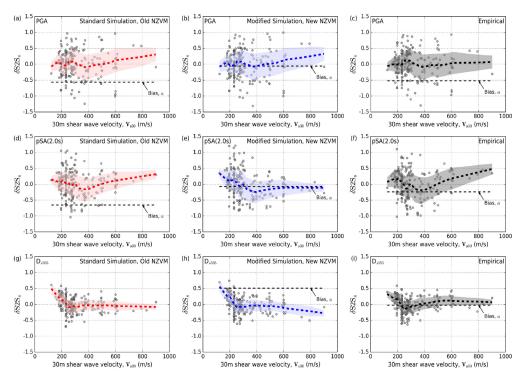
Figure 15. Comparison of systematic site-to-site residual, $\delta S2S_s$, against V_{s30} for: (a) Standard simulation PGA, (b) Modified simulation PGA, (c) empirical prediction PGA, (d) Standard simulation pSA(2.0 s), (e) Modified simulation pSA(2.0 s), (f) empirical prediction pSA(2.0 s), (g) Standard simulation D_{s595}, (h) Modified simulation D_{s595}, and (i) empirical prediction D_{s595}. The locally weighted scatterplot smoothing regression trend lines are represented as the thick dashed lines with shaded regions corresponding to 16-84th percentile ranges, while the associated model prediction biases, a, are represented as the thin dashed line. For Standard simulation D_{s595}, (g), the model prediction bias is a = 1.94 and is therefore outside of the axes shown.

trend at low V_{s30}. This was also identified in Lee et al. (2020) with the Standard simulation and remains in the Modified simulation as no direct modification has been made on this aspect. This is attributed to the fact that the acceleration amplitudes have been modified by the period-dependent V_{s30}-based site amplification while the duration of the motion (reflected by D_{s595}) has had no modification (i.e. the site amplification is applied to the FAS amplitudes) despite it being well understood that softer sites result in longer durations of motion (e.g. Afshari and Stewart, 2016). In addition, the Boore and Thompson (2014) path duration model used in the Modified simulation is independent of site conditions and therefore does not address this issue either. Overall, the lack of site effect duration contributions in the HF simulations leads to its underestimation, and this is more significant for softer sites at shorter distances, where the "site contribution" to duration is relatively larger.

Spatial dependence. Figure 16a and b illustrate the spatial variation of $\delta S2S_s$ for pSA(2.0 s) for the Modified Simulation and empirical prediction, respectively, showing the values at station locations and a surface developed using Kriging. Spatial plots of $\delta S2S_s$ for other
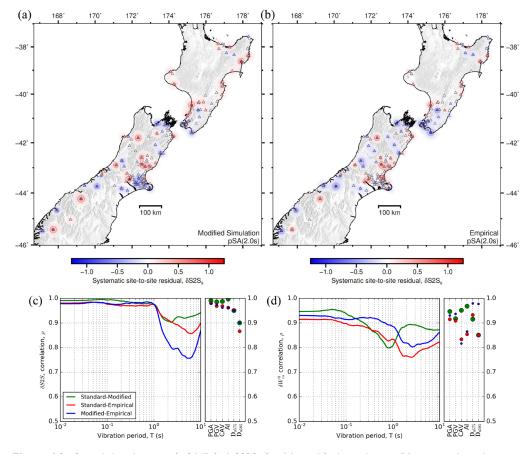
Figure 16. Spatial distribution of pSA(2.0 s) δS2S$_s$ for (a) modified simulation, (b) empirical prediction, for the 212 stations across NZ, (c) a summary of δS2S$_s$ correlation between prediction methods for all IMs and (d) a summary of δW$_{es}^0$ correlation between prediction methods for all IMs.

IMs are included in Electronic Supplement F. The spatial distributions between the two methods are relatively similar but small localized differences are noticeable. Figure 16c summarizes the correlation of δS2S$_s$ between the different prediction methods for all IMs. The correlation between all methods at short periods is very high as the site effects are accounted for by V_{s30}-based amplification in all methods. At longer periods, the correlations reduce as a result of different methodological treatment of site effects. The largest difference, between Modified simulation and empirical predictions, occurs because the Modified simulation does not have any direct dependence on V_{s30} since no site amplification is applied (although V_{s30} is used in the 3D crustal velocity model's formulation of the geotechnical layer which is applied outside of explicitly modeled sedimentary basins). Figure 16d summarizes the correlation of δW$_{es}^0$ between the different prediction methods for all IMs. Although the correlation between methods is different, the Standard–Empirical and Modified–Empirical correlations have similar shape. The size and spatial distribution of δS2S$_s$ values and similarities in treatment of site effects among all predictions (indicated by the high correlations in Figure 16c and d) suggest that explicit site response is likely needed to better model site-specific effects, although several

improvements can also be made to the HF simulations to better consider site-specific characteristics (e.g. site-specific 1D velocity models and the near-surface diminution factor κ_0).

Station subcategory analysis. In addition to identifying significant trends through spatial plots, grouping stations with similar site characteristics into subcategories and analyzing them collectively can provide insight into their systematic trends. Subcategories considered are (1) stiff alluvial gravel sites which have estimated V_{s30} that are considered too low considering the age of the deposits, or distance from any coastline or water feature (39 sites); (2) stiff rock sites which have estimated V_{s30} that are considered too low (34 sites); (3) sites with shallow soil overlying stiff rock (10 sites); and (4) sites which show strong basin amplification in observed ground motions, but their sedimentary basin is not explicitly modeled in the NZVM (11 sites). Each of the 212 stations was manually classified considering publicly available data on the site characteristics (e.g. previous V_{s30} estimates, depth-to-rock estimates, site period (Kaiser et al., 2017), Google satellite imagery, and surface geology and topography maps). Supplemental Figure B.1 in Electronic Supplement B presents a plot of the station locations corresponding to the subcategories. This analysis is only presented here for the Modified simulation as it has been assessed to be the simulation method that provides the better prediction of ground-motion IMs. Equivalent analyses for the Standard simulation and empirical prediction are included in Electronic Supplement G, as well a comparison between prediction methods.

The results for all the subcategories are shown in Figure 17. To analyze the subcategories, the sum of model prediction bias and systematic site-to-site residual (i.e. $a + \delta S2S_s$) for each station in that subcategory are plotted (gray), along with the average of those stations (black, termed the "subset bias"), and the global model prediction bias for all sites (a) as a benchmark for comparison (blue). Shaded regions and error bars indicate the 90% confidence interval of the respective mean bias estimates.

Figure 17a and b present the subcategory analyses for the stiff gravel, low V_{s30} sites; and stiff rock, low V_{s30} sites, respectively. The subset bias for both subcategories is significantly more negative than the global bias at all pSA periods, with largest difference at the periods corresponding to greatest HF empirical site amplification (see Figure 5b). This indicates that the sites in this subcategory are disproportionately overpredicted, and this is possibly a primary contributor to the global overprediction at short periods. The low inferred V_{s30} values would lead to overamplification from the empirical site amplification factor, which contributes to the triangular-shaped feature in the global model prediction bias. The Foster et al. (2019) V_{s30} model was found to predict estimates that are considered too low for these sites, most likely because they reflect conditions not well represented in the limited data used in the Foster et al. (2019) model. Specifically, some alluvial soil classifications are dominated by the large amount of lower V_{s30} data values from the Christchurch central business district, which is adjacent to the eastern coastline of the South Island, and the lack of explicit age considerations and the broad deposition classifications. This issue was identified at several sites located on the Canterbury alluvial plains (e.g. GDLC, CACS, RKAC, ROLC, SWNC, TPLC, and DFHS sites) which have measured V_{s30} values between roughly 400 and 550 m/s, while surrounding stations which used the Foster et al. (2019) model were prescribed lower V_{s30} ranging from approximately 200–350 m/s. Rock sites also often had low inferred V_{s30} values. This was partially caused by the use of the Foster et al. (2019) model in its published raster-based form which does not accurately represent boundaries of surficial geology units. Therefore interpolation for V_{s30} values between raster grid points may be incorrect if a rock–soil boundary exists between the grid points. While
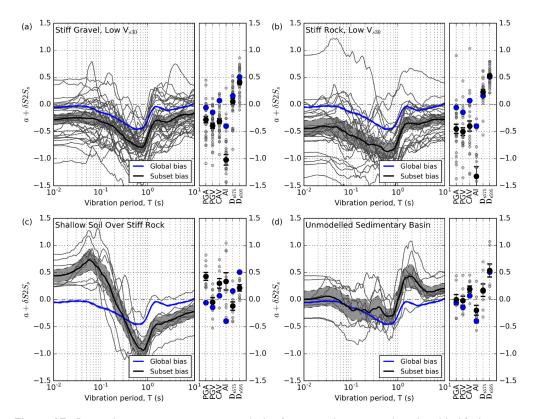
Figure 17. Bias and systematic site-to-site residuals of station subcategories based on Modified simulation prediction (a) stiff gravel sites with V_{s30} that is too low, (b) stiff rock sites with V_{s30} that is too low, (c) shallow soil overlying stiff rock sites, and (d) sites located in a sedimentary basin that is not modeled in the NZVM. Blue lines show the global model prediction bias, a, for the Modified simulation while the gray lines show $a + \delta S2S_s$ for each station in the subcategory, and the black line is the subset average (of the gray lines). Blue- and gray-shaded regions reflect the 90% confidence intervals of the a and $a + \delta S2S_s$ estimates, respectively.

direct querying of the code associated with the Foster et al. (2019) model would generally alleviate this issue, the surficial geology unit boundaries (modeled as polygon objects) are still often simplified, not drawn accurately, or sometimes not modeled, which can lead to errors in assigned values too.

Figure 17c presents the subcategory analysis for the shallow soil overlying stiff rock sites. The short-period pSA of this subcategory is underpredicted compared to the global bias. This is inferred to be a result of the HF simulation using the quarter wavelength (QWL) method with a generic 1D velocity model which does not capture the resonance effects of such shallow impedance contrasts. An alternative, which was previously alluded to, would be to carry out explicit physics-based site response analysis with a site-specific velocity profile in a post hoc manner (e.g. De la Torre et al., 2020; Hartzell et al., 2002; Jeong and Bradley, 2017a, 2017b; Roten et al., 2012).

Figure 17d presents the subcategory analysis for the sites located in a sedimentary basin that is not explicitly modeled in the NZVM v2.02 and exhibits strong basin amplification.
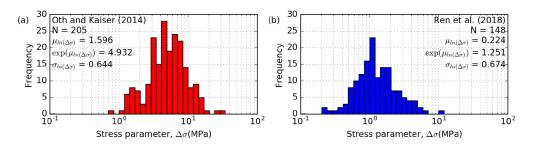
Figure 18. Distribution of inferred stress parameter of all events considered from spectral inversion studies. (a) Oth and Kaiser (2014). (b) Ren et al. (2018).

The largest difference between this subset bias and the global bias occurs at moderate periods, between T = 1.0–3.0 s, which is expected to be the range of periods corresponding to the basin responses at the sites. Over this period range, the subset bias is most underpredicted as the basin effects would not be produced in the simulations without an explicit sedimentary basin model. Therefore, inclusion of more embedded sedimentary basin models in the NZVM would improve the prediction at these sites.

The subset of 32 measured sites was also investigated and identified to have a subset bias that is systematically 0.1–0.2 natural log units higher than the global bias at T ⩽ 1.0 s. This decreases the overprediction bias between T = 0.1–1.0 s but increases the bias toward underprediction below T = 0.1 s. An updated site database with higher quality and quantity of measured V_{s30} values would lead to more informed inferences regarding the improvement which can be achieved with measured V_{s30}. However, the results presented in this study are informative as regional models of V_{s30} are necessary for forward prediction in regional seismic hazard analyses.

Discussion

Throughout this article, several further pathways for improved simulation prediction were identified. In this section, further discussion is presented on the underlying theoretical concepts, their potential implementation, and expected outcomes. In addition, ideas which were not explicitly highlighted in the prior analysis are also discussed.

Spatially varying stress parameter

Analysis of the spatial variation of δB_e across NZ for short-period pSA provided evidence toward a spatially varying stress parameter for the HF simulations. While a spatial model of stress parameter in NZ does not exist, comparisons of δB_e from the simulations against inferred stress parameters from spectral inversion studies provide a basis for a preliminary study to quantify the potential predictive improvement. Two independent stress parameter studies have been recently carried out in NZ: Oth and Kaiser (2014) for the 2010–2011 Canterbury earthquake sequence and Ren et al. (2018) for the 2016 Kaikōura earthquake sequence. As the studies make different modeling assumptions (such as source spectra model and reference site conditions), the estimated stress parameter values cannot be directly combined (Atkinson and Beresnev, 1997) and are therefore compared to the simulations separately.
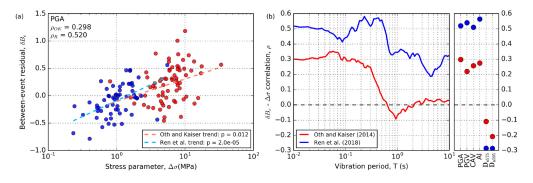
Figure 19. Correlation statistics of between-event residual and inferred stress parameter from Oth and Kaiser (2014) and Ren et al. (2018) for (a) PGA as an example and (b) summary for all IMs.

Of the 479 earthquake events simulated in this study, 158 were also considered in Oth and Kaiser (2014) and 79 in Ren et al. (2018). Figure 18 presents histograms of the inferred stress parameters of the two studies and corresponding summary statistics. Comparison between the Modified simulation δB_e in this study and the inferred stress parameters from these other two studies shows positive correlations at short-period pSA. Figure 19a presents this comparison for PGA where the positive correlation can be seen, while Figure 19b provides the summary of correlation coefficients for each study across all periods. Overall, the short-period correlation coefficient is roughly $\rho_{OK} = 0.3$ for Oth and Kaiser (2014) and $\rho_R = 0.5$ for Ren et al. (2018). The decrease in correlation between 0.3 and 0.9 s is due to the stress parameter dependence of Fourier amplitudes decreasing below the source corner frequency and the effect of the LF–HF simulation transition frequency near 1 Hz. While some studies have equated ground-motion between-event variability with stress parameter variation in a random vibration theory construct (i.e. Cotton et al., 2013), the result shown here with $\rho < 1$ illustrates the nature of this idealization.

Preliminary simulations of the same events in this study were run with revised stress parameters considering its correlation with δB_e (i.e. Figure 19) and it was found, as expected, to significantly reduce between-event variability. However, the direct use of δB_e for calculation of stress parameter is not possible for forward prediction of future earthquake ground motions. One possibility would be to develop a spatially varying coefficient model of stress parameter across NZ based on the geostatistical Kriging of δB_e presented here which can be used in ground-motion simulations for active shallow crustal earthquakes (to provide mean and standard deviations of stress parameter).

HF duration modification for soft soil sites

Systematic underprediction of significant durations at low V_{s30} sites was highlighted (from comparison of $\delta S2S_s$ and V_{s30}) and assumed to be due to the lack of site consideration in the HF simulation duration, which currently consists of only source and path components. The source component is simply the inverse of the FAS corner frequency and the path component depends only on R_{rup} for both GP10 and BT14 models used in the Standard and Modified simulations, respectively. The BT14 model was developed using the NGA-West2 database and therefore corresponds to a duration that is representative of the database's "average" site (since there is no site dependence). However, the percentage of

records in the NGA-West2 database from sites with $V_{s30} \leqslant 250m/s$ is relatively small (approximately 7.5%, especially compared to 34.8% of such sites in this study) and therefore the BT14 model gives path durations which are more representative of a stiff soil or rock site. The residual analysis of Boore and Thompson (2014) also identified this relative underprediction at low V_{s30} (and some weak M_w and Z_{TOR} dependence) but those authors decided that the advantages of small adjustments to the model to account for such biases did not outweigh the simplicity of the model for applications. The range of V_{s30} for which there were no biases corresponded well to the reference output velocities commonly used in "stochastic" simulations (i.e. $V_{s30} \geqslant 500m/s$) and their recommendations were to use the reference motions as inputs into local site response modeling for sites with low V_{s30} values, which echo the sentiments of this study. In addition, it is also plausible to consider a V_{s30}-based empirical modification factor to the HF simulation duration, analogous to the empirical site amplification for amplitudes.

V_{s30} model sensitivity

The V_{s30} used for empirical site amplification was shown to be a significant contributor to model prediction bias and uncertainty. While this study used a combination of measured values and estimated values from the Foster et al. (2019) model, it is acknowledged that there are other V_{s30} models available which can provide values at many ground-motion recording stations, such as Kaiser et al. (2017). However, Kaiser et al. (2017) is primarily based on geologic categories, similar to the Foster et al. (2019) model, and high-quality estimates of V_{s30} from that model are also included in the Foster et al. (2019) model. Preliminary simulations using V_{s30} values from Kaiser et al. (2017) resulted in minor differences in model prediction bias and uncertainty compared to simulations using the Foster et al. (2019) model. One subcategory where the Kaiser et al. (2017)-based results perform better is with stiff rock sites (particularly BB stations) where high V_{s30} values are prescribed, whereas Foster et al. (2019) generally predict values that are considered as too low. The benefit of Foster et al. (2019) is its ability to prescribe V_{s30} at any arbitrary location in NZ, which is important for many seismic hazard analysis applications. To improve such models, subsequent research toward the V_{s30} estimation method and increasing the number of V_{s30} measurements are clearly necessary.

Examination of topographic effects

As the LF and HF simulations both do not consider explicit topographic representation in the wave propagation or near-surface site response, biases associated with topographic effects may exist. This possibility is examined through comparisons of $\delta S2S_s$ and relative elevation (the difference between the elevation at the site and the average surrounding elevation), a simple proxy for topographic character (Rai et al., 2016). Therefore, 250 m and 1250 m relative elevation (H_{250} and H_{1250}, respectively, where the distance is the diameter over which elevation is averaged) were calculated from digital elevation model rasters with a 25 m grid resolution. Comparisons of $\delta S2S_s$ against H_{250} and H_{1250} for PGA, pSA(2.0 s), and D_{s595} for the Standard and Modified simulations, and empirical prediction are presented in Electronic Supplement E. For PGA and D_{s595}, there does not appear to be any trend with either relative elevation metrics. For empirically predicted pSA(2.0s), there is a positive trend in $\delta S2S_s$ across the H_{1250} range of the sites considered. For both simulations, there are systematically negative $\delta S2S_s$ at negative H_{250} and H_{1250} which may reflect the absence of topographic deamplification effects in the simulations. However, it is also noted

that there is significant scatter in the $\delta S2S_s$, which may result from biases related to other site effects, and most of the data (74.1%) lies in the range of $-15 < H_{1250} < 15$ (and roughly equivalent ranges for H_{250}) which generally corresponds to no evident topographic effects. Therefore, while some biases may exist due to topographic effects, more rigorous analyses are required to make stronger inferences. Topographic effects may be considered in simulations either explicitly (through appropriate meshing in finite element methods), through additional 2D or 3D site response analyses which model the topographic features, or using empirical topographic amplification factors.

Conclusion

This article has provided a comprehensive validation of the Graves and Pitarka (2010, 2015, 2016). Hybrid BB ground-motion simulation method in a NZ context with small M_w active shallow crustal point source ruptures using an extensive set of 5218 ground motions recorded at 212 sites from 479 earthquakes. The simulations used recent models of crustal velocity and near-surface site characteristics, and a Modified simulation method. Ground-motion predictions using the Standard and Modified simulations, and commonly used empirical predictions for benchmarking purposes, were compared against observed ground motions to quantify their predictive capability and identify biases and potential improvements.

Through the validation considering the entire dataset, it was found that the modifications in the Modified simulations reduced the overprediction bias of pSA across all periods and drastically reduced the underprediction bias of significant durations. The spatial distribution and high correlation of δB_e between methods indicated that modeling of the spatial variation of source parameters, such as the stress parameter, could improve the prediction. Likewise for $\delta S2S_s$, the spatial distribution and high correlation indicated that changes to the modeling of near-surface site response, such as explicit physics-based wave propagation methods, could improve the prediction. This advanced approach to site response would also improve the underprediction of significant durations at soft soil sites (relative to stiffer sites, based on V_{s30}), where site contributions to ground-motion duration are significant, which was identified in comparisons between $\delta S2S_s$ and V_{s30}. Finally, a station subcategory analysis identified specific types of sites which appeared to be systematically biased, such as inland stiff gravel sites and stiff rock sites, which have estimated V_{s30} values from Foster et al. (2019) that are expected to be too low, resulting in overamplification from the V_{s30}-based empirical site amplification. Improvements to the Foster et al. (2019) model to better characterize such sites would lead to improved prediction.

To extend the findings of this study, future work should consider moderate and large M_w earthquakes, which would require additional attention to source modeling as the point source approximation would no longer be valid, and subduction earthquakes where the fundamental source rupture and wave propagation processes may differ from active shallow crustal earthquakes. Additional improvements can be made to crustal velocity modeling to include explicit modeling of more sedimentary basins in the NZVM with the aim of complete NZ-wide coverage with simple topographically inferred models, followed by more refined models incorporating available subsurface data. Inclusion of small-scale crustal velocity heterogeneities will also be important for appropriately simulating high frequencies deterministically and would increase simulated ground-motion durations through wave scattering. Finally, the uncertainty in simulations needs to be explicitly quantified for use in probabilistic seismic hazard analyses.

ORCID iDs

Brendon A Bradley (iD) https://orcid.org/0000-0002-4450-314X
Peter J Stafford (iD) https://orcid.org/0000-0003-0988-8934
Adrian Rodriguez-Marek (iD) https://orcid.org/0000-0002-8384-4721

Data and resources

Earthquake source descriptions used in this study were obtained from the GeoNet New Zealand earthquake centroid moment tensor catalog (https://github.com/GeoNet/data/tree/master/moment-tensor), recorded ground motions were obtained from the GeoNet file transfer protocol (ftp:// ftp.geonet.org.nz/strong/), the 3D crustal velocity models were created using the NZVM code (https://github.com/ucgmsim/Velocity-Model), estimated V_{s30} were obtained from https://github. com/fostergeotech/Vs30_NZ, and measured V_{s30} values were provided by Dr Liam Wotherspoon. The ground-motion simulations were computed on NeSI high-performance computing resources using the workflow developed by the QuakeCoRE Technology Platform 4 (https://github.com/ ucgmsim/slurm_gm_workflow). Linear mixed-effects regression was carried out using the lme4 package on RStudio. Figures were prepared using Generic Mapping Tools (https://www.generic-mapping-tools.org/), Python (https://www.python.org/), and Matplotlib (https://matplotlib.org/).

Supplemental material

Supplemental material for this article is available online.

References

Afshari K and Stewart JP (2016) Physically parameterized prediction equations for significant duration in active crustal regions. *Earthquake Spectra* 32(4): 2057–2081.
Al Atik L, Abrahamson N, Bommer JJ, Scherbaum F, Cotton F and Kuehn N (2010) The variability of ground-motion prediction models and its components. *Seismological Research Letters* 81(5): 794–801.

Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BSJ, Wooddell KE, Graves RW, Kottke AR, Boore DM, Kishida T and Donahue JL (2014) NGA-West2 database. *Earthquake Spectra* 30(3): 989–1005.

Atkinson GM and Assatourians K (2015) Implementation and validation of EXSIM (A stochastic finite-fault ground-motion simulation algorithm) on the SCEC Broadband Platform. *Seismological Research Letters* 86(1): 48–60.

Atkinson GM and Beresnev I (1997) Don't call it stress drop. *Seismological Research Letters* 68(1): 3–4.

Bates DM, Mächler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1–48.

Bellagamba X, Lee R and Bradley BA (2019) A neural network for automated quality screening of ground motion records from small magnitude earthquakes. *Earthquake Spectra* 35(4): 1637–1661.

Bijelić N, Lin T and Deierlein GG (2018) Validation of the SCEC Broadband Platform simulations for tall building risk assessments considering spectral shape and duration of the ground motion. *Earthquake Engineering & Structural Dynamics* 47(11): 2233–2251.

Boore DM and Bommer JJ (2005) Processing of strong-motion accelerograms: Needs, options and consequences. *Soil Dynamics and Earthquake Engineering* 25(2): 93–115.

Boore DM and Thompson EM (2014) Path durations for use in the stochastic-method simulation of ground motions. *Bulletin of the Seismological Society of America* 104(5): 2541–2552.

Boore DM, Joyner WB and Fumal TE (1993) Estimation of response spectra and peak accelerations from western North American earthquakes: An interim report. Technical report, US Geological Survey, Pasadena, CA. https://doi.org/10.3133/ofr93509

Boore DM, Stewart JP, Seyhan E and Atkinson GM (2014) NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra* 30(3): 1057–1085.

Bora SS, Scherbaum F, Kuehn N and Stafford P (2016) On the relationship between Fourier and response spectra: Implications for the adjustment of empirical ground-motion prediction equations (GMPEs). *Bulletin of the Seismological Society of America* 106(3): 1235–1253.

Bradley BA (2013) A New Zealand-specific pseudospectral acceleration ground-motion prediction equation for active shallow crustal earthquakes based on foreign models. *Bulletin of the Seismological Society of America* 103(3): 1801–1822.

Bradley BA (2015) Systematic ground motion observations in the Canterbury earthquakes and region-specific non-ergodic empirical ground motion modeling. *Earthquake Spectra* 31(3): 1735–1761.

Bradley BA, Pettinga D, Baker JW and Fraser J (2017a) Guidance on the utilization of earthquake-induced ground motion simulations in engineering practice. *Earthquake Spectra* 33(3): 809–835.

Bradley BA, Razafindrakoto HN and Polak V (2017b) Ground-motion observations from the 14 November 2016 Mw 7.8 Kaikōura, New Zealand, earthquake and insights from broadband simulations. *Seismological Research Letters* 88(3): 740–756.

Brune JN (1970) Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research* 75(26): 4997–5009.

Campbell KW and Bozorgnia Y (2010) A ground motion prediction equation for the horizontal component of cumulative absolute velocity (CAV) based on the PEER-NGA strong motion database. *Earthquake Spectra* 26(3): 635–650.

Campbell KW and Bozorgnia Y (2012) A comparison of ground motion prediction equations for Arias intensity and cumulative absolute velocity developed using a consistent database and functional form. *Earthquake Spectra* 28(3): 931–941.

Campbell KW and Bozorgnia Y (2014) NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra* 30(3): 1087–1115.

Chiou BSJ, Youngs R, Abrahamson N and Addo K (2010) Ground-motion attenuation model for small-to-moderate shallow crustal earthquakes in California and its implications on regionalization of ground-motion prediction models. *Earthquake Spectra* 26(4): 907–926.

Cotton F, Archuleta R and Causse M (2013) What is sigma of the stress drop? *Seismological Research Letters* 84(1): 42–48.

Crempien JG and Archuleta RJ (2015) UCSB method for simulation of broadband ground motion from kinematic earthquake sources. *Seismological Research Letters* 86(1): 61–67.

De la Torre CA, Bradley BA and Lee RL (2020) Modeling nonlinear site effects in physics-based ground motion simulations of the 2010–2011 Canterbury earthquake sequence. *Earthquake Spectra* 36: 856–879.

Dreger DS and Jordan TH (2015) Introduction to the focus section on validation of the SCEC broadband platform V14.3 simulation methods. *Seismological Research Letters* 86(1): 15–16.

Eberhart-Phillips D, Reyners M, Bannister S, Chadwick M and Ellis S (2010) Establishing a versatile 3-D seismic velocity model for New Zealand. *Seismological Research Letters* 81(6): 992–1000.

Ely GP, Jordan T, Small P and Maechling PJ (2010) A V_{s30}-derived near-surface seismic velocity model. In: *Proceedings of the abstract S51A-1907 at 2010 fall meeting, AGU*, San Francisco, CA, 13–17 December.

Foster KM, Bradley BA, McGann CR and Wotherspoon LM (2019) A V_{s30} map for New Zealand based on geologic and terrain proxy variables and field measurements. *Earthquake Spectra* 35: 1865–1897.

Frankel A (2009) A constant stress-drop model for producing broadband synthetic seismograms: Comparison with the next generation attenuation relations. *Bulletin of the Seismological Society of America* 99(2A): 664–680.

Galasso C, Zhong P, Zareian F, Iervolino I and Graves RW (2013) Validation of ground-motion simulations for historical events using MDoF systems. *Earthquake Engineering & Structural Dynamics* 42(9): 1395–1412.

Goulet CA, Abrahamson NA, Somerville PG and Wooddell KE (2015) The SCEC broadband platform validation exercise: Methodology for code validation in the context of seismic-hazard analyses. *Seismological Research Letters* 86(1): 17–26.

Graves RW and Pitarka A (2010) Broadband ground-motion simulation using a hybrid approach. *Bulletin of the Seismological Society of America* 100(5A): 2095–2123.

Graves RW and Pitarka A (2015) Refinements to the Graves and Pitarka (2010) broadband ground-motion simulation method. *Seismological Research Letters* 86(1): 75–80.

Graves RW and Pitarka A (2016) Kinematic ground motion simulations on rough faults including effects of 3D stochastic velocity perturbations. *Bulletin of the Seismological Society of America* 106(5): 2136–2153.

Graves RW, Jordan TH, Callaghan S, Deelman E, Field E, Juve G, Kesselman C, Maechling P, Mehta G, Milner K, Okaya D, Small P and Vahi K (2011) CyberShake: A physics-based seismic hazard model for Southern California. *Pure and Applied Geophysics* 168(3–4): 367–381.

Hartzell S, Leeds A, Frankel A, Williams RA, Odum J, Stephenson W and Silva W (2002) Simulation of broadband ground motion including nonlinear soil effects for a magnitude 6.5 earthquake on the Seattle fault, Seattle, Washington. *Bulletin of the Seismological Society of America* 92(2): 831–853.

Jeong S and Bradley BA (2017a) Amplification of strong ground motions at Heathcote Valley during the 2010–2011 Canterbury earthquakes: Observation and 1D site response analysis. *Soil Dynamics and Earthquake Engineering* 100: 345–356.

Jeong S and Bradley BA (2017b) Amplification of strong ground motions at Heathcote Valley during the 2010–2011 Canterbury earthquakes: The role of 2D nonlinear site response. *Bulletin of the Seismological Society of America* 107(5): 2117–2130.

Kaiser A, Van Houtte C, Perrin N, Wotherspoon L and 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): 39–49.

Lee RL, Bradley BA, Stafford PJ, Graves RW and Rodriguez-Marek A (2020) Hybrid broadband ground motion simulation validation of small magnitude earthquakes in Canterbury, New Zealand. *Earthquake Spectra* 36(2): 673–699.

Maufroy E, Chaljub E, Hollender F, Bard P-Y, Kristek J, Moczo P, De Martin F, Theodoulidis N, Manakou M, Guyonnet-Benaize C, Hollard N and Pitilakis K (2016) 3D numerical simulation

and ground motion prediction: Verification, validation and beyond—Lessons from the E2VP project. *Soil Dynamics and Earthquake Engineering* 91: 53–71.

Maufroy E, Chaljub E, Hollender F, Kristek J, Moczo P, Klin P, Priolo E, Iwaki A, Iwata T, Etienne V, De Martin F, Theodoulidis NP, Manakou M, Guyonnet-Benaize C, Pitilakis K and Bard P-Y (2015) Earthquake ground motion in the Mygdonian Basin, Greece: The E2VP verification and validation of 3D numerical simulation up to 4 Hz. *Bulletin of the Seismological Society of America* 105(3): 1398–1418.

New Zealand Standards (2004) *Structural Design Actions—Part 5: Earthquake Actions—New Zealand* (Standard NZS 1170.5:2004). Wellington, New Zealand: Standards New Zealand.

Olsen K and Takedatsu R (2015) The SDSU broadband ground-motion generation module BBtoolbox version 1.5. *Seismological Research Letters* 86(1): 81–88.

Oth A and Kaiser AE (2014) Stress release and source scaling of the 2010–2011 Canterbury, New Zealand earthquake sequence from spectral inversion of ground motion data. *Pure and Applied Geophysics* 171(10): 2767–2782.

Rai M, Rodriguez-Marek A and Chiou BSJ (2017) Empirical terrain-based topographic modification factors for use in ground motion prediction. *Earthquake Spectra* 33(1): 157–177.

Rai M, Rodriguez-Marek A and Yong A (2016) An empirical model to predict topographic effects in strong ground motion using California small- to medium-magnitude earthquake database. *Earthquake Spectra* 32(2): 1033–1054.

Razafindrakoto HN, Bradley BA and Graves RW (2018) Broadband ground-motion simulation of the 2011 Mw 6.2 Christchurch, New Zealand, earthquake. *Bulletin of the Seismological Society of America* 108(4): 2130–2147.

Ren Y, Zhou Y, Wang H and Wen R (2018) Source characteristics, site effects, and path attenuation from spectral analysis of strong-motion recordings in the 2016 Kaikōura earthquake sequence. *Bulletin of the Seismological Society of America* 108(3B): 1757–1773.

Ristau J (2008) Implementation of routine regional moment tensor analysis in New Zealand. *Seismological Research Letters* 79(3): 400–415.

Ristau J (2013) Update of regional moment tensor analysis for earthquakes in New Zealand and adjacent offshore regions. *Bulletin of the Seismological Society of America* 103(4): 2520–2533.

Ristau J (2018) Overview of moment tensor analysis in New Zealand. In: D'Amico S (ed.) *Moment Tensor Solutions*. Cham: Springer, pp. 281–305.

Rodriguez-Marek A, Montalva GA, Cotton F and Bonilla F (2011) Analysis of single-station standard deviation using the KiK-net data. *Bulletin of the Seismological Society of America* 101(3): 1242–1258.

Roten D, Olsen KB and Pechmann JC (2012) 3D simulations of M 7 earthquakes on the Wasatch fault, Utah, Part II: Broadband (0–10 Hz) ground motions and nonlinear soil behavior. *Bulletin of the Seismological Society of America* 102(5): 2008–2030.

Savran WH and Olsen K (2019) Ground motion simulation and validation of the 2008 Chino Hills earthquake in scattering media. *Geophysical Journal International* 219(3): 1836–1850.

Stafford PJ (2014) Crossed and nested mixed-effects approaches for enhanced model development and removal of the ergodic assumption in empirical ground-motion models. *Bulletin of the Seismological Society of America* 104(2): 702–719.

Taborda R, Azizzadeh-Roodpish S, Khoshnevis N and Cheng K (2016) Evaluation of the southern California seismic velocity models through simulation of recorded events. *Geophysical Journal International* 205(3): 1342–1364.

Thomson EM, Bradley BA and Lee RL (2019) Methodology and computational implementation of a New Zealand Velocity Model (NZVM2.0) for broadband ground motion simulation. *New Zealand Journal of Geology and Geophysics* 63: 110–127.

Wirth EA, Frankel AD and Vidale JE (2017) Evaluating a kinematic method for generating broadband ground motions for great subduction zone earthquakes: Application to the 2003 Mw 8.3 Tokachi-oki earthquake. *Bulletin of the Seismological Society of America* 107(4): 1737–1753.

Withers KB, Olsen KB, Shi Z and Day SM (2019) Validation of deterministic broadband ground motion and variability from dynamic rupture simulations of buried thrust earthquakes. *Bulletin of the Seismological Society of America* 109(1): 212–228.

Yenier E and Atkinson GM (2015) An equivalent point-source model for stochastic simulation of earthquake ground motions in California. *Bulletin of the Seismological Society of America* 105(3): 1435–1455.