Evaluation of Empirical Ground-Motion Models for the 2022 New Zealand National Seismic Hazard Model Revision

Robin L. Lee[*1] , Brendon A. Bradley[1] , Elena F. Manea[2,3] , Jesse A. Hutchinson[4] , and Sanjay S. Bora[2]

ABSTRACT

This article presents an evaluation of empirical ground-motion models (GMMs) for active shallow crustal, subduction interface, and subduction slab earthquakes using a recently developed New Zealand (NZ) ground-motion database for the 2022 New Zealand National Seismic Hazard Model revision. This study considers both NZ-specific and global models, which require evaluation to inform of their applicability in an NZ context. A quantitative comparison between the models is conducted based on intensity measure residuals and a mixed-effects regression framework. The results are subsequently investigated to assess how the models are performing in terms of overall accuracy and precision, as well as to identify the presence of any biases in the model predictions when applied to NZ data. Many models showed reasonable performance and could be considered appropriate for inclusion within suites of models to properly represent ground-motion predictions and epistemic uncertainty. In general, the recent models that are NZ-specific or developed on large international databases performed the best. This evaluation of models helped inform suitable GMMs for the ground-motion characterization model logic tree. In addition, spatial trends in systematic site-to-site residuals to the west of the Taupō Volcanic Zone demonstrated the need for backarc attenuation modifications for slab earthquakes.

KEY POINTS

- Several New Zealand (NZ)-specific and global empirical ground-motion models (GMMs) were evaluated against the recent NZ ground-motion database.
- The quantitative evaluation informed suitable GMMs for the ground-motion characterization model logic tree for the 2022 New Zealand National Seismic Hazard Model revision.
- Spatial trends in site-to-site residuals indicated the need for backarc attenuation adjustments for slab GMMs.

Supplemental Material

INTRODUCTION

Ground-motion characterization modeling is an important component of seismic hazard analysis, with the purpose of predicting ground motions generated at a geographic location by a given earthquake source. To achieve this, a suite of credible ground-motion models (GMMs) is required. Empirical GMMs have been conventionally used in such applications. In the recent decades, several hundreds of empirical GMMs have been published (Douglas and Edwards, 2016). Empirical GMMs have become increasingly complex in their mathematical formulation and computational implementation over many

decades of development, and utilize databases of recorded ground motions that are continuously increasing in size. These databases may comprise ground motions from earthquakes specific to a particular region or across many regions that would generally be used in the development of regional or global models, respectively.

Many major international modeling efforts have been recently undertaken as a part of the Next Generation Attenuation (NGA) series, in which several teams of experienced empirical ground-motion modelers each developed models starting with a consistent international ground-motion database (GMDB). The projects to date include NGA-West (Power *et al.*, 2008) and NGA-West2 (Bozorgnia *et al.*, 2014)

for active shallow crustal earthquakes, NGA-Sub (Bozorgnia et al., 2022) for subduction interface and slab earthquakes, and NGA-East (Goulet et al., 2021) for stable continental earthquakes.

New Zealand (NZ) has a long history of empirical GMM development, although efforts have been sparse and intermittent, primarily due to data paucity limitations at large magnitudes (**M**) and short source-to-site distances (R_{rup}). The earliest models were developed by Matuschka (1980) and Peek (1980), whereas models commonly used in modern hazard assessments were developed by McVerry et al. (2006) and Bradley (2013). The McVerry et al. (2006) model was used in the previous published version of the National Seismic Hazard Model (NSHM; Stirling et al., 2012). Both McVerry et al. (2006) and Bradley (2013) were used in an interim update of seismic hazard in Canterbury (Gerstenberger et al., 2016) following the 2010/2011 Canterbury earthquake sequence. The use of only one or two empirical GMMs has been a major limitation of the previous NZ NSHMs and probabilistic seismic hazard analyses (PSHAs) with respect to appropriate consideration of epistemic uncertainty.

With few NZ-specific empirical GMMs developed recently, it is imperative that global models are considered to appropriately account for epistemic uncertainty in NZ PSHA. Because global models tend to be developed on international datasets, their predictive capability of ground motions from NZ earthquakes must be quantified through evaluation against observed data. Van Houtte (2017) presented an evaluation of various models that were available (at the time of the study) against the Van Houtte et al. (2017) strong-motion database. Van Houtte (2017) considered six models for active shallow crustal earthquakes, which are also included in the model set considered in this study, as well as four models for subduction zone (interface and slab) earthquakes, three of which are considered in this study. Of the subduction zone models considered in Van Houtte (2017), three models were developed between 2003 and 2006, and only one had been developed in the last decade. Subduction interface and slab earthquakes were also analyzed together (rather than as separate tectonic classes) due to data paucity, as only six interface earthquakes were included. Results from the study found several crustal models to perform similarly well, including both NZ-specific and global models. However, subduction zone models had a wide range of performance, and guidance on their use simply defaulted to equal logic tree weighting, given the limitations of the evaluation.

Since Van Houtte (2017) was published, several developments have occurred that present an opportunity to more rigorously evaluate the performance of empirical GMMs in NZ. Specifically, a larger GMDB for NZ has been developed by Hutchinson et al. (2022), and several key empirical GMMs have been developed for subduction interface and slab earthquakes, mainly from the NGA-Sub project. This article

provides a rigorous evaluation of credible empirical GMMs in an NZ context, intended to inform seismic hazard modeling efforts for the 2022 NZ NSHM revision (Bradley et al., 2022; Gerstenberger et al., 2022; Brdaley et al., 2024; Gerstenberg et al., 2024). The final logic tree and weighting used in the 2022 NZ NSHM revision is not included in this article, because this evaluation was not the sole determinant of the corresponding decisions. The holistic assessment and process, including consideration of all factors, are included and detailed in Bradley et al. (2024) and Gerstenberger et al. (2024).

EMPIRICAL GMMS CONSIDERED

The aim of this study is to assess the performance of empirical GMMs, both NZ-specific and global, against NZ data. Although several hundreds of models exist (Douglas and Edwards, 2016), an evaluation of all models would be prohibitively time consuming. Hence, a set of criteria was adopted to trim the list of models and determine a collection of candidate models that were considered the most appropriate. In general, the following aspects were considered:

1. Recent NZ-specific models were included.
2. Credible global models developed by experienced modeling teams and based on large GMDBs were included. More recent models that supersede older models were typically considered.
3. Region-specific foreign models were considered if the region they were developed for is of similar tectonic character to NZ.
4. Models should include appropriate consideration of key parameter scenarios that are likely in NZ with respect to applicable **M**, R_{rup}, nonlinear site effects, and others.
5. Models that provide prediction of the RotD50 component (Boore, 2010) were preferred.
6. Models that provide prediction of pseudospectral acceleration (PSA) for periods up to 10 s were preferred. If models do not achieve this criterion, they are evaluated up to their maximum predicted period.

It is important to reiterate that the suites of models included are not intended to be an exhaustive list of all published models. In addition, preference for models that are not NZ-specific was given to models that are derived from international datasets rather than from region-specific, foreign datasets (as emphasized earlier).

For active shallow crustal earthquakes, a total of eight models are considered. The McVerry et al. (2006) and Bradley (2013) models were selected as the relevant existing NZ-specific models. Because there were insufficient data to develop completely new NZ-specific models, the McVerry et al. (2006) and Bradley (2013) crustal models used the Abrahamson and Silva (1997) and Chiou et al. (2010) models as base models, respectively, with modifications to improve

Model	Abbreviation	Period Range	Component	M	Distance (km)	V_{S30} (m/s)
McVerry et al. (2006)	McV06	0.075–3 s	Geomean	5.25–7.5	0–400	NZS1170.5:2004*
Bradley (2013)	B13	0.01–10 s	Geomean	3.5–8.5	0–300	180–1500
Abrahamson et al. (2014)	ASK14	0.01–10 s	RotD50	3.0–8.5	0–300	180–1500
Boore et al. (2014)	BSSA14	0.01–10 s	RotD50	3.0–8.5	0–400	150–1500
Campbell and Bozorgnia (2014)	CB14	0.01–10 s	RotD50	3.3–8.5	0–300	150–1500
Chiou and Youngs (2014)	CY14	0.01–10 s	RotD50	3.5–8.5	0–300	180–1500
Atkinson (2022)	A22	0.01–10 s	RotD50	4.5–8.5	0–400	150–1000
Stafford (2022)	S22	0.01–10 s	RotD50	4.5–8.4	0–300	180–1500

In addition to the period range shown, all models also provide predictions of peak ground acceleration. The **M** range shown is indicative of the widest range considering all faulting styles.
*Site classes based on NZS1170.5:2004 (New Zealand Standards, 2004).

their fit to NZ data. Although McVerry et al. (2006) and Bradley (2013) provide prediction of the geometric mean (geomean) component of ground motion, they are still included in this study. It is observed that the RotD50 component is similar to the geomean component for response spectral periods shorter than 0.1 s and up to approximately 3%–4% larger for longer periods (Boore and Kishida, 2017). This difference is considered negligible for this application in comparison to uncertainties in observation metadata and prediction differences between alternative models, and does not influence the outcomes of this study. However, a more rigorous consideration of the conversion between geomean and RotD50 components will be investigated in the future studies. Four NGA-West2 models were also selected, specifically the Abrahamson et al. (2014), Boore et al. (2014), Campbell and Bozorgnia (2014), and Chiou and Youngs (2014) models, because they are currently considered the most robust global models. In addition, the central branches of two backbone models developed as a part of the 2022 NZ NSHM revision are also included, the Atkinson (2022) and Stafford (2022) models. Table 1 provides model descriptions and key parameter ranges for these selected active shallow crustal models. The model name abbreviations provided in Table 1 are used herein for brevity.

For subduction interface and slab earthquakes, a total of 10 models are considered for each of these tectonic classes. First, the McVerry et al. (2006) model is included as the relevant existing NZ-specific model. Because of insufficient NZ data, the McVerry et al. (2006) model used the Youngs et al. (1997) models as the base models that were modified to improve their fit to NZ data. The Zhao et al. (2006) model (that also predicts the geomean component) was selected because of its extensive use, both internationally and in NZ-based studies. Notably, the Zhao, Liang, et al. (2016) interface and Zhao, Jiang, et al. (2016) slab models were not used, as previous experience with the model indicated erroneous predictions for NZ ground motions. The Abrahamson et al. (2016) (BC Hydro) and Abrahamson et al. (2018) (BC Hydro Update) models developed for the western United States were also selected, because the tectonic

character of this region is considered to be relatively similar to NZ. The models were developed by experienced teams of ground-motion modelers, and they are state-of-practice models at present. The three main models from the NGA-Sub project have been considered as the most recent significant global models. However, in this study, each regionalization is counted as a separate "model" to better track the different predictions. Where available, the global and NZ regionalizations are used—specifically, the Abrahamson and Gülerce (2020) global and NZ models, the Kuehn et al. (2020) global and NZ models, and the Parker et al. (2022) global model have been included. Hence, we count this to be five NGA-Sub models. Finally, the central branches of the Atkinson (2022) backbone models for subduction interface and slab earthquakes developed as a part of the 2022 NZ NSHM revision are also included. Tables 2 and 3 provide model descriptions and key parameter ranges for the selected interface and slab models, respectively. The model name abbreviations provided in Tables 2 and 3 are used herein for brevity.

More detailed information on each of the models considered can be found in their respective publications. In addition, many models have been thoroughly compared to understand modeling similarities and differences in Gregor et al. (2014) for NGA-West2 models and Gregor et al. (2022) for NGA-Sub models.

GMDB

For evaluation of the predictive capability of the empirical GMMs, a high-quality database of ground-motion records with source, site, and path metadata is necessary. This study adopts the GMDB v.1.0 developed by Hutchinson et al. (2022). Although several aspects of quality control were undertaken in the GMDB development, a preliminary evaluation of the empirical GMM performance utilizing the full database identified significant biases and imprecisions that were systematically attributed to particular ground-motion record, earthquake, or site data sources. This caused issues with the proper and consistent evaluation of GMM performance. Therefore,

TABLE 2
Interface Model Descriptions and Key Parameter Ranges

Model	Abbreviation	Period Range	Component	M	Distance (km)	V_{S30} (m/s)
McVerry et al. (2006)	McV06 SI	0.075–3 s	Geomean	5.25–7.5	0–400	NZS1170.5:2004*
Zhao et al. (2006)	Z06 SI	0.05–5 s	Geomean	5–8.5[†]	0–300[†]	Site class[‡]
Abrahamson et al. (2016)	A16 SI	0.02–10 s	RotD50	6–8.8[†]	0–300[†]	150–1500[†]
Abrahamson et al. (2018)	A18 SI	0.01–10 s	RotD50	5–9.5	0–1000	150–1500[†]
Abrahamson and Gülerce (2020) Global	AG20 SI Global	0.01–10 s	RotD50	6–9.5	0–500	150–1500
Abrahamson and Gülerce (2020) NZ	AG20 SI NZ	0.01–10 s	RotD50	6–9.5	0–500	150–1500
Kuehn et al. (2020) Global	KBCG20 SI Global	0.01–10 s	RotD50	5–9.5	10–1000	150–1500
Kuehn et al. (2020) NZ	KBCG20 SI NZ	0.01–10 s	RotD50	5–9.5	10–1000	150–1500
Parker et al. (2022) Global	PSBAH22 SI Global	0.01–10 s	RotD50	4.5–9.5	20–1000	150–2000
Atkinson (2022)	A22 SI	0.01–10 s	RotD50	4.5–8.5	0–400	150–1000

In addition to the period range shown, all models also provide predictions of peak ground acceleration.
*Site classes based on NZS1170.5:2004 (New Zealand Standards, 2004).
[†]Approximate applicable parameter range inferred from the details in the respective publication, because they were not explicitly stated.
[‡]Site classes based on V_{S30} criteria in Zhao et al. (2006).

TABLE 3
Slab Model Descriptions and Key Parameter Ranges

Model	Abbreviation	Period Range	Component	M	Distance (km)	V_{S30} (m/s)
McVerry et al. (2006)	McV06 SS	0.075–3 s	Geomean	5.25–7.5	0–400	NZS1170.5:2004*
Zhao et al. (2006)	Z06 SS	0.05–5 s	Geomean	5–8.5[†]	0–300[†]	Site class[‡]
Abrahamson et al. (2016)	A16 SS	0.02–10 s	RotD50	5–7.9[†]	0–300[†]	150–1500[†]
Abrahamson et al. (2018)	A18 SS	0.01–10 s	RotD50	5–9.5	0–1000	150–1500[†]
Abrahamson and Gülerce (2020) Global	AG20 SS Global	0.01–10 s	RotD50	5–8	0–500	150–1500
Abrahamson and Gülerce (2020) NZ	AG20 SS NZ	0.01–10 s	RotD50	5–8	0–500	150–1500
Kuehn et al. (2020) Global	KBCG20 SS Global	0.01–10 s	RotD50	5–8.5	10–1000	150–1500
Kuehn et al. (2020) NZ	KBCG20 SS NZ	0.01–10 s	RotD50	5–8.5	10–1000	150–1500
Parker et al. (2022) Global	PSBAH22 SS Global	0.01–10 s	RotD50	4.5–8.5	35–1000	150–2000
Atkinson (2022)	A22 SS	0.01–10 s	RotD50	4.5–8.5	0–400	150–1000

In addition to the period range shown, all models also provide predictions of peak ground acceleration.
*Site classes based on NZS1170.5:2004 (New Zealand Standards, 2004).
[†]Approximate applicable parameter range inferred from the details in the respective publication, because they were not explicitly stated.
[‡]Site classes based on V_{S30} criteria in Zhao et al. (2006).

the following additional quality criteria had to be applied to produce a dataset appropriate for evaluation:

- A source model with a rigorous source-specific finite fault inversion (e.g., 2010 Darfield, 2011 Christchurch, 2016 Kaikōura earthquakes) or centroid moment tensor solution (Ristau, 2013) was required.
- Only records from strong-motion accelerometer channels (HN and BN) were used. Records from broadband seismometer channels (HH and BH) were not used.
- Records were screened based on minimum M and maximum R_{rup} to better align the database with the applicable parameter ranges of the GMMs, in addition to an M-dependent R_{rup} filter to mitigate against poor signal-to-noise ratio and instrument triggering issues. This filter was initially based on M-dependent peak ground velocity thresholds that were subsequently modified to better achieve an appropriate trade-off between data quality and quantity.

In addition to these quality criteria, a minimum of three records per event and per site were enforced. Table 4 presents the minimum M and the maximum R_{rup} limits applied, as well as the resulting quantity of earthquakes, recording stations, and ground motions that make up the datasets for each tectonic class (crustal, interface, and slab). It is observed that there was no explicit tectonic class for volcanic earthquakes in the adopted database; hence, they are included within the considered classes based on the classification scheme used (Hutchinson et al., 2022). This would result in most of the volcanic earthquakes being classified as crustal earthquakes. However, this is likely to comprise a small quantity of earthquakes, as shown in subsequent map plots in which there are few crustal earthquakes in the Taupō Volcanic Zone. This is a limitation that could be improved upon in the future studies. There is also no station backarc flag in the database, and hence the analyses assume all stations are in the forearc region unless otherwise stated, because this is subsequently examined in a later section.

Ground-Motion Database Adopted Parameter Ranges and Quantities for Each Tectonic Class

Tectonic Class	Minimum M	Maximum R_{rup} (km)	Earthquakes	Stations	Records
Crustal	3.5	300	655	306	12,432
Interface	4.5	500	83	221	1,800
Slab	4.5	500	115	226	2,432
Total			853	340	16,664

Figure 1 presents $\mathbf{M}$–R_{rup} scatterplots of the adopted datasets as well as histograms showing the distribution of the two parameters, in which the frequency count is by the number of records (as opposed to the number of earthquakes for $\mathbf{M}$). The gray points shown in the scatterplots are records removed following the application of the additional quality criteria observed previously. The red points show the adopted datasets. The locations of earthquakes and stations are shown in Figure 2. Most crustal earthquakes are located in the central and northern South Island regions. Many of these earthquakes are attributed to the 2010/2011 Canterbury (Bradley and Cubrinovski, 2011; Bradley, 2012, 2013), Marlborough (Holden *et al.*, 2013), and 2016 Kaikōura (Bradley *et al.*, 2017) earthquake sequences. Interface earthquakes are generally located in the shallower regions of the Hikurangi and Puysegur subduction zones, whereas slab earthquakes are distributed across both shallow and deep parts of the two subduction zones (Gledhill *et al.*, 2011).

The intensity measures (IMs) that this study will consider are peak ground acceleration (PGA) and 5%-damped PSA for vibration periods between 0.01 and 10 s. Each record has the minimum usable frequency (f_{min}) determined from a ground-motion quality classification neural network (Dupuis *et al.*, 2023), which is used to determine the maximum period of PSA that each record can be used for. It is important to mention that the lowest f_{min} that could be predicted by the specific version of the neural network used is 0.1 Hz, which would correspond to a vibration period of 10 s. The neural network rarely predicts this lower limit value due to the labeled data it was trained on, the continuous nature of the variable, and the neural network architecture. To include PSA(10s) in this study, any record with $f_{min} \leq 0.12$ Hz was used for the prediction of PSA(10s), despite only corresponding to a vibration period of 8.33 s. Therefore, any analyses for PSA(10s) should not be used quantitatively or considered to be rigorous; rather, they should be used as a qualitative indication of model prediction performance at periods close to 10 s.

MODEL EVALUATION FRAMEWORK
A robust statistical framework is necessary to provide a quantitative evaluation of each model's predictive capability in terms of accuracy, precision, and bias. Many different analysis frameworks have been presented and used in past studies, for example, goodness-of-fit, log-likelihood, and residual regression testing (e.g., Scherbaum *et al.*, 2009; Mak *et al.*,

2017, etc.). This study couples ground-motion IM residuals with a mixed-effects regression framework. The residual for each IM is defined as,

$$\Delta_{es} = \ln \text{IM}_{\text{obs},es} - f_{es}, \tag{1}$$

in which Δ_{es} is the total residual; $\ln \text{IM}_{\text{obs},es}$ is the natural logarithm of the observed IM; and f_{es} is the natural logarithm of the predicted (median) IM via an empirical GMM, which is a function of the earthquake rupture, e, and site location, s. With this definition, a positive residual indicates that the GMM-based IM is underpredicted, and a negative residual indicates that the IM is overpredicted. All residuals are in natural log units. Mixed-effects regression analysis is applied to the residuals across the entire dataset to partition the residuals into various components associated with ground-motion variability. Subsequent interpretation of the residual components provides a means to identify systematic biases and trends. Following the notation of Al Atik *et al.* (2010), the total residual can be written as,

$$\Delta_{es} = a + \delta B_e + \delta S2S_s + \delta W^0_{es}, \tag{2}$$

in which a is the model prediction bias (fixed effect), δB_e is the between-event residual (random effect) with zero mean and variance τ^2, $\delta S2S_s$ is the systematic site-to-site residual (random effect) with zero mean and variance ϕ^2_{S2S}, and δW^0_{es} is the "remaining" within-event residual with residual variance ϕ^2_{SS}. Equation (2) illustrates that Δ_{es} has mean a and total variance $\sigma^2 = \tau^2 + \phi^2_{S2S} + \phi^2_{SS}$, assuming that δB_e, $\delta S2S_s$, and δW^0_{es} are independent random variables. The δW^0_{es} represents factors not systematically accounted for by the δB_e or $\delta S2S_s$, or not accounted for in the models themselves.

COMPARISON OF MODEL PERFORMANCE ON DATA
This section provides a comparison of empirical GMM predictions against the observed ground-motion data from the NZ GMDB through analysis of residuals.

Model prediction bias and total standard deviation
Model prediction bias (a) and total standard deviation (σ) for the predicted IMs are presented in Figure 3. Figure 3a shows that model prediction biases are small for most crustal models across the period range considered, with values generally
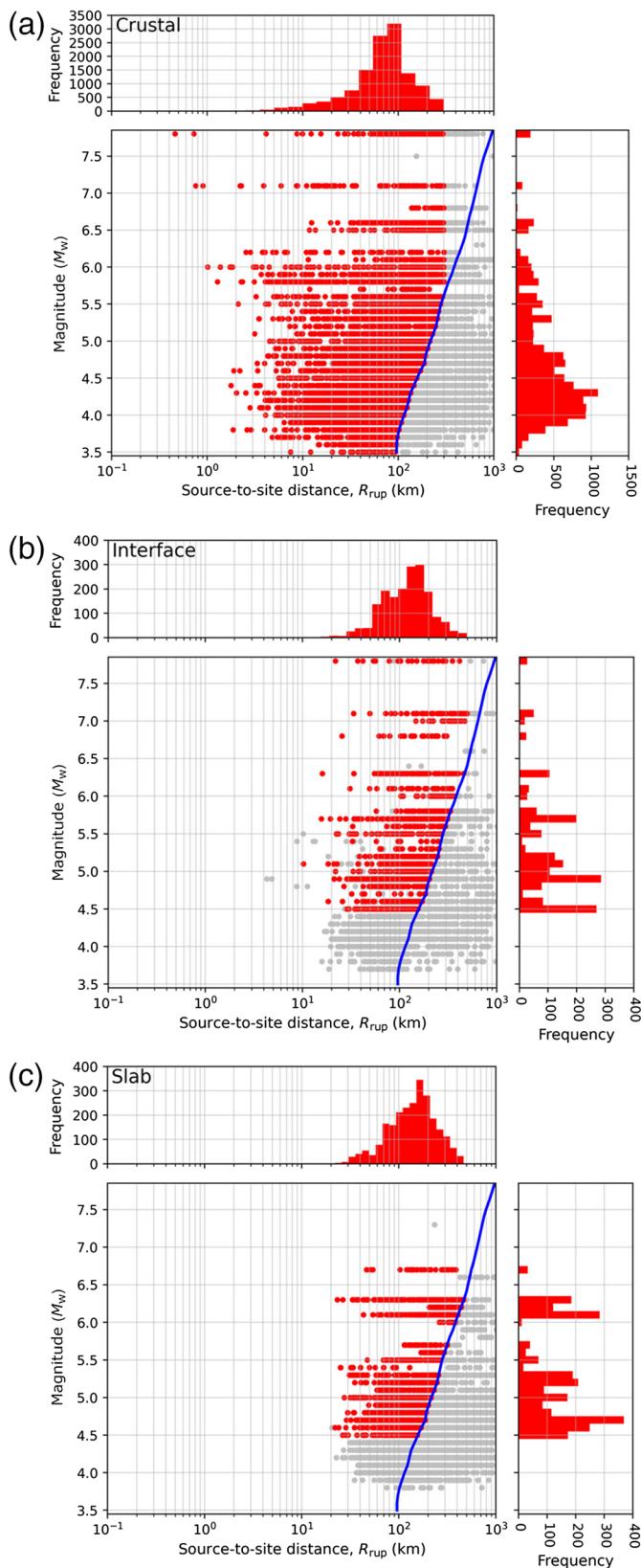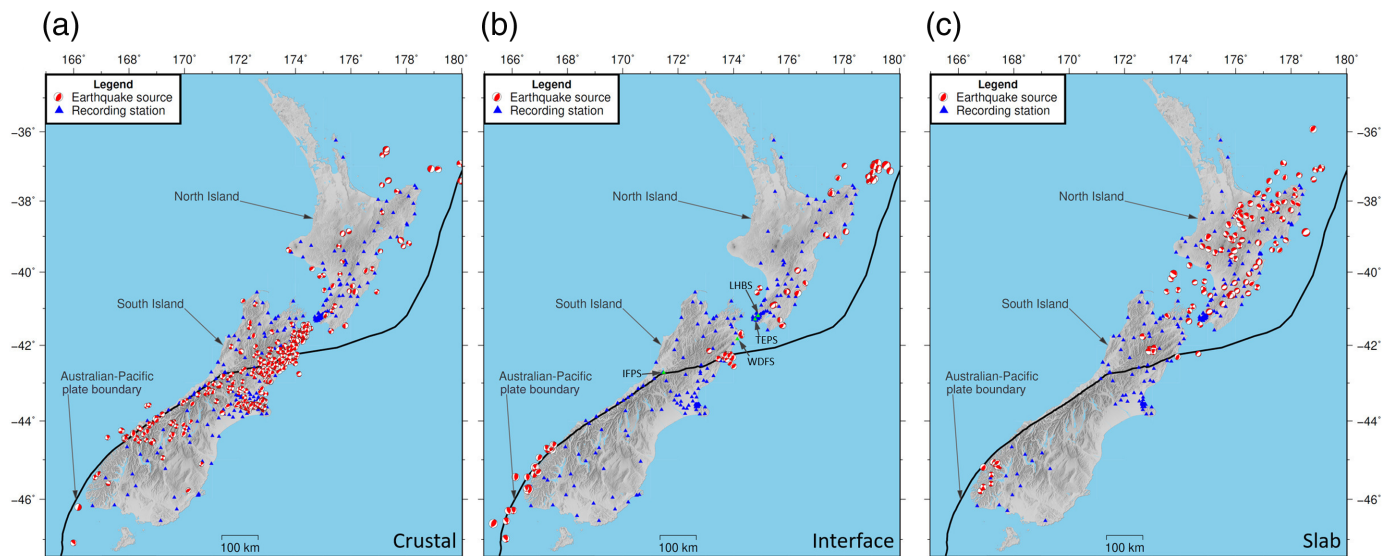
Figure 1. Scatterplots and histograms of earthquake magnitude and ground-motion source-to-site distance distributions from the adopted datasets for: (a) crustal, (b) interface, and (c) slab earthquakes. The magnitude histogram frequency is counted by record. The blue line indicates the magnitude-dependent source-to-site distance filter applied. The color version of this figure is available only in the electronic edition.

between −0.2 and 0.2. The exceptions are McV06, which significantly overpredicts (mostly below the bottom y-axis limit, around −2.0), and the CY14, A22, and S22 models, which underpredict at longer periods ($T \geq 3$ s). The similarity of S22 and CY14 over long periods is expected, given that the S22 backbone model uses CY14 as the host model. The overprediction by McV06 is primarily a result of the database being mostly comprised of small **M** earthquakes, for which the model is strictly not applicable. However, the database is not altered to maintain a like-for-like comparison; rather, the bias will be compared with δB_e subsequently to identify the predictive performance within its applicable **M** range. This general result of consistent performance between the B13, A22, S22, and NGA-West2 models indicates that, to first order, each model seems to perform relatively accurately and may be appropriate candidate models for application in NZ. Van Houtte (2017) found that the B13 and NGA-West2 models tended to slightly overpredict at short periods and slightly underpredict at long periods (see fig. 9a in Van Houtte, 2017). Differences between the results of this study and Van Houtte (2017) may have arisen due to different earthquakes, stations, and records used in the analysis, as well as their characterization.

Figure 3b shows that for short periods, σ associated with B13, BSSA14, A22, and S22 models are the lowest, followed by ASK14, and, finally, the McV06, CY14, and CB14 models have the largest σ. At moderate-to-long periods ($T \geq 1.0$ s), the models converge to similar σ values between 0.6 and 0.8. The gray shaded area indicates an approximate range of published model σ values (i.e., total standard deviations from the model development) based on the models considered and the scenarios in the GMDB. Overall, the prediction σ from the evaluation appears mostly consistent with the upper values of published model σ at short periods and the lower values of model σ at long periods. The exception is McV06, which has a relatively large prediction σ up to its maximum predicted period of 3.0 s.

Figure 3c shows that interface model prediction biases have a larger range than that of crustal models, generally having values between −0.5 and 0.5, with the exception of AG20 SI NZ, which is consistently below −0.5 (overpredicting), and McV06 SI, which dips under −0.5 at moderate periods. The overprediction by McV06 SI is less severe compared to McV06 for crustal earthquakes, as the minimum **M** used from the observed GMDB is 4.5 for interface earthquakes (compared to 3.5 for crustal). McV06 SI is also fundamentally different from McV06 for crustal earthquakes, because they use different base models. Although the range of bias values is larger, which may suggest larger epistemic uncertainty in their prediction compared to crustal models, the model prediction biases of most models are still considered to be relatively consistent in a ground-motion context. For comparison, Van Houtte (2017) considered four subduction zone models and found their model prediction biases to have a larger range, from −1.0 to 1.0 across the period range considered, although this

Figure 2. Maps showing the location of earthquake sources and ground-motion recording stations considered across New Zealand for (a) crustal, (b) interface, and (c) slab earthquakes. The locations of the four sites in Figure 6 are shown in panel (b). The Australian-Pacific tectonic plate boundary is shown by the thick black line. The color version of this figure is available only in the electronic edition.

included predictions of both interface and slab earthquakes using the appropriate forms of each model, because Van Houtte (2017) analyzed the two tectonic classes together (see fig. 9b in Van Houtte, 2017).

Figure 3d shows that the σ of all interface models are relatively similar. At short periods ($T \leq 1.0$ s), σ generally ranges between 0.8 and 0.9, which is comparable to crustal models over the same period range. However, the σ of interface models are slightly larger than the crustal models at moderate-to-long periods ($T \geq 1.0$ s). Overall, the prediction σ from the evaluation appears consistent with the upper values of the published model σ based on the models considered and the scenarios in the GMDB (i.e., the gray shaded area).

Figure 3e shows that slab model prediction biases have a larger range than that of both crustal and interface models, generally having values between −1.0 and 1.0, which may suggest that slab models currently have the largest epistemic uncertainty of the tectonic classes considered. McV06 SS is an exception, with a larger overprediction (with a model prediction bias between −1 and −1.5). Again, this is due to extrapolation below the McV06 SS model's applicable **M** range. As mentioned previously, Van Houtte (2017) analyzed the prediction of interface and slab earthquakes together, and found model prediction biases that ranged from −1.0 to 1.0, which is similar to the analysis for slab models in this study. Aside from McV06 SS, Z06 SS has the next most significant overprediction at all periods, whereas PSBAH22 SS Global has the largest underprediction at all periods. A22 also has similar underpredictions to PSBAH22 SS Global over long periods. Most models experience slight dips in model prediction bias at periods between roughly 0.4 to 4.0 s.

The σ for slab models, shown in Figure 3f, indicate that slab models are the least precise of the different tectonic classes, with values at short periods generally between 0.9 and 1.0. At short periods, the prediction σ from the evaluation is similar to, or slightly above, the upper values of published model σ
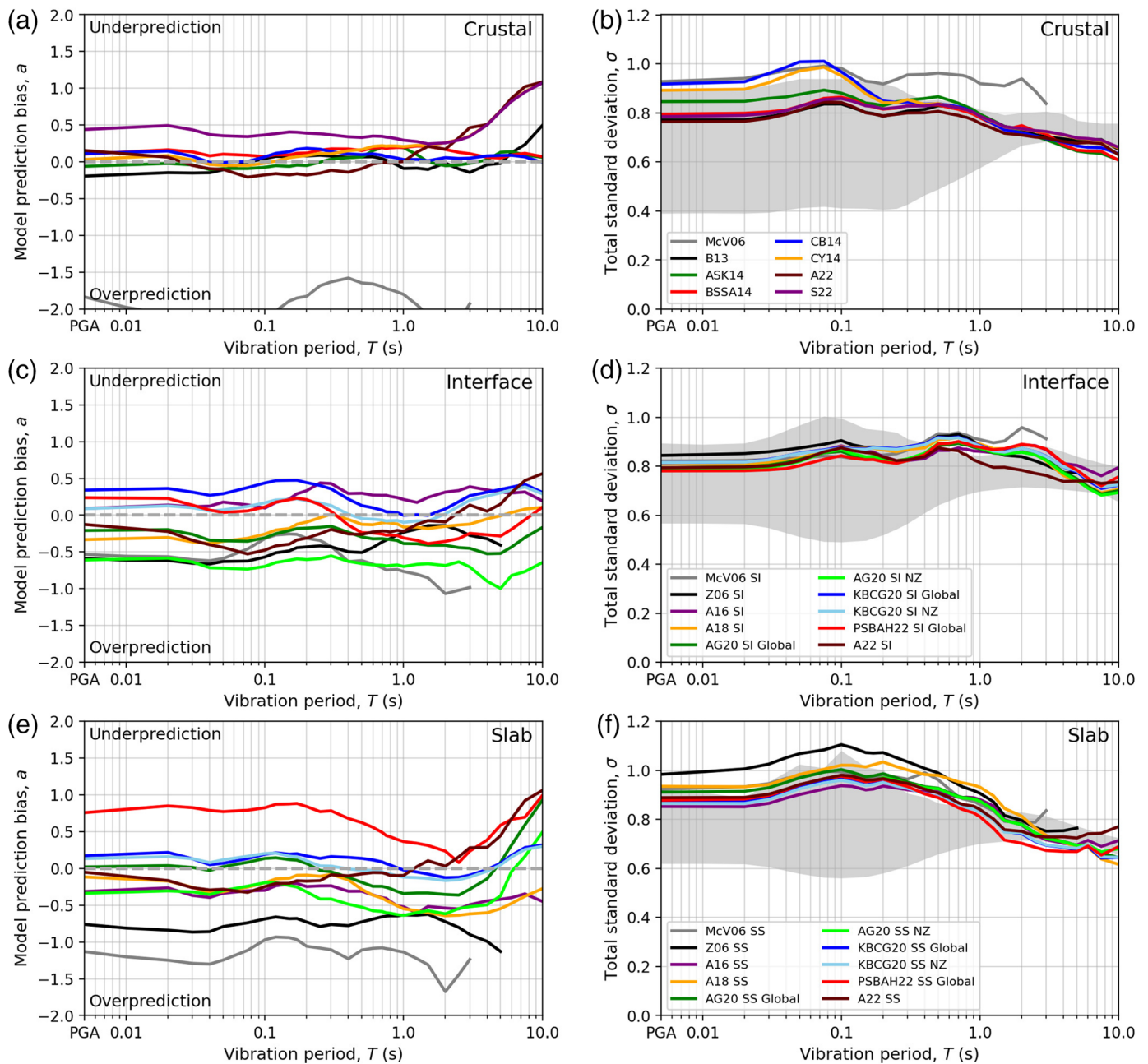
based on models considered and scenarios in the GMDB. At longer periods, the prediction σ is similar to, or slightly below, the lower values of published model σ.

Between-event standard deviation

Figure 4 presents the between-event standard deviation, τ, for crustal, interface, and slab tectonic classes. For crustal models, τ has large variability between models at short periods but converges to similar τ at long periods. The relative trends are similar to those in σ, which indicates that source-specific biases may be the primary cause of differences in crustal model precision. Notably, the B13, A22, and S22 models have the lowest τ at short periods, which may be due to their NZ-specific development compared with global (ergodic) models that represent earthquakes worldwide. The McV06 model typically has larger τ, despite also being a model developed specifically for NZ. This may be due to limitations related to data (from the NZ strong-motion station network up to the end of 1995), adopted functional form and input parameters, among other modeling decisions. Interface models all appear to have relatively consistent τ—the most consistent of all tectonic classes. The exception is A22, which has lower τ at periods $1 \leq T \leq 4$ s. Slab models have some significant differences in τ between models across all periods, but without any apparent trend.

Systematic site-to-site and remaining within-event standard deviations

Figure 5 presents the systematic site-to-site standard deviation, ϕ_{S2S}, and the remaining within-event standard

Figure 3. Summary evaluation results for (a) crustal model prediction bias, (b) crustal total standard deviation (σ), (c) interface model prediction bias, (d) interface σ, (e) slab model prediction bias, and (f) slab σ. The gray shaded area indicates an approximate range of published model σ based on their development for the models considered and scenarios in the ground-motion database for each tectonic class. The color version of this figure is available only in the electronic edition.

deviation, ϕ_{SS}. For ϕ_{S2S}, all models appear to be consistent with each other (i.e., in a model-to-model comparison) within each tectonic class. This is likely due to most models' use of 30 m time-averaged shear-wave velocity (V_{S30}) as the primary site characterization parameter. Crustal and interface models appear to have values that are within expectations for empirical GMMs. However, ϕ_{S2S} for slab models at short periods appear to have particularly large values: between 0.6 and 0.8. This indicates that these slab models have higher variability in prediction accuracy on a site-to-site basis. Subsequent analyses show that the likely cause of this increased ϕ_{S2S} is the omission of backarc attenuation effects in the slab models. For ϕ_{SS}, all models have values generally between 0.4 and 0.5.

Between-event and systematic site-to-site residuals

Figure S1, available in the supplemental material to this article, presents the δB_e as a function of vibration period for all earthquakes within the crustal, interface, and slab tectonic classes for each respective model. For crustal earthquakes, the larger
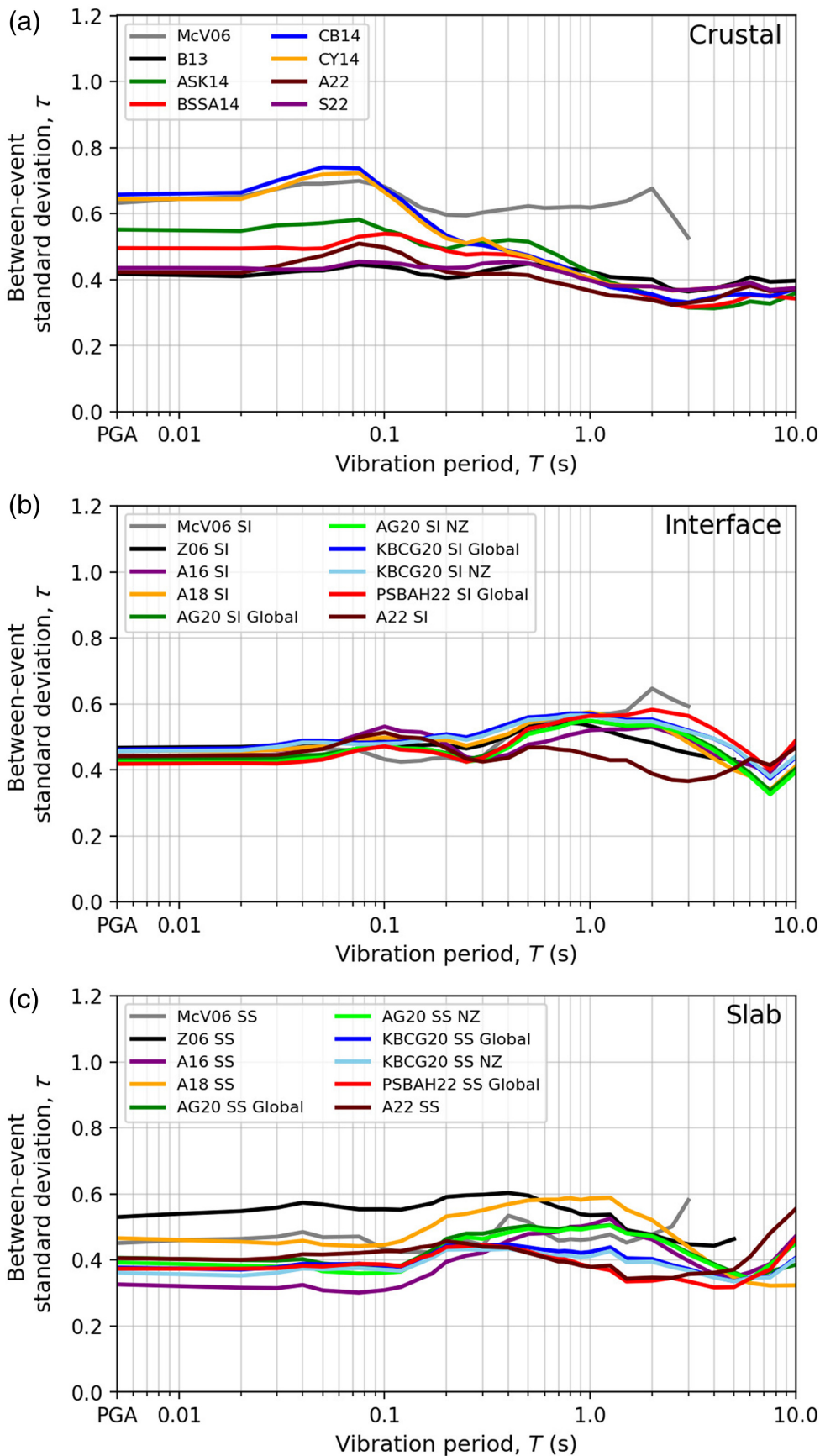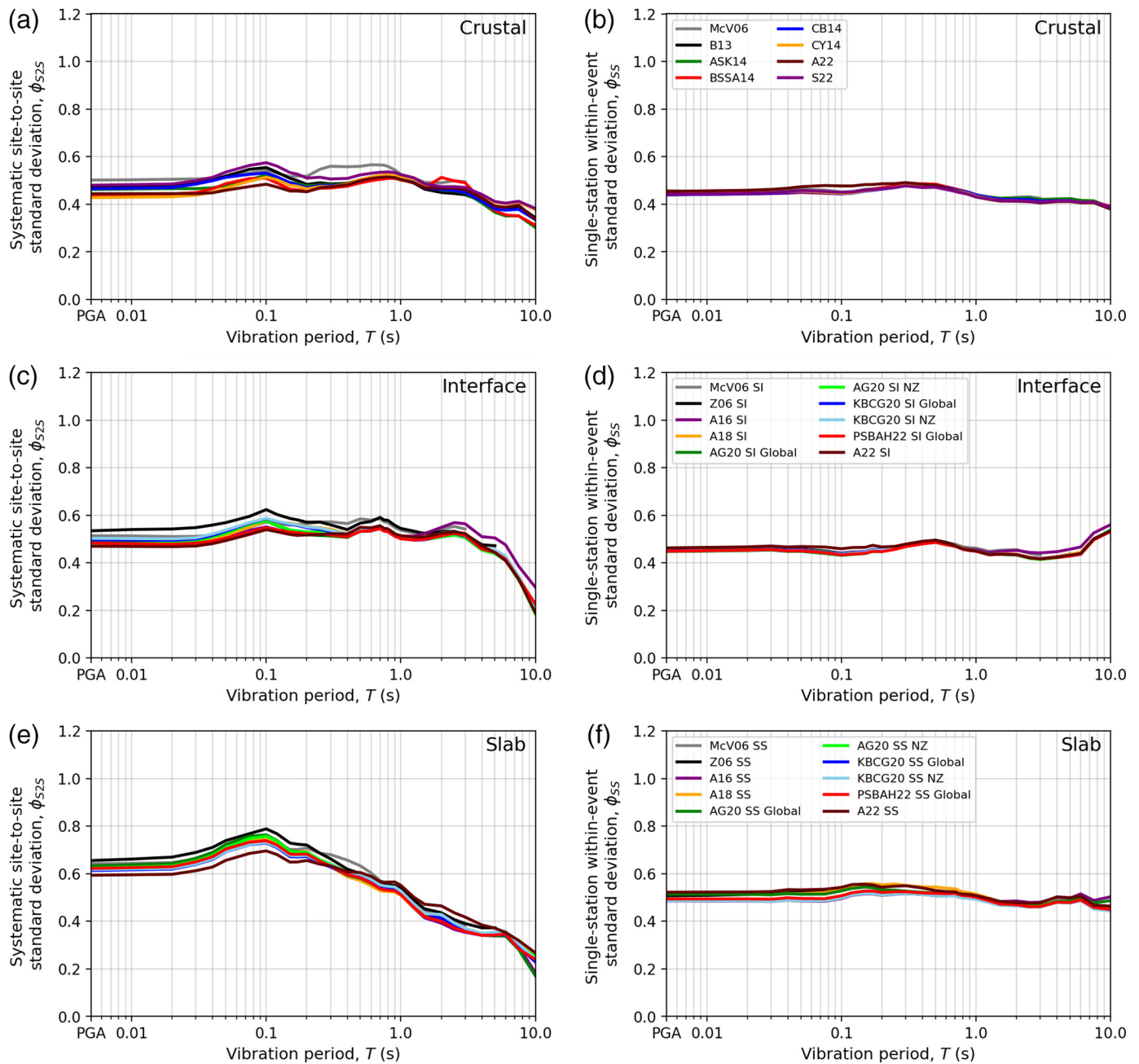
Figure 4. Between-event standard deviation (τ) for (a) crustal models, (b) interface models, and (c) slab models. The color version of this figure is available only in the electronic edition.

scatter of δB_e at short periods for the CB14 and CY14 models that resulted in the relatively large τ can be explicitly seen. Interface earthquakes have values of δB_e that appear relatively consistent between all models. Finally, for slab models, several earthquakes have particularly large positive values of δB_e for the Z06 SS and A18 SS models. The same earthquakes appear to have smaller positive values of δB_e for the other slab models; hence, the Z06 SS and A18 SS models are likely less capable of modeling important aspects of those particular earthquakes. All of these earthquakes with large positive values of δB_e for the Z06 SS and A18 SS models are located across the North Island, but no other common characteristic has been identified so far.

Figure S2 presents the $\delta S2S_s$ as a function of vibration period for all stations within the crustal, interface, and slab tectonic classes for each respective model. For stations within the crustal set, no stations appear to have $\delta S2S_s$ that are considered outliers. For stations within the interface set, there are a few stations with large negative $\delta S2S_s$ (i.e., below −1.5). For stations within the slab set, there are a few stations with large positive $\delta S2S_s$ (i.e., greater than 1.5) that are located at the bottom of the South Island near the Puysegur subduction zone. There are also several stations with large negative $\delta S2S_s$ (i.e., below −1.5) at short periods; these are located in the northwest area of the North Island, around Waikato and Auckland (GRZ, HIZ, KUZ, TKHS, TLZ, and TOZ). It is subsequently shown that the lack of backarc attenuation through the Taupō
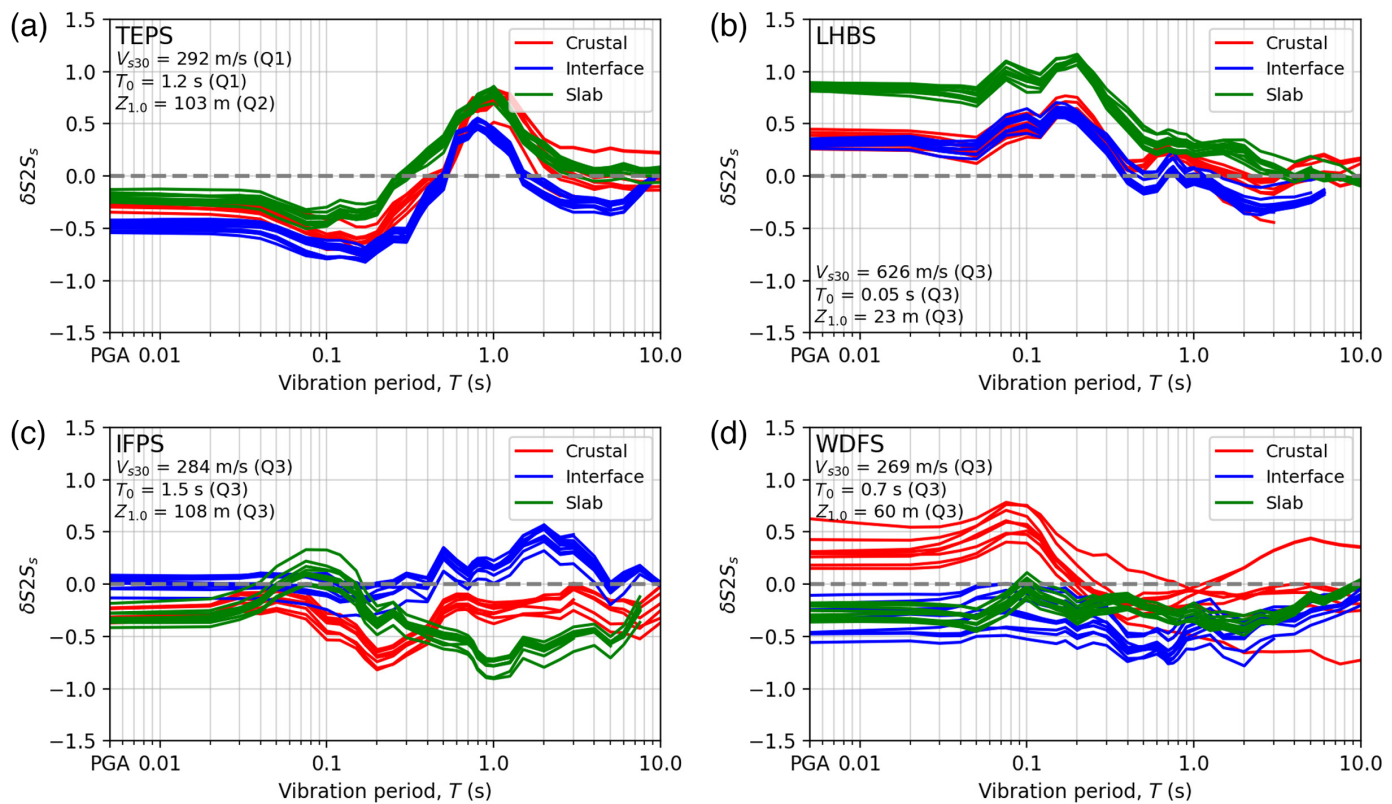
Figure 5. Within-event standard deviations (a) crustal systematic site-to-site standard deviation (ϕ_{S2S}), (b) crustal remaining within-event standard deviation (ϕ_{SS}), (c) interface ϕ_{S2S}, (d) interface ϕ_{SS}, (e) slab ϕ_{S2S}, and (f) slab ϕ_{SS}. The color version of this figure is available only in the electronic edition.

Volcanic Zone in the models likely led to relative overprediction at these stations.

$\delta S2S_s$ determined from the analyses of the different tectonic classes were also compared with one another for the same site to understand whether similarities or differences exist in the $\delta S2S_s$ between tectonic classes. In most cases, the $\delta S2S_s$ were found to be similar in values and/or relative period-dependent trends. Figure 6a,b presents examples in which similarities are present, with each line representing a model and the colors denoting the tectonic class (i.e., 8 red lines for the eight crustal models, 10 blue lines for the 10 interface models, and 10 green lines for the 10 slab models). Site characteristics V_{S30}, fundamental site period (T_0), and depth to 1.0 km/s shear-wave velocity horizon ($Z_{1.0}$)

are included in each plot, along with corresponding quality flags for parameter estimates based on the NSHM site characterization database (Wotherspoon *et al.*, 2022). Q1 corresponds to the highest quality, and Q3 corresponds to the lowest quality. For detailed information on the site characteristics and quality flags, the reader is referred to Wotherspoon *et al.* (2022). On the contrary, some sites showed significant differences between the $\delta S2S_s$ determined from the various

Figure 6. Systematic site-to-site residuals ($\delta S2S_s$) from crustal, interface, and slab models for (a) TEPS, (b) LHBS, (c) IFPS, and (d) WDFS. Site characteristics for each site are included within their respective plots. The locations of the four sites are shown in Figure 2b. The color version of this figure is available only in the electronic edition.

tectonic classes, as illustrated in Figure 6c,d. Further investigation into the cause of differences is necessary, because it may be due to poor or erroneous recording setup (e.g., incorrect instrument installation and calibration, or digital processing), poor ground-motion record distributions, or real physical ground-motion phenomena (e.g., differences in site effects and how they are modeled).

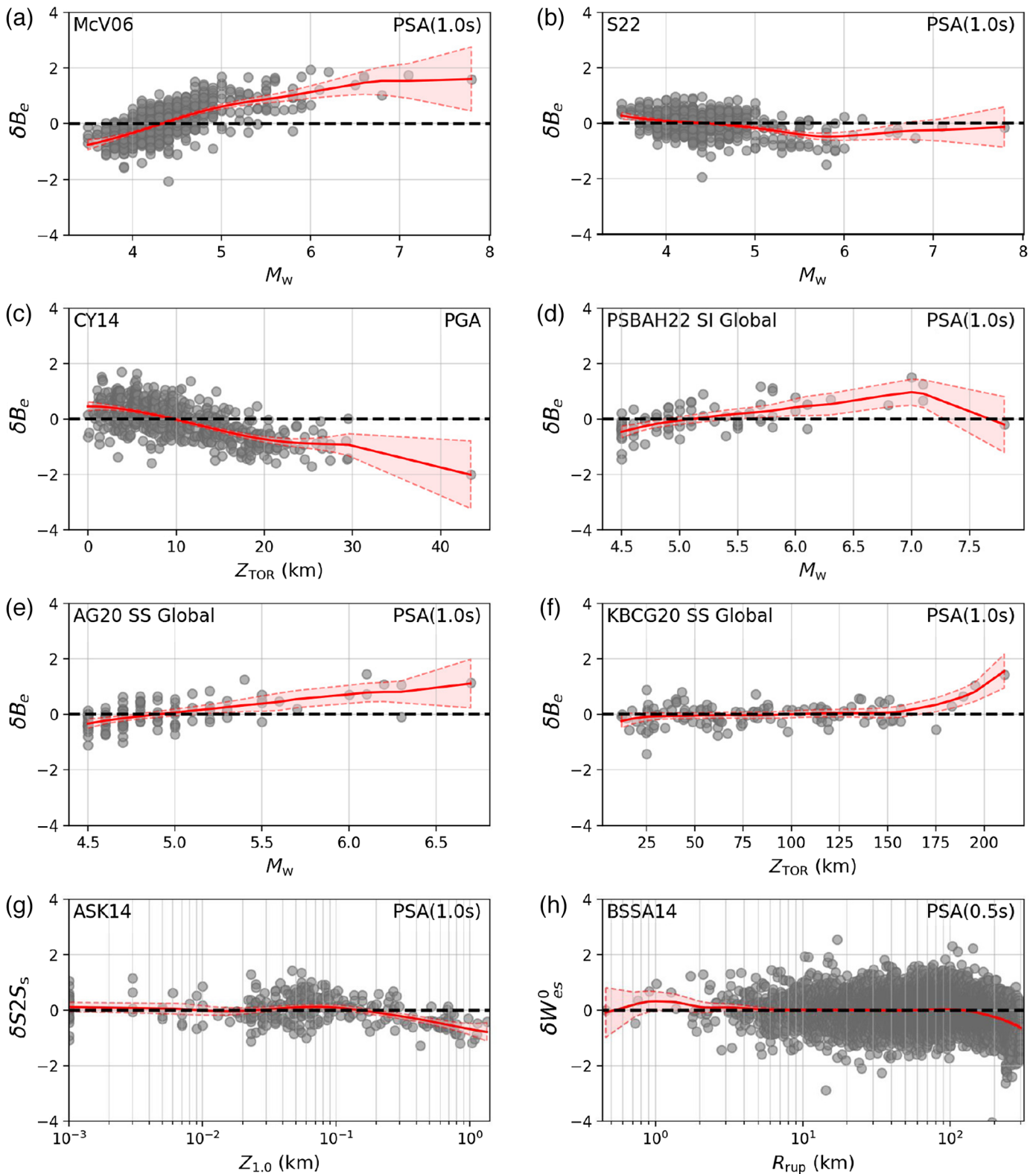### Parameter dependence of residuals

To understand biases in parameter scaling, the partitioned residuals are plotted against their appropriate predictor parameters. Specific IMs considered are PGA, PSA(0.5s), PSA(1.0s), and PSA(3.0s). Locally weighted regression lines are included within each scatterplot, along with 95% confidence intervals. Several trends and systematic features are present; however, only those with relevant implications for hazard modeling or that provide direct feedback toward GMDB improvement are discussed herein. A comprehensive set of plots is provided in the supplemental material to this article for source, site, and path parameters.

As a summary of the subsequent subsections, parameter biases seem to be most prevalent within source parameters, whereas site and path parameters show lesser biases. The A22 and S22 models tend to be relatively well centered within their applicable parameter ranges, which could be expected, given that their adjustments to NZ conditions were based on the same GMDB. Parameter dependence is similar among the A16, A18, and AG20 (both global and NZ regionalization) models within each tectonic class (i.e., interface and slab). Likewise,
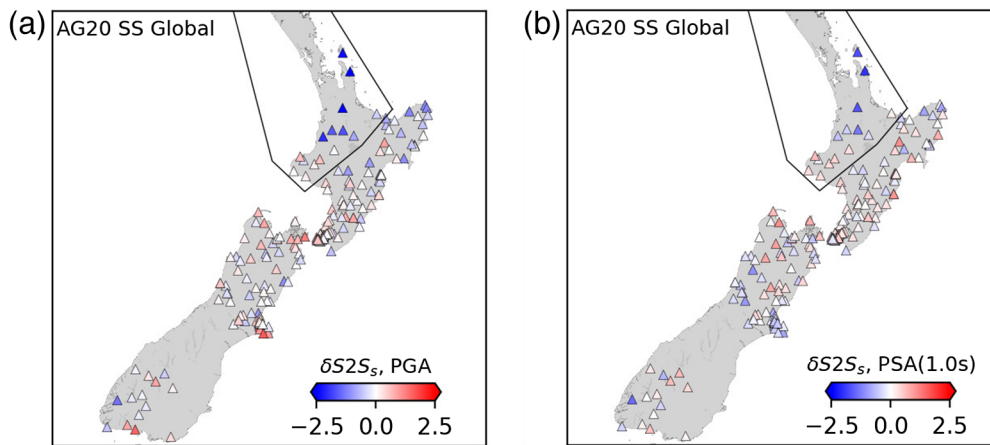
parameter dependence is similar between the KBCG20 global and NZ regionalizations within each tectonic class.

Source parameter dependence. Dependence of δB_e on **M** and depth to top of rupture (Z_{TOR}) were examined. For crustal models, it was previously observed that McV06 had a large negative model prediction bias of around -2.0, and S22 had a relatively consistent model prediction bias of around 0.5 across most of the spectral period range. However, both of these model prediction biases are partially offset toward more accurate prediction once the δB_e within their applicable **M** ranges are considered. This is illustrated in Figure 7a for McV06, using PSA(1.0s) as an example, in which the δB_e trendline ranges from approximately 1.0 to 1.5 between **M** 5.25 and 7.5. With consideration of both model prediction bias and the δB_e within the applicable **M** range, McV06 would still have the largest overprediction of the crustal models ($a \approx -0.5$), although not as large as suggested by the model prediction bias alone. Similarly, for S22, the δB_e between **M** 4.5 and 8.4 is typically between -0.5 and 0, as shown in Figure 7b. This would offset the model prediction bias ($a \approx 0.5$) and indicate that the S22 model is relatively unbiased within its applicable **M** range.

Figure 7. Selected residual parameter dependence plots for (a) δB_e versus **M** for McV06 PSA(1.0s), (b) δB_e versus **M** for S22 PSA(1.0s), (c) δB_e versus Z_{TOR} for CY14 peak ground acceleration (PGA), (d) δB_e versus **M** for PSBAH22 SI Global PSA(1.0s), (e) δB_e versus **M** for AG20 SS Global PSA(1.0s), (f) δB_e versus Z_{TOR} for KBCG20 SS Global PSA(1.0s), (g) $\delta S2S_s$ versus $Z_{1.0}$ for ASK14 PSA(1.0s), and (h) δW_{es}^0 versus R_{rup} for BSSA14 PSA(0.5s). The color version of this figure is available only in the electronic edition.

In terms of trends with Z_{TOR} for crustal models, the ASK14 and CB14 models show a negative trend between 0 and 20 km for PGA, whereas CY14 shows a negative trend down to 30 km for PGA. The latter case is shown as an example in Figure 7c. However, it is difficult to draw any strong inferences from this due to the large uncertainties and biases in the source depth

Figure 8. Spatial map plots of systematic site-to-site residual ($\delta S2S_s$) for AG20 SS Global (a) PGA and (b) PSA(1.0s). Black outline in the North Island indicates the adopted backarc polygon. The color version of this figure is available only in the electronic edition.

determination from which Z_{TOR} is calculated in the database (Ristau, 2013). Earthquake locations should be improved using more accurate location determination methods, and tectonic classifications should also be checked, because there are some peculiar earthquakes within the crustal set (e.g., a crustal earthquake with Z_{TOR} of 43 km, which is below common estimates of crustal seismogenic depth).

For interface models, positive **M**-dependence up to **M** 7 is identified for PSA(0.5s), PSA(1.0s), and PSA(3.0s) and for all the interface models, with the exception of A22 SI. Figure 7d for PSBAH22 SI Global PSA(1.0s) is presented as an example of this. This suggests that the **M**-scaling of the models may have inaccuracies when compared to NZ data. Surprisingly, this trend is not present for the PGA for any model. Trends at large **M** cannot be inferred, because few large earthquakes are present in the database. Notably, the largest **M** interface earthquake is the **M** 7.8 2009 Dusky Sound earthquake that occurred in the Puysegur Subduction Zone, which is typically of less interest than the Hikurangi Subduction Zone in the seismic hazard context due to proximity to major population densities and critical infrastructure considerations.

For slab models, there appears to be slight positive **M**-dependence for PSA(0.5s), PSA(1.0s), and PSA(3.0s) and for all models, with the exception of the PSBAH22 SS Global and A22 SS models. Figure 7e for AG20 SS Global PSA(1.0s) is presented as an example of this. This again suggests there may be differences in **M**-scaling between the models and NZ data. In terms of trends with Z_{TOR}, for PGA and all response spectral periods examined, there does not appear to be any strong trend except for systematically positive values of δB_e for Z_{TOR} at depths greater than 160 km. This is close to the deepest Z_{TOR} that most models are generally applicable for, which may be the cause of this bias, because there may be weaker constraints at those deep Z_{TOR}. Figure 7f for KBCG20 SS Global PSA(1.0s) is presented as an example of this.
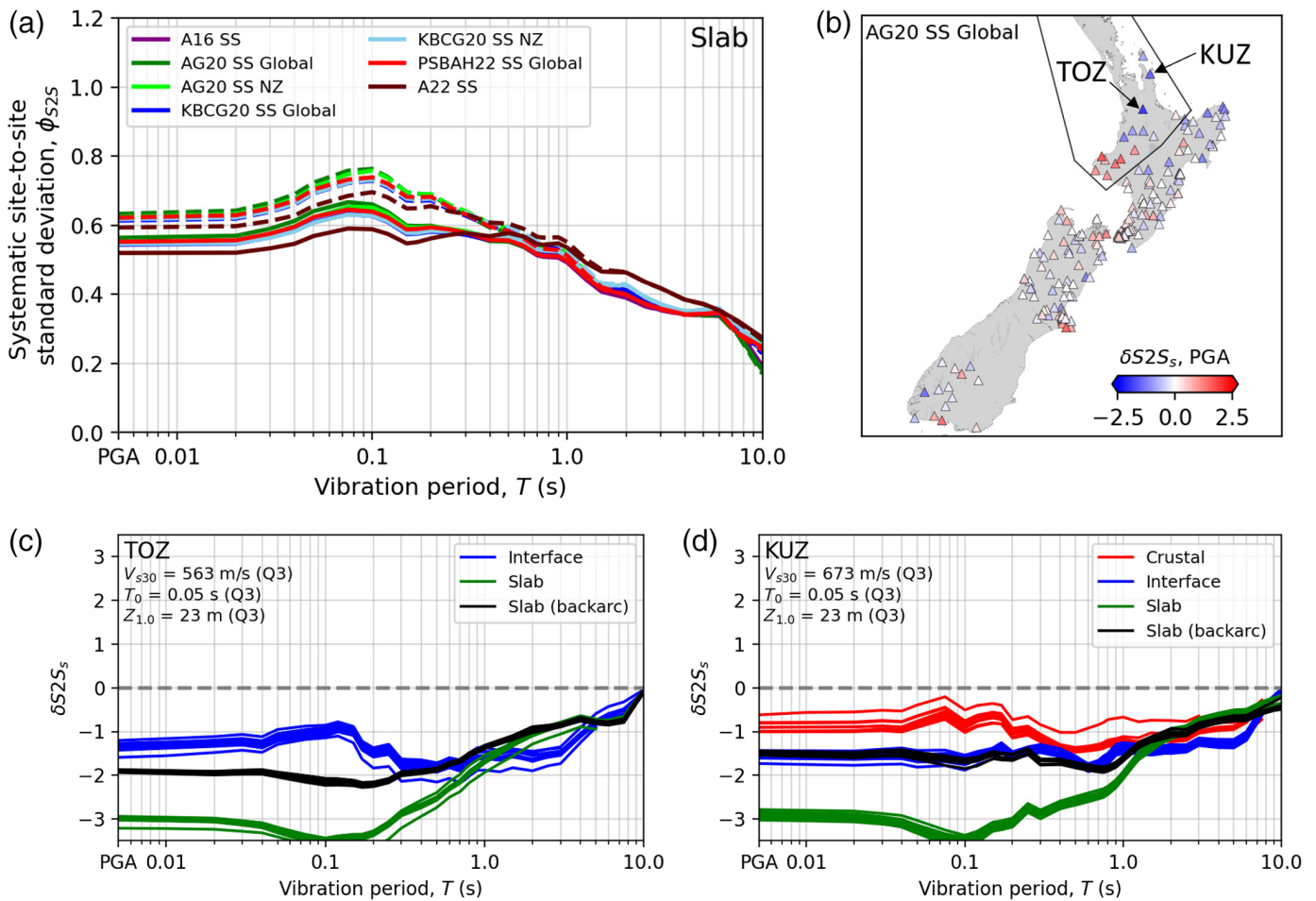
Site parameter dependence. Dependence of $\delta S2S_s$ on V_{S30}, $Z_{1.0}$, and depth to 2.5 km/s shear-wave velocity horizon ($Z_{2.5}$) were examined. Across all the tectonic types and response spectral periods examined, there does not appear to be any noteworthy trend with V_{S30}. For crustal models, plots of $\delta S2S_s$ against $Z_{1.0}$ show systematically negative $\delta S2S_s$ in which $Z_{1.0}$ is greater than 300 m. Figure 7g for ASK14 PSA(1.0s) is presented as an example of this. This may suggest that there is a systematic difference in deep sedimentary basin response between the models and NZ data, or a different intrinsic correlation between $Z_{1.0}$ and other site parameters in the GMMs (e.g., V_{S30}). However, it is observed that the data quality of NZ-specific $Z_{1.0}$ and $Z_{2.5}$ values is highly variable, and therefore such trends are difficult to draw strong inferences from. Many $Z_{1.0}$ values are extracted from a regional velocity model in which some sedimentary basin models are explicitly included but many are not; hence, $Z_{1.0}$ values may not be accurate, especially for lower quality estimates. Likewise, $Z_{2.5}$ values are mostly determined using a coarse travel-time tomography velocity model (Eberhart-Phillips *et al.*, 2020) in the absence of more accurate data-driven estimates. A better characterization of the two parameters is likely required to properly understand the model dependence of these parameters (as proxies for basin characteristics) relative to NZ data.

Path parameter dependence. Dependence of δW_{es}^0 on R_{rup} was examined. δW_{es}^0 from a few models were found to systematically deviate from zero at large R_{rup}. The most significant cases (of those examined) are crustal models for PSA(0.5s) with a slight negative curve at distances greater than 200 km. Figure 7h for BSSA14 PSA(0.5s) is presented as an example of this. This is likely a result of being near or beyond the upper limit of R_{rup} applicability, which is 300 km for many of these models. This is unlikely to be an issue, because large R_{rup} scenarios typically have less contributions to seismic hazard. There are also some records with large δW_{es}^0 within each of the tectonic class analyses (greater than 3 in absolute value), which should be further investigated, although poor record quality or recording error are likely to be the cause.

Spatial trends of residuals

Spatial plots of δB_e and $\delta S2S_s$ can provide insight into whether any regional trends are present. An exhaustive set of spatial

Figure 9. Updated results following implementation of the backarc attenuation modification (a) ϕ_{S2S} for slab models without (dashed) and with (solid) backarc attenuation modification, (b) a spatial map plot of $\delta S2S_s$ for AG20 SS Global PGA, (c) $\delta S2S_s$ for all models at station TOZ, and (d) $\delta S2S_s$ for all models at station KUZ. The locations of TOZ and KUZ are shown in panel (b). In panel (a), the dashed and solid lines indicate the slab models without and with the backarc attenuation modification, respectively. The color version of this figure is available only in the electronic edition.

plots is provided in the supplemental material to this article due to space constraints, and only the key trend identified is discussed here. From plots of $\delta S2S_s$ associated with slab models across NZ, a group of large negative $\delta S2S_s$ values were found to be present throughout and to the west of the Taupō Volcanic Zone in the north-west region of the North Island. The large negative $\delta S2S_s$ values indicate that the IMs are systematically overpredicted. The extent of overprediction is larger for short periods and smaller for long periods. This is illustrated in Figure 8 for PGA (more negative $\delta S2S_s$) and PSA(1.0s) (less negative $\delta S2S_s$) predicted using the AG20 SS Global model as an example. All other slab models also display these features. Given the location and distribution of these sites, and the period-dependence of the $\delta S2S_s$, this overprediction is likely due to the absence of increased backarc attenuation for waves traveling through the Taupō Volcanic Zone in slab models. This prompted the addition of the Abrahamson et al. (2016) backarc attenuation term into the final suite of slab models selected for the NSHM (i.e., AG20 SS Global, PSBAH22 SS Global, KBCG20 SS Global, and A22 SS) and a backarc polygon (indicated by the black-outlined polygon in Fig. 8) to help provide more realistic median predictions (Bradley et al., 2022, 2024). The backarc flag is set to true for stations in the backarc polygon.

A subsequent iteration of the GMM evaluation was conducted with the backarc attenuation modifications implemented to quantify its effect. Figure 9 provides a summary of the updated results with a plot of ϕ_{S2S}, a spatial plot of PGA $\delta S2S_s$ (AG20 SS Global) to contrast with Figure 8a, and plots of $\delta S2S_s$ as a function of vibration period from all models, including slab models with backarc attenuation modification, for two stations in the backarc region (TOZ and KUZ). The model prediction bias (not shown) at short periods increases by less than 0.1 with this change, because only a small percentage of records are affected by this change. In contrast, the ϕ_{S2S} values for the six slab models that include the backarc attenuation term (i.e., the final suite of slab models for the NSHM, any associated NZ regionalizations, and A16 SS) have decreased, indicating the subset of sites affected are more

accurately predicted now. This is illustrated in Figure 9a, in which the dashed and solid lines are the ϕ_{S2S} when the models exclude and include the backarc attenuation term, respectively. The reductions in ϕ_{S2S} are also reflected in reductions in the corresponding σ (not shown). Comparing the spatial map plots in Figures 9b to 8a, the extent of overprediction at the backarc sites is now reduced. With the backarc polygon adopted, there is a trade-off of larger underprediction in the Taranaki region on the west of the North Island though, where the $\delta S2S_s$ values are now larger positive values. Finally, the $\delta S2S_s$ at two stations are used to illustrate the improvement across the considered period range, in which a comparison between slab models with and without the backarc attenuation term indicates the improvement is largest at short periods, which expectedly coincides with the largest negative $\delta S2S_s$. Although the backarc modification has drawbacks, it was deemed the most appropriate solution given time constraints, and more rigorous treatment of backarc effects will be explored in future studies (Bradley et al., 2022, 2024).

DISCUSSION AND LIMITATIONS

The most significant limitation of this study is the lack of large **M** earthquakes (and, to a lesser extent, short R_{rup} records) used in the evaluation, despite their tendency to dominate the seismic hazard in most engineering applications. With an NZ-only GMDB, this limitation is unavoidable. Hence, this evaluation was primarily conducted to determine whether:

1. A model is considered for inclusion in the NSHM, given the limitations of the observed rupture scenarios and uncertainty in metadata values (from which the McVerry et al. [2006] and Zhao et al. [2006] models were excluded).
2. A model that has been superseded by more recent developments provides similar or different predictions (from which the Abrahamson et al. [2016] and Abrahamson et al. [2018] models were excluded).
3. The NZ-specific regionalizations of the NGA-Sub project (Abrahamson and Gülerce, 2020; Kuehn et al., 2020) provided any superior prediction relative to their global models.

For the third item, it was found with respect to the NZ data that the performance of the NZ-specific regionalizations of the Kuehn et al. (2020) and Abrahamson and Gülerce (2020) models was not better than their global counterparts. Other aligned studies (Lee et al., 2022; Bradley et al., 2022, 2024) compared the models between one another and against NGA-Sub data for scenarios that are not (or poorly) represented in the NZ data. It was found that the NZ-specific regionalizations of Kuehn et al. (2020) and Abrahamson and Gülerce (2020) models performed less favorably. This is likely due to the regionalizations being based on the Van Houtte et al. (2017) strong-motion database that was smaller and also had relatively poor metadata quality, and adjustments to the unique subduction zone characteristics in NZ may not be as robust as the global or

Cascadia-specific developments. Therefore, NZ regionalizations of the NGA-Sub models were excluded.

Additional comments need to be made regarding the GMDB version adopted in this study. The GMDB had recently achieved v.1.0 status at the time when the analyses were conducted, in which its quality was considered adequate for research applications. However, this study initially revealed that the quality of particular data sources was still not adequate for strong-motion prediction applications; hence, only a small fraction of available data was used in this study. Several studies are currently underway to improve the various data sources that feed into the database. Earthquake source descriptions will be improved, including more accurate location, magnitude, focal mechanism, finite fault dimensions, and aftershock flags. Ongoing site characterization studies at ground-motion recording stations will also help characterize the accuracy and precision of ground-motion predictions. Improvements to record reliability are being investigated so that broadband instrument channels can be included. A few broadband channels have been flagged as potentially having erroneous conversion factors, instrument configuration, or installation issues. Interim updates of the GMDB (e.g., v.3.2) have since been released, but none have had the technical scrutiny and quality assurance and control as the GMDB v.1.0 adopted in this study, which remains as the only stable version so far. In particular, reprocessing of f_{min} using an updated neural network seems to have predicted values that are too low for many records, thus compromising the quality of some PSA calculations. Within the limitations and shortcomings of more recent GMDB versions, the empirical GMMs were assessed, with a subsequent iteration of evaluation, to have performed similarly within the parameter domain considered credible. Hence, the outcomes of this study remain the same.

A key difference in the use of the GMMs between this evaluation and the PSHA calculation for the NSHM are the basin depth parameters, $Z_{1.0}$ and $Z_{2.5}$. In this evaluation, the site-specific values from the GMDB were used. However, the PSHA calculation was conducted on a nationwide scale for a suite of constant V_{S30} values. Therefore, corresponding $Z_{1.0}$ and $Z_{2.5}$ values were determined from appropriate correlations with V_{S30} (Chiou and Youngs, 2014; Campbell and Bozorgnia, 2014). This was necessary, because high-quality estimates of $Z_{1.0}$ and $Z_{2.5}$ at every possible location in NZ are not feasible compared to high-quality estimates made at the relatively small number of instrumental locations in the GMDB. Of the final suite of GMMs selected for the NSHM, only crustal models were affected. A subsequent iteration of this evaluation was conducted, making use of $Z_{1.0}$ and $Z_{2.5}$ calculated from correlations with V_{S30} to understand the extent of this difference. Small differences were found in model prediction bias and standard deviations (σ and ϕ_{S2S}) at moderate-to-long periods. However, they remained within the range of values observed previously when site-specific $Z_{1.0}$ and $Z_{2.5}$ from the GMDB were used. Ultimately, the differences were small enough such that this

evaluation is considered to extend well to the NSHM implementation of GMMs.

Analysis methods and interpretation of results can also be improved, particularly with respect to identifying the cause of observed residual biases and trends. Importantly, this work will serve as a basis for subsequent regional examination of ground-motion phenomena toward more advanced prediction methods, such as nonergodic analysis. This will be critical for the improvement of ground-motion prediction accuracy and precision that are limited in ergodic models (Lavrentiadis *et al.*, 2022). A deeper investigation of specific data subsets may also yield insights into such causes. For example, the quality of $Z_{1.0}$ values are known to be highly variable and can form discrete categories, depending upon whether the values are measured or estimated and, in the latter case, further categorized in terms of how the estimate is obtained. Equivalent sentiments also apply for V_{S30}. An analysis of systematic differences between Hikurangi and Puysegur subduction zone earthquakes would also be informative, as different subducting interfaces are understood to likely produce systematically different earthquakes.

CONCLUSIONS

This study has provided an evaluation of candidate empirical GMMs for active shallow crustal, subduction interface, and subduction slab earthquakes using an NZ GMDB developed for the 2022 NZ NSHM revision (Hutchinson *et al.*, 2022). The evaluation provided a quantitative comparison between the models and investigated the presence of any biases in the model predictions when applied to NZ data. This informed decisions regarding appropriate models to include in the NSHM and provided some insights toward logic tree weighting in ground-motion characterization modeling. Ultimately, many models showed reasonable performance and were considered appropriate to include within suites of models to properly represent ground-motion predictions and epistemic uncertainty. In general, the recent models that are NZ-specific (Bradley, 2013; Atkinson, 2022; Stafford, 2022) and global models developed on large international databases (Abrahamson *et al.*, 2014; Boore *et al.*, 2014; Bozorgnia *et al.*, 2014; Chiou and Youngs, 2014; Abrahamson and Gülerce, 2020; Kuehn *et al.*, 2020; Parker *et al.*, 2022) performed the best. In addition, spatial trends in systematic site-to-site residuals established the need for backarc attenuation modifications for slab models. Finally, the evaluation process has been integral to informing improvements to the GMDB throughout the duration of the project and has also identified several future improvements. As previously mentioned, the final logic tree and weighting used in the 2022 NZ NSHM revision were not included in this article, because this evaluation was not the sole determinant of the corresponding decisions, and the reader is directed to Bradley *et al.* (2024) and Gerstenberger *et al.* (2024) for those details.

Although a comprehensive evaluation of models was conducted and presented in this study, there were still several limitations and shortcomings that were acknowledged. To this end, there are many improvements and pathways for further investigations that would lead to more informed decisions in hazard modeling. Studies to develop and implement more rigorous modeling of backarc attenuation will be essential for more accurate ground-motion characterization. As data quality and reliability improve, resulting in more usable data, iterations of this evaluation with more robust and statistically rigorous analyses will be possible. In addition, a more exhaustive examination of model predictions for significant scenarios not covered by the GMDB and an investigation of model standard deviations will be critical for more informed PSHA calculations too.

DATA AND RESOURCES

DECLARATION OF COMPETING INTERESTS

The authors acknowledge that there are no conflicts of interest recorded.

ACKNOWLEDGMENTS

REFERENCES

Abrahamson, N. A., and Z. Gülerce (2020). Regionalized ground-motion models for subduction earthquakes based on the NGA-sub database, *Technical Rept. 2020/25, PEER*, doi: 10.55461/SSXE9861.

Abrahamson, N. A., and W. J. Silva (1997). Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seismol. Res. Lett.* **68,** no. 1, 94–127.

Abrahamson, N., N. Gregor, and K. Addo (2016). BC Hydro ground motion prediction equations for subduction earthquakes, *Earthq. Spectra* **32,** no. 1, 23–44.

Abrahamson, N., N. Kuehn, Z. Gülerce, N. Gregor, Y. Bozorgnia, G. Parker, J. Stewart, B. Chiou, I. Idriss, K. Campbell, *et al.* (2018). Update of the BC hydro subduction ground-motion model using

the NGA-subduction dataset, *Technical Rept. 2018/02*, PEER, doi: 10.55461/OYCD7434.

Abrahamson, N. A., W. J. Silva, and R. Kamai (2014). Summary of the ASK14 ground motion relation for active crustal regions, *Earthq. Spectra* **30,** no. 3, 1025–1055.

Al Atik, L., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010). The variability of ground-motion prediction models and its components, *Seismol. Res. Lett.* **81,** no. 5, 794–801.

Atkinson, G. (2022). Backbone ground-motion models for crustal, interface and slab earthquakes in New Zealand, *Technical Rept. 2022/48*, GNS Science, doi: 10.21420/QMJ6-P189.

Boore, D. M. (2010). Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion, *Bull. Seismol. Soc. Am.* **100,** no. 4, 1830–1835.

Boore, D. M., and T. Kishida (2017). Relations between some horizontal-component ground-motion intensity measures used in practice, *Bull. Seismol. Soc. Am.* **107,** no. 1, 334–343.

Boore, D. M., J. P. Stewart, E. Seyhan, and G. M. Atkinson (2014). NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes, *Earthq. Spectra* **30,** no. 3, 1057–1085.

Bozorgnia, Y., N. A. Abrahamson, S. K. Ahdi, T. D. Ancheta, L. A. Atik, R. J. Archuleta, G. M. Atkinson, D. M. Boore, K. W. Campbell, B. S. J. Chiou, *et al.* (2022). NGA-Subduction research program, *Earthq. Spectra* **38,** no. 2, 783–798.

Bozorgnia, Y., N. A. Abrahamson, L. A. Atik, T. D. Ancheta, G. M. Atkinson, J. W. Baker, A. Baltay, D. M. Boore, K. W. Campbell, B. S.-J. Chiou, *et al.* (2014). NGA-West2 research project, *Earthq. Spectra* **30,** no. 3, 973–987.

Bradley, B., S. Bora, R. Lee, E. Manea, M. Gerstenberger, P. Stafford, G. Atkinson, G. Weatherill, J. Hutchinson, C. de la Torre, *et al.* (2024). Summary of the ground-motion characterization model for the 2022 New Zealand National Seismic Hazard Model, *Bull. Seismol. Soc. Am.*, in revision.

Bradley, B. A. (2012). Strong ground motion characteristics observed in the 4 September 2010 Darfield, New Zealand earthquake, *Soil Dynam. Earthq. Eng.* **42,** 32–46.

Bradley, B. A. (2013). A New Zealand-specific pseudospectral acceleration ground-motion prediction equation for active shallow crustal earthquakes based on foreign models, *Bull. Seismol. Soc. Am.* **103,** no. 3, 1801–1822.

Bradley, B. A., and M. Cubrinovski (2011). Near-source strong ground motions observed in the 22 February 2011 Christchurch earthquake, *Seismol. Res. Lett.* **82,** no. 6, 853–865.

Bradley, B. A., S. Bora, R. L. Lee, E. F. Manea, M. C. Gerstenberger, P. J. Stafford, G. M. Atkinson, G. Weatherill, J. Hutchinson, C. de la Torre *et al.* (2022). Summary of the ground-motion characterisation model for the 2022 New Zealand National Seismic Hazard Model, *Technical Rept. 2022/46*, GNS Science, doi: 10.21420/9BMK-ZK64.

Bradley, B. A., H. N. Razafindrakoto, and V. Polak (2017). Ground-motion observations from the 14 November 2016 Mw 7.8 Kaikoura, New Zealand, earthquake and insights from broadband simulations, *Seismol. Res. Lett.* **88,** no. 3, 740–756.

Campbell, K. W., and Y. Bozorgnia (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra, *Earthq. Spectra* **30,** no. 3, 1087–1115.

Chiou, B., R. Youngs, N. Abrahamson, and K. Addo (2010). Ground-motion attenuation model for small-to-moderate shallow crustal earthquakes in California and its implications on regionalization of ground-motion prediction models, *Earthq. Spectra* **26,** no. 4, 907–926.

Chiou, B. S.-J., and R. R. Youngs (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra, *Earthq. Spectra* **30,** no. 3, 1117–1153.

Douglas, J., and B. Edwards (2016). Recent and future developments in earthquake ground motion estimation, *Earth Sci. Rev.* **160,** 203–219.

Dupuis, M., C. Schill, R. Lee, and B. Bradley (2023). A deep-learning-based model for quality assessment of earthquake-induced ground-motion records, *Earthq. Spectra* doi: 10.1177/87552930231195113.

Eberhart-Phillips, D., S. Bannister, M. Reyners, and S. Henrys (2020). New Zealand wide model 2.2 seismic velocity and Qs and Qp models for New Zealand [Dataset], *Zenodo* doi: 10.5281/zenodo.3779523.

Gerstenberger, M., S. Bora, B. Bradley, C. DiCaprio, R. Van Dissen, G. Atkinson, C. Chamberlin, A. Christophersen, K. Clark, G. Coffey, *et al.* (2022). New Zealand National Seismic Hazard Model 2022 revision: Model, hazard and process overview, *Technical Rept. 2022/57*, GNS Science, Lower Hutt, New Zealand, doi: 10.21420/TB83-7X19.

Gerstenberger, M., S. Bora, B. Bradley, C. DiCaprio, R. Van Dissen, G. Atkinson, C. Chamberlin, A. Christophersen, K. Clark, G. Coffey, *et al.* (2024). The 2022 Aotearoa New Zealand National Seismic Hazard Model: Process, overview and results, *Bull. Seismol. Soc. Am.* doi: 10.1785/0120230182.

Gerstenberger, M. C., D. A. Rhoades, and G. H. McVerry (2016). A hybrid time-dependent probabilistic seismic-hazard model for Canterbury, New Zealand, *Seismol. Res. Lett.* **87,** no. 6, 1311–1318.

Gledhill, K., J. Ristau, M. Reyners, B. Fry, and C. Holden (2011). The Darfield (Canterbury, New Zealand) Mw 7.1 earthquake of September 2010: A preliminary seismological report, *Seismol. Res. Lett.* **82,** no. 3, 378–386.

Goulet, C. A., Y. Bozorgnia, N. Kuehn, L. Al Atik, R. R. Youngs, R. W. Graves, and G. M. Atkinson (2021). NGA-east ground-motion characterization model Part I: Summary of products and model development, *Earthq. Spectra* **37,** no. 1_suppl, 1231–1282.

Gregor, N., N. A. Abrahamson, G. M. Atkinson, D. M. Boore, Y. Bozorgnia, K. W. Campbell, B. S.-J. Chiou, I. Idriss, R. Kamai, E. Seyhan, *et al.* (2014). Comparison of NGA-West2 GMPEs, *Earthq. Spectra* **30,** no. 3, 1179–1197.

Gregor, N., K. Addo, N. A. Abrahamson, L. Al Atik, G. M. Atkinson, D. M. Boore, Y. Bozorgnia, K. W. Campbell, B. S. Chiou, Z. Gülerce, *et al.* (2022). Comparisons of the NGA-subduction ground motion models, *Earthq. Spectra* **38,** no. 4, 2580–2610.

Holden, C., A. Kaiser, R. Van Dissen, and R. Jury (2013). Sources, ground motion and structural response characteristics in Wellington of the 2013 Cook Strait earthquakes, *Bull. N. Z. Soc. Earthq. Eng.* **46,** no. 4, 188–195.

Hutchinson, J., B. Bradley, R. Lee, L. Wotherspoon, M. Dupuis, C. Schill, J. Motha, A. Kaiser, and E. Manea (2022). 2021 New Zealand strong ground motion database, *Technical Rept. 2021/56*, GNS Science, doi: 10.21420/Z20E-5507.

Kuehn, N., Y. Bozorgnia, K. Campbell, and N. Gregor (2020). Partially non-ergodic ground-motion model for subduction regions using the NGA-subduction database, *Technical Rept. 2020/04, PEER*, doi: 10.55461/NZZW1930.

Lavrentiadis, G., N. A. Abrahamson, K. M. Nicolas, Y. Bozorgnia, C. A. Goulet, A. Babič, J. Macedo, M. Dolšek, N. Gregor, A. R. Kottke, *et al.* (2022). Overview and introduction to development of non-ergodic earthquake ground-motion models, *Bull. Earthq. Eng.* **21,** 5121–5150.

Lee, R. L., B. A. Bradley, E. F. Manea, and J. Hutchinson (2022). Evaluation of empirical ground-motion models for New Zealand application, *Technical Rept. 2021/61*, GNS Science, doi: 10.21420/W2M5-YC09.

Mak, S., R. A. Clements, and D. Schorlemmer (2017). Empirical evaluation of hierarchical ground-motion models: Score uncertainty and model weighting, *Bull. Seismol. Soc. Am.* **107,** no. 2, 949–965.

Matuschka, T. (1980). Assessment of seismic hazards in New Zealand, *Ph.D. thesis*, University of Auckland.

McVerry, G. H., J. X. Zhao, N. A. Abrahamson, and P. G. Somerville (2006). New Zealand acceleration response spectrum attenuation relations for crustal and subduction zone earthquakes, *Bull. N. Z. Soc. Earthq. Eng.* **39,** no. 1, 1–58.

New Zealand Standards (2004). Structural design actions - Part 5: Earthquake actions - New Zealand, Standards New Zealand, Wellington, New Zealand, available at https://www.standards.govt.nz/shop/nzs-1170-52004-includes-amdt-1/ (last accessed March 2023).

Parker, G. A., J. P. Stewart, D. M. Boore, G. M. Atkinson, and B. Hassani (2022). NGA-subduction global ground motion models with regional adjustment factors, *Earthq. Spectra* **38,** no. 1, 456–493.

Peek, R. (1980). Estimation of seismic risk for New Zealand, *Technical Rept. 80-21*, Department of Civil Engineering Research Report, University of Canterbury, ISSN: 0110-3326.

Power, M., B. Chiou, N. Abrahamson, Y. Bozorgnia, T. Shantz, and C. Roblee (2008). An overview of the NGA project, *Earthq. Spectra* **24,** no. 1, 3–21.

Ristau, J. (2013). Update of regional moment tensor analysis for earthquakes in New Zealand and adjacent offshore regions, *Bull. Seismol. Soc. Am.* **103,** no. 4, 2520–2533.

Scherbaum, F., E. Delavaud, and C. Riggelsen (2009). Model selection in seismic hazard analysis: An information-theoretic perspective, *Bull. Seismol. Soc. Am.* **99,** no. 6, 3234–3247.

Stafford, P. (2022). A model for the distribution of response spectral ordinates from New Zealand crustal earthquakes based upon adjustments to the Chiou and Youngs (2014) response spectral model, *Technical Rept. 2022/15*, GNS Science, doi: 10.21420/5098-0S19.

Stirling, M., G. McVerry, M. Gerstenberger, N. Litchfield, R. Van Dissen, K. Berryman, P. Barnes, L. Wallace, P. Villamor, R. Langridge, *et al.* (2012). National seismic hazard model for New Zealand: 2010 update, *Bull. Seismol. Soc. Am.* **102,** no. 4, 1514–1542.

Van Houtte, C. (2017). Performance of response spectral models against New Zealand data, *Bull. N. Z. Soc. Earthq. Eng.* **50,** no. 1, 21–38.

Van Houtte, C., S. Bannister, C. Holden, S. Bourguignon, and G. McVerry (2017). The New Zealand strong motion database, *Bull. N. Z. Soc. Earthq. Eng.* **50,** no. 1, 1–20.

Wotherspoon, L., A. Kaiser, E. Manea, and A. Stolte (2022). National Seismic Hazard Model: Site characterisation database summary report, *Technical Rept. 2022/28*, GNS Science, Lower Hutt, New Zealand, doi: 10.21420/363X-CK83.

Youngs, R., S.-J. Chiou, W. Silva, and J. Humphrey (1997). Strong ground motion attenuation relationships for subduction zone earthquakes, *Seismol. Res. Lett.* **68,** no. 1, 58–73.

Zhao, J. X., F. Jiang, P. Shi, H. Xing, H. Huang, R. Hou, Y. Zhang, P. Yu, X. Lan, D. A. Rhoades, *et al.* (2016). Ground-motion prediction equations for subduction slab earthquakes in Japan using site class and simple geometric attenuation functions, *Bull. Seismol. Soc. Am.* **106,** no. 4, 1535–1551.

Zhao, J. X., X. Liang, F. Jiang, H. Xing, M. Zhu, R. Hou, Y. Zhang, X. Lan, D. A. Rhoades, K. Irikura, *et al.* (2016). Ground-motion prediction equations for subduction interface earthquakes in Japan using site class and simple geometric attenuation functions, *Bull. Seismol. Soc. Am.* **106,** no. 4, 1518–1534.

Zhao, J. X., J. Zhang, A. Asano, Y. Ohno, T. Oouchi, T. Takahashi, H. Ogawa, K. Irikura, H. K. Thio, P. G. Somerville, *et al.* (2006). Attenuation relations of strong ground motion in Japan using site classification based on predominant period, *Bull. Seismol. Soc. Am.* **96,** no. 3, 898–913.